Modelling the response of Antarctic marine species to environmental changes. Methods, applications and limitations.

Charlène Guillaumot

Modelling the response of Antarctic marine species to environmental changes. Methods, applications and limitations.

Thesis submitted by Charlène GUILLAUMOT
in fulfilment of the requirements of the PhD Degree
in Sciences (ULB - "Docteur en Sciences")
and in Life Science (UBFC - "Docteur en Sciences de la vie")

Academic year 2020-2021

Supervisors: Professor Bruno DANIS (Université Libre de Bruxelles)
Laboratoire de Biologie Marine

and Doctor Thomas SAUCEDE (Université de Bourgogne Franche-Comté)
Biogéosciences

Thesis jury:

Mrs Nathalie GYPENS (Université Libre de Bruxelles, Chair)
Mr Bruno DANIS (Université Libre de Bruxelles, Secretary, Thesis supervisor)
Mr Thomas SAUCEDE (Université de Bourgogne Franche-Comté, Thesis co-supervisor)
Mrs Nicole HILL (University of Tasmania, Institute for Marine and Antarctic Studies)
Mrs Sophie MORMEDE (SoFishconsulting)
Mr Jean-Olivier IRISSON (Sorbonne University, Station marine de Villefranche-sur-Mer)
Mr Nicolas NAVARRO (Université de Bourgogne Franche-Comté)

MODELLING THE RESPONSE OF ANTARCTIC MARINE SPECIES TO ENVIRONMENTAL CHANGES

METHODS, APPLICATIONS AND LIMITATIONS

GUILLAUMOT Charlène

PhD thesis in Sciences

Université Libre de Bruxelles, Marine Biology Laboratory
Université de Bourgogne Franche-Comté, Biogéosciences

Lemaire Channel, Antarctic Peninsula

REMERCIEMENTS / ACKNOWLEDGEMENTS

Je voudrais commencer par remercier mon jury de thèse qui a bien voulu accepter mon invitation pour relire et commenter ce travail. Cela me fait énormément plaisir que ce soit vous tous, car vous faites partie des chercheurs que j'ai pu rencontrer au cours de mes études ou de ma thèse et pour qui j'ai ressenti une très grande admiration. Merci pour votre temps !!

Merci beaucoup à toi Bruno pour m'avoir accueilli chaleureusement au labo et pour m'avoir donné l'opportunité d'avoir pu participer au concours FRIA, pour y décrocher ce, enfin… ces contrats de travail ! et pour l'opportunité d'avoir pu participer à l'expédition Belgica121…une expérience unique ! J'en garde un souvenir grandiose… !! Merci également pour la liberté que tu as su m'accorder pour mener cette thèse, pour ta générosité et ta sympathie.

Un énorme merci à Thomas. Tu es vraiment un encadrant hors paire qui sait offrir le temps, l'énergie et le sérieux nécessaires pour empiler le travail bien fait, et tu sais trop bien transmettre la motivation de continuer à vouloir encore faire plus et mieux. J'ai vraiment eu beaucoup de chance de t'avoir eu comme mentor pour découvrir le monde de la recherche depuis mes débuts en Master et d'autant plus de chance d'avoir pu continuer avec toi en thèse !! Merci pour tout.

Merci beaucoup à Philippe pour le temps que tu as pris pour relire et relire mes projets FRIA, FNRS et FRIA^2 ! et m'entrainer à l'oral pour la préparation des concours. Je pense que ça aurait été quasiment impossible que ça marche sans tes pointilleux apports ! Et merci à Serge Aron qui m'a soutenue pour le concours, et à qui je dois aussi une fière chandelle.

Merci également au labo BIOMAR pour votre présence à ces séances d'entrainement, et pour tous ces moments passés à vos côtés ! Je pense à Antonio, Mathieu, Phil', Mishal, Aurélie, Saloua, Laurence, Chantal, Quentin (et Sophie !!), Thierry (l'ancien Clément), Clément (le nouveau Thierry), les 'ptites dernières' Marine et Mathilde, Sarah (merci pour toute l'énergie que tu sais si généreusement donner pour qu'on se sente intégré !!) et mon très cher collègue m'sieur Moreau Camille, pour tout ce temps vigoureusement passé à démonter mon travail (je cite « Tes modèles ? bah, tu cliques sur un bouton, tu rajoutes une virgule, et paf ça fait des Chocapics ! » ; « t'as de la chance de ne pas être là aujourd'hui sinon je répondrai à ton mail en disant ce que j'en pense de vos modèles !! ») ou encore à m'encourager à la réalisation de cette thèse « Charlène, sérieux…ne rédige jamais une thèse… !! ah, j'avais oublié…tu dois en rédiger une aussi ! ».

Merci à ensuite un ensemble de personnes avec qui j'ai pu travailler, communiquer, apprendre, planifier et sympathiser : Alexis Martin, Marc Eléaume, Salomé Fabri-Ruiz, Antonio Agüera, Jean Artois, Loïc Michel, Jean-Christophe Poggiale, MaRRko Jusup, Tin Klanjšček, Thomas Desvignes, Juliette Auvignet, Simon Morley, Starrlight Augustine, Rémi Laffont, Jonathan Flye-Sainte-Marie, Eric Dabas, Jean-François Cudennec, Yoann Thomas, Dina Lika, Anton van de Putte, Henrik Christiansen, Valérie Dulière, Katrijn Baetens, Zambra Lopéz-Farran, Karin Gérard**,** Damien Fourcy, Joni Belmaker, Yezi Buba… aussi, les étudiants de Master au top: Laura, Louise, Valentin, Margot ; l'ensemble de l'équipe de l'expédition B121: Ben, Katy, Ryan, Franz, Francesca, Henri et de nouveau Henrik, Quentin, Camille, Thomas et Bruno !, et aux belles rencontres pendant les voyages divers : Stéphanie, Claire, Sophia, Phoebe, Karen et sa désormais inséparable bouteille de crème de cassis ! Côté technico-pratique**,** merci également à l'intendance des super-ordinateurs belges et dijonnais (VUB ICT Service Desk et Ben Pohl à Dijon).

Parmi ces collaborations, j'aimerai plus chaleureusement encore remercier Bas Kooijman, qui a passé des heures à répondre à mes mails, ou à me recevoir avec mes piles de questions sur divers projets DEB, toujours avec un enthousiasme sans faille, et qui sait toujours trouver les mots simplement mais très efficacement pour remotiver quand ça bloque! [Among these collaborations, I would like to warmly and more especially thank Bas Kooijman, who has spent hours to answer my emails, and has always welcomed me -and my lists of questions !- with enthusiasm. Thank you so much, you always found simple but right words to efficiently motivate me when I was stuck !]

Reste maintenant à remercier tous mes proches et amis, à commencer par tous les copains de la plongée (vive le CSL !!), qui m'ont apporté énormément en parallèle de mes études. Aussi bien avec les cours de bio du CODEP21 (pensée pour Seb, et toute la super troupe !!) que pour toutes les sorties diverses qui rôdent, qui vident la tête, qui rechargent les batteries et qui font tout simplement trop plaisir !! Un merci plus spécial et chaleureux à mon Tout Ptit et à mon Gégé…

Un merci chaleureux aussi à Nouméa…même si à toi seule tu pourrais largement réécrire le ptit Larousse… !! sache qu'une chose est sûre, c'est qu'en ta personne même tu sais définir le mot amitié… merci pour ta présence à l'épreuve de la distance et du temps ! Merci aussi à tes proches fantastiques : Eric, Iza (ma tortue ninja préférée !!) et Antoine !

Merci à Monique… Monique c'est beaucoup de choses. C'est la folie du devant du jardin qui est surprise quand on lui demande si on peut essayer un sous-marin dans sa piscine, mais qui surprend quand elle débouchonne une bouteille de rouge sur un parking au milieu de la nuit… c'est la courageuse qui n'hésite pas à se battre jusqu'au sang avec une soupe mais qui prend ses jambes à son cou quand il faut pousser une maison avec un unibrassiste !! et…c'est une fierté et une chance d'avoir pu croiser sa route ! Merci à vous pour tous ces moments fous passés au dessus, en dessous de l'eau, en maillot de bain dans la tiny ou emmitouflées dans la voiture les hivers après des séances sociales plus qu'enrichissantes ! pour les éclats de rire liés à des problématiques diverses de rotules, de doigt arthrosé, de camions citernes de vin de Bourgogne, de jardin prospère, d'immuno-dépresseur immergeable, de system-direct ou de coucouillotte franc-comtoise… !! et plus sérieusement merci énormément pour l'oreille attentive, les conseils précieux que vous savez tant bien donner… j'ai vraiment eu une chance folle de vous avoir rencontré, de passer ces années à vos côtés (oui je sais, plus ou moins !). Merci énormément pour tout ça.

Merci à Thomas, sacrée luciole pour toutes ces soirées fabuleuses à parler de politique belge (merci…ou pas !!) ! Merci à Vincent… pour cette tiny magnifique ! Merci à Jenny pour les super dessins des chapitres de cette thèse ! Merci à la famille Mouton. Vous êtes franchement une famille géniale, je vous ai toujours regardé avec admiration. Merci à mon Timy, à Elise.
Merci à mon Mathieu, pour tous ces souvenirs déjà récoltés (à 4h du matin à escalader une montagne pour se faire un ptit dej' de fou, à 45 m sous l'eau à s'engueuler à cause d'un mec qui t'as écrasé pour mieux voir l'hélice de l'épave ou dans des eaux turquoises à rêver et à chercher les pierres bleues de l'Ingrid !) et pour tous les souvenirs à venir !!!! Merci pour ton soutien, ta patience, la force avec laquelle tu me fais avancer, les milliards d'activités qu'on organise et pour tous les moments de rigolade !! Hâte de construire la souite !!

Merci à mes grands-parents, qui malgré le fait qu'ils ne savaient pas toujours ce que je faisais « Oh ma ptite fille a eu sa thèse sur les sous marins !!! », « tu vas récolter des étoiles de mer bleues pendant ton voyage dans le froid là ?? », ont toujours porté le soutien, l'amour et la fierté qui font chaud au cœur. « Tu ne peux pas t'imaginer à quel point tu vas pouvoir faire tout ce que je n'ai pas pu faire. C'est la vengeance du destin ! je suis tellement fière. Tu as le plus beau métier du monde ».

Et enfin, merci à mes parents. Si j'en suis là, c'est pour vous et grâce à vous. Soutien inébranlable, renfort, réconfort et source de motivation. Faut pas grand chose de plus pour faire pousser des ailes…et comme le dit la mamie « oh, tu iras loin…si tu t'arrêtes pas en ch'min ! »
Cette thèse en fait, je l'ai surtout faite pour vous rendre fiers. C'était ma motivation première. J'espère vraiment qu'aujourd'hui, quand vous verrez ce manuscrit, ce sera le cas (même si c'est un charabia pas possible !!). Je vous suis tellement reconnaissante pour tous les sacrifices que vous avez faits pour moi, la période de mes études n'a pas été toujours facile… hein maman, quand on faisait le concours de celle qui arrivait à ne pas pleurer quand je devais repartir à Marseille… !! mais malgré ça, vous étiez toujours derrière moi. Toujours les bons mots pour que je trouve la force d'avancer. Toujours ce grain de fierté dans vos yeux même si vous ne saviez pas tout le temps ce que je faisais. Vous êtes géniaux, merci... (pour tout !!)

Bon, allez… c'est pas le tout de s'attarder sur l'émotionnel… y'a 600 pages qui vous attendent maintenant !!! hop hop hop !

PHD THESIS RESEARCH

➔ Chapter 1: Mechanistic models

• [1] **Guillaumot C**, Saucède T, Morley S, Augustine S, Danis B, Kooijman S (2020a). Can DEB models infer metabolic differences between intertidal and subtidal morphotypes of the Antarctic limpet *Nacella concinna* (Strebel, 1908)? ***Ecological Modelling*** **(IF: 2.5)**. 430: 109088. https://doi.org/10.1016/j.ecolmodel.2020.109088.

• [2] Arnould-Pétré M, **Guillaumot C**, Danis B, Féral J-P, Saucède T (2020). Individual-based model of population dynamics in *Abatus cordatus*, a sea urchin of the Kerguelen (Southern Ocean), under changing environmental conditions. ***Ecological Modelling*** **(IF: 2.5)**. 440, 109352. https://doi.org/10.1016/j.ecolmodel.2020.109352.

➔ Chapter 2: Correlative models

• [3] **Guillaumot C,** Danis B, Saucède T (2021). Species Distribution Modelling of the Southern Ocean benthos : methods, main limits and some solutions. ***Antarctic Science*** **(IF: 1.4)**, 1-24. doi:10.1017/S0954102021000183.

• [4] **Guillaumot C**, Artois J, Saucède T, Demoustier L, Moreau C, Eléaume M, Agüera A, Danis B (2019). Broad-scale species distribution models applied to data-poor areas. ***Progress in Oceanography*** **(IF: 4.1)**, 175, 198-207. https://doi.org/10.1016/j.pocean.2019.04.007.

• [5] **Guillaumot C**, Danis B, Saucède T (2020b). Selecting environmental descriptors is critical to modelling the distribution of Antarctic benthic species. ***Polar Biology*** **(IF: 1.7)**, 1-19. https://doi.org/10.1007/s00300-020-02714-2.

• [6] **Guillaumot C,** Moreau C, Danis B, Saucède T (2020c). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. ***Progress in Oceanography*** **(IF: 4.1)**, 188, 102438. https://doi.org/10.1016/j.pocean.2020.102438.

➔ Chapter 3: Integrated approaches

• [7] **Guillaumot C** / López-Farrán Z (co-firstauthorship), Vargas-Chacoff L, Paschke K, Dulière V, Danis B, Poulin E, Saucède T, Gerard K (2021). Current and predicted invasive capacity of *Halicarcinus planatus* (Fabricius, 1775) in the Antarctic Peninsula. ***Global Change Biology*** **(IF: 8.6)**, 00:1–18. DOI: 10.1111/gcb.15674.

• [8] Fabri-Ruiz S, **Guillaumot C,** Agüera A, Danis B, Saucède T (2021). Using correlative and mechanistic niche models to assess the sensitivity of the Antarctic echinoid *Sterechinus neumayeri* (Meissner, 1900) to climate change. ***Polar Biology*** **(IF: 1.7)**. https://doi.org/10.1007/s00300-021-02886-5.

• [9] **Guillaumot C**, Buba Y, Belmaker J, Fourcy D, Danis B, Dubois P, Saucède T (submitted). Simple or hybrid ? The performance of next generation ecological models to study the response of Southern Ocean species to changing environmental conditions. ***Diversity and Distributions*** (**IF: 3.9)**.

➔ Chapter 4: Dispersal models : lagrangian approach

• [10] Dulière V / **Guillaumot C** (co-firstauthorship), López-Farrán Z, Lacroix G, Saucède T, Danis B, Baetens K (submitted). Potential impact of ballast water exchanges on the introduction of invasive species in Marine Protected Areas of the Western Antarctic Peninsula. ***Diversity and Distributions*** **(IF: 3.9)**.

• [11] Christiansen H, Van de Putte A, **Guillaumot C**, Barrera-Oro E, Volckaert FAM, Young EF (final stages, draft version included in the manuscript). Integrated assessment reveals large scale connectivity of a historically overexploited fish in the Southern Ocean.

➔ PhD supplementary material

• [12] Agüera A, Ahn I-Y, **Guillaumot C**, Danis B (2017). A Dynamic Energy Budget (DEB) model to describe *Laternula elliptica* (King, 1832) seasonal feeding and metabolism. ***PLOS One*** **(IF: 2.7)**.12(8), e0183848. doi: https://doi.org/10.1371/journal.pone.0183848.

• [13] **Guillaumot C**, Martin A, Saucède T, Eléaume M. (2018) Methods for improving species distribution models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau. ***Marine Ecology Progress Series*** **(IF: 2.3)**. *594*, 149-164. doi: 10.3354/meps12538.

• [14] López-Farrán Z, Frugone MJ, Gerard K, Vargas-Chacoff L, Poulin E, **Guillaumot C**, Dulière V (final stages, draft version). Can the Patagonian crab *Halicarcinus planatus* (Fabricius, 1775) reach Antarctic coasts? Study of the dispersal potential of its larvae using a Lagrangian approach.

OTHER RESEARCH STUDIES, INTERNATIONAL COLLABORATIONS

• [15] Danis B, Christiansen H, **Guillaumot C**, Heindler FM, Jossart Q, Moreau C, Pasotti F, Robert H, Wallis B, Saucède T (submitted as a datapaper). The Belgica121 expedition to the Western Antarctic Peninsula: a high resolution biodiversity census. ***Biodiversity Data Journal*** (**IF: 1.3)**.

• [16] Danis B, Wallis B, **Guillaumot C**, Moreau C, Pasotti F, Heindler F, Robert H, Christiansen H, Jossart J, Saucède T and (submitted). Nimble vessel cruises as an alternative for Southern Ocean biodiversity research: preliminary results from the Belgica121 expedition. ***Antarctic Science*** **(IF: 1.4)**.

• [17] Moreau C, Jossart Q, Danis B, Eléaume M, Christiansen H, **Guillaumot C**, Downey R, Saucède T (2020). The overlooked diversity of Southern Ocean sea stars (Asteroidea) reveals original evolutionary pathways. ***Progress in Oceanography*** **(IF: 4.1)**, 102472. https://doi.org/10.1016/j.pocean.2020.102472.

• [18] Saucède T, **Guillaumot C**, Michel L, Fabri-Ruiz S, Bazin A, Cabessut M, García-Berro A, Mateos A, Mathieu O, De Ridder C, Dubois P, Danis B, David B, Díaz A, Lepoint G, Motreuil S, Poulin E & Féral J-P (2019). Modelling species response to climate change. Case study of echinoids on the Kerguelen Plateau. In: Welsford, D., J. Dell and G. Duhamel (Eds). The Kerguelen Plateau: marine ecosystem and fisheries. ***Proceedings of the Second Symposium. Australian Antarctic Division, Kingston, Tasmania, Australia***. ISBN: 978-1-876934-30-9, pp 95-116, doi: 10.5281/zenodo.3249143.

• [19] Féral J-P, Poulin E, Gonzalez-Wevar CA, Améziane N, **Guillaumot C**, Develay E, Saucède T (2019) Long-term monitoring of coastal benthic habitats in the Kerguelen Islands: a legacy of decades of marine biology research. In: Welsford, D., J. Dell and G. Duhamel (Eds). The Kerguelen Plateau: marine ecosystem and fisheries. ***Proceedings of the Second Symposium. Australian Antarctic Division, Kingston, Tasmania, Australia***. ISBN: 978-1-876934-30-9, pp 383-402, doi : 10.5281/zenodo.3249143

• [20] Moreau C, **Guillaumot C** et al. (2018) Antarctic and sub-Antarctic Asteroidea database. ***Zookeys* (IF: 1.1)**, (747), 141. doi: 10.3897/zookeys.747.22751.

PEER-REVIEW ARTICLES REALISED BEFORE THE PHD

• [21] **Guillaumot C**, Fabri-Ruiz S, Martin A, Eléaume M, Danis B, Féral J-P, Saucède T. (2018). Benthic species of the Kerguelen Plateau show contrasting distribution shifts in response to environmental changes. ***Ecology and Evolution* (IF: 2.4)**. 8(12), 6210-6225. http://dx.doi.org/10.1002/ece3.4091.

• [22] Pagano M, Rodier M, **Guillaumot C**, Thomas Y, Henry K, Andréfouët S (2017). Ocean-lagoon water and plankton exchanges in a semi-closed pearl farming atoll lagoon (Ahe, Tuamotu archipelago, French Polynesia). ***Estuarine, Coastal and Shelf Science* (IF: 2.3)**. 191: 60-73. https://doi.org/10.1016/j.ecss.2017.04.017.

• [23] **Guillaumot C**, Martin A, Fabri-Ruiz S, Eléaume M, Saucède T (2016) Echinoids of the Kerguelen Plateau – occurrence data and environmental setting for past, present, and future species distribution modelling. ***Zookeys* (IF: 1.1).** 630: 1-17. https://doi.org/10.3897/zookeys.630.9856.

EXPEDITION REPORTS

• Danis B, Christiansen H, **Guillaumot C**, Heindler F, Houston R, Jossart Q, Lucas K, Moreau C, Pasotti F, Robert H, Wallis B, Saucède T (2019). Report of the Belgica121 expedition to the West Antarctic Peninsula. 96 pp. Available at http://belgica120.be/index.php/report/ and http:// doi.org/10.5281/zenodo.4551452.

DATABASES REALISED DURING THE PHD

•**Guillaumot C**, Raymond B, Danis B (2018) Marine environmental data layers for Southern Ocean species distribution modelling. Australian Antarctic Data Centre - doi:10.26179/5b8f30e30d4f3 ; available at https://australianantarcticdivision.github.io/blueant/articles/SO_SDM_data.html.

• Asteroidea database| http://ipt.biodiversity.aq/resource?r=asteroidea_southern_ocean, 28/06/2017.

• Particulate Carbon Export Flux, maps. *Australian Antarctic Data Center*, Particulate_carbon_export_flux_layers, doi:10.4225/15/58fff5231f00a, 27/04/2017.

DATABASES REALISED BEFORE THE PHD

• Kerguelen environmental datasets | **Guillaumot C**, Martin A, Fabri-Ruiz S, Eléaume M, Saucède T (2016) Environmental parameters (1955-2012) for echinoids distribution modelling on the Kerguelen Plateau. *Australian Antarctic Data Centre.* doi:10.4225/15/578ED5A08050F 20/07/2016

• Echinoid database of the Kerguelen Islands | http://ipt.biodiversity.aq/resource.do?r=echinoids_kerguelen_plateau_1872_2015, 07/2016.

R PACKAGE (created before & updated during the PhD)

•**Guillaumot C**, A Martin, M Eléaume, Danis B, T Saucède (2016) 'SDMPlay': Species Distribution Modelling Playground, CRAN. https://cran.r-project.org/web/packages/SDMPlay 04/08/2016.

ORAL PRESENTATIONS (during the PhD)

* underlined name: presenter

• [1] **Guillaumot C**, Saucède T, Danis B. Les modèles de niche écologique, outils pour évaluer la sensibilité des espèces marines antarctiques face aux changements environnementaux: potentiel, limites et méthodes. **JCAD, Dijon, France**, *December 2020* (lightning talk 5 min in French).

• [2] **Guillaumot C**, Kooijman S, Saucède T, Danis B. Application des modèles de budget énergétique (Dynamic Energy Budget, DEB) à des cas d'étude polaires. **CNFRA, La Rochelle, France**, *September 2020* (15 min talk in French).

• [3] **Guillaumot C**, Kooijman S, Danis B, Saucède T. Application of Dynamic Energy Budget (DEB) models to Antarctic case studies. **SCAR Symposium, Hobart, Tasmania**, *August 2020* (presentation published online, in English).

• [4] Arnould-Pétré M, **Guillaumot C**, Danis B, Féral J-P, Saucède T. Individual-based model of population dynamics in *Abatus cordatus*, a sea urchin endemic to the Kerguelen Plateau, under changing environmental conditions. **SCAR Symposium, Hobart, Tasmania**, *August 2020* (presentation published online, in English).

• [5] López-Farrán Z, Frugone MJ, Vargas-Chacoff L, **Guillaumot C**, Gerard K, Poulin E, Dulière V. Assessment of the capacity of *Halicarcinus planatus* to arrive and settle as a potential invasor of Antarctic shallow ecosystems. **SCAR Symposium, Hobart, Tasmania**, *August 2020* (abstract published online, in English).

• [6] López-Farrán Z, Frugone MJ, Vargas-Chacoff L, **Guillaumot C**, Gerard K, Poulin E, Dulière V *Halicarcinus planatus*, la primera especie exótica encontrada en la Península Antártica: Evaluación de su potencial invasor. **Reunión Anual Conjunta 2019, Valdivia, Chile**, *November 2019* (15min talk in English).

• [7] López-Farrán Z, **Guillaumot C,** Dulière V, Paschke K, Gerard K, Vargas-Chacoff L, Poulin E. *Halicarcinus planatus*, el cangrejo subantártico con potencial para establecerse en Antártica, evaluación del escenario actual y futuro. **IX Congreso Chileno de Investigaciones Antárticas, Olmue, Chile**, *October 2019* (15min talk in English).

• [8] **Guillaumot C**, Artois J, Saucède T, Danis B. Broad-scale species distribution models applied to data-poor areas. **ISEM, Salzburg, Austria**, *October 2019* (15min talk in English).

• [9] López-Farrán Z, Frugone MJ, Vargas-Chacoff L, **Guillaumot C**, Gerard K, Poulin E, Dulière V. Assessment of the capacity of *Halicarcinus planatus* as a potential invasor of Antarctic shallow ecosystems. 10 minutes presentation in English. **IMARCO 2019, Aveiro, Portugal,** *September 2019* (15min talk in English).

• [10] **Guillaumot C**, Saucède T, Danis B. Ecological niche modelling as a tool to assess the sensitivity of Antarctic marine species to environmental changes: potential, limitations and methods. **Forum des Jeunes Chercheurs, Dijon, France,** *June 2019* (10min talk in French).

• [11] **Guillaumot C**, Saucède T, Danis B. Inferring the response of Southern Ocean benthic species to environmental changes using Dynamic Energy Budget models. **DEB Symposium, Brest, France,** *April 2019* (15min talk in English).

• [12] **Guillaumot C**, Saucède T, Danis B. Inferring the responses of Southern Ocean benthic species to environmental changes using Dynamic Energy Budget models. **Zoology congress, Anvers, Belgium,** *December 2018* (12min talk in English).

• [13] **Guillaumot C**, Agüera A, Eléaume M, Danis B. Physiological modelling in Antarctica: dynamic inferences in a changing environment. **SCAR Symposium, Davos, Switzerland**, *June 2018* (12min talk in English).

• [14] **Guillaumot C**, Martin A, Eléaume M, Danis B, Saucède T. Modelling species distribution shifts with environmental changes in data-poor areas. An example from the Kerguelen Plateau. **MEASO, Hobart, Tasmania**, *April 2018* (12min talk in English).

• [15] **Guillaumot C**, Martin A, Eléaume M, Saucède T. Modelling species distribution: influences of temporal, spatial, and sampling heterogeneities in data-poor areas. An example from the Kerguelen Plateau. **SCAR Symposium, Leuven, Belgium**, *July 2017* (12min talk in English).

• [16] **Guillaumot C**, Martin A, Eléaume M, Saucède T. SDMPlay: a pedagogic package for a first approach in species distribution modelling. **Semin'R workshop, Paris, France,** *June 2017* (15min talk in French).

• [17] **Guillaumot C**, Agüera A, Danis B. Physiological performances of Southern Ocean key species. **DEB Symposium, Tromso, Norway,** *June 2017* (10min talk in English).

• [18] **Guillaumot C**, Agüera A, Danis B. Projecting species energetic performance in a spatially-explicit context: Trait Distribution Modelling of a key antarctic species. **CNFRA Workshop, Paris, France,** *May 2017* (10min talk in English).

ORAL COMMUNICATIONS (before the PhD)

• [19] **Guillaumot C**, Martin A, Eléaume M, Saucède T. Modelling species present distribution in data-poor areas and predicting the impact of climate change. **Société Française d'Ecologie (SFE), Marseille, France,** *October 2016* (12min talk in English).

• [20] **Guillaumot C**, Martin A, Eléaume M, Saucède T. Temporal, spatial, and sampling heterogeneities in species distribution modelling. The case study of the data-poor area of the Kerguelen Plateau. **Biodiversity and Ecosystem Scenarios Network (ScenNet), Montpellier, France**, *August 2016* (12min talk in English).

• [21] **Guillaumot C**, Martin A, Eléaume M, Saucède T. Prédire l'actuel et anticiper les changements futurs en évaluant l'influence des hétérogénéités spatiale et temporelle sur la modélisation de la distribution d'espèces. Le cas d'étude des échinides du Plateau des Kerguelen. **CNFRA, Lyon, France,** *May 2016* (10min talk in French).

POSTER PRESENTATIONS (during the PhD)

* underlined name: presenter

• [1] Christiansen H, Van de Putte A, **Guillaumot C**, Barrera-Oro E, Volckaert F, Young EF. Large scale connectivity of the marbled rockcod *Notothenia rossii* revealed through population genomics and modelling. **SCAR Symposium, Hobart, Tasmania**, *August 2020* (abstract published online, in English).

• [2] Arnould-Pétré M, **Guillaumot C**, Féral J-P, Danis B, Saucède T. Modelling the distribution and survival of the echinoid *Abatus cordatus* in the Kerguelen Islands. **10th Echinoderm conference, Moscou, Russia,** *September 2019.*

• [3] Verdon V, **Guillaumot C**, Saucède T. Combining correlative and mechanistic modelling approaches to characterise the limits and sensitivity of the ecological niche of two benthic Antarctic species**. MME Lyon, France,** *July 2019.*

• [4] López-Farrán Z, Gérard K, Poulin E, Paschke K, **Guillaumot C**, Dulière V, Vargas-Chacoff L. What are the possibilities of *Halicarcinus platanus*, a subantarctic crab, to survive in the West Antarctic Peninsula? **III International Conference Island Biology 2019, La Réunion,** *July 2019.*

• [5] López-Farrán Z, Frugone MJ, Vargas-Chacoff L, **Guillaumot C**, Gérard K, Poulin E, Dulière V. Assessment of the capacity of *Halicarcinus platanus* larvae to reach the South Shetland Islands through passive dispersal. **Liège Colloquium, Belgium,** *May 2019.*

• [6] Arnould-Pétré M, **Guillaumot C**, Danis B, Saucède T. Modelling the response of a sea urchin endemic to the Kerguelen Islands, *Abatus cordatus*, to environmental changes. **DEB Symposium, Brest, France,** *April 2019.*

• [7] **Guillaumot C**, Saucède T, Danis B. Assessing the response of a simple bentho-pelagic network to environmental changes: case study from the Southern Ocean. **DEB Symposium, Brest, France,** *April 2019.*

• [8] **Guillaumot C**, Saucède T, Danis B. Strict extrapolation in modelling Southern Ocean species distribution. **Zoology congress, Anvers, Belgium,** *December 2018.*

• [9] **Guillaumot C**, Agüera A, Danis B, Deregibus D, Quartino ML, Saravia LA. Dynamic growth model of Antarctic macroalgae in a fast-changing environment. **SCAR, Davos, Switzerland,** *June 2018.*

• [10] **Guillaumot C**, Artois J, Saucède T, Demoustier L, Moreau C, Eléaume M, Agüera A, Danis B. Modelled distributions of benthic species of the Southern Ocean in a fast-changing environment. **MEASO, Hobart, Tasmania,** *April 2018.*

• [11] **Guillaumot C**, Martin A, Eléaume M, Saucède T. Modelling species distribution: influences of temporal, spatial, and sampling heterogeneities in data-poor areas. An example from the Kerguelen Plateau. **Kerguelen Symposium, Hobart, Tasmania,** *November 2017*.

• [12] Agüera A, **Guillaumot C**, Danis B. Trait distributions of key marine species from the Western Antarctic Peninsula. **DEB Symposium, Tromso, Norway,** *June 2017.*

• [13] **Guillaumot C**, Agüera A, Danis B. Projecting species energetic performance in a spatially-explicit context: Trait Distribution Modelling of antarctic benthic species. **SOOS Workshop (Southern Ocean Observing System network), British Antarctic Survey, Cambridge, United Kingdom,** *May 2017.*

POSTER PRESENTATIONS (before the PhD)

• [14] **Guillaumot C**, Martin A, Eléaume M, Saucède T. Temporal, spatial, and sampling heterogeneities in species distribution modelling: the case study of the data-poor area of the Kerguelen Plateau. **SCAR, Kuala Lumpur, Malaysia,** *August 2016.*

TABLE OF CONTENTS

INTRODUCTION

1. MODELLING IN ECOLOGY

1.1 Models and their application in ecology

1.1.1 Some generalities on models

Models are "purposeful representations of a system, hypothesis or experiment and include any useful form of abstraction to assist thinking" (Starfield et al. 1990). They are used in a substantial panel of scientific contexts (e.g. in cosmology to study astronomical object movements or compositions, in oncology to predict the effect of a therapy, in biochemistry to determine molecular structures, in climatology to forecast weather, in mechanics to design technologies, in epidemiology to anticipate the spread of a disease, in archaeology to rebuild an artifact...), and play a crucial role to analyse complex situations that are difficult to describe (Frigg and Hartmann 2020).
Models are built using observations, and can be represented with mathematical equations, computer codes (Kennedy and O'Hagan 2001, Hucka et al. 2003), matrices, networks (Keller et al. 2006, Kuperstein et al. 2015), schematic diagrams or images (Ludvigsen et al. 2006, Bryson et al. 2017, Fisher et al. 2018). Whenever conceptualizing models, it is essential to be aware that they are mere simplifications of real processes, and by definition are wrong, as they cannot encompass the complexities of the studied system (Knutti 2010). Models do not aim at perfectly representing the overall processes, but should be useful enough to enable a part of their understanding (Grimm 1994). G. Box, a British statistician, used to write in his studies that "all models are wrong; some are useful… the practical question is how wrong do they have to be to not be useful". This sentence illustrates the trade-off between model overfit and their explanatory power (Box 1979).

The way models are designed depends on the final objectives of the model and respects a balance between generality, realism and precision (Levins 1966, Fig. 0.1). An infinity of models can be therefore generated to target a single question.

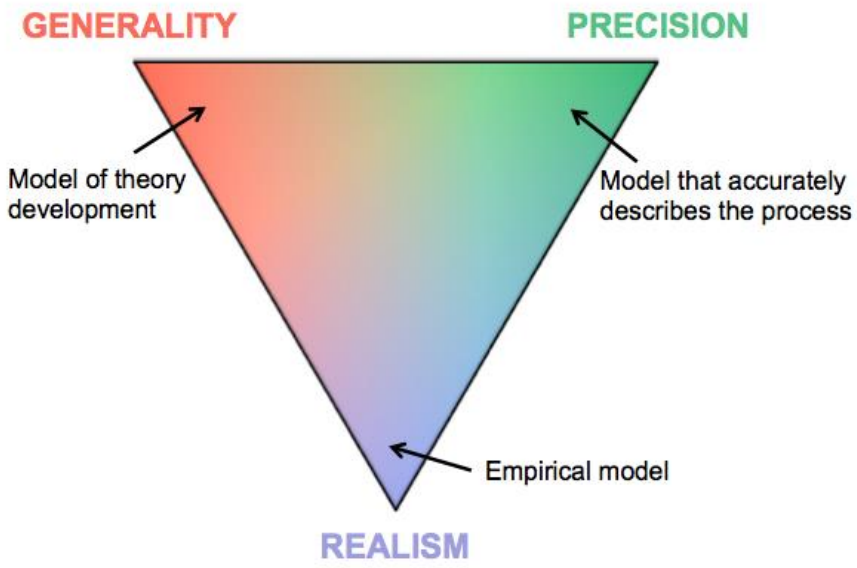


Figure 0.1. Trade-off between model properties when designing a model. The balance between generality, precision and realism depends on the questions the modeller addresses. It is also dependent on data availability. This scheme highlights the fact that a broad range of models can be generated to represent a given system.

Models are efficient to enable rapid explorations of mechanisms (Goodman and Gotlib 1999, Villari et al. 2016), to test and validate hypotheses (Eren et al. 2016), to identify key interactions (Lorenzo-Trueba et al. 2010, Gaichas et al. 2016), or to provide testable predictions, that may corroborate (or not) with experimental observations and help prioritize new experiments or model improvements (Grimm 1994, Marini et al. 2010, Enderling et al. 2019).

After designing a model, a crucial step is the evaluation of its relevance, by assessing whether the model is suited to describe the studied process and efficient enough to provide accurate predictions (Tropsha et al. 2003, Ivanescu et al. 2016). Though essential, the evaluation step is sometimes neglected or fails, because of limited available independent observations or inappropriate methods (Fielding and Bell 1997, Steyerberg et al. 2003, Hijmans 2012), which highlights the importance of adapting methodologies for this validation step (Robinson et al. 2011, Muscarella et al. 2014). Robustness analysis can complete model evaluation by assessing "whether a result depends on the essential of the model or on the details of the simplifying assumptions" (Railsback and Grimm 2019). This analysis aims at "breaking" the model, by forcefully changing its parameters, structure and/or representation of processes in order to evaluate the assumptions that mostly drive model stability (Grimm and Berger 2016). The relevance of such evaluation procedures is crucial and validation results need to be associated with model predictions to enable complete and accurate interpretations (Guisan et al. 2013, Yates et al. 2018).

1.1.2 Ecological modelling in marine environments

A recent review on marine ecology seascape analyses (Kavanaugh 2018) wrote "The technological advancement and proliferation of space-, air- and water-based ocean sensing systems, together with increased sophistication in geospatial tools and mathematical simulation models […] have allowed to collect, integrate, analyse and visualise vast quantities of marine data that have revealed unimaginable structural complexity and interconnectedness across the seafloor, sea surface and throughout the water column." These sentences illustrate the huge amount of data collected during the past decades to understand ecological processes, the development of new technologies to analyse them and the strong desire of disentangling the way elements are interacting between each other in natural systems (Borgman et al. 2007, Aanensen et al. 2009, Hallgren et al. 2016). The use of ecological models to represent in a simpler way these complex ecological systems and to facilitate their understanding by simplifying existing interactions between components is thus fully appropriate (Holling 1966, Wu and David 2002, Elsawah et al. 2015).

Ecological models can be used to describe on-going processes but can also be predictive-based (Jørgensen and Bendoricchio 2001, Austin 2002). They can be applied to various systems and fields, from the scale of a water pond to an entire ocean or continent (Hecnar and M'Closkey 1996, Hassall et al. 2011, Xavier et al. 2015) or from the scale of a cell to an entire ecosystem (Klanjšček et al. 2013, Blanchard et al. 2017, Dahood et al. 2019). They can be used to predict species distribution in space or time under contrasting environmental conditions (Peterson et al. 2011), to assess energetic shifts or individuals survival when facing environmental change or toxicant exposure (Jager et al. 2016, Lenz et al. 2019, Muller et al. 2019), to predict population dynamics in space or time (Martin et al. 2013, Goedegebuure et al. 2018), to evaluate marine individuals dispersal in oceans by simulating particle trajectories in marine currents with lagrangian approaches (Hays et al. 2010, Thomas et al. 2015).

Representing natural systems is however a difficult exercise (Wu and David 2002), as systems are composed of many actors and factors, with variable and complex interactions (Fig. 0.2), influenced by intra-individual complexity, at multiple scales, with contrasted environmental conditions and

habitats (Jørgensen 1995, Levin 1999, Johnston et al. 2007). Figure 0.3 illustrates well this complexity, with the example of zooplankton density in the water column, forced by several nested biotic and abiotic factors and challenged by cross-scale interactions that may interact together, and that may change according to the way the system is studied (Seidl et al. 2017).

Figure 0.2. Diversity and complexity of marine benthic communities in the Southern Ocean. Each species interacts with its neighbours and is influenced by the coupled actions of physical, chemical and biological processes of the surrounding environment. © J. Stark, MEASO 2018.

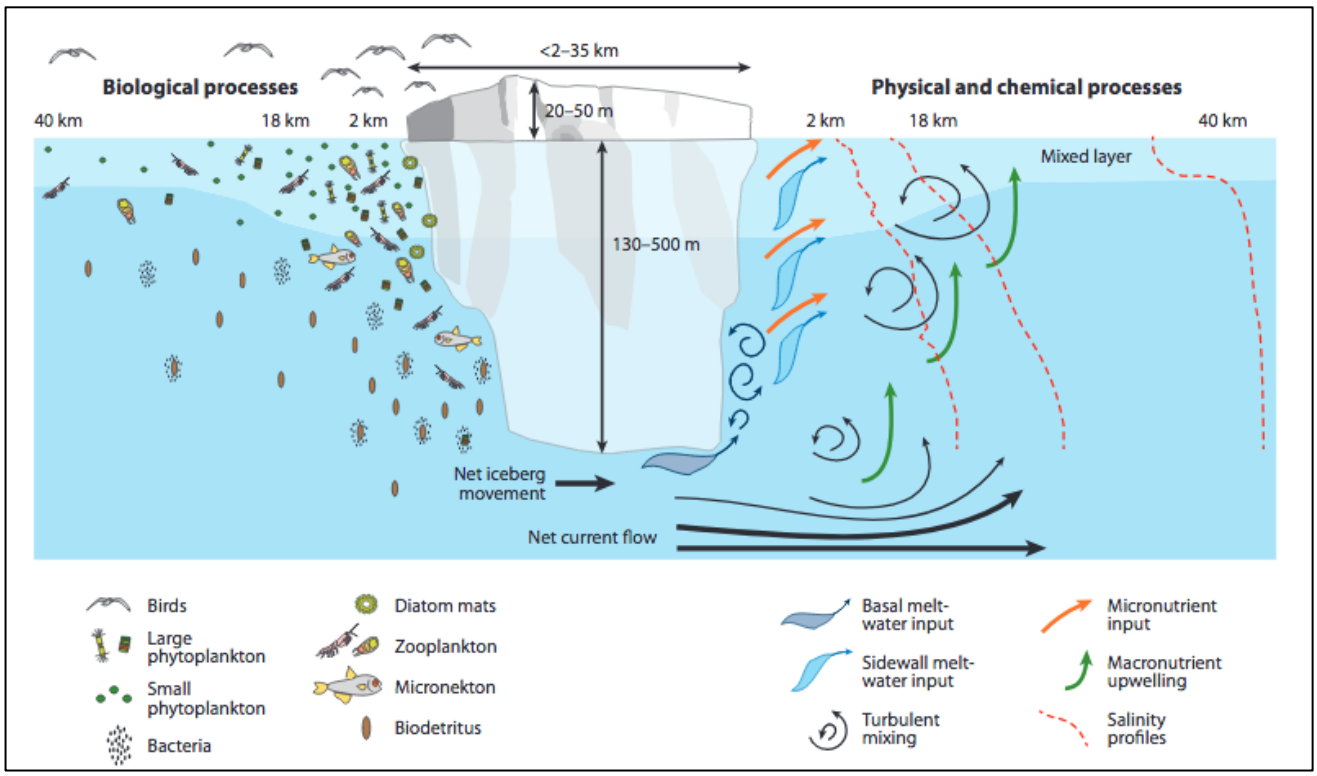


Figure 0.3. Effects of coupled large-scale climate, local physical forcing and environmental chemical properties on biological processes in the vicinity of a free-drifting iceberg in the northwest Weddell Sea. Global climate forcing induces regional ice melting, causing shifts in water column stratification and water movement at local scales, which may affect the survival, behaviour and dynamics of planktonic communities at regional scales. This illustrates marine systems' complexity and inter-scale interactions. Figure extracted from Smith et al. (2013).

Understanding such complex processes requires a huge amount of time and studies (Sagarin and Pauchard 2012) and implies to study each biological pattern at different scales (Anderson 2018), as each system generally shows variability on a range of spatial, temporal and organizational scales

(Levin 1992). For ecological models, the choice of grain size and spatial extent to represent a system therefore constitutes a strong assumption as it conditions the accuracy in describing the system (i.e. estimation of species richness, evaluation of environmental variability, detection of rare species; Whittaker et al. 2001, Keil et al. 2015).
Because there is no general scale to best represent ecological processes (Wiens 1989, Levin 1992, Blackburn and Gaston 2002), representing ecological systems as a combination of several simple systems at different scales and levels (Gonzalez et al. 2016, Boyd et al. 2018) or by a hierarchical approach (Wu 1999, Wu and David 2002) can constitute alternatives to improve the overall understanding (Fig. 0.4, Fig. 0.5). Combining these representations by multi-scale analyses also constitutes a powerful method to more accurately represent biodiversity patterns (Gonzalez et al. 2016, Anderson 2018).

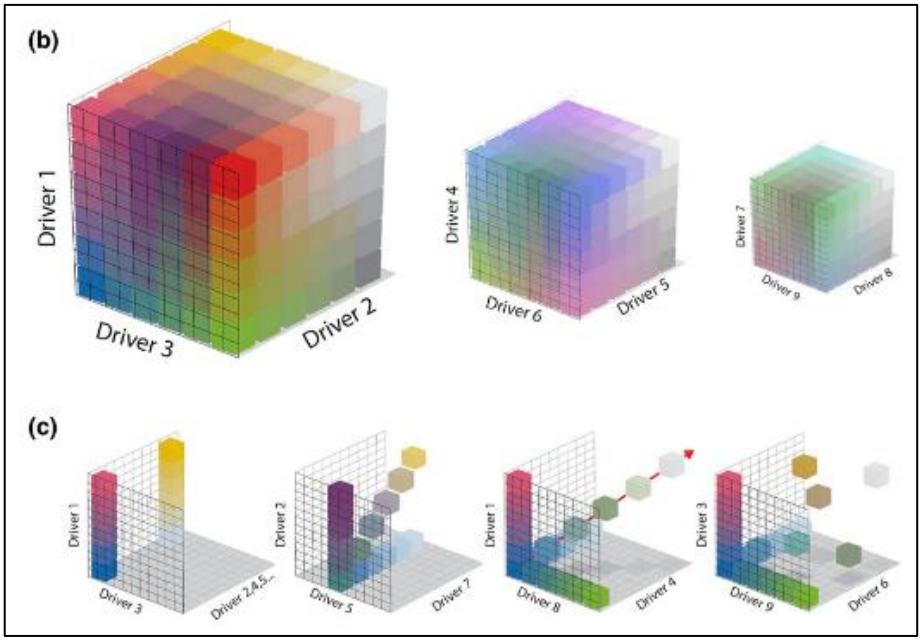


Figure 0.4. Theoretical scheme of an experimental design, that aims at isolating the most relevant key drivers to optimise the understanding of an ecological process. Main influencing drivers are identified (e.g. Fig. 0.2), (b) full-factorial designs are created to study interactions and effects and (c) subsets are defined to isolate processes that best explain the research questions. Extracted from Boyd et al. (2018).

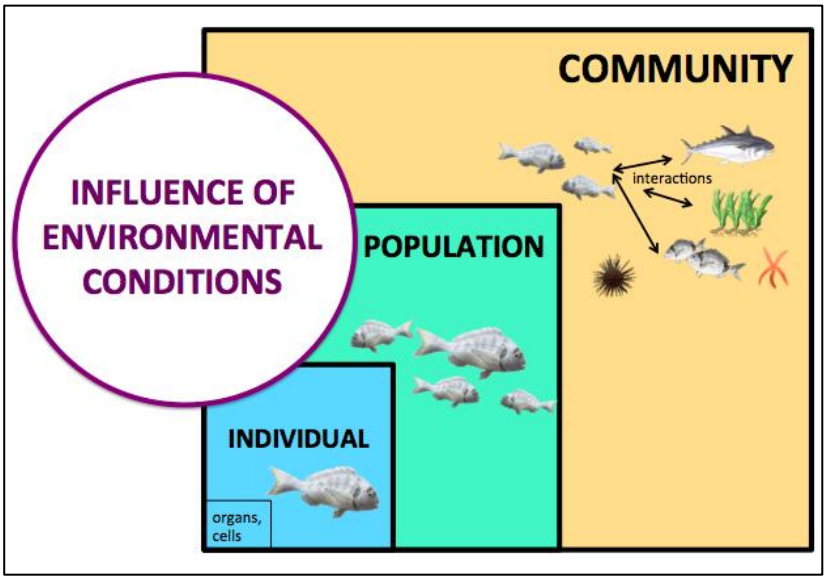


Figure 0.5. Analysis of a marine community, using a modelling approach, with the individual being the central foundation of the model, from which processes will be downscaled to organs or cells or upscaled to population or community levels (inspired from Railsback and Grimm 2019).

1.2 Ecological modelling and the species niche

1.2.1 Ecological niche theory

The niche theory was initiated by ecologists to analyse the complex question of 'which set of environmental factors allow a species to exist in a given geographic region or biotic community and respectively, what effects does this species have on its environment?' (Peterson et al. 2011). The niche concept was defined, used and developed in several founding works, leading to several definitions.

- Grinnell (1917) defines the species niche as the climatic and habitat requirements (environmental factors expressed geographically) that enable the species to survive and reproduce. Grinnell (1904) pioneered the concept that two species should have some contrasting traits related to their fitness to coexist if they want to coexist. Grinnell was also among the first to discuss niche organisation within communities, with saturated/unsaturated communities, containing "empty niches". This concept is still discussed nowadays (Peterson et al. 2011).
- Elton (1927) adopted a contrasting definition that brought new advances in the use of the ecological niche concept. The niche was in his works defined as the functional role of the species in its community, in other words, as its local effect in the "food cycle". Grinnell's and Elton's definitions are contrasting in terms of the considered geographical scales to define the niche concept, but these two definitions are interestingly complementary to more accurately understanding the species geographic distribution.
- It was Hutchinson, in 1957, who made the link between these concepts, by defining the ecological niche as "an hypervolume of environmental variables, every point of which corresponds to a state of the environment which would permit the species to exist indefinitely". Hutchinson (1957) also defined the concepts of fundamental and realised niches, the fundamental niche being the set of environmental states which enables the species to exist; and the realised niche a subset of the fundamental niche that corresponds to the ensemble of environmental conditions for which the species survives and reproduces, adding into consideration the influence of biotic interactions (competition, predation, parasitism, symbiosis…). Scale was not considered in the pioneer theoretical works of Hutchinson, it was several years later (Hutchinson 1978), that he described niches based on case-studies. Hutchinson (1957) did not consider the influence of biogeographic barriers neither. The realised niche is not limited by potential geographical barriers, nor by species dispersal capabilities in Hutchinson's definition.
- Some following studies then revised the definition of the niche concept (Leibold 1995, Chase and Leibold 2003). From these, new concepts such as the "potential niche" (i.e. the intersection between the fundamental niche space and the available environmental space) or the "occupied niche" (i.e. a subset of the fundamental niche that takes into consideration both biogeographical barriers and biotic interactions) were introduced (Jackson and Overpeck 2000, Soberón and Nakamura 2009) and adopted by the community (Pearman et al. 2008, Barve et al. 2011, Saupe et al. 2012).

Following these concepts, statistical and computing approaches have been developed to go beyond the niche description and generate models that provide an effective way of describing the different types of species niches (Guisan and Zimmermann 2000, Pulliam 2000, Pearson and Dawson 2003, Soberón and Peterson 2005, Soberón 2007, Soberón and Nakamura 2009).

1.2.2 Niche modelling approaches

Niche models link modelling techniques and niche theory with the aim of explaining as accurately as possible the conditions that drive species distribution and help fulfill their best fitness, based on

statistical approaches, experimental works and/or *in situ* observations (Guisan and Zimmermann 2000, Peterson 2006, Kearney and Porter 2009).
Methodological issues to design these models have been widely discussed, such as the influence of contrasting spatial or temporal scales, the geographical influence of dispersal, biotic interaction knowledge, biotic interaction changes according to spatial scales, shift of equilibrium between species occurrence and sampling effort, or the nature of occurrence records used to calibrate models (Araújo and Guisan 2006, Jiménez -Valverde et al. 2008, Godsoe 2010, Sillero 2011, Anderson 2013, Pittman 2017, Fig. 0.6). In parallel, numerous methods have been developed to address these issues (Soberón and Peterson 2005, Soberón 2007, Godsoe 2010, Peterson and Soberón 2012, Real et al. 2016, Soberón and Arroyo-Peña 2017).

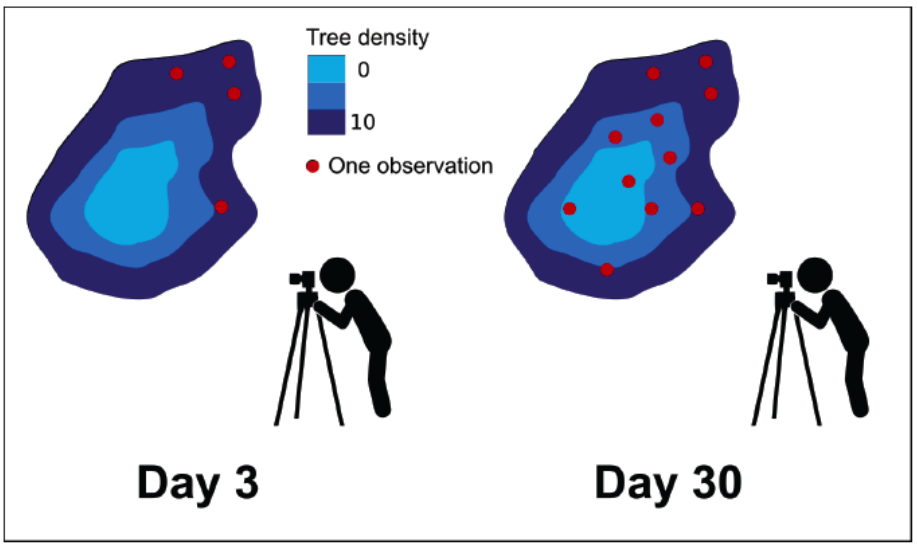


Figure 0.6. Schematic representation of the equilibrium bias, that compromises the definition of occurrence occupied space according to sampling effort. Figure extracted from J. Artois PhD thesis (2019).

One of the most recently developed representations of ecological niches is the BAM diagram (Soberón 2007, Peterson et al. 2011, Sillero 2011, Saupe et al. 2012). This theoretical framework hypothesizes that three main conditions determine the distribution of a species: biotic factors (B), abiotic conditions (A) and regions that are accessible through dispersal (M, movement). FN is the fundamental niche, corresponding to the ensemble of environmental conditions suitable to the species distribution. RN, in the center of the BAM diagram, is the realised niche, that is the real space occupied by the species, restrained by A, B and M. G_i is the invadable area, abiotically suitable but that has not been explored by the species yet. Biotic interactions and dispersal barriers are theoretically a constraint. G_i could play the role of a potential refuge.

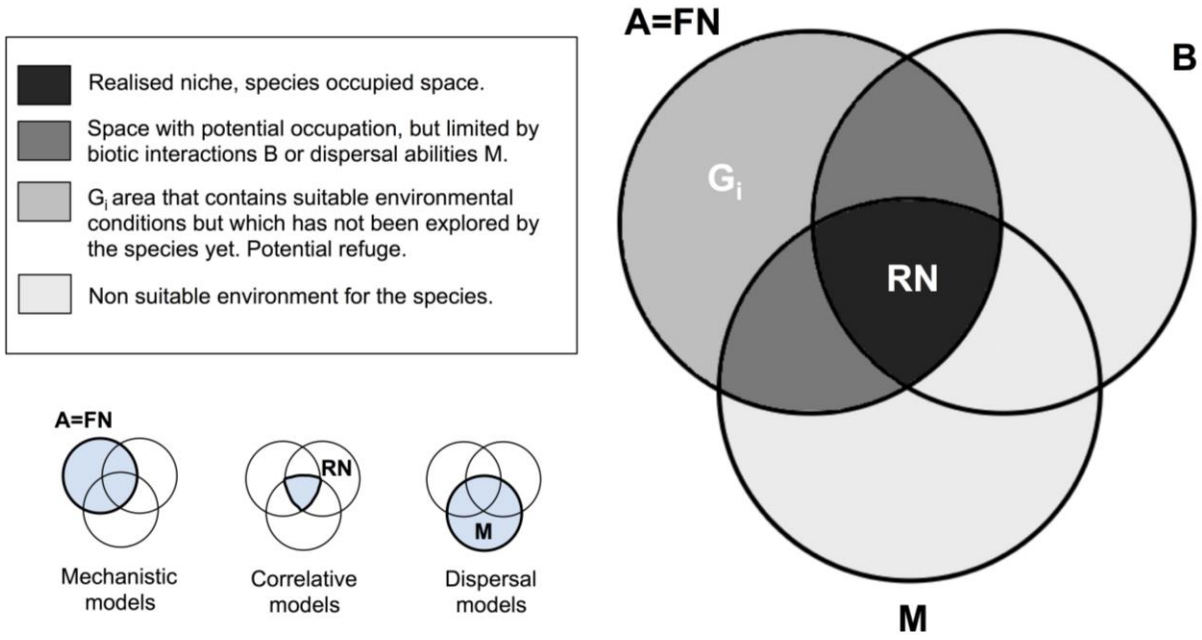


Figure 0.7. Representation of the BAM diagram. B is the portion of the environment restrained by suitable biotic interactions; A is the part of the environment that contains suitable abiotic conditions for the development and survival of the species, which corresponds to the fundamental niche FN; M is the region that is accessible to the species during a considered amount of time, not limited by dispersal nor geographic barriers. The intersection of B, A and M is the realised niche RN. G_i corresponds to an area that contains suitable environmental conditions but which has not been explored by the species yet. This area is the focus of modelling approaches (interpolations). Areas that are assessed by mechanistic, correlative and dispersal models are illustrated in the left bottom corner of the panel. Adapted from Sillero (2011), Saupe et al. (2012).

The BAM diagram can take different shapes, according to the respective size of the different B,A and M areas. This has been discussed by Peterson et al. (2011) and Saupe et al. (2012) (Fig. 0.8).

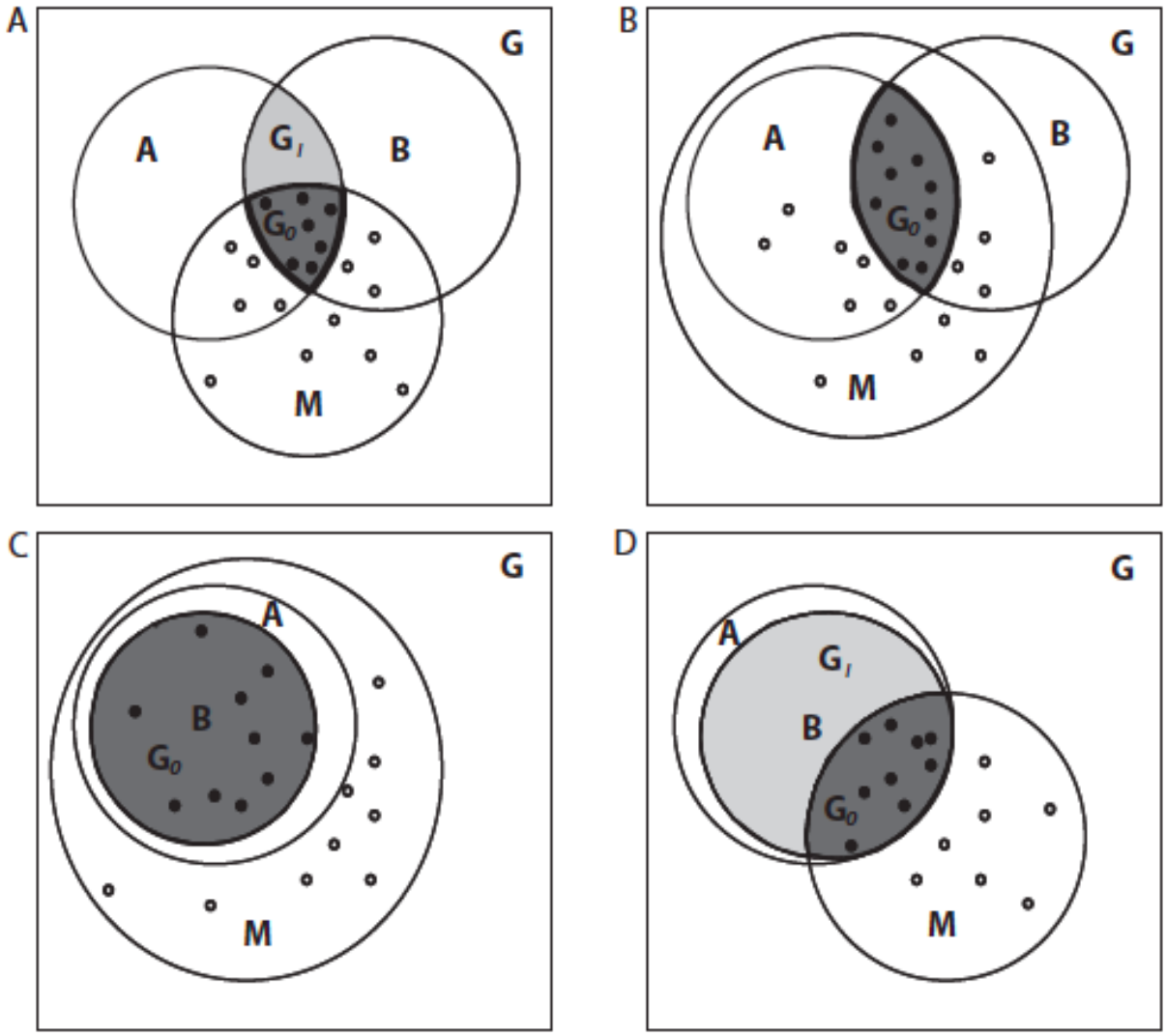


Figure 0.8. Examples of different configurations of the BAM diagram, figure from Peterson et al. (2011). G_o is the realised niche, defined as RN in Figure 0.7. Panel A shows an intuitive hypothetical and theoretical configuration. Panel B shows a situation in which all of the abiotic suitable area A is accessible, so the invadable distributional area G_i is null. Panel C shows a situation in which A and B are almost coincident, and the entire area is accessible to the species, so neither biotic nor movement considerations reduce the distributional potential of the species, solely the environmental conditions are limiting the distribution. Panel D depicts a situation similar to C, except that substantial restrictions of dispersal exist, such that not all suitable potential distributional areas are inhabited. In all panels, open circles denote absences of the species, solid circles denote presences of the species, light stippling indicates G_i, and darker stippling indicates G_o = RN.

1.2.3 Modelling the fundamental niche

Different physiological models and approaches.

Physiological models describe the rates at which an individual organism feeds, assimilates or utilises energy for metabolic processes (such as maintenance, growth or reproduction) during its lifetime and depending on the surrounding environmental conditions (van der Meer 2006). Physiological models therefore explore the influence of environmental conditions on species physiological performances. They establish a causal relationship between species distribution and environmental variables, characterise the range of suitable abiotic conditions for the species to reproduce or survive, and consequently constitute a good proxy to characterise species fundamental niche (Kearney and Porter 2004, Sillero et al. 2011).

One of the most integrative theories of dynamic energy budgets is the DEB theory, developed in the late 1980's (Kooijman 1993, Kooijman 2000, Kooijman 2010), which inspired at least the development of 26 other popular empirical models (Comments on DEB 3, Kooijman 2010).

The Dynamic Energy Budget (DEB) theory: general principle.

The DEB theory defines individuals as dynamic systems and provides a mathematical framework for the life cycle of an organism, from the start of the embryo development to the death by ageing. It describes the physiological processes with four primary state variables: reserve, structure, maturity and reproduction buffer (the latter for adults only), directly linked to mass and energy flows and influenced by two forcing environmental variables: temperature and food resources availability (Fig. 0.9, Kooijman 2010). DEB theory relies on key concepts such as consistency with biological and ecological principles, as well as first laws of thermodynamics for conservation of mass, energy and time (Jusup et al. 2017) and assumes that the various energetic processes, such as assimilation and maintenance rates are dependent either on surface area or on body volume (van der Meer 2006).

The DEB model considers that consumed products are assimilated at a rate ($\dot{p}_A$) into a reserve pool, following a functional response of Holling's type II in the simplest case. This initiated energy is then mobilized at a rate ($\dot{p}_C$) from the reserves and allocated to maintenance ($\dot{p}_M$), structural growth ($\dot{p}_G$), maturity maintenance ($\dot{p}_J$), maturation of immature individuals ($\dot{p}_R$) or reproduction of mature individuals ($\dot{p}_R$) following a so-called κ-rule that controls energy acquisition and priority with assumptions related to empirical observations (van der Meer 2006, Sousa et al. 2008, Kooijman 2010) (Fig. 0.9). Priority is always given to somatic maintenance, followed by structural growth, maturity maintenance and reproduction. If the energy utilization rate from the reserves is not sufficient to pay for the somatic maintenance costs, the individual is assumed to die.

Biomass is modelled by the reserve and structure compartments. The non-structural complexification of the individual is symbolized by a cumulative investment of energy into maturity. The level of energy accumulated in this maturity compartment triggers metabolic switches such as the transition between the different life stages (e.g. ability to feed, to reproduce).

The development cycle of each species is divided into three life stages: (1) development starts at the embryo stage, when the organism is not able to feed nor to reproduce and is composed mainly of reserve and a negligible amount of structure; (2) the organism comes at the juvenile stage once the threshold for "birth" is passed as the organism starts feeding, however it is still not able to reproduce; (3) the adult stage is reached at "puberty", when the organism acquires the ability to reproduce. At this time, the maturity compartment stops receiving energy, organism complexification has reached its maximum, and this flow of energy is rather directed to reproduction (Kooijman 2010).

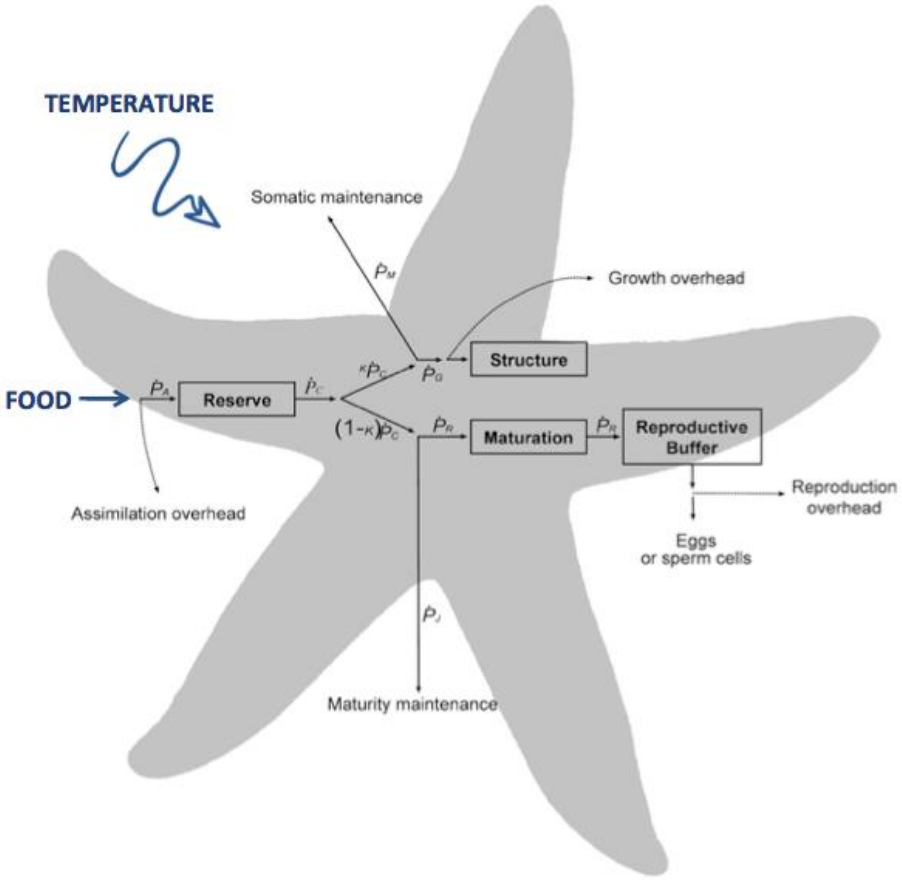


Figure 0.9. Conceptual scheme of the basic parameters and theoretical compartments of the DEB theory. Food initiates energy availability in the reserve compartment, energy is then allocated to the different metabolic processes. Temperature influences metabolism following Arrhenius principle (Kooijman 2010). Figure modified from Monaco et al. (2013).

Each DEB parameter (Table 0.1) is linked to specific physiological processes (van der Meer 2006) and the combination of these parameters covers the different energetic processes of the organism (feeding, digestion, storage, maintenance, growth, development, reproduction, ageing) (Marques et al. 2018).

Table 0.1. The 14 main DEB parameters and their units (Kooijman 2010, Marques et al. 2018).

DEB parameter	Notation	Unit
Specific searching rate	$\{\dot{Fm}\}$	$cm^{-2}.d^{-1}$
Assimilation efficiency	κ_X	-
Maximal specific assimilation rate	$\{\dot{pAm}\}$	$J.cm^{-2}.d^{-1}$
Energy conductance	$\dot{v}$	$cm.d^{-1}$
Fraction of energy allocated to somatic maintenance and growth	κ	-
Reproduction efficiency	κ_R	-
Volume specific somatic maintenance cost	$[\dot{pM}]$	$J.cm^{-3}.d^{-1}$
Surface specific somatic maintenance cost	$\{\dot{pT}\}$	$J.cm^{-2}.d^{-1}$
Maturity maintenance rate coeff	$\dot{kJ}$	d^{-1}
Specific cost for structure	$[E_G]$	$J.cm^{-3}$
Maturity at birth	E_H^b	J
Maturity at puberty	E_H^p	J
Gombertz stress coefficient	s_G	-
Weibull ageing acceleration	$\ddot{ha}$	d^{-2}

DEB theory: model implementation.

The calibration of a DEB model is fully documented on the DEB portal: (https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/, accessed November 2019), where Matlab codes and tutorials are provided. A new platform, *AMPeps* also helps complete these codes following a step-by-step tutorial (http://www.bio.vu.nl/thb/deb/deblab/add_my_pet/AmPeps.html).

Once the DEB model is created, the codes are checked by the administrators of the DEB community and shared in the Add-my-Pet collection (https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/species_list.html).

This collection of DEB models is growing fast, with more than 2765 species modelled in February 2021 (Fig. 0.10). For its creation, the model requires a set of zero-variate (single data) and uni-variate data (x~y relationship data) that can be extracted from the literature or obtained from experiments purposely designed for implementing the DEB model. These data should be recorded at different life stages of the individual. Food and temperature conditions at which data were recorded should be informed in the model (Table 0.2).

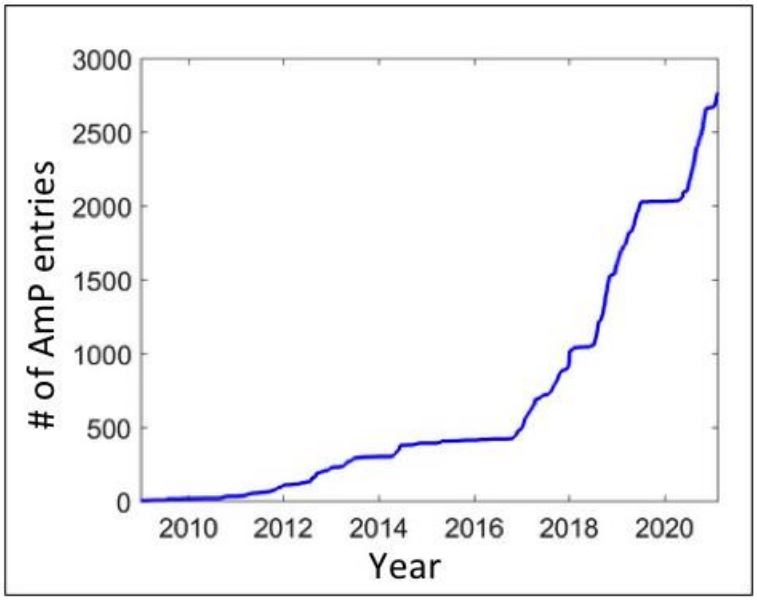


Figure 0.10. Number of DEB models built and published in the Add-my-Pet (AmP) collection: 2765 modelled species on 8th February 2021. Source: https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/about.html.

Table 0.2. Example of observations used to calibrate the DEB model of the Antarctic sea star *Odontaster validus,* from Agüera et al. (2015).

Table 2. Zero-variate data used for the estimation of the DEB model parameters. All values are given at a temperature of 271.5 K. This is the temperature at which these values where measured. MRE: mean absolute relative error.

Variable		Obs	Pred	Units	MRE	Reference
Age at birth[1]	a_b	33	27	d	0.19	[31]
Age at metamorphosis[2]	aj	165	162	d	0.02	[31]
Length at birth[1]	L_b	0.04	0.03	cm	0.24	[31]
Length at metamorphosis[2]	L_j	0.14	0.16	cm	0.17	[31]
Length at puberty[3]	L_p	2	2.15	cm	0.08	[26]
Maximum length[4]	L_i	7	7.10	cm	0.01	Pearse et al. *Unp. data*
Maximum length at f_r [5]	L_i	6	5.63	cm	0.05	[25]
Egg dry weight	Wd_0	1.1	1.18	µg	0.07	[46]
Weight at puberty[3]	W_p	2.95	2.84	g	0.04	[26]
Maximum weight[4]	W_i	100	105	g	0.05	Pearse et al. *Unp. data*
Maximum weight at f_r [5]	W_i	50	49.5	g	0.01	[25]
Gonadosomatic Index[6]	GSI	0.10	0.08	-	0.16	Pearse et al. *Unp. Data*
Pyloric Index[7]	PI	0.30	0.30	-	0.00	Pearse et al. *Unp. Data*

[1] birth is set at the moment the animal starts or is able to feed.
[2] moment at which the brachiolaria larvae is ready to settle for metamorphosis
[3] start of first gametogenesis
[4] maximum size reached by the species when there is no food limitation
[5] observations of maximum size at McMurdo field station
[6] maximum gonad index for an animal of the maximum size, gonad index as gonad weight/total wet weight.\]
[7] pyloric index for the food conditions in the laboratory as pyloric caeca weight/(total wet weight–gonad weight)

DEB theory: Parameter estimation.

DEB parameter estimation follows the covariation method (Lika et al. 2011a, 2011b), based on simultaneous minimizations of a weighted sum of squared deviations between observations and model predictions (i.e. a loss function), using the Nelder-Mead simplex method, updated and explained in Marques et al. (2018, 2019). The loss function that is minimized is:

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{w_{ij}}{n_j} \frac{(d_{ij-}p_{ij})^2}{\bar{d}_i^2 + \bar{p}_i^2}$$

where i scans datasets and j points in this dataset. d_{ij} and p_{ij} are respectively the data and the predictions and $\bar{d}_i$ and $\bar{p}_i$ their average values in set i. w_{ij} are the attributed coefficients (see below), n is the number of data sets, n_i denotes the data in a dataset, n_j the data in data-points.

Because it is assumed that certain observations have been made with greater confidence and accuracy than others, the procedure associates to each data-point a weight coefficient, on the basis of this prior knowledge. Complementary to that, the model structure is initiated with pseudo-data, being a set of potential parameters that describe a generalised animal, taxonomically close to the study species. Whether available, species-specific observations replace pseudo-data. Otherwise, pseudo-data are kept but associated with lower weight coefficients (i.e. lower confidence of the data) (Lika et al. 2011a). The covariation method has therefore similarities with a Bayesian estimation, but is not embedded in a maximum likelihood context, since the stochastic element is not modelled (Kooijman et al. 2008).

The goodness of fit of each prediction is quantified by the relative error (RE). The mean relative error (MRE) quantifies the overall model performance. RE corresponds to the sum of the absolute differences between observed and predicted values, divided by the predicted values. Contrarily to the loss function, the MRE does not take into consideration the weights of the different data (Marques et al. 2018). MRE values can have values from 0 to infinity, with 0 value meaning that predictions match data exactly.

DEB theory: model outputs and applications.

In complement of the estimated DEB parameters (Table 0.1), several compound parameters can be calculated to further describe species physiological traits (some examples are given in Table 3.3 in Kooijman 2010 or in Table 1 in Petter et al. 2014). This makes DEB applicable in an important number of fields, that continue to increase thanks to an intensive work on model compilation, validation and code sharing (van der Meer et al. 2014).

DEB theory has been widely used in aquaculture, fisheries and biotechnology (i.e. growth rate estimations, stock assessment) (Hanegraaf 1997, Ren et al. 2010, Serpa et al. 2013, van der Meer and Kooijman 2014), reconstruction of feeding history (Jusup et al. 2014, Agüera et al. 2017), description of species traits under contrasting environmental conditions (Pecquerie et al. 2009, Petter et al. 2014, Marn et al. 2017) or climate change conditions (Jager et al. 2016, Thomas et al. 2016, Ren et al. 2020), understanding of species distribution (Montalto et al. 2015, Schwarzkopf et al. 2016, Tagliarolo et al. 2016), comparison of species metabolic properties (Marques et al. 2018, Marn et al. 2019), ecotoxicology (Bodiguel et al. 2009, Jager and Zimmer 2012, Martin et al. 2014, Baas and Kooijman 2015, Sussarellu et al. 2016), or for the adaptation of experimental designs (Nisbet et al. 2000, Ashauer et al. 2016).

DEB model equations also provide life-history information for given environmental conditions, which makes DEB theory appropriate to study population dynamics (growth, reproduction mortality), that can be upscaled at the community or ecosystem levels by assuming body-size relationships or matrix population models (Klanjšček et al. 2006, Maury and Poggiale 2013, Guiet et al. 2016). Symmetrically, the DEB model can be downscaled to study sub-organismal processes, such as

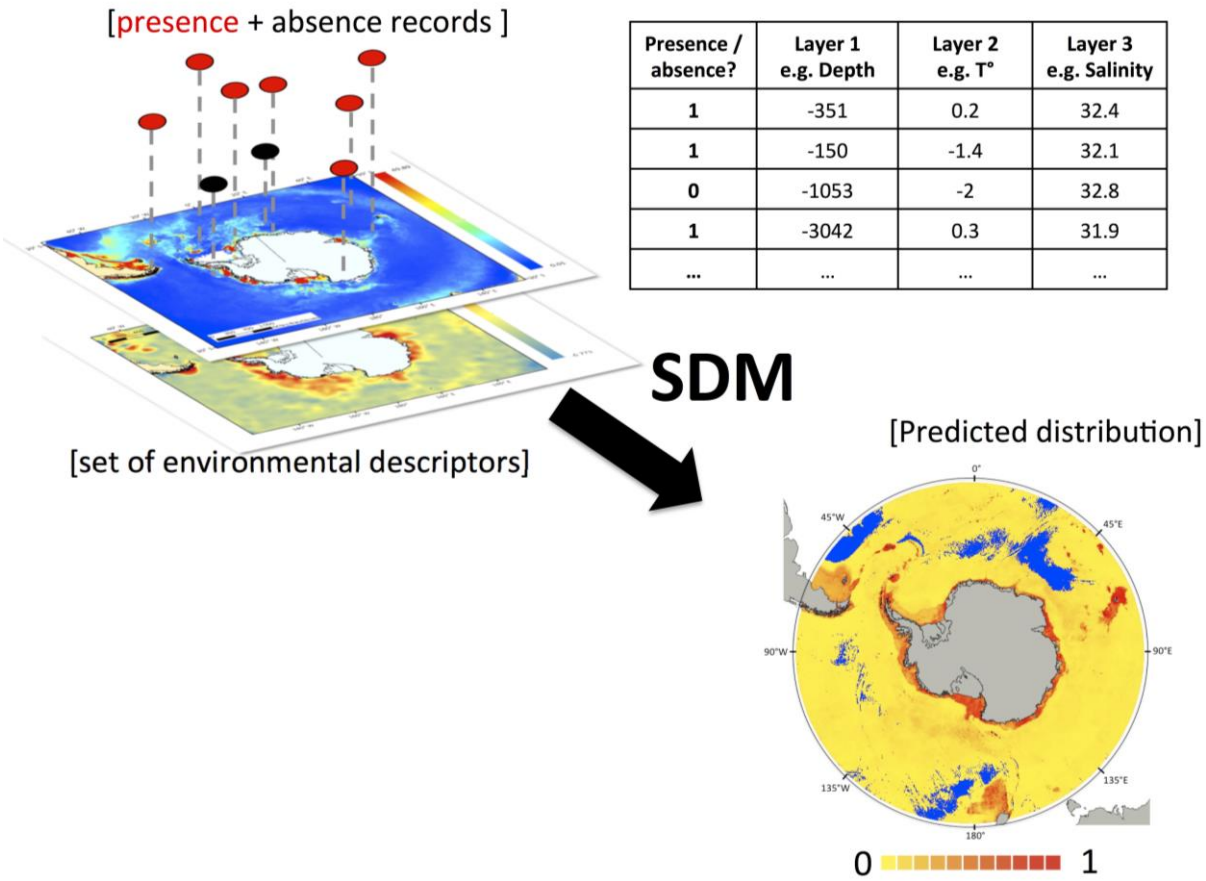
studying the impact of toxic compounds or damaging agents on these processes, using the principle of Synthetizing Units (Jager and Kooijman 2005, Muller 2011, Muller et al. 2019).

1.2.4 Modelling the realised niche

Principle and relation to niche theory.
Species Distribution Modelling (SDM) is also known as ecological niche modelling, habitat suitability modelling or climate envelope modelling (Austin 2002, Pearson 2007, reviewed in Sillero 2011). The acronym SDM is the most frequently used term in ecological modelling when referring to correlative models, that aims at predicting the distribution of a species (Pearson 2007). Soberón and Peterson (2005) and Soberón (2010) however distinguish SDM from ecological niche models (ENM), with the latter rather described as correlative models based on ecological niche theory, that provide an approximation of the species niche by forecasting the environmental conditions that are suitable for a species to survive or reproduce, rather than to the species distribution by itself. Considering the lack of consensus terminology (Sillero 2011), the term SDM refers in this manuscript to an ecological niche model that helps representing species realised niche.

SDM is based on a statistical relationship between occurrence records and environmental data (Elith et al. 2006, Elith and Leathwick 2009, Peterson et al. 2011). Environmental conditions at the location of available presence-only (or presence-absence) data are extracted to generate a matrix used to build the SDM (Fig. 0.11). The complexity of the relationship between occurrence records and environmental conditions is conditioned by the chosen mathematical representation of the SDM (i.e. the model algorithm: linear or polynomial relationships, classification trees, entropy minimisation) (Fig. 0.11) (Elith and Leathwick 2009, Anderson 2013). Model outputs that represent the probability distribution of the species are projected on a geographic and/or climatic/environmental space to identify areas where the environment fulfills the required environmental conditions (Anderson 2013).

Associated to prediction maps, several model outputs can be generated by SDMs: partial dependence plots, that describe how the range of values of each environmental descriptor is associated to model predicted suitability; descriptors contribution to the model or interactions between these descriptors within the model (see examples in Guillaumot et al. 2018b).



Presence / absence?	Layer 1 e.g. Depth	Layer 2 e.g. T°	Layer 3 e.g. Salinity
1	-351	0.2	32.4
1	-150	-1.4	32.1
0	-1053	-2	32.8
1	-3042	0.3	31.9
...	...	...	...

Figure 0.11. General principle underlying the construction of a Species Distribution Model (SDM) that determines the correlation between occurrence records (presence/absence) using a set of environmental descriptors. The chosen algorithm (black arrow) can be chosen to integrate more or less complex relationships between data and environment. SDM output is a map that provides the probability of the species to be distributed in the area (0: non suitable; 1: highly suitable).

A short history of the development of correlative approaches.

Understanding species distribution has long been a major issue in ecology (von Humboldt 1807, de Candolle 1855), and before using modelling approaches, many pioneering works aimed at explaining species distribution patterns as a response to environmental factors with experiments and observations (Salisbury 1926, Good 1947, Holdridge 1967, McArthur 1972, Box 1981). In the 1990s, predictive habitat distribution models were first introduced as efficient tools to test for biogeographical hypotheses, improve information provided by atlases, set up conservation priorities or assess the impact of environmental changes on species distribution (reviewed in Guisan and Zimmermann 2000). These models were initially based on simple algorithms, describing processes with empirical or static approaches. More complex methods (e.g. individual-based, stochastic forest path models,…) with new algorithms were progressively developed to push forward these theoretical limitations (Decoursey 1992, Korzukhin et al. 1996, Lischke et al. 1998). This development increased in parallel with the rise of new powerful statistical techniques (e.g. Bayesian approaches) and the improvement of Geographic Information Systems (GIS) (Guisan and Zimmermann 2000). Methodological works have consequently flourished in the literature with topics such as model verification, evaluation, calibration, and sensitivity that took the lead in study titles (Leohle 1983, Oreskes et al. 1994, Araújo and Guisan 2006).

More recently, following the development of computer sciences and calculation performances, 'machine-learning' approaches have been developed (Araújo and Guisan 2006, Elith et al. 2006). They were proved highly powerful to accurately model more complex relationships between occurrences and environmental conditions (Elith and Leathwick 2009, Lorena et al. 2011, Merow et al. 2014), and can be enriched by more information such as species dispersal abilities or inter-specific interactions (see section 1.2.6 ; Gobeyn et al. 2019).

Applications.
SDM is widely used in various fields of ecology, from conservation, biogeography, and palaeoecology, to global change biology (Pearson 2007). SDM has already been applied for several aims such as predicting the potential of alien species to invade new environments (Thuiller et al. 2005, Václavík and Meentemeyer 2012, Byrne et al. 2016), exploring speciation mechanisms (Graham et al. 2004, Kozak and Wiens 2006), testing evolution hypotheses (Kozak et al. 2008, Culumber and Tobler 2016) or discovering new species (de Siqueira et al. 2009), delimiting species distribution (Raxworthy et al. 2007, Williams et al. 2009), assessing the impact of land cover change (Pearson et al. 2004) or environmental shifts on species' distribution (Araújo and New 2007, Meier et al. 2011, Weinert et al. 2016), guiding the reintroduction of endangered species (Pearce and Lindenmayer 1998, Maes et al. 2019) or supporting diverse conservation planning, decisions or strategies such as providing a frame to observe and simulate the consequences of such decisions (Addison et al. 2013, Syfert et al. 2014, Ferrari et al. 2018) or guiding field survey to find populations of known species (Bourg et al. 2005, Guisan et al. 2006). Recent and innovative developments include the application of 3-D approaches to marine pelagic case studies (Bertrand et al. 2016, Duffy and Chown 2017, Freer 2018), the integration of high resolution oceanographic data with SDM (Pearman et al. 2020) or the coupling of SDM with extra knowledge or other models (such as mechanistic and/or dispersal models, see section 1.2.6 for further details). Stack-SDMs and Joint-SDMs constitute an important step towards estimating species richness by stacking several SDM predictions of different species and spatially aligning the cells with presence-absence, competition or interaction matrices to describe communities composition in space (Pollock et al. 2014, Distler et al. 2015, Harris 2015, Tikhonov et al. 2019, Zurell et al. 2020).

1.2.5 Modelling dispersal vectors and biogeographic barriers: Lagrangian models

Species distribution patterns do not only depend on abiotic conditions and biotic interactions, they are also determined by the possibility of adult individuals and propagules to access and settle in suitable areas (Anderson 2013, Caccavo et al. 2018, González-Wevar et al. 2019). Evaluating the connectivity between these areas has therefore important implications for the study of species distribution and population dynamics. It can be defined as the spatial movement of individuals, gene flow or transfer of information between individuals (Kool et al. 2013). Connectivity is important in marine environments, where oceanographic features such as currents, eddies, marine fronts, up and downwellings, play a crucial role in population structuring (Selkoe et al. 2008). These oceanographic features strongly complexify distribution patterns and studying the link between species biogeography, spatial distances, genetic differentiation or population structures becomes irrelevant without any complete analysis of species dispersal fluxes (Young et al. 2015).

Physical oceanography includes the study of several processes, widening from small scale water turbulence to global climate changes (Chelton 1994). The study of water movement constitutes a relevant approach to the analysis of larval dispersal, which is difficult to directly observe or measure in the water column, given that larvae are generally small compared to the vast ocean scale and that dispersal can sometimes occur during long periods (Helmuth et al. 1994, Matschiner et al. 2009).

The use of transport models has been widely applied in several contexts, such as the definition of connectivity networks that can be helpful for the definition of marine protected areas (Gaines et al. 2003, Berglund et al. 2012, Burgess et al. 2014, Thomas et al. 2014), the conservation of coral reefs (Treml et al. 2008, Wood et al. 2013), the sustainability of fisheries (Gilbert et al. 2010, Scales et al. 2018), the spread of invasive species (Brandt et al. 2008, Brickman 2014) or aquaculture parasites (Salama and Rabe 2013), the identification of barriers to larval dispersal (Lett et al. 2008, Thomas et al. 2014) and more recently the tracking of plastic debris (Zambianchi et al. 2017, Liubartseva et al. 2018).
Among transport models, the Lagrangian approach aims at following a particle from an initial position and along its entire trajectory (Bennett 2006). Such models are the combination of (1) an oceanographic model, that takes into consideration bathymetry, water current direction and speed, tidal motion, water stratification... in relation with atmospheric forcing (temperatures, winds, atmospheric pressure) (Huthnance 1991, Robinson and Golnaraghi 1994) and (2) biological properties of the dispersed individual such as size, development rate, buoyancy properties, ability to swim or orientate, or behaviour in the water column (e.g. nycthemeral movement to escape predation) (McManus and Woodson 2012, Van Sebille et al. 2018).

1.2.6 Integrated approaches

Principle and motivations
There is a strong trend for developing new methods to improve the ability of models to describe species distribution. The integration/coupling of several methods has been long recognized as a promising approach to improve model performance and gain in modelling capabilities and analytical power (Vincenot et al. 2011, Gutt et al. 2012). Indeed, integrated models are more efficient as they can represent a complex system using several accurate and precise submodels (Gray and Wotherspoon 2012) which can be totally merged (Vincenot et al. 2017).

Review of the different applications.
Integrated models have been developed in different fields: engineering, environmental science, microbiology, oceanography, demography, epidemiology (Bobashev et al. 2007, Emrich et al. 2008, Vincenot and Moriya 2011, Bradhurst et al. 2015), economics/management, health science (Martin and Schlüter 2015, Drogoul et al. 2016), in order to answer to contrasted objectives.
In ecology, there is also a broad range of applications: (1) merging SDM with models that dynamically describe landscapes has shown better realism and better predictive performance compared to traditional SDMs (Pagel and Schurr 2012, Zurell et al. 2016), as the equilibrium between occurrences and environmental conditions is dynamically updated (Brotons et al. 2012). (2) Close to these, Eulerian-Lagrangian approaches in oceanography can be coupled with biogeochemical models to understand spatial patterns and tracers dynamics in moving fluids (Chenillat et al. 2015), or to physiological models to simulate the growth and survival of organisms while they are drifting or migrating (Goodwin et al. 2006, Berline et al. 2013, Rivière et al. 2019). (3) Adding a dispersal information to SDMs can also improve species potential habitat predictions, as the environmental information is complemented by dispersal matrices characterising areas that are suitable for colonization (Engler and Guisan 2009, Anderson 2013, Normand et al. 2013). (4) Integrating population dynamics information (e.g. carrying capacity of the habitat, mean survival or fecundity rates of each stage class, population connectivity) strongly improves the ability of SDMs to assess species potential distribution (Keith et al. 2008, Anderson et al. 2009, Nenzén et al. 2012) in spatially or climatically contrasting areas (Parrott et al. 2012, Girard et al. 2015, Strauss et al. 2017). (5) SDMs can also be linked to phylogenetic analyses, to analyse species distribution in link with evolutionary connectivity (Morales-Castilla et al. 2017, Pardo-Gandarillas et al. 2018) or (6) to biotic

interactions, to improve process-based understanding (Meier et al. 2010, Dormann et al. 2018) or to better characterise the behaviour of a species within its community in future environmental conditions (Schweiger et al. 2008, Wisz et al. 2013, Bebber and Gurr 2019). (7) A last example is the combination of SDMs with physiological information (e.g. using a mechanistic model) that was proved efficient to improve predictions compared to simple correlative SDMs (Buckley et al. 2010, Elith et al. 2010, Singer et al. 2016, Pertierra et al. 2020). Whereas SDMs explain the statistical correlation between occurrence records and habitat suitability (Elith and Leathwick 2009) and assess the main ecological drivers of species distribution (Elith et al. 2006, Peterson et al. 2011), they are limited to a static description of the species distribution, and cannot accurately perform in non-equilibrium states, which limits their use for future projections (Loehle and Leblanc 1996, Schouten et al. 2020). Integrating physiological information enables to explicitly include processes in the analysis, offering the opportunity to describe the process-based causes of the species distribution (Kearney and Porter 2009, Dormann et al. 2012a), even in non-equilibrium states (Kearney et al. 2008, Keith et al. 2008).

Overall, evidence is accumulating that species' responses to climate changes are best predicted by modelling the interaction of physiological limits, biotic processes and the effects of dispersal-limitation (Fordham 2013, Tingley et al. 2014, Gotelli and Stanton-Geddes 2015). Combining simple model results with information from experiments or observed functional traits facilitates interpretation and strengthen conclusions (Dormann et al. 2018, Benito Garzón et al. 2019).

2. THE SOUTHERN OCEAN AS AN APPLICATION FRAMEWORK

2.1 The Southern Ocean

Oceanographic features.

The Southern Ocean (SO) here defined as waters south of 45°S latitude (Breitzke 2014), covers 8% of the world ocean surface (Barnes and Peck 2008) and plays a crucial role in the global ocean circulation (Schlosser et al. 1991, Doney and Hecht 2002). Huge water masses are put into movement due to contrasts between water densities (shifts in temperature and salinity values), playing a key role in the physico-chemical conditions of the whole world ocean, by connecting water masses all together (i.e. the 'thermohaline circulation') (Wunsch 2002, Jacobs 2004, Fig. 0.12).

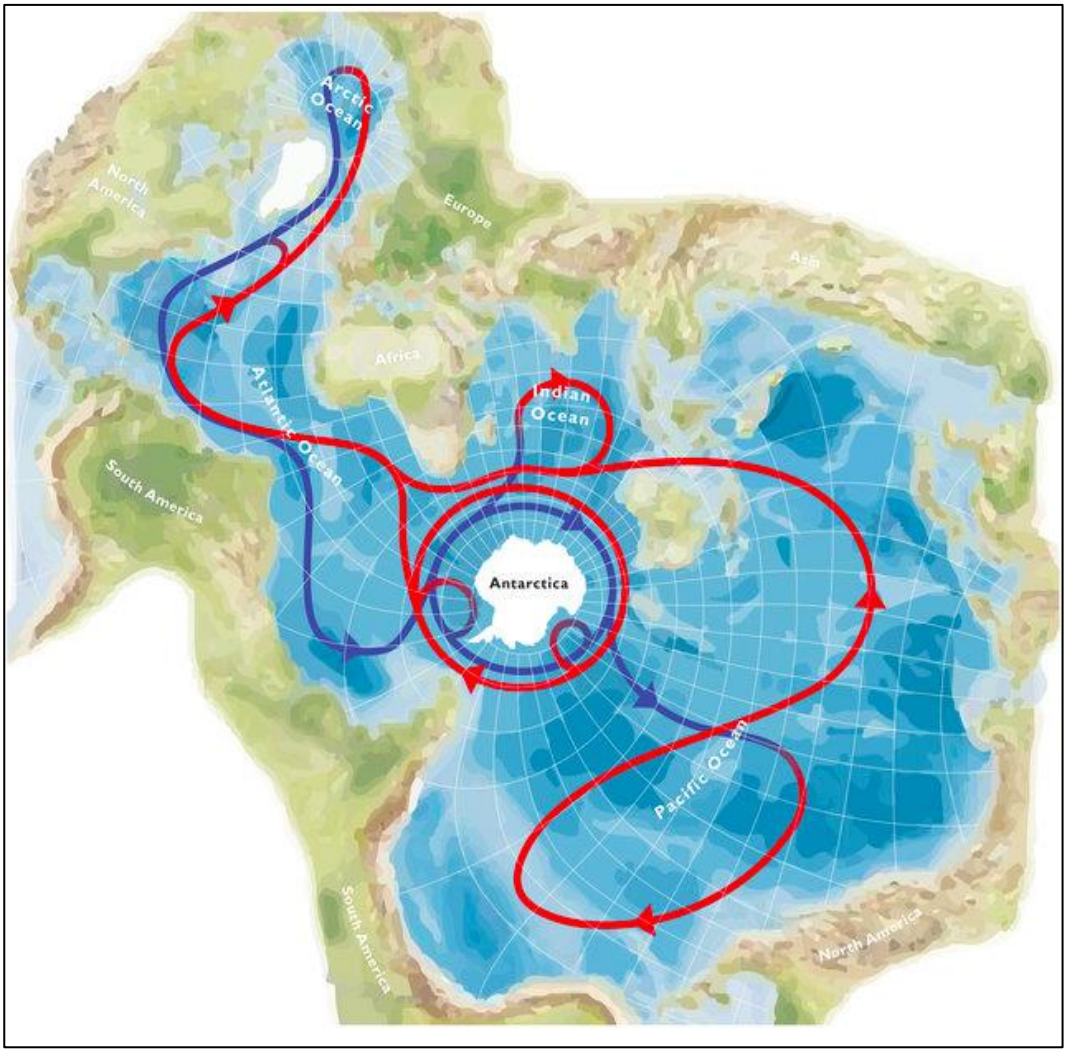


Figure 0.12. Spilhaus projection representing the Southern Ocean compared to all other oceans. The global thermohaline circulation is represented by red arrows for upper-layer flows (surface water currents) and blue arrows for lower-layer flows (deep water currents). The overturning of this 'conveyor belt' lasts between 1,000 and 2,000 years (Döös et al. 2012). Figure from Meredith (2019).

The SO is strongly structured by a major eastward flowing current, the Antarctic Circumpolar Current (ACC) that flows at ≈130.10^6 m^3/s on average (Rintoul et al. 2001, Fig. 0.13). The ACC reaches the highest width in the Atlantic sector (over 1,000 km) and is narrowed in the region of

the Drake Passage, between South America and the Western Antarctic Peninsula (Knox 2007). The ACC is associated with circumpolar marine fronts, separated by sharp changes in water densities, among which the Polar Front and the Sub-Antarctic Front are the strongest (Orsi et al. 1995, Rintoul et al. 2001). The Polar Front separates the northern and the southern parts of the ACC, and therefore represents a significant biogeographical barrier to the dispersal of Antarctic marine benthic faunas northward (Clarke et al. 2005, Sanches et al. 2016). The ACC simultaneously promotes the eastward dispersal of marine organisms (plankton larvae and propagules) around Antarctica (Fell 1962, Olbers et al. 2004).

Close to the Antarctic coasts, at about 60-65°S, the Antarctic divergence marks a rupture between the ACC and a westward coastal current, and corresponds to an area where deep waters, less salty but richer in nutrients, upwell to the surface (Gordon 1971). In the embayments of the Weddell and Ross seas are found cyclonic gyres that also have a strong influence on deep water properties and a substantial role in atmospheric interactions (Rintoul et al. 2001, Vernet et al. 2019, Fig. 0.13).

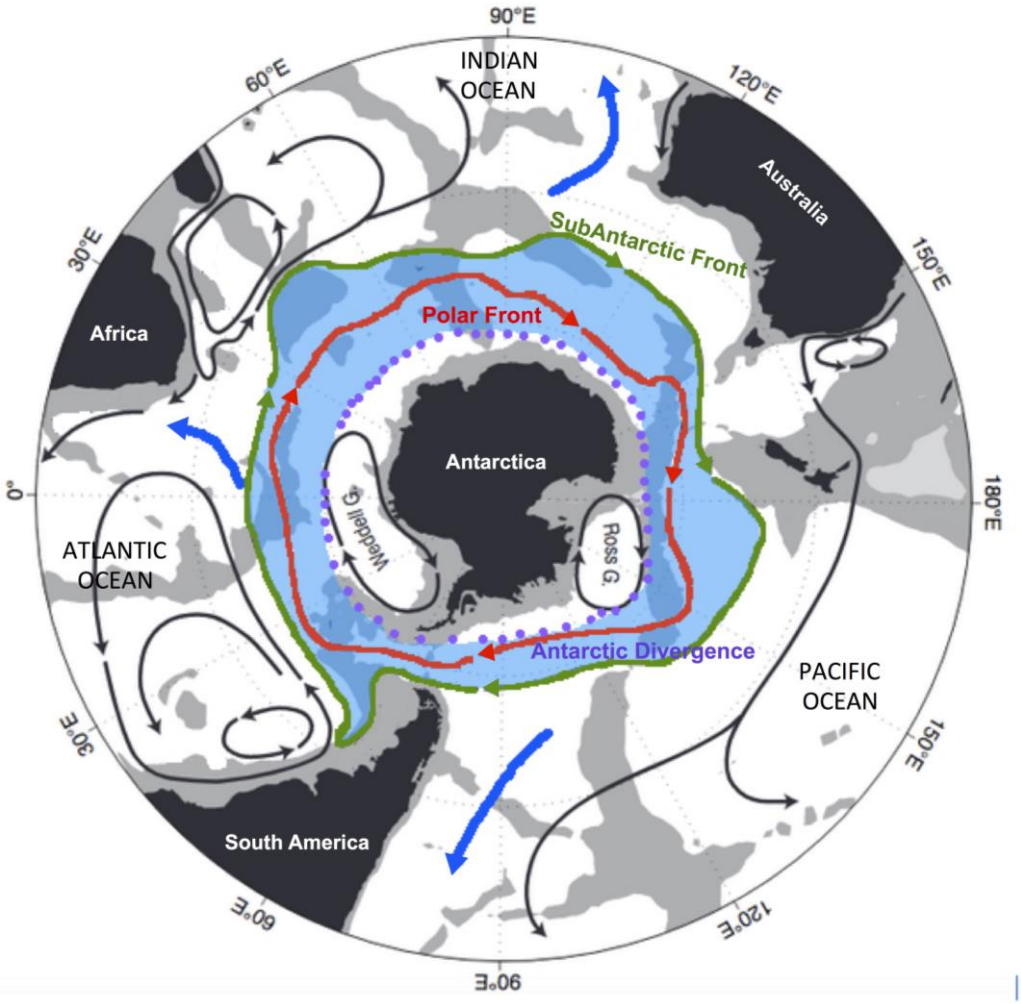


Figure 0.13. Main currents and marine fronts of the Southern Ocean system. The Antarctic Circumpolar Current flows eastward in the blue area, contained between the Sub-Antarctic Front in the north and the Antarctic Divergence in the south. Modified from https://geographyeducation.org/2017/01/07/the-worlds-newest-official-ocean/ (accessed January 2020) and Rintoul et al. (2001).

Bathymetry.

Similarly to other oceans, the SO is dominated by deep-sea habitats (> 3,000m depth) but Antarctica contrasts with other continents by a deeper continental shelf (averaging 400-900 m

depth) in comparison with other continental shelves (200 m on average) (Uri et al. 1992). The deepening of the Antarctic continental shelf is mainly explained by glacial isostasy: the continental shelf subsides due to the pressure exerted by ice loads on top of the Antarctic continent (average ice thickness of 2,100 m) and the lithosphere gets deformed (Okuno et al. 2012).

2.2 Biodiversity of the Southern Ocean

Biogeographic constraints, past isolation and species endemicity.
The geographic isolation of the Antarctic continental shelf fauna, along with the specific environmental conditions of the SO have led to a substantial proportion of endemic taxa (Clarke et al. 2005, Brandt and Gutt 2011), with levels reaching between 50 and 80% of shelf communities (Griffiths et al. 2009). Endemism is strongly varying according to SO regions and levels are comparable to other large and isolated regions such as New Zealand (Griffiths et al. 2009).
The SO is also characterised by a substantial species richness (Fig. 0.14), higher than in the Arctic for example (Gray 2001), due to the broader area that covers the SO (Dayton 1990), and a higher number of habitats and biogeographic provinces (Rosenzweig 1995). This higher species richness is also explained by the isolation of the Antarctic continent 20 million years ago (Crame 2000) that favoured allopatric speciation events (González-Wevar et al. 2012, Poulin et al. 2014, González-Wevar et al. 2018a) and by the many physical barriers present in the SO compared to the Arctic (i.e. currents, depth, ice coverage, fronts, geomorphological features) (Gray 2001, Clarke et al. 2005, Venables et al. 2012). Finally, the combined impacts of long-term gradual cooling and past glacial-interglacial cycles that occurred during the SO history have also led to diversification in biogeographic regions (González-Wevar et al. 2012, Strugnell et al. 2012, Fabri-Ruiz et al. 2020).
The SO marine communities are also characterised by the absence of some taxa (durophagous species, barnacles, most cartilaginous fish) (Clarke and Johnston 2003, Clarke et al. 2004) that are known in the fossil records of the SO, but that went extinct probably due to major cooling events during the Cenozoic era (Griffiths et al. 2013, Crame 2018).

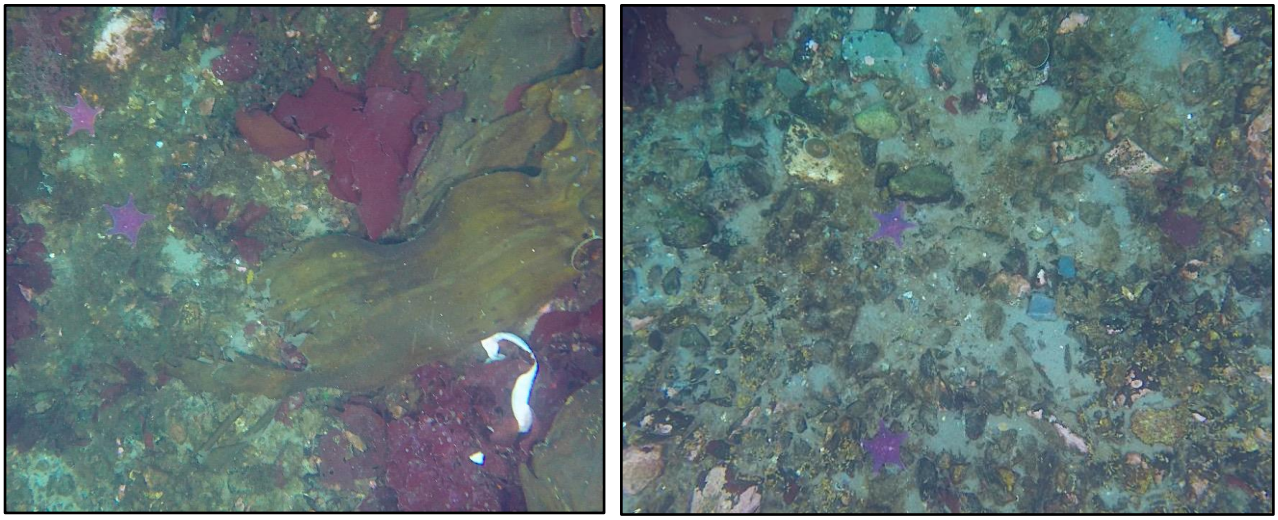


Figure 0.14. Pictures of seafloor communities at Useful Island (Gerlache Strait, Western Antarctic Peninsula), 15 m depth, March 2018. Rocky shallows to muddy substrate with gravels, with regular but shallow iceberg disturbance. Left picture: high macroalgae coverage, purple sea stars *Odontaster validus* on the left hand corner of the picture and white worm *Parborlasia corrugatus* on the bottom right corner. Right picture: Some *Nacella concinna* limpets are also present on rocks. © B121 Expedition.

Influence of environmental conditions on marine communities.
Spatial and temporal variabilities in marine community structure and diversity are mainly explained by the influence of current speed, ice dynamics, sediment properties, and food availability (i.e. chlorophyll-a concentration or sediment organic content) (Grange and Smith 2013, Cummings et al. 2018). Ice dynamics strongly influence marine species abundance (Gutt 2001, Palma et al. 2007, Lagger et al. 2017, Braeckman et al. 2021), as iceberg scouring directly impacts benthic communities down to 250 m depth (Barnes and Peck 2008, Barnes and Souster 2011, Barnes et al. 2014). Moreover, sea-ice duration and extent along with glacier and ice shelves melting guide variations in water mixed layer depth, light availability (Vernet et al. 2008, Venables et al. 2013, Schofield et al. 2018), modify wind impact on marine habitats (Saba et al. 2014) and lead to significant inputs of fresh water and sediment supplies (Lien et al. 1989, Dierssen et al. 2002, Moline et al. 2008, Monien et al. 2017, Barnes et al. 2018) that fertilize water in nutrients (Saba et al. 2014, Hendry et al. 2018, Moffat and Meredith 2018). These events lead to seasonal primary production blooms that contrast with low energy systems that dominate the major part of the year (McClintock 1994). Low temperatures below the 0°C threshold are also very frequent in coastal habitats (Jacobs et al. 1979, Ryan et al. 2004) and explain the important physiological and plasticity adaptations encountered in SO marine communities (see next paragraph), as observed for marine species of higher latitudes (Clarke 1980, Brey and Clarke 1993, Albers et al. 1996).

Physiological peculiarities of Southern Ocean organisms.
Southern Ocean species are characterised by a metabolism with a low protein production, as low temperatures induce increased synthesis costs (Marsh et al. 2001, Robertson et al. 2001, Fraser et al. 2004, 2007, Pörtner et al. 2007, Peck 2016). This leads to slower larval development and growth rates (Peck et al. 2007, Peck 2016, 2018), between 4 to 18 times slower than tropical water counterparts (Kaiser et al. 2013), which induces longer lifespans and generation times (Johnson et al. 2001, Higgs et al. 2009, Peck 2018). Another novel adaptation to the cold is the production of antigel proteins observed in certain fish species (Scott et al. 1986, Cziko et al. 2014). As SO species are highly adapted to these cold conditions, their ability to acclimate to elevated temperatures is often poor compared to marine groups elsewhere (Peck et al. 2014), with most of the Antarctic marine species having suitable temperature envelopes between 5 to 12°C above the minimum sea temperature of -2°C (Peck et al. 2004).
Due to low and seasonal food availability, most of the SO species have also adapted their mode of acquisition and allocation of energy according to food availability (Lawrence and McClintock 1994). Some species were also shown to have a substantial trophic plasticity and were proved capable of modifying the range of consumed prey according to environmental or community shifts (Calizza et al. 2018, Michel et al. 2019).

Reproduction.
Southern Ocean species present two reproductive behaviours that mainly differ in terms of nutrition and dispersal strategy (Moreau et al. 2017). Broadcast spawners (that disperse eggs in the water column) produce eggs that are generally 2 to 5 times bigger than those of species of lower latitude but in less important number (Bosch and Pearse 1990, Arntz et al. 1994, Leis et al. 2013, Peck 2018). This implies a greater reserve load enabling either an increase survival if metamorphosis occurs quickly, or the capacity to drift over periods of several months in order to coordinate their settlement close to the summer period when food is abundant (White 1998, Stanwell-Smith and Clarke 1998, Stanwell-Smith et al. 1999, Chiantore et al. 2002). This long range dispersal of pelagic larvae facilitates the geographic spreading of many species (Shilling and Manahan 1994, Poulin et al. 2002, Young et al. 2015) and played a key role in the evolutionary history of SO benthic invertebrates (Thatje 2012).

The second main reproductive strategy that characterises SO species is the brooding behaviour, where youngs are carried by parents during a long period of time (Pearse et al. 1991, Poulin et al. 2002, Moreau et al. 2017). The unusually high number of SO benthic marine species with non-pelagic development is explained by adaptation to current environmental conditions (protection of the offspring) and the result of population selection, as a consequence of repetitions in population fragmentation over time with isolated units forming new species (Poulin et al. 2002, Pearse et al. 2009). This results in a lower dispersal capacity of species that promotes geographic isolation of populations between provinces (Moreau et al. 2017, Halanych and Mahon 2018).

2.3 Climate change in the Southern Ocean

Observed and predicted environmental changes.
As in other parts of the world, the SO is facing environmental changes with important regional contrasts (Meredith and King 2005, Martinson et al. 2008, Convey et al. 2009). While sea ice has significantly been increasing in the Ross Sea both in concentration, extent (Comiso and Nishio 2008) and duration (Stammerjohn et al. 2012), the Western Antarctic Peninsula has shown important temperature warming during the twentieth century, with particularly pronounced events during winter, and observed a rise of +3°C in atmospheric temperatures since 1951 (King et al. 2003, Vaughan et al. 2003, Meredith and King 2005, Henley et al. 2019). Ocean warming has also been observed, with a rise of water temperature of +0.17°C in depths between 700 m and 1,100 m between the 1950s and the 1980s (Gille 2002). These warmings influence atmospheric variability (Meredith and King 2005), and temperatures of water masses connected with the world ocean's deep seas (Fig. 0.12; Sallée et al. 2018). It also resulted in increased surface water freshening close to glacier meltwater sources (Schloss et al. 2012, Bers et al. 2013), led to changes in duration and extent of ice cover since the 1970's (Ducklow et al. 2013, Turner et al. 2016, Schofield et al. 2017) and changes in glacier retreat (Padman et al. 2012, Cook et al. 2016) (Fig. 0.15).

In the near future, meta-analyses of several global climate models are predicting continuing atmospheric and oceanic warmings of several degrees (Walsh 2009, Bracegirdle and Stephenson 2012, Mayewski et al. 2015). These climate models (CMIP5, Coupled Model Intercomparison Project), are developed by the Intergovernmental Panel on Climate Change (IPCC) to predict water temperature of the entire water column south of the Polar Front by the end of the century (IPCC, 2014). They describe four RCP scenarios (Representative Concentration Pathways, 5th report 2013), that base the assumptions on different greenhouse gases emissions in the atmosphere in coming decades, between moderate (RCP 4.5) to business-as-usual (RCP 8.5) scenarios (Turner et al. 2014, Liu and Curry 2010).

Consequences of climate change on Southern Ocean marine communities.
Impacts on organisms of these cascading environmental changes have already been recognized (see Convey and Peck 2019 for a review) and include studies on fish (Bilyk and DeVries 2011, Strobel et al. 2012), molluscs (Clark et al. 2008, Peck et al. 2007, Reed and Thatje 2015), echinoderms (Peck et al. 2009a, Morley et al. 2016), isopods (Young et al. 2006, Janecki et al. 2010), foraminifera, nematoda, amphipoda (Ingels et al. 2012) and sponges (Fillinger et al. 2013). Warming temperatures directly reduce species survival (Peck et al. 2009b, Morley et al. 2009a, 2010, Peck 2011, Navarro et al. 2020), as biological functions such as feeding, rasping, swimming activities or even respiration ability, that are important for long-term survival are tightly constrained by the elevation of temperature (Peck et al. 2004, Morley et al. 2009b). Moreover, ocean warming

reduces oxygen availability for marine organisms, as oxygen concentration is lower in warmer waters (Benson and Krause 1984, Peck and Uglow 1990). These changes affect species ability to produce energy to maintain (in this context of increased metabolic rates) without using anaerobic processes that induce toxic end products (Peck 2005, Pörtner et al. 2007).

In addition, the combined effect of altered sea ice dynamics and increased meltwater runoffs, with wind patterns and oceanographic conditions sometimes have a unprecedented impacts on plankton communities, with declining habitat suitability (Whitehouse et al. 2008), induced shifts in dominating species within planktonic communities, likely to modify relative species abundances (Whitehouse et al. 2008, Montes-Hugo et al. 2009, Schloss et al. 2012, Schofield et al. 2017), consequently altering community assemblages (Moline et al. 2004, Ashton et al. 2017), functions (Braeckman et al. 2021) and predator-prey interactions up in food-webs (Michel et al. 2019). More consequent impacts of global change are therefore expected in shallow marine communities and coastal habitats (Kidawa and Janecki 2011, Grange and Smith 2013, Obryk et al. 2016) compared to deeper ones (Gutt et al. 2015).

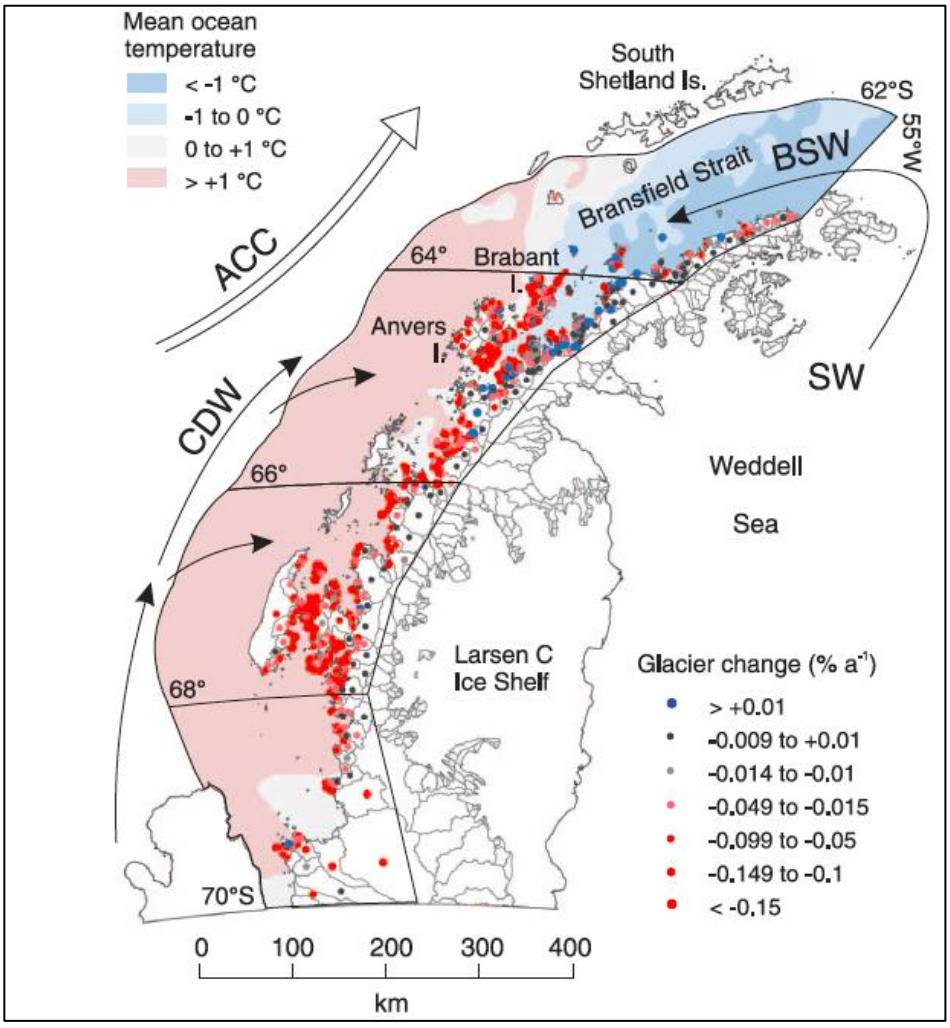


Figure 0.15. Mean ocean temperatures and overall glacier area changes, from 1945 to 2009 along the Western Antarctic Peninsula. Mean *in situ* ocean temperature at 150 m depth (shaded) and glacier change (points). For each of the 674 glaciers along the west coast, the point shows overall change between its earliest and latest recorded ice-front position, relative to basin size (% relative change rate a^{-1}). A similar spatial pattern is found for changes in absolute area loss per glacier. The point symbols are layered in the same order as in the legend (i.e. blue above red). Ocean circulation and water masses are also shown schematically: CDW (Circumpolar Deep Water), Shelf Water (SW), BSW (Bransfield Strait Water), and ACC (Antarctic Circumpolar Current). From Cook et al. (2016).

2.4 Science in the Southern Ocean

Contrarily to the northern pole, there are no permanent inhabitants nor native human populations in Antarctica and the surrounding islands, but temporary visitors of research stations ruled by several countries (Fig. 0.16).

The distribution of research stations strongly conditions our knowledge of marine life that is contrasting between SO regions (Clarke et al. 2007). To cover these gaps, recent organisations have promoted connectivity between international scientific programs, data accessibility and data cross-checking (Schiaparelli et al. 2013). These programs, such as the International Polar Year (IPY 2007-2008), the Census of Antarctic Marine Life (CAML 2005-2010) or the Scientific Committee on Antarctic Research, Evolution and Biodiversity in Antarctica (SCAR-EBA 2006-2013) (non-exhaustive list), were often associated with numerous field campaigns that contributed to considerably filling knowledge gaps, also increasing the sampling of the benthos and data accessibility (De Broyer et al. 2014). The recent development of underwater imagery in polar environments also helped to significantly improve data collection (Piepenburg et al. 2017). For several years, many programs have also settled long term high frequency observatories of marine life to characterise marine biodiversity and monitor potential shifts in community structures through time, in link with recorded environmental changes (e.g. Potter Cove in King George Island since 1993; MORSea in the Ross Sea since 1994; REVOLTA program in Adélie Land since 2009; PROTEKER program in the Kerguelen Islands since 2011).
In addition to these programs, online platforms that gather samples and their associated metadata were developed (RAMS: Registry of Antarctic Marine Species), OBIS (Ocean Biogeographic Information System), GBIF (Global Biodiversity Information Facility), SCAR-MarBin (Scientific Committee on Antarctic Research, Marine Biodiversity Information Network), which promoted free and open access to raw biodiversity data, in order to improve the accuracy of SO biogeographic and ecological studies (Pierrat 2011, De Broyer et al. 2014, Fabri-Ruiz 2018).

Despite this progress, the amount and quality of collected data are still limited in comparison to the extent of the SO (De Broyer et al. 2014). Sampling is concentrated nearby stations and generally performed in summer, as it is challenging to sample during the austral winter due to ice coverage (Griffiths 2010, Henley et al. 2019). Experiments in research stations are possible and performed since several decades (Féral and Magniez 1988, Peck et al. 2014, Suckling et al. 2015) but the possibility to settle long term experiments is often constrained by the harsh conditions (remoteness, cold, wind, ice coverage) (Kaiser et al. 2013).
When studying species distribution, sampled data are restrained to presence-only records, without the possibility to trust absence records given that they are not kept in the large biodiversity databases and that the entire sampled are not always characterised on board, depending on the expertise of the research team participating to the survey (Pierrat et al. 2012, Fabri-Ruiz et al. 2019). Important uncertainties are finally present in these data platforms, as they consist in a collection of several historical databases, and may contain inconsistencies between georeferencing systems or taxonomic definitions through time (Newbold 2010) or unchecked identification errors while sampling, above all for deep species, more recently studied (Brandt et al. 2007).

Figure 0.16. Political map of Antarctica, from https://www.geographicguide.com/antarctica-political-map.htm, accessed November 2019.

2.5 Tourism in the Southern Ocean

In parallel to science, tourism has been developing rapidly in the Antarctic region (Wace 1990, Lamers et al. 2008, Bender et al. 2016, McCarthy et al. 2019, Hughes et al. 2019). It began in the 1970's with the first aviation journeys for commercial purposes (Headland 1994) to reach a number of visitors of more than 55,400 people in 2018-2019 (IAATO 2019). The more recent promotion of cruise ships has provoked an exponential rise during the last few years, with a total number of visitors between summers 2014-15 and 2018-19 shifting from 36,700 to 55,400 people (Hughes and Convey 2010, IAATO 2019) and predicted to importantly increase in the coming years (Kruczek et al. 2018).

The combination of climate changes and tourism development in the SO region increases the risk that non-native species will access and survive in the area (Walther et al. 2009, McCarthy et al. 2019), which would constitute one of the most critical global threats to native biodiversity (Sax et al. 2005). Species can be introduced by the release of ballast waters, the fouling on ship hulls, floating anthropogenic debris, kelp rafts, or human visits, mainly from the Patagonian Peninsula, where cruise ship departures are the most frequent (Barnes 2002, Lewis et al. 2003, Tavares and De Melo 2004, Lewis et al. 2005, Lee and Chown 2007, Fraser et al. 2018). With the impressive number of tourist visits, along with the scientific activity (4,000 scientists working in Antarctica during the summer and 1,000 in winter; Hughes and Convey 2014) the arrival of propagules in Antarctic communities is growing (Tavares and De Melo 2004, Lee and Chown 2007, Hellmann et al. 2008, Galera et al. 2018, Avila et al. 2020). Consequently, records of terrestrial exotic species in Antarctica are increasing over recent decades (Smith and Richardson 2011) including the invasive grass *Poa annua* (Molina-Montenegro et al. 2012, Chwedorzewska et al. 2015), seeds of the toad rush *Juncus bufonius* (Cuba-Díaz et al. 2013), the invasive mosquito *Trichocera maculipennis* (Potocka and Krzemińska 2018), and several South-American invertebrates (e.g. insects, worms, freshwater crustaceans; Hughes and Worland 2010, Hughes et al. 2015). In marine habitats, alien species have also been reported in shallow areas of the South Shetland Islands (e.g. decapods, bivalves, macroalgae) and East Antarctica (i.e. bryozoans, hydrozoans, bivalves and tunicates) (Fraser et al. 2018, McCarthy et al. 2019, Avila et al. 2020, Cárdenas et al. 2020) but also from Sub-Antarctic waters and SO deep seas, such as anomuran king crabs (Thatje and Fuentes 2003, Thatje et al. 2005a, Aronson et al. 2014, 2015).

To date, there is no evidence for any exotic marine species having established in Antarctica, due to ecological and physiological constraints (Convey and Peck 2019). However, as climate keeps warming, the potential for successful marine invasions and settlement into Antarctica is expected to increase substantially (Richardson et al. 2000, Hellmann et al. 2008, Galera et al. 2018). Consequences of such invasions on native marine communities will have severe impacts on community assemblages, as observed in other regions of the world (shifts in competition, increase of predation pressure, colonisation of associated parasites that may infect other species, Falk-Petersen et al. 2011, David et al. 2017, Britton et al. 2018, Bevins 2019). Although the effects of invasive species are impossible to measure, the return of durophagous predators that became extinct million years ago (Aronson and Blake 2001, Zachos et al. 2008, Hansen et al. 2013) such as decapods, chondrichthyans and teleosteans in Antarctic shallow waters is widely feared, because they will fragilize benthic communities, modifying trophic relationships, and homogenizing the Antarctic ecosystem (Aronson et al. 2007, 2014).

2.6 Conservation of the Southern Ocean marine life

The Antarctic Treaty, signed in 1959 by a current number of 54 parties, regulates international relations in link with Antarctica, with an ensemble of 15 articles that rule politics, war, access and trade for all countries (Antarctic Treaty 1959). The treaty, which will end in 2048, is complemented by the CCAMLR's work (Commission for the Conservation of Antarctic Marine Living Resources). CCAMLR was established by an international convention in 1982 (https://www.ccamlr.org). The main aim of the commission is to manage marine communities in response to an increasing commercial interest in Antarctic fisheries (such as krill and fish resources) since the past few decades (Nicol and Foster 2003, Brooks 2013). This management does not exclude harvesting but agrees on the establishment of a set of conservation measures to carry out harvesting of marine living resources in a sustainable manner by taking account of the effects of fishing on other components of the ecosystem. The conservation of Antarctic marine life also includes the establishement and monitoring of marine protected areas (MPAs) (https://www.ccamlr.org/en/organisation/home-page, accessed November 2019) and the update of a list of Vulnerable Marine Species and Ecosystems (https://www.ccamlr.org/en/compliance/vulnerable-marine-ecosystems-vmes; Thompson et al. 2016).

The convention area ruled by the CCAMLR represents around 10% of the total surface of Earth's oceans and almost 70% of the SO, with a surface of more than 35 million km^2 (Fig. 0.17). Among the CCAMLR managed area, two MPAs have been established so far (Fig. 0.17): the South Orkney Island (in 2009) and the Ross Sea region (in 2016). In complement, several countries have declared national MPAs around sub-Antarctic islands outside of CCAMLR jurisdiction: Heard and McDonald Islands (in 2002, extended in 2014; Australia), Crozet and Kerguelen Islands (in 2006, extended in 2017; France), South Georgia and South Sandwich Islands (in 2012, United Kingdom) and Prince Edward Islands (in 2013, South Africa). In total, about 11.98% of the SO is protected with MPAs, with 4.61% being encompassed by no-take areas (Brooks et al. 2020) and negotiations are in process to extend protection to East Antarctica, to the Weddell Sea and the Antarctic Peninsula regions (Fig. 0.17).

Modelling approaches are broadly used for designing management decisions. Among them, SDMs are widely applied to define niche occupation of vulnerable species that are a priority to conservation (e.g. sea birds and mammals), to rank areas by importance of species richness or to model catch and effort data (Candy 2004, Ballard et al. 2012, Baird and Mormède 2014). The software MARXAN (Ardron et al. 2008, Ball et al. 2009, Teschke et al. 2017) is commonly used, as an efficient and useful conservation planning software for the decision making process through the identification of the most priority areas to be protected (Loos 2006, Klein et al. 2008). MARXAN is fully adapted to solve complex solutions for seascapes or landscapes zoning (Smith et al. 2009, Watts et al. 2009) and its application for SO case studies consequently follows the popularity it has gained in the other regions of the world (Zacharias et al. 2006, Teschke et al. 2015, 2017). Constantly updated (Watts et al. 2009), MARXAN is flexible to integrate different types of data, such as SDM outputs (Marshall et al. 2014), but is highly sensitive to their initial calibration (Loiselle et al. 2003, Wilson et al. 2005). Food-web models and bioregion statistical clustering approaches are also developed at local or broad scales (Pinkerton and Bradford-Grieve 2010, Sharp et al. 2010, Koubbi et al. 2011a, 2016, Martin et al. 2019, Fabri-Ruiz et al. 2020). Population dynamics models can also be used to define fisheries stocks (Mormède et al. 2014a, 2014b).

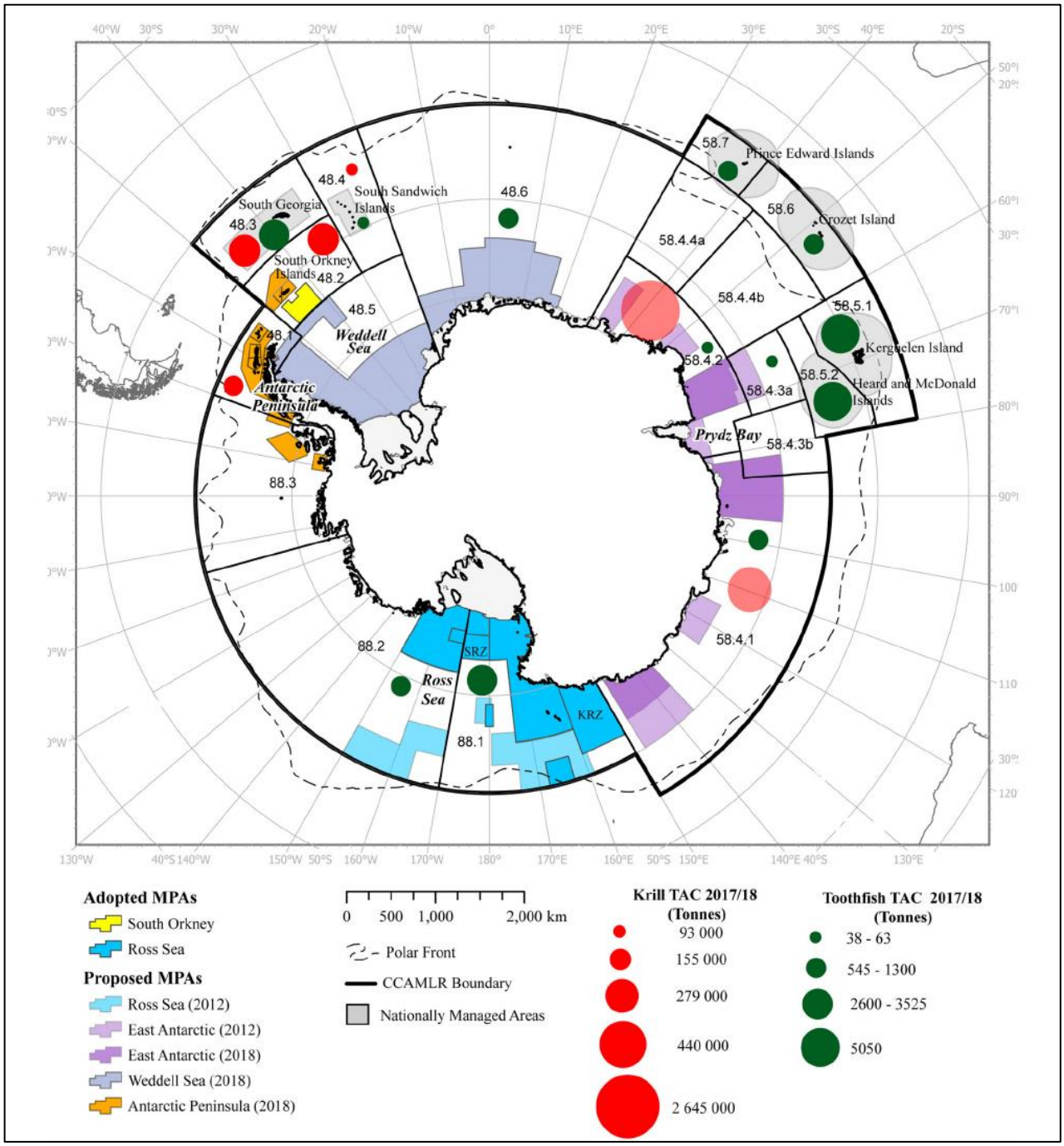


Figure 0.17. Proposed and adopted MPAs, management areas, and fisheries in the CCAMLR area. CCAMLR boundary indicated by thick black line with management area delineations labelled numerically. CCAMLR's adopted MPAs and MPA proposals from 2012 to 2018, including the South Orkney Islands Southern Shelf MPA (yellow), Ross Sea MPA (blue), East Antarctic (violet), Weddell Sea (purple) and the western Antarctic Peninsula (orange). Total Allowable Catch (TAC) for toothfish (blue) and krill (red) in the CCAMLR management area; circles proportional to respective TAC (tonnes in 2017/18), transparency indicates underutilization. Shaded circles around subantarctic islands reflect delineated exclusive economic zone boundaries generated prior to the signing of the CCAMLR Convention. Shaded squares indicate toothfish management areas around South Georgia and South Sandwich Islands. Figure does not include subantarctic MPAs which fall outside of CCAMLR's jurisdiction. From Brooks et al. (2020).

3. MODELLING THE ECOLOGICAL NICHES OF ANTARCTIC MARINE LIFE

3.1 State of the art

Everywhere on Earth, the observed impact of environmental changes on terrestrial and marine life is significant, and predictions of the increase of these changes and associated consequences are even further pessimistic. Scientific researches focussing on the consequences of these changes on living populations are consequently growing. Among regions and impacted ecosystems, the SO has encountered impressive environmental changes over the past fifty years that are also expected to increase in the future. In this context, and following the opportunity of technological improvements to access polar environments, research activities in Antarctica have substantially increased.

Modelling approaches constitute one essential tool nowadays in research to help understand ecological processes by synthetically representing complex systems. Methods and applications have been widely published worldwide. Models integrate data, imply experiments to validate hypotheses and perform predictive simulations. Applying ecological models to the SO has regularly been done for many years (De Broyer et al. 2014). Physiological models were used, with the pioneer studies applied to pelagic species: the Antarctic krill *Euphausia superba* (Groeneveld et al. 2015, Jager and Ravagnan 2015), and the salp *Salpa thompsoni* (Henschke et al. 2018) in order to investigate the effect of environmental changes on individual metabolic activity. These works were rapidly followed by the first application to benthic species, with the work of Agüera et al. (2015) that built a DEB model for the Antarctic sea star *Odontaster validus*, to describe its life cycle and better understand its adaptations to environmental conditions. More recently, DEB has been also applied to SO marine mammals (Goedegebuure et al. 2018) for evaluating population densities and structure. Regarding correlative approaches, models were mainly used in studies on commercial species such as pelagic fish or crustaceans (Loots et al. 2007, Cheung et al 2008, Pinkerton et al. 2010, Koubbi et al. 2011b, Basher and Costello 2016, Freer et al. 2019), top predators (Thiers et al. 2017) or bottom fisheries (Hibberd 2016), phytoplankton (Pinkernell and Beszteri 2014), sea birds (Krüger et al. 2018) and sea mammals (Southwell et al. 2005, Murase et al. 2013, Bombosch et al. 2014), by sometimes gathering occurrence records by GPS trackers fixed on animals' backs (Nachtsheim et al. 2017). The development of SDMs for marine invertebrate studies is more recent (Gutt et al. 2012), with the analyses of the potential distribution of sea urchins (Gutt et al. 2012, Pierrat et al. 2012, Fabri-Ruiz et al. 2019), sea stars (Byrne et al. 2016), crinoids (Hemery et al. 2011), cephalopods (Xavier et al. 2015) or barnacles (Gallego et al. 2017). Finally, dispersal models have been used to localize primary production hot spots (Piñones et al. 2011), study species or larvae spatial connectivity (Ashford et al. 2012, Piñones et al. 2013, La Mesa et al. 2015, Ashford et. 2017) or study the formation or retention of plankton, krill swarms in the context of fisheries sustainability or efficiency (Huntley and Niiler 1995, Fach et al. 2002, Hofmann et al. 2004, Thorpe et al. 2004, Hill et al. 2006, Young et al. 2014). Often, these models are combined with phylogeography studies (Young et al. 2015) or with species ecological or physiological information to fill knowledge gaps (Ashford et al. 2010, La Mesa et al. 2015).

All these works have faced methological challenges when implementing models, including the poor quality and availability of environmental descriptors that reduce the capacity to accurately integrate the variability and complexity of natural systems; the spatial aggregation and limited number of occurrence records that bias model predictions, influence the performance of model evaluation and

reduce the quality of the description of the species occupied space; the choice of the boundaries of the projection area, balanced between research objectives and data availability; or data gaps that limit the implementation of physiological models or the biological properties of the lagrangian approaches... Models should be also adapted to the physiological peculiarities of SO marine species (low adaptation to temperature increase, brooding reproductive behaviour with parental care or broadcaster species that disperse larvae that can drift in the water column during several months…). These points are typical of the broad-scale SO region and it is necessary to conduct some analyses to evaluate their real influence on ecological models, which has never been done so far.

3.2 Research objectives and motivations

In this context, this PhD thesis aims at analysing the potential and limits of ecological models applied to SO case studies. The BAM diagram scheme, presented earlier, is used to structure our study in several steps. Thus, using marine benthic species examples, we evaluate models that represent species fundamental niche (physiological, DEB models), models that study species realised niche (correlative models, SDMs) and models that focus on dispersal capacities (lagrangian approaches).

The first objective is to assess the quality of each of these ecological models, generated with such datasets, and to analyse their limits. Some correction methods, inspired from methods used in other regions of the world, are proposed to improve the performance of models (i.e. improve model evaluation procedures when using aggregated datasets, reduce model extrapolation, reduce the influence of spatial aggregation on predictions,...) and to provide guidelines for model implementation (choice of environmental descriptors, choice of SDM algorithm, cross-checking and preparation of datasets,...).

The second objective is to test the performance of “integrated approaches” compared to “simple” ones. These integrated approaches combine several types of information or models (e.g. combination of SDMs with physiological information, with dispersal capacities, with phylogenetic analyses…). Integrated approaches are widely used in other regions of the world and prove a better performance to describe species occupied space and environmental preferences compared to simple approaches. However, they have not been tested for SO case studies yet. This PhD thesis proposes some analyses related to comparisons and integrations of SDMs and physiological models (DEB) and SDMs combined with experimental data. All codes to generate these models are provided for future applications.

The third objective of this study is, after dealing with these corrections and methodological adaptations, to discuss about the capacities of ecological models applied to SO case studies and the remaining limits. What can we learn from these models (ecologically-wise)? Can we accurately represent the different parts of the species ecological niche? What is still uncertain? What should be improved to generate more relevant models?

This PhD thesis is declined into four chapters that present an ensemble of peer-reviewed articles submitted or published in international journals.

THESIS SUMMARY !

STUDY
species distribution
species sensitivity to climate change

COMMON QUESTIONS IN ANTARCTIC SCIENCES !

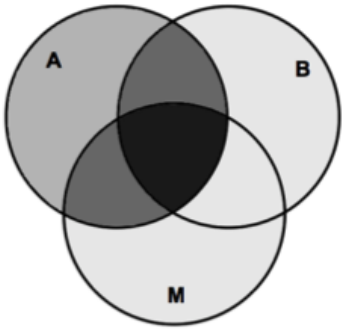
In order to study species distribution and predict the effects of climate change, this PhD thesis uses a tool :

ECOLOGICAL MODELLING

STUDY SPECIES DISTRIBUTION

NICHE THEORY

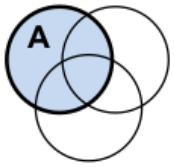
'BAM diagram'



B: positive and negative interactions between species

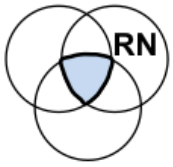
A: environmental conditions suitable for species survival and reproduction

M: dispersal abilities



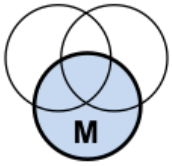
Mechanistic models

Study the fundamental niche: the ensemble of abiotic conditions that are suitable for species survival



Correlative models

Study the realised niche: the area where abiotic conditions are favourable + biotic interactions are positive + dispersal is possible



Dispersal models

Study the geographic barriers and the dispersal abilities of species

MAIN QUESTIONS OF THE PHD THESIS

- What are the peculiarities of Southern Ocean marine case studies ?
- Can we produce relevant and accurate models at the scale of the Southern Ocean?
- What are the limits of these modelling approaches ?
- Which methodological improvements can be applied to generate better models ?
- Are combined modelling approaches more relevant than simple approaches to study species distribution?
- What are the peculiarities of Southern Ocean species?
- What are the expected distribution in future environments facing climate change ?
- What are we able to accurately model so far in Southern Ocean studies?
- What are the main perspectives in Southern Ocean modelling?

CHAPTER 1 MECHANISTIC MODELS

This chapter focusses on the Dynamic Energy Budget (DEB) approach and studies the capacity of DEB models to accurately describe the physiology and population dynamics of Southern Ocean marine organisms.

•The first study used the example of the limpet species *Nacella concinna (*Strebel, 1908). This species is known to have distinct intertidal and subtidal morphotypes that are genetically similar but differ in morphology and physiology. This species case study was used (1) to evaluate the potential of the DEB approach, and assess whether a DEB model could be built separately for the intertidal and subtidal morphotypes, based on a field experiment and data from the literature and (2) to analyse whether models were contrasting enough to reflect the two morphotypes' respective physiology and morphology.

•The second part of this chapter studied population dynamics modelling. Using an Individual Based Modelling approach (IBM), DEB models can be upscaled at the population level to simulate the response of populations to variations in food resources and temperatures. The DEB-IBM approach was applied to an endemic sea urchin of the Kerguelen Plateau, *Abatus cordatus* (Verrill, 1876) and modelled population changes through time, according to changes in food and temperature conditions, under present and future scenarios.

•A last study, presented in the appendix section, used DEB modelling for better understanding the role of low temperature and seasonal food availability conditions on the life cycle and reproduction strategy of an Antarctic bivalve, *Laternula elliptica* (King, 1832). The DEB model was also used to describe the effect of varying environmental conditions on energy allocation, using an available time-series dataset.

•**Guillaumot C**, Saucède T, Morley SA, Augustine S, Danis B and Kooijman S (2020). Can DEB models infer metabolic differences between intertidal and subtidal morphotypes of the Antarctic limpet *Nacella concinna* (Strebel, 1908)? *Ecological Modelling*. 430. 109088.

•Arnould-Pétré M, **Guillaumot C**, Danis B, Féral J-P and Saucède T (2020). Individual-based model of population dynamics in a sea urchin of the Kerguelen Plateau (Southern Ocean), *Abatus cordatus*, under changing environmental conditions. *Ecological Modelling*. 440, 109352.

•[*Appendix section*] Agüera A, Ahn I-Y, **Guillaumot C** and Danis B (2017). A Dynamic Energy Budget (DEB) model to describe *Laternula elliptica* (King, 1832) seasonal feeding and metabolism. *PLOS One*. 12(8), e0183848.

Can DEB models infer metabolic differences between intertidal and subtidal morphotypes of the Antarctic limpet *Nacella concinna* (Strebel, 1908)?

Guillaumot Charlène[1,2], Saucède Thomas[2], Morley Simon A.[3], Augustine Starrlight [4], Danis Bruno[1], Kooijman Sebastiaan[5]

1 Laboratoire de Biologie Marine, Université Libre de Bruxelles, Avenue F.D.Roosevelt, 50. CP 160/15. 1050 Bruxelles, Belgium
2 UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, EPHE, 6 bd Gabriel F-21000 Dijon, France
3 British Antarctic Survey, Natural Environment Research Council, Cambridge, CB30ET UK
4 Kvaplan-Niva, Fram High North Research Centre for Climate and the Environment, 9296 Tromsø, Norway
5 Department of Theoretical Biology, VU University Amsterdam, de Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

Ecological Modelling, 430 (2020). Accepted April 13rd 2020.

Abstract

Studying the influence of changing environmental conditions on Antarctic marine benthic invertebrates is strongly constrained by limited access to the region, which poses difficulties to performing long-term experimental studies. Ecological modelling has been increasingly used as a potential alternative to assess the impact of such changes on species distribution or physiological performance.

Among ecological models, the Dynamic Energy Budget (DEB) approach represents each individual through four energetic compartments (i.e. reserve, structure, maturation and reproduction) from which energy is allocated in contrasting proportions according to different life stages and to two forcing environmental factors (food resources and temperature).

In this study, the example of an abundant coastal limpet, *Nacella concinna* (Strebel 1908), was studied. The species is known to have intertidal and subtidal morphotypes, genetically similar but physiologically and morphologically contrasting.

The objectives of this paper are (1) to evaluate the potential of the DEB approach, and assess whether a DEB model can be separately built for the intertidal and subtidal morphotypes, based on a field experiment and data from literature and (2) to analyse whether models are contrasting enough to reflect the known physiological and morphological differences between the morphotypes.

We found only minor differences in temperature-corrected parameter values between both populations, meaning that the observed differences can be only explained by differences in environmental conditions (i.e. DEB considered variables, food resources and temperature, but also other variables not considered by DEB). Despite the known morphological difference between the populations, the difference in shape coefficients was small.

This study shows that even with the amount of data so far available in the literature, DEB models can already be applied to some Southern Ocean case studies, but, more data are required to accurately model the physiological and morphological differences between individuals.

Keywords

Ecological modelling, Southern Ocean, marine benthic species, model relevance, model accuracy

ACKNOWLEDGEMENTS
We are thankful to Philippe Pernet for *N. concinna* pictures, and Jonathan Flye-Sainte-Marie, Jean-François Cudennec for sclerochronology protocol advices, Eric Dabas, Rémi Laffont for sclerochronology lab trials.
This work was supported by a "Fonds pour la formation à la Recherche dans l'Industrie et l'Agriculture" (FRIA) and "Bourse Fondation de la mer" grants to C. Guillaumot. S.A. Morley was supported by Natural Environment Research Council core funding to the British Antarctic Survey.
This is contribution no. 40 to the vERSO project (www.versoproject.be), funded by the Belgian Science Policy Office (BELSPO, contract n°BR/132/A1/vERSO). Research was also financed by the "Refugia and Ecosystem Tolerance in the Southern Ocean" project (RECTO; BR/154/A1/RECTO) funded by the Belgian Science Policy Office (BELSPO), this study being contribution number 14.

AUTHORS' CONTRIBUTION
C. Guillaumot: Conceptualization, Methodology, Writing
T. Saucède: Supervision, Validation, review & editing
S.A. Morley: Data acquisition, Validation, review & editing
S. Augustine: Methodology, Validation, review & editing
B. Danis: Supervision, Validation, review & editing
S.A.L.M. Kooijman: Data curation, Methodology, Validation, review & editing

1. INTRODUCTION

Antarctic regions have faced strong environmental change since the twentieth century (recently reviewed in Henley et al. 2019), with a strong warming in some regions, such as in the Western Antarctic Peninsula (King et al. 2003, Vaughan et al. 2003, Meredith and King 2005), leading to important shifts in sea ice regimes and seasonality, including the duration and extent of sea ice cover (Stammerjohn et al. 2012, Turner et al. 2016, Schofield et al. 2017). The increase in the rate of glacier melting has been reported as a cause of important disturbance of the physical (currents, salinities) and biological environment (phytoplankton blooms, communities) (Meredith and King 2005, Schloss et al. 2012, Bers et al. 2013). Such changes have a direct impact on marine communities and particularly in coastal marine areas (both intertidal and subtidal)(Barnes and Peck 2008, Smale and Barnes 2008, Barnes and Souster 2011, Waller et al. 2017, Stenni et al. 2017, Gutt et al. 2018), which are places of complex land-sea interface and ecological processes. The multiple effects of ice retreat and meltwater on nearshore marine habitats have contributed to the expansion of intertidal zones and habitat alteration due to seawater freshening and stratification, shifting near-shore sedimentation, changes in water properties and current dynamics.

However, studying Antarctic marine life is challenging. Not only do the environmental conditions make the region difficult to access and work in, but substantial financial and technical constraints make field sampling and experiments difficult to organise (e.g. cold, ice, duration of daylight; Kaiser et al. 2013, Kennicutt et al. 2014, 2015, Xavier et al. 2016, Gutt et al. 2018). However, conducting physiological studies of Antarctic marine organisms has become urgent as we aim to assess their sensitivity and potential response (resilience, distribution shift or local extinction) to environmental change, a key issue for the conservation of marine life and special protected areas (Kennicutt et al. 2014, 2015, 2019 https://www.ccamlr.org/en/organisation/home-page).

An alternative to completing studies in these environments is the use of modelling approaches. Ecological modelling is used to describe species distribution and assess their climate envelopes (Elith et al. 2006, Peterson et al. 2011), study species tolerances to toxicants and to environmental change (Jager et al. 2014, Petter et al. 2014, Baas and Kooijman 2015) and model species energetic performance (Serpa et al. 2013, Thomas et al. 2016). Among these ecological models, the Dynamic Energy Budget (DEB) theory (Kooijman, 2010) has become increasingly popular. DEB parameters have been so far estimated for more than 2,000 animal species and collected in the 'Add-my-Pet' (AmP) collection (http://www.bio.vu.nl/thb/deb/deblab/add_my_pet/). It constitutes one of the most powerful approaches to characterise individual metabolic performances (Nisbet et al. 2012, Kearney et al. 2015, Jusup et al. 2017) and can be calibrated for data-poor animals (Mariño et al. 2019). DEB models rely on thermodynamic concepts (Jusup et al. 2017) and study how energy flows are driven within individuals during their entire life cycle (Kooijman 2010). Each individual is divided into four energetic compartments: reserve E, structure V, maturation E_H and reproduction E_R from which the energy is allocated in contrasting proportions according to the different life stages and two forcing environmental factors (i.e. food resources and temperature).
DEB models can be built with data coming from experiments and/or literature, to quantify age, length, weight of the different life stages and provide information on reproduction, growth and metabolic rates to calibrate the model (van der Meer 2006, Marques et al. 2014).

Application of DEB models to Antarctic species is increasing. They can be easily extracted from the AmP collection, using the software AmPtool. The Matlab command "select_eco('ecozone', {'MS'})" presently gives a list of 37 species, where MS stands for "Marine, Southern Ocean". Command "select_eco('ecozone', {'TS'})" gives another 3 species for the terrestrial Antarctic environment, among which the mite *Alaskozetes antarcticus*. Among the most common and well studied Southern Ocean benthic invertebrates are the sea star *Odontaster validus* (Agüera et al. 2015), the bivalve *Laternula elliptica* (Agüera et al. 2017), the bivalve *Adamussium colbecki* (Guillaumot 2019a) and the sea urchins *Sterechinus neumayeri* (Stainthorp and Kooijman 2017) and *Abatus cordatus* (Arnould-Pétré et al. 2020 - Chapter 1). DEB models have also been developed for some pelagic species such as the Antarctic krill *Euphausia superba*, the salp *Salpa thompsoni* (Jager and Ravagnan 2015, Henschke et al. 2018) and are under development for marine mammals such as the elephant seal *Mirounga leonina* (Goedegebuure et al. 2018).

Antarctic species have a range of notable physiological traits when compared to their temperate counterparts. Among others, they are physiologically adapted to constant cold temperatures (Peck et al. 2009b, Morley et al. 2009b, 2014), shifting day length also imposes a marked seasonal feeding behaviour (McClintock 1994, Clarke et al. 2008, Halanych and Mahon 2018), and they exhibit slow metabolic and growth rates, explaining their longer lifespans and higher longevities compared to species in other regions (Peck and Brey 1996, Peck 2002).

The limpet *Nacella concinn*a (Strebel, 1908) (Mollusca: Patellogastropoda) is a common and abundant gastropod of shallow marine benthic communities. Distributed all along the Western Antarctic Peninsula (González-Wevar et al. 2011, phylogeny recently reviewed in González-Wevar et al. 2018b), it has widely been studied for decades (Shabica 1971, 1976, Walker 1972, Hargens and Shabica 1973, Houlihan and Allan 1982, Peck 1989, Clarke 1989, Cadée 1999, Ansaldo et al. 2007, Fraser et al. 2007, Markowska and Kidawa 2007, Morley et al. 2011, 2014, Suda et al. 2015, Souster et al. 2018). The limpet is found from intertidal rocky shores down to over 100 meters depth (Powell 1951, Walker 1972). It has a 2-5 cm long shell (Fig. 1.1), that grows only a few millimeters a year with a seasonal pattern. It is sexually mature after four to six years and has a life span of up to 10 years (Shabica 1976, Picken 1980, Brêthes et al. 1994). The limpet mainly feeds on microphytobenthos and microalgae (Shabica 1976, Brêthes et al. 1994). It spawns free-swimming planktonic larvae once a year, when water temperature rises in the austral summer (Shabica 1971, Picken 1980, Picken and Allan 1983). Larvae drift in the water column and metamorphose after more than two months (Stanwell-Smith and Clarke 1998).
N. concinna does not have a homing behaviour (Stanwell-Smith and Clarke 1998, Weihe and Abele 2008, Suda et al. 2015) and intertidal individuals can either migrate to subtidal areas in winter to escape freezing air temperatures that may drop below -20°C (Walker 1972, Branch 1981, Brêthes et al. 1994) or shelter in rock cracks and crevices in the intertidal area. In the latter case, they do not become dormant but have a limited access to microphytobenthos, as recently observed around Adelaide Island (Obermüller et al. 2011).

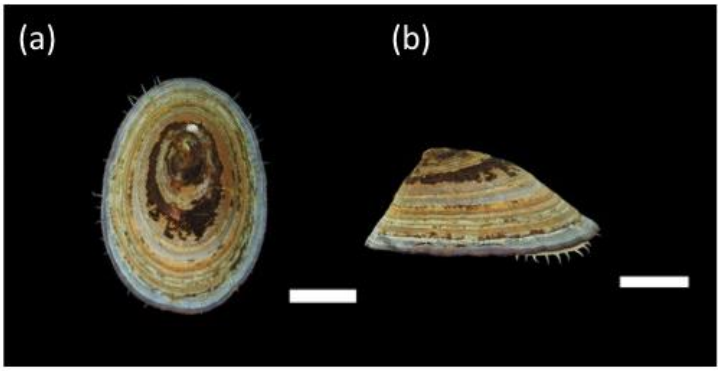


Figure 1.1. *Nacella concinna* in apical view (a) and lateral view (b). Scale bar: 1 cm. Source: Q. Jossart, B121 expedition.

Two morphotypes of *N. concinna* have been distinguished, an intertidal and a subtidal type, with the intertidal type having a taller, heavier and thicker shell compared to the subtidal one that is characterised by a lighter and flatter shell (Beaumont and Wei 1991, Hoffman et al. 2010). Initially, Strebel (1908) and Powell (1951) referred to these two morphotypes as the 'polaris' (intertidal) and 'concinna' types (below 4m depth). From that point, the potential genetic differentiation between the two morphotypes has been investigated, some of the studies concluding an absence of genetic distinction (Wei 1988, Beaumont and Wei 1991, Nolan 1991) while contrarily, de Aranzamendi et al. (2008) reported significant differences based on inter-simple sequence repeat (ISSR) markers. More recently, this last method was questioned (Hoffman et al. 2010) and several studies using different markers and populations (Chwedorzewska et al. 2010, Hoffman et al. 2010, Gonzalez-Wevar et al. 2011) have concluded an absence of genetic differentiation between the two morphotypes.

Apart from the absence of genetic differences, intertidal and subtidal populations strongly contrast in morphology and physiology, which has been explained by the prevalence of habitat

heterogeneity and strong environmental gradients along rocky shore habitats, a common feature also observed in other gastropods (Johannesson 2003, Butlin et al. 2008, Hoffman et al. 2010). For instance, in *N. concinna*, the higher shell thickness observed in the shallow morphotype was hypothesised to play a role in resistance against crushing pack ice (Shabica 1971, Morley et al. 2010). Intertidal morphotypes are further resistant to air exposure thanks to higher shells, bigger inner volumes relative to their shell circumference, a combination that makes them more efficient than subtidal individuals, able to store more water and oxygen, reducing desiccation risks and delaying the metabolic switch to anaerobic fermentation (Nolan 1991, Weihe and Abele 2008). The subtidal morphotype has also proved to be less resistant to cold than the intertidal population (Waller et al. 2006), due to extra production of mucus and stress proteins in intertidal morphotypes (Clark et al. 2008, Clark and Peck 2009, Obermüller et al. 2011) and due to diverse metabolic processes that contrast between both populations (reviewed in Suda et al. 2015).

The development of ecological models enables precise models to be built, that highlight subtle differences in parameters between ecologically similar or closely related species (Freitas et al. 2010, Holsman et al. 2016, Marn et al. 2019, Lika et al. 2020). The idea of building individual-specific models for understanding of physiological processes is not new (Bevelhimer et al. 1985, DeAngelis et al. 1994) and grew from the development of computational ecology that resulted in the possibility of generating "individual-oriented" models (IOM's) (Hogeweg and Hesper 1990, DeAngelis et al. 1994). The IOM theory relies on the principle that "no two biological organisms are exactly alike, even when they have identical genes". A group of organisms within a population can have contrasting size or physiological performances according to, for example, food conditions or competition. Modelling each individual, separately, therefore constitutes a powerful approach to enhance the understanding of the entire community (DeAngelis et al. 1994).

In this study, due to the known morphological and physiological differences between the morphotypes, we first separately built independent DEB models for the intertidal and subtidal morphotypes of the limpet *N. concinna,* based on field experiment and literature data, to assess the potential differences between the models. Secondly, we analyse whether the two model outputs suggest contrasting physiologies between the morphotypes, using a method recently developed in DEB theory, that tries to reduce differences in parameter values that are still consistent with the data (Lika et al. 2020). Using this method -the augmented loss function- we try to merge the information of the two species models into a single one. If DEB parameters of the two species can be merged, it means that the physiological differences between these two species are not strongly different.

These results finally help assess DEB model accuracy giving the amount of data available to build the models in the context of Antarctic case studies and help evaluate which type of information is necessary to gather in order to fill model gaps. Finally, the study evaluates if such models are valuable for studying Southern Ocean organisms in the context of altered environments.

2. MATERIAL AND METHODS

2.1. DEB Model description

DEB models are based on an ensemble of rules that allocate energy flows to four main compartments (reserve E, structure V, maturity E_H, reproduction E_R) according to a set of priorities and the level of complexity (i.e. maturity) gained by the organism through time (Fig. 1.2, Kooijman 2010). Maturity is treated as information, having mass nor energy. Food is first of all ingested and assimilated ($\dot{p}_A$) and energetically stored into a reserve compartment (E). A fraction of the energy that is mobilised from reserve, $\dot{p}_C$, is divided into two branches according to the 'kappa-rule': a part of the energy contained in the reserve compartment ($\kappa . \dot{p}_C$) is allocated to somatic maintenance and structure growth, whereas the second part (1- κ). $\dot{p}_C$ contributes to maturity (before the 'puberty' threshold) or reproduction (after the 'puberty' threshold).
The energy is allocated within and in between these branches by the establishment of some priorities, where somatic maintenance ($\dot{p}_M$) has priority over growth and maturity maintenance ($\dot{p}_J$) has priority over maturity and reproduction. During its lifetime, the organism allocates energy to

maturity which symbolizes its complexity and reaches some life stages at some defined thresholds (E_H^b, birth, when the organism is capable to feed; E_H^j, metamorphosis; E_H^p, puberty, when it can reproduce). After reaching sexual maturity, the energy that was formerly allocated to maturity is attributed to the reproduction buffer and the available energy is allocated to the development of gametes.
Different types of DEB models have been developed and coded for parameter estimation, see frequently updated https://github.com/add-my-pet/DEBtool_M page (Marques et al. 2018, 2019). Here, the *abj* model was used for *N. concinna*. This model considers that growth acceleration occurs between birth and metamorphosis (Kooijman 2010, Mariño et al. 2019).

The DEB model is forced by food availability and temperature. Temperature acts on metabolic rates following the Arrhenius principle (see Kooijman 2010, Jusup et al. 2017 for details). A temperature correction factor is applied to each rate that takes into account the lower and higher optimal boundaries of the individual tolerance range. Food available for ingestion is represented by the functional response *f* comprised between [0,1], where 0 is starvation condition and 1 very abundant food.

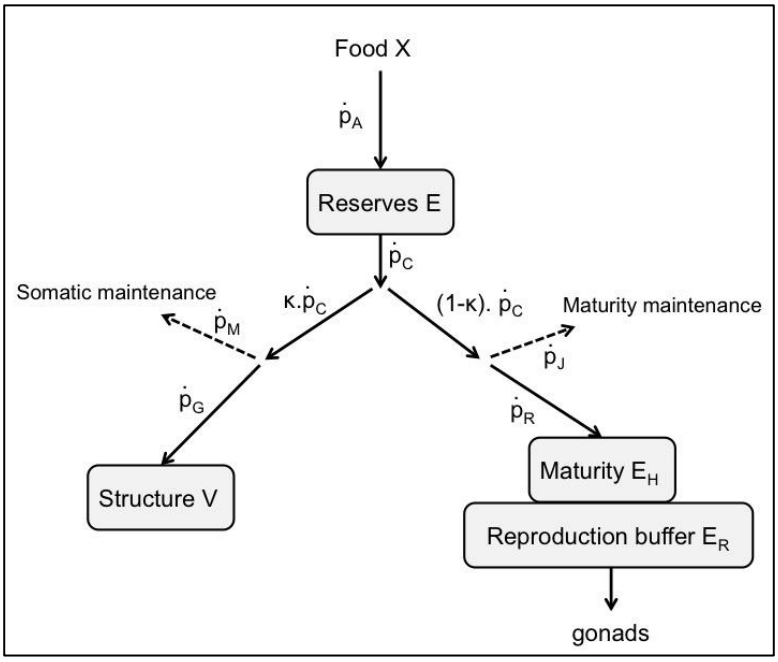


Figure 1.2. Schematic representation of the standard DEB model, with energy fluxes (arrows, in J.d^{-1}) that connect the four compartments (boxes). Energy enters the organism as food (X), is assimilated at a rate of $\dot{p}_A$ into the reserve compartment (E). The mobilization rate $\dot{p}_C$, regulates the energy leaving the reserve to cover somatic maintenance $\dot{p}_M$, structural growth $\dot{p}_G$, maturity maintenance $\dot{p}_J$, maturity $\dot{p}_R$ (sexually immature individuals) and reproduction $\dot{p}_R$ (mature individuals). κ. $\dot{p}_C$ is the proportion of the mobilized energy diverted to $\dot{p}_M$ and $\dot{p}_G$, while the remaining part (1- κ). $\dot{p}_C$ is used for $\dot{p}_J$ and $\dot{p}_R$.

The parameters of the DEB model can be estimated from multiple data on the eco-physiology of a species. The ones studied in this work are presented in Table 1.1.

Table 1.1. List of the main DEB parameters, definition and units.

Parameters	Description	Units
Primary DEB parameters		
$\{\dot{p}_{Am}\}$	surface-area-specific maximum assimilation rate	J.cm^{-2}d^{-1}
$\dot{v}$	energy conductance (velocity)	cm.d^{-1}
κ	fraction of mobilised reserve allocated to soma	-
$[\dot{p}_M]$	specific volume-linked somatic maintenance rate: $\dot{p}_M$ / V	J.cm^{-3}d^{-1}
$[E_G]$	volume-specic costs of structure; better replaced by $[E_V]= \kappa_G$, where κ_G is the fraction of growth energy fixed in structure: $[E_V]=[E_G]$	J.cm^{-3}
E_H^b	maturity at birth	J
E_H^j	maturity at metamorphosis	J
E_H^p	maturity at puberty	J
$\ddot{h}_a$	Weibull ageing acceleration for animals	d^{-2}
s_G	Gompertz stress coefficient	-
δ_M	shape (morph) coefficient: L=Lw	-
δ_{M_larvae}	shape (morph) coefficient of the larvae	-
Other parameters		
z	zoom factor to compare body sizes inter-specifically; *z* = 1 for L_m = 1 cm	-
s_M	Acceleration factor at *f* =1, it is equal to the ratio of structural length at metamorphosis and birth.	-
$[E_m]$	$[E_m]= \{\dot{p}_{Am}\}/\dot{v}$; ratio of specific assimilation over energy conductance	J.cm^{-3}

2.2. Data collection and DEB calibration

DEB models were calibrated using zero-variate data (single data points at defined life stages, such as length or weight at sexual maturity, number of eggs produced per female) and uni-variate data (relationships between two variables such as oxygen consumption and temperature, length~weight relationship, weight or size~time relationships)(van der Meer 2006, Guillaumot 2019b). Data that were collected from the literature (Table 1.2), paying attention to the different taxonomic names adopted for the species through time (see http://www.marinespecies.org/aphia.php?p=taxdetails&id=197296, accessed December 2018); to the sampling area to enable the two morphotypes to be distinguished (intertidal/subtidal) and to the environmental conditions under which each dataset was recorded (available food resources and temperature).

Data from the literature were supplemented by experiments led by S. Morley at Rothera Station (Adelaïde Island, Western Antarctic Peninsula) in January-February 2018 (details in Appendix 1.1). Individual shells were brought back to Europe and processed with imagery to collect growth ring data (Appendix 1.2).

Some data are shared between the intertidal and subtidal morphotypes due to a lack of information on the morphotypes physiology in the literature (Table 1.2). The characteristics of the first developmental life stage, when the larvae become able to feed (i.e. age, length and weight at birth) and the pace of development (i.e. age at puberty, maximal observed age) are assumed to be identical.

Each data set was characterised by the corresponding temperature and food resources present in the field. Food resources were represented in the model by a scaled functional response *f* constrained between 0 and 1, with 0 meaning no food availability and 1 maximal food abundance. *f* parameters were differentiated between the different stations along with temperatures. Food is very abundant in the field for the limpet and *f* parameters were therefore kept fixed with values ≥ 0.9. Food availability from the Rothera Station was described by pictures taken in the field and was estimated at *f*=1. Signy and Anvers Islands *f* was set at 0.9 because physiological traits (growth rate, maximal size) are very close (but slightly lower) than Rothera's observations, but no precise information is available for food conditions in the different publications for these stations.

Table 1.2. Zero and uni-variate data used to build the intertidal and subtidal models. AFDW stands for 'Ash Free Dry Weight'.

INTERTIDAL GROUP			**SUBTIDAL GROUP**	
Zero-variate data, (unit)	**Value**	**Reference**	**Value**	**Reference**
Age at birth *ab (days)*	10	Peck et al. (2016)	Same as intertidal	
Age at puberty *ap (years)*	4	Shabica (1976)	Same as intertidal	
Maximal observed age *am (years)*	14	Shabica (1976)	Same as intertidal	
Length at birth *Lb (cm)*	0.0228	Peck et al. (2016)	Same as intertidal	
Length at puberty *Lp (cm)*	1.54	S. Morley experiment (2018)	1.59	Picken (1980)
Maximal observed shell length *Li (cm)*	5.8*	Shabica (1976)	5.52**	S. Morley experiment (2018)
Wet weight of the egg *Ww0 (g)*	$5.8.10^{-6}$***	Peck et al. (2016)	Same as intertidal	
AFDW at puberty *Wdp (g)*	0.0236	Shabica (1976)	0.057	S. Morley experiment (2018)
Uni-variate data, (unit)		**Reference**		**Reference**
Length ~ AFDW *LWd_signy (cm, g)*		Nolan (1991), Signy Island		Nolan (1991), Signy Island
Length ~ AFDW *LWd (cm, g)*		S. Morley experiment (2018)		S. Morley experiment (2018)
Length ~ Gonado somatic index *LGSI (cm, -)*		S. Morley experiment (2018)		S. Morley experiment (2018)
Length ~ Oxygen consumption *LJO (cm, µmol/h)*		S. Morley experiment (2018)		S. Morley experiment (2018)
Temperature ~ Oxygen consumption *TJO (K, µL/h)*		Peck (1989)		Peck (1989)
Time ~ Length *tL (d, cm)*		S. Morley experiment (2018)****		S. Morley experiment (2018)****

* Max sized collected individual on the field during Belgica121 expedition (Danis et al. 2019)
** Shabica (1976) indicates an observed value of 5.8 cm and S. Morley measurements indicate a ratio between intertidal/subtidal lengths of the morphotypes of 1.05**.** The unknown subtidal Li value was calculated as 5.8/1.05= 5.52 cm.
*** based on egg diameter of 221 µm
**** imagery and growth ring measurements, see Appendix 1.2

2.3. DEB parameter estimation and goodness of fit

Sets of zero and uni-variate data, supplemented by pseudo-data were used to estimate the DEB primary parameters. Pseudo-data are extra data coming from different taxa that help calibrate the model estimation similarly to a prior element (Lika et al. 2011a). This procedure has similarities with Bayesian estimation, but are not embedded in a maximum likelihood context, since the

stochastic component is not modelled. Before parameter estimation, each data set can be subjectively linked by a weight coefficient to quantify the realism of reducing variation in parameter values. Selected weight coefficients are always selected small enough in order to hardly affect parameter estimation if the information contained in the real data set is sufficient.
The DEB parameters estimation is done by simultaneously estimating each parameter using these empirical and pseudo-data by minimizing a loss function, using the Nelder-Mead simplex method, updated and explained in Marques et al. (2018, 2019). The loss function that is minimized is

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{w_{ij}}{n_j} \frac{(d_{ij}-p_{ij})2}{\overline{d_i}^2+\overline{p_i}^2}$$

where i scans datasets and j points in this dataset. d_{ij} and p_{ij} are respectively the data and the predictions and $\overline{d}_i$ and $\overline{p}_i$ their average values in set i. w_{ij} are the attributed coefficients, n is the number of data sets, n_i denotes the data in a dataset, n_j the data in data-points.

The value of the loss function is evaluated for each parameter trial. The goodness of fit of each prediction was quantified by the relative error (RE). The mean relative error (MRE) quantifies the overall model performance. RE corresponds to the sum of the absolute differences between observed and predicted values, divided by the predicted values. Contrarily to the loss function, the MRE does not take into consideration the weights of the different data (Marques et al. 2018). MRE values can have values from 0 to infinity, with 0 value meaning that predictions match data exactly.

2.4. Merging parameters

The augmented loss function approach developed by Lika et al. (2020) is a new extension that enables to compare small variations in parameter values between (close) species. The second term (in bold) of the following equation is the new extension of the 'symmetric bounded (sb)' loss function:

$$F_{sb} = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{w_{ij}}{n_j} \frac{(d_{ij}-p_{ij})2}{\overline{d_i}^2+\overline{p_i}^2} + \boldsymbol{\sum_{k=1}^{N} \frac{w_k\, var(\theta_k)}{mean(\theta_k)^2}}$$

where **w**'s are weights, **d**'s data, **p**'s predictions, θ's parameters, **j** scans data-points with a data-set of $\mathbf{n}_i$ points ($\mathbf{n}_i$ = 1 is allowed), **i** scans the data-sets and **k** the parameters.
In this second term, when w_k=0, the parameter θ_k between species are different, but when increasing w_k, the parameter θ_k tends to be similar between species. Therefore, the augmented loss function method uses this mathematical principle to spot potential differences between parameters of different species. First, the set of DEB parameters are separately estimated for each species and weight coefficients are set to zero. Then, for each parameter, the weight coefficient will be step-wise increased, making the loss function shift as a result. If a maximal weight value is reached without sharp changes in the loss function value along the weight increase, it means that the parameter value has a minimum variance between species. Contrarily, if the loss function value presents a sharp increase due to the change in weight coefficient, it means that the studied parameter should present contrasting values between the related species.

By applying this method to the case study of an intertidal and subtidal morphotype of the limpet *N. concinna*, we aim to evaluate whether there are any differences between both morphotypes caused by differences in parameters, or whether these differences are explained only by differences in environmental conditions (i.e. food resources and temperature). Initially, the sets of parameters have been estimated separately for both morphotypes and all weight coefficients are set to zero. By step-wise increasing the weight coefficient for a particular shared parameter, the overall loss function may increase and a common merged DEB parameter is reached. If a common value of the DEB parameter can be found without important increase in MRE or loss function values, it means that the intertidal and subtidal morphotypes do not significantly differ for this parameter. A similar procedure is applied for each DEB parameter separately and iteratively. In order to have a quick idea of replicability in the results, the procedure was replicated five times,

contrasting in different orders of DEB parameters merging (Appendix 1.3). The order of permutation of merged parameters of these five replicates was chosen randomly among the 11! possible solutions. Changes in MRE and loss function values at each weight modification were reported and the predictions of the intertidal, subtidal and merged models were compared.

3. RESULTS

3.1. Parameters of DEB models

DEB predictions for the separate intertidal and subtidal models are accurate, with MRE values lower than 0.2 (Table 1.3). Average MRE value of the AmP collection is close to 0.06. Relative Errors are quite low, with the highest values obtained for length~GSI data (RE= 0.6089 and 0.8702 for intertidal and subtidal models respectively) and time~length relationships, obtained from the sclerochronology measurements, that are highly variable between each measured shell (respectively RE= 0.3645 and 0.5924 for intertidal and subtidal models) (Fig. 1.3, Table 1.3, Appendix 1.2).

In view of the substantial morphological difference between the populations, we expected to see a clear difference in the shape coefficient δ_M. We found a slightly larger value of δ_M for the intertidal morphotype, meaning that for the same shell length, it has slightly more structure, compared to the subtidal one.

Subtidal morphs have a lower energy conductance $\dot{v}$ as well as double the value of maximum surface area specific assimilation rate $\{\dot{p}_{Am}\}$ with respect to the intertidal morphs. The ratio of specific assimilation over energy conductance $[E_m]= \{\dot{p}_{Am}\}/\dot{v}$, determines the maximum reserve capacity of a species.
The fraction of mobilised reserve allocated to soma κ is also bigger (0.9368 for subtidal vs. 0.9084 for intertidal type), and the intertidal individuals also present a lower value for somatic maintenance rate $[\dot{p}_M]$ compared to the subtidal ones. This highlights contrasts between the morphotypes in energy allocated to maturation along the first life stages (E_H^b, E_H^j) and more available energy for growth for the intertidal morphotype that has lower values of somatic maintenance. Intertidal morphotypes seem to accelerate metabolism with a two-fold difference in acceleration factor s_M between intertidal and subtidal types (respectively 7.862 and 4.049). The maturity threshold to reach puberty, E_H^p is also lower for the intertidal morphotype than the subtidal.

The MRE values of the merged models stay below 0.25 and the value of the loss function for the merged situation is only a little larger than the sum of both populations, reflecting that a substantial reduction in the total number of parameters by almost a factor 2 hardly affects the goodness of fit (Table 1.3, Appendix 1.3).
DEB parameters of the merged models are quite close to the values of the intertidal and subtidal models, with $[\dot{p}_M]$, δ_M, E_H^p, E_H^b, $[E_G]$ and $\dot{v}$ merged values being almost exactly in between the values of the intertidal and the subtidal morphotypes. Parameters κ, z, E_H^j, $\{\dot{p}_{Am}\}$ and $\ddot{h}_a$ are closer to the intertidal predictions.
Univariate predictions are also extremely close between the two models and the merged model (Fig. 1.3), with only a small difference for the subtidal model for which the GSI~length predictions are higher than the intertidal and merged predictions, mainly due to errors in predictions and scatter in the data. This higher potential of energy allocation to reproduction can, however, be linked to the higher E_H^p values estimated for the subtidal type (Table 1.3).

Table 1.3. Summary of goodness of fit, DEB model estimates at a reference temperature of T_{ref}= 20°C. RE: Observed and predicted values for zero-variate data, relative error (RE) for the uni-variate data. See Fig. 1.3 for comparisons for uni-variate predictions between models. MRE= Mean Relative Error. For the merged model, the MRE values respectively correspond to the mean relative error of model prediction for data of both intertidal and subtidal populations. All DEB parameters indicated were allowed to vary during covariance estimation. The *abj* parameters that are not mentioned in that table were kept constant with the standard initial values.

	INTERTIDAL	**SUBTIDAL**	**MERGED**
MRE	0.166	0.192	0.196 0.227
Loss function	0.2441	0.2345	0.7936
Parameters			
z (-)	0.3055	0.4317	0.2579
$\{\dot{p}_{Am}\}$ (J/d. cm^{-2})	8.361	19.07	8.859
$\dot{v}$ (cm/d)	0.0501	0.0426	0.0499
K (-)	0.9084	0.9368	0.9256
$[\dot{p}_M]$ (J/d. cm^{-3})	19.62	31.68	24.62
[E_G] (J. cm^{-3})	3956	3949	3952
E_H^b (J)	0.00174	0.00115	0.0014
E_H^j (J)	0.8749	0.0779	0.9206
E_H^p (J)	75.23	121.4	94.66
$\ddot{h}_a$ (1/d^{-2})	5.003.10^{-8}	8.335.10^{-8}	4.24.10^{-8}
s_G (-)	10^{-4}	10^{-4}	10^{-4}
δ_M (-)	0.4517	0.3866	0.4247
δ_{M_larvae} (-)	0.7167	0.7125	0.7215
s_M (-)	7.862	4.0491	8.5372
Zero-variate			
	Data // prediction// RE	**Data // prediction// RE**	**prediction// RE**
ab (d)	10 10.62 0.0619	10 10.59 0.0586	10.61 0.0609
ap (y)	4 3.54 0.1141	4 3.75 0.0607	3.66 0.0845
am (y)	14 14 9.4.10^{-5}	14 13.99 4.8.10^{-4}	14 1.6.10^{-4}
Lb (cm)	0.0228 0.02279 2.4.10^{-4}	0.0228 0.0228 6.05.10^{-4}	0.0228 1.424.10^{-6}
Lp (cm)	1.54 1.225 0.2045	1.59 1.81 0.1384	1.49 0.0323
Li (cm)	6.5 5.319 0.1816	5.52 4.515 0.1827	5.184 0.2024
Ww0 (g)	5.8.10^{-6} 5.8.10^{-6} 0.0181	5.8.10^{-6} 5.7.10^{-6} 0.0157	5.72.10^{-6} 0.0138
Wdp (g)	0.0236 0.0263 0.1181	0.057 0.05649 0.0089	0.0396 0.6762

Uni-variate			
	RE	**RE**	**RE**
LWd_signy (cm, g)	0.1443	0.1698	0.1274
LWd (cm, g)	0.1469	0.1834	0.216
LGSI (cm, -)	0.6089	0.8702	0.5835
LJO (cm, µmol/h)	0.2567	0.2831	0.2487
TJO (K, µL/h)	0.1034	0.1216	0.0876
tL (d, cm)	0.3645	0.5924	0.4097

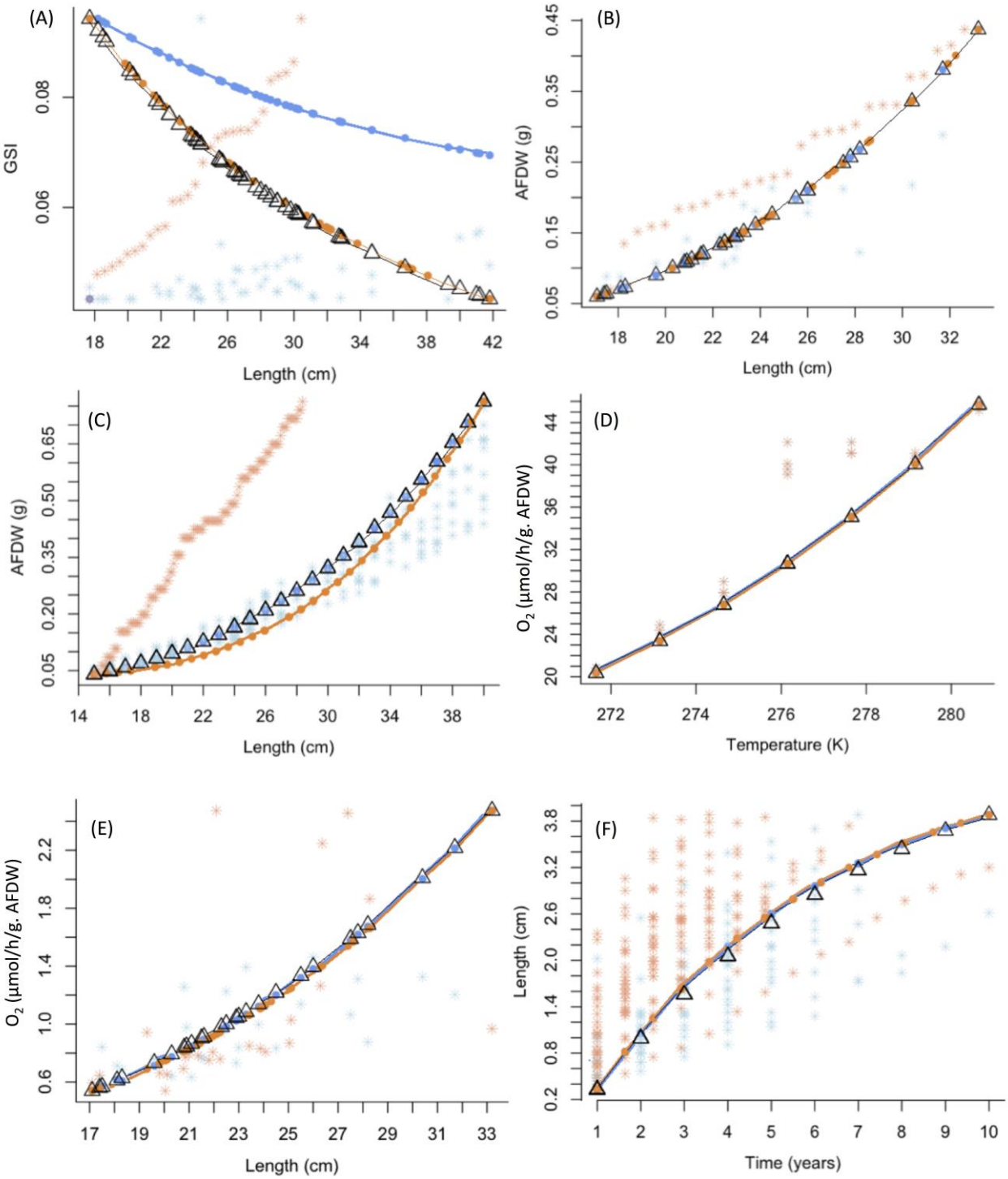
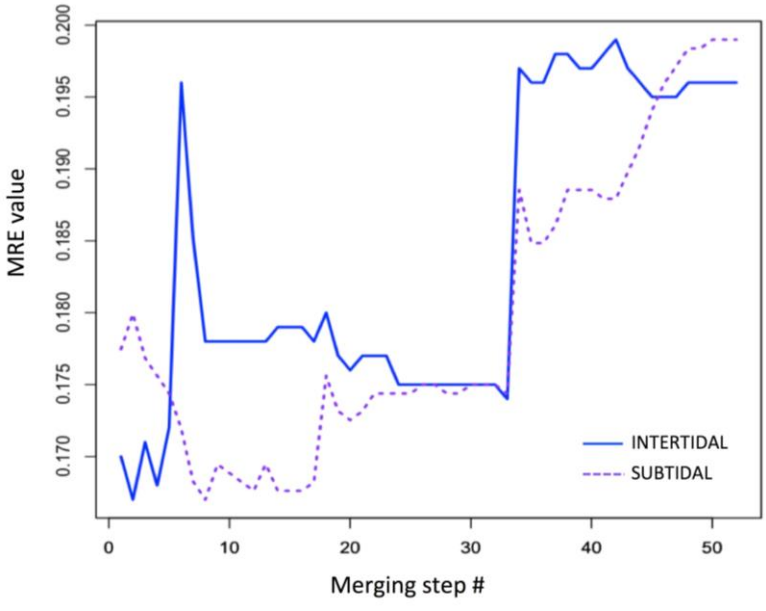


Figure 1.3. Comparison of model predictions (uni-variate data). Blue dots joined by lines: subtidal model predictions, blue stars: subtidal data (observations); orange dots joined by lines: intertidal model predictions, orange stars: intertidal data (observations); black triangle joined by lines: merged model predictions. Prediction points may overlap (D).

3.2. Merging process

Along the merging procedure, the loss function and MRE values of the model at each step of the merging procedure are observed, one 'step' corresponding to the interative increase of the weight coefficient of the studied parameter (i.e. merging step, Fig. 1.4). Changes in MRE values are not that important between the initial step and the final step of the merging procedure (Fig. 1.4, Appendix 1.4) (respectively from 0.170 to 0.196 and from 0.192 to 0.227 for the MRE intertidal and MRE subtidal values), meaning that merging parameters is possible. E_H^p and δ_M seem to be the parameters that are the most influencing the model during the merging procedure for both the intertidal and subtidal models and $[\dot{p}_M]$ seems to further influence the intertidal model.



Merged parameter	Merging step #	Merged parameter	Merging step #
z	2-4	E_H^b	26
$[\dot{p}M]$	5-12	$[E_G]$	27-29
δ_{M_larvae}	13-15	E_H^j	30-32
κ	16-20	E_H^p	33-41
$\ddot{h}_a$	21-22	δ_M	42-52
$\dot{v}$	23-25		

Figure 1.4. Evolution of Mean Relative Error (MRE) values along the merging of the different parameters.

MRE intertidal in solid blue line, MRE subtidal in dashed purple line. Example of Trial #5 (merging of z, $[\dot{p}_M]$, δ_{M_larvae}, κ, $\ddot{h}_a$, $\dot{v}$, E_H^b, $[E_G]$, E_H^j, E_H^p, δ_M).

4. DISCUSSION

4.1. DEB models relevance

DEB models are powerful tools enabling predictions of the individuals energetic scope for survival, growth and reproduction, given the considered environmental conditions (Kooijman 2010, Jusup et al. 2017). These mechanistic approaches have been of interest for several years to the marine Antarctic community (Gutt et al. 2012, Constable et al. 2014a, Gutt et al. 2018), and have been increasingly developed during recent years (e.g. Agüera et al. 2015, 2017, Goedegebuure et al. 2018, Henschke et al. 2018).

This study is based on the example of the limpet *Nacella concinna* and uses data from literature supplemented by experiments conducted in Antarctica in February 2018, to build the DEB models of the intertidal and subtidal morphotypes of the species. The separately produced models were accurate, with a reduced error between observations and model predictions, except for some scatter among data such as Length~GSI relationship. Such accuracy was mainly possible thanks to the important amount of uni-variate data that were provided by the complementary experiments conducted in Rothera, which filled knowledge gaps about reproduction, collected more precise length weight relationships to observe the morphological contrasts between intertidal and subtidal individuals and collected more precise information on the limpet's metabolic performance through its development.

Rates (and ages) depend on temperature. Here we correct for differences in temperature using an Arrhenius relationship. However, in order to meaningfully compare differences in parameters between species living in different habitats, it is useful to standardize all parameters to a common reference temperature: T_{ref}= 20°C. This is the standard for presenting and comparing DEB parameters across the 2,000 different species in AmP. When comparing DEB parameters estimations of *N. concinna* to those of their temperate counterpart *Patella vulgata* (Kooijman et al. 2017) at T_{ref}= 20°C, we notice clear differences between the species in term of metabolic strategies, although the limpets morphology and therefore size and volume are close between the two species (close length and predicted shape coefficient δ_M). For *N. concinna,* predicted κ is much higher and close to 1 (0.9256 *vs* 0.617 for *P. vulgata*), meaning that almost all the energy available in the reserve compartment is allocated to somatic maintenance and growth, and only a small amount is available for reproduction. This is clearly visible with the ultimate rate of reproduction more than 40 times lower for the Antarctic limpet compared to the temperate one. The capacity to assimilate resources $\{\dot{p}_{Am}\}$ was estimated to be 10 times higher for *P. vulgata*, explaining the 2.5-fold lower growth rate for *N. concinna*. The two metabolisms also contrast by the fact that *P. vulgata* is predicted to store more reserves than *N. concinna* in similarly abundant food conditions. These results are consistent with published experiments, where it was shown that rasping rates (i.e. feeding potential) were higher for temperate and tropical species than for *N. concinna* (Morley et al. 2014) and that development rates of Antarctic marine molluscs are much slower than at higher temperatures (Peck et al. 2007, 2016), which could be partially due to the increased costs of protein production in the cold (Marsh et al. 2001, Robertson et al. 2001, Pörtner et al. 2007).
Such examples of comparison of energetic performance between these two species highlight the performance of DEB models to be efficiently applied for Antarctic case studies and powerful and accurate enough to enhance physiological contrasts even between closely related species; as previously discussed in other works (van der Veer et al. 2006, Gatti et al. 2017, Marques et al. 2018, Marn et al. 2019).

4.2. Comparison between morphotypes

In a second step, we evaluated if known contrasts in physiological traits between the morphotypes could be highlighted by the modelling approach. By simply comparing the two single models, we observed minimal energetic contrasts between the intertidal and subtidal morphotypes (small differences in assimilation rate and ability to store reserves, Table 1.3, Fig. 1.3). By using the augmented-loss-function method, we tried to merge the models into a single one, parameter by parameter, to evaluate the contrasts in parameters between the types (Lika et al. 2020). Results show that models were merged without generating significant changes in MRE and loss function values (Fig. 1.4, Appendix 1.4). Predictions of the uni-variate data are really similar between the three models (Fig. 1.3), with only minor differences in temperature-corrected parameter values between both populations, meaning that the observed differences are best explained by differences in environmental conditions (temperature and food availability).
Despite the known physiological contrasts in the field, the available data did not allow the models to capture these physiological differences between the morphotypes, using only the available data. Scatter distribution of the data used to calibrate the model (Fig. 1.3) can hide metabolic differences, which calls for more experiments to describe the physiology of the different morphotypes. Using more complete datasets, for which all parameters are independent between intertidal and subtidal morphotypes, may also help to further constrain the differences. In our case study, several zero-variate data are shared between the intertidal and subtidal models, among which age, length and weight at birth, that control the very beginning of the development. The observed results of a two-fold difference in metabolism acceleration of intertidal morphotypes compared to subtidal ones ($s_M \approx 8 \text{ and } 4$ for intertidal and subtidal) is in fact an artefact caused by common parameters related to birth and puberty stages (age, length, weight). Indeed, specific assimilation at birth for the subtidal is two times larger than that for the intertidal, which indicates that subtidal individuals develop faster. However, according to available data provided in the model, puberty is reached at the same time for both types. E_H^p consequently needs to be smaller for the intertidal type to reach puberty at the same age a_p and length L_p, explaining the observed contrasts between the intertidal and subtidal groups.

Improving the completeness of these models would therefore be necessary to enable further detailed conclusions.
A common approach in biology is to focus on differences between individuals, populations and species. Here we adopt a contrasting strategy in which we force models to determine in what manner the populations are similar in order to quantify in what manner they differ. This work is a first step to compare the energetics of both populations, and we discovered how (given the data) they seem more metabolically similar than what their appearance would suggest as first. We also highlight some artefacts that come from the quality of the data and the scatter therein. New data (so new knowledge) that fill current knowledge gaps will yield further insight into how the metabolisms of these populations have diverged to adapt to differences in environment. The current work is a contribution to understanding the relationship between observations (data) and metabolism for these two populations.

4.3. Models drawbacks and improvements

Apart from data availability, a drawback of our model construction is the lack of information about environmental properties that makes comparisons between estimations of the two morphotypes quite difficult to perform. In the models, we just considered an average temperature for intertidal or subtidal habitats from where the limpets come from, but do not add any supplementary detail on environmental contrasts between these habitats nor in the difference of food availability between the morphotypes. However, contrasting environmental pressures (desiccation, salinity, hydrodynamism) and habitat characteristics (immersion time, substratum type, and surrounding physico-chemical factors) contribute to contrasting adaptative strategies among which morphological adaptation is really important for limpets, but have not been integrated into our DEB models (because it requires more data we do not have) (Vermeij 1973, Branch 1981, Denny & Blanchette 2000, Sa Pinto et al. 2008, Bouzaza and Mezali 2013, Grandfils 1982, Gray & Hodgson 2003, Espinosa et al. 2009). Desiccation is one of the strongest hypothesis to explain the morphological differences between the intertidal and subtidal morphotypes (Mauro et al. 2003, Bouzaza & Mezali 2018). The presence of high upstream shifted apex form for the intertidal morphotypes, more exposed to desiccation, could help to store more water and absorb more oxygen, as described for *Patella ferruginea* (Branch 1985, Paracuellos et al. 2003). Similarly, shell volumes are bigger for the intertidal type and help reduce water loss (Vermeij 1973, Wolcott 1973, Branch 1975, Branch 1981) but also infer resistance to the effects of ice damage (Morley et al. 2010). Differences in the energetic responses of the two morphotypes of *N. concinna* to the difference in mean intertidal (0.45°C) and subtidal (-0.1°C), or the much greater difference in maximum (12.3 versus 1.7°C respectively; Morley et al. 2012), could be a proximate cause of the morphological differences. Taking into consideration differences between environments is therefore important but strongly lacking in the analysis presented here.

In our study, field data show a slight difference in shell length of +5% and a small difference also in the predicted shape coefficient of 0.45 against 0.39 (Table 1.3) for respectively the intertidal and subtidal individuals. This indicates very small differences in inner volumes between the studied populations as calculated by the DEB, meaning that the DEB model does not adequately reflect the difference in morphology between the intertidal and subtidal morphotypes. In the raw data, shell heights present a 33% difference between intertidal and subtidal individuals (Appendix 1.1) but shell length was used, rather than shell height, in the model to characterise the growth structure of the species. Fine tuning the models with extra shape information could have helped to bring further contrasts between the two models, but also requires much more information on shell growth.

Moreover, the difference in food availability and quality was hypothetized between the morphotypes when calibrating the model, despite food abundance and quality knowledge being responsible for strong contrasts in DEB model outputs (Kooijman 2010, Thomas et al. 2011, Saraiva et al. 2012, Sarà et al. 2013). During winter time, the intertidal type seems to have supplementary access to ice-algae and microphytobenthos in rock crevices, whereas the subtidal type mainly grazes on the diatoms films growing on encrusing red algae (Appendix 1.1, Obermüller et al. 2011). But food abundance and quality were assumed for the construction of the models, as

no data accurate enough were available to characterise the feeding behaviour of the limpets. Moreover, in the case of intertidal type, no clear hypothesis is available for their behaviour during winter period, as several authors hypothesize either a migration into the subtidal or a dormance period hidden into crevices during the period where ice is covering their habitat (Brêthes et al. 1994, Obermüller et al. 2011). However, this information would be essential to explain how these individuals energetically behave during this period.

4.4. Potential of the approach

This study showed that it is feasible to build a DEB model for a marine Antarctic species, with few available data. Adding extra information from sampling and experiments during a single expedition in the field considerably increased the accuracy of the model and highlighted some small differences in energy allocation priorities, maintenance costs and reproductive potential between the intertidal and subtidal morphotypes. But the method is then limited by model calibration and data availability since it could not prove that these contrasts are explained by anything else but environmental conditions.

Such DEB models would already be sufficient to (1) describe the performance of the species physiological traits in spatially or temporally contrasting environmental conditions (Kearney et al. 2012, Teal et al. 2012), (2) to be upscaled to the population level to assess population structure and density dynamics (Klanjšček et al. 2006, Arnould-Pétré et al. 2020 - Chapter 1), or (3) to be integrated into a dynamic network by adding knowledge about interaction with other species (Ren et al. 2010, 2012). Adding some data from extra experiments would easily enable further development of these models for ecophysiological or ecotoxicological applications (Muller and Nisbet 1997, Pouvreau et al. 2006, Peeters et al. 2010, Sarà et al. 2011), or to improve knowledge about development stages, behaviour or reproduction (Pecquerie et al. 2009, Rico-Villa et al. 2010, Kooijman et al. 2011).

In this study, we wanted to explore whether the amount of data that was available to build these models were sufficient to see the known physiological and morphological differences between the two morphotypes, and results show that more data are necessary.

To conclude, we advise the use of DEB approach for ecological modelling for Antarctic case studies but modellers should be aware of the necessity to calibrate models with accurate data to fine tune results. Among these data, the description of the species habitat is complex information to be integrated into a model and most of the time only partial information is available. Working in narrow scale areas where habitat is known and described and where experiments can be run might be a good option.
Our study also hightlights the interest of DEB models to reuse data from experiments from historical published works from Antarctic campaigns and highlights the importance of precisely documenting the associated metadata (notably the description of the environment and the conditions in which the limpets are living), data that is not always available.

APPENDIX 1.1. Experimental design S. Morley February 2018, Rothera Station

Collection of limpets

Intertidal *N. concinna* were collected from East Beach at low tide and subtidal *N. concinna* by SCUBA divers at 30m depth off the wharf at the British Antarctic Survey Research Station, Rothera Point, Adelaide Island (67°34.25′S, 68°08.00′W). Representative habitats are shown in Figure S1.1.

Figure S1.1. Left panel, image of the intertidal *Nacella concinna* habitat at low tide. Right panel, representative image of the *N. concinna* habitat at 30m. In the right panel the urchin *Sterechinus neumayeri* and the limpet *N. concinna* are clearly visible on the rock in the foreground.

Measurement of routine metabolic rate

After collection animals were transported in seawater to the Rothera flow through aquarium where they were maintained without supplementary feeding for 10-20 days to allow the majority of their last meal to be processed and the peak in specific dynamic action to have passed before routine metabolic rate was measured.

Routine metabolic rate was measured in closed cell respirometers, following the methodology of Peck (1989), except that oxygen concentration was measured with a Fibox-3 oxygen meter (Presens GmbH, Regensberg, Germany; e.g. Morley et al. 2009a). Oxygen sensitive foils were calibrated before each measurement using 5% w/w sodium dithionite for 0% and fully aerated water for 100%. During trials oxygen concentration was not allowed to fall below 70% of air saturation, which is above the threshold for oxy-regulation of *N. concinna* (Morley et al. 2009a). Two empty chambers (controls) were run with each trial to account for background oxygen consumption, which was routinely less than 10% of the animal's consumption. After each trial the volume of each limpet was measured (through displacement) and this was subtracted from the volume of the respirometer to measure the volume of water within each respirometer.

To calculate the oxygen consumption of organic tissue per gram, wet weights of whole animals and wet weights of all tissues, minus the shell, were measured. Tissue was then dried in an oven at 60°C for 24 hours and then reweighed every 24 hours until a constant dry weight, ± 0.010g, was achieved. The dried tissue was then ashed in a furnace at 475°C for 24 hours and ash free dry mass was measured.

Supplementary results

Table S1.1A. Size, dry mass, Ash Free Dry Mass and routine metabolic rate of *Nacella concinna* collected from the intertidal and 30m depth in January 2018.

Shore Height	Shell Length/mm	Shell height / mm	Tissue Wet Mass/g	Ash Free Dry Mass/g	MO_2/ $\mu mol.O2hr^{-1}$ .g $AFDM^{-1}$
Intertidal	23.0	7.9	1.39	0.23	31.92
Intertidal	27.1	11.4	1.89	0.34	6.09
Intertidal	27.1	9.7	1.45	0.20	8.05
Intertidal	19.2	6.4	0.59	0.11	16.58
Intertidal	22.8	8	1.08	0.18	7.20
Intertidal	22.6	7.7	0.88	0.16	6.66
Intertidal	28.8	11.7	2.53	0.35	18.18
Intertidal	21.3	6.1	0.73	0.12	9.57
Intertidal	25.7	11.3	1.70	0.23	5.65
Intertidal	21.5	7.4	0.93	0.14	8.62
Intertidal	16.6	5.3	0.39	0.08	9.38
Intertidal	24.8	9.3	1.41	0.18	7.94
Intertidal	22.1	7.5	1.01	0.16	9.99
Intertidal	30.5	15.3	2.85	0.55	7.45
Intertidal	31.4	12.3	3.02	0.47	10.65
Intertidal	23.2	7.8	0.97	0.20	5.62
Intertidal	20.2	5.8	0.52	0.08	4.39
Intertidal	38.1	19.7	2.06	0.29	6.34
Intertidal	26.0	10.8	1.81	0.27	9.80
Intertidal	19.9	6.4	0.49	0.08	9.80
Intertidal	30.2	11.6	2.32	0.40	17.98
Intertidal	28.7	10.4	2.56	0.40	7.20
Intertidal	22.0	8.4	0.88	0.12	10.09
Intertidal	20.3	6.6	0.69	0.09	8.19
Intertidal	27.0	10.9	1.79	0.25	5.96
Intertidal	21.2	6.6	0.72	0.10	9.03
Intertidal	16.2	5	0.35	0.06	8.74
Intertidal	25.4	7.9	1.11	0.19	7.13
Intertidal	25.6	9.6	1.60	0.32	7.69

30m	23.3	5.8	0.95	0.16	11.29
30m	26.0	6.8	1.28	0.14	9.81
30m	28.2	7.2	1.22	0.09	9.63
30m	17.5	4.3	0.35	0.05	14.93
30m	20.9	5.8	0.77	0.11	9.60
30m	21.5	5.8	0.60	0.09	7.47
30m	24.5	7.3	1.15	0.18	4.22
30m	27.5	6.8	1.57	0.18	6.56
30m	18.3	4.6	0.33	0.03	14.63
30m	20.3	4.7	0.43	0.07	6.12
30m	23.8	6.5	0.92	0.15	7.21
30m	26.0	7.5	1.34	0.16	11.78
30m	20.8	6.2	0.62	0.08	18.07
30m	23.0	6.7	0.76	0.10	8.03
30m	27.8	6.6	1.59	0.24	6.83
30m	30.4	9.1	1.62	0.18	8.90
30m	22.5	4.9	0.58	0.07	21.05
30m	25.5	5.6	1.02	0.16	8.04
30m	21.1	4.9	0.65	0.07	5.48
30m	22.5	5.4	0.60	0.11	11.36
30m	31.7	10.9	2.26	0.26	5.53
30m	33.2	6.9	3.25	0.41	9.02
30m	22.9	6.4	0.76	0.08	7.49
30m	18.1	4.1	0.42	0.05	10.57
30m	17.1	4.6	0.25	0.02	10.25
30m	22.3	6	0.85	0.11	8.67
30m	21.6	5.3	0.60	0.07	8.77
30m	17.4	3.8	0.23	0.04	10.90
30m	19.6	6	0.47	0.07	8.82

Table S1.1B. Size at first reproduction for both intertidal and subtidal (30m depth) *Nacella concinna* collected from Rothera Point, Adelaide Island. Sex was determined where possible; otherwise individuals were classified as immature. There was only one ripe male in January when measurements were made.

Shore Height	Shell Length/mm	Tissue wet mass/g	Gonad wet mass/ g	Somatic wet mass/ g	Gonad somatic Index	sex	Note
intertidal	18.8	0.38	0.01	0.38	0.02	F	
intertidal	26.1	1.63	0.20	1.43	0.12	M	
intertidal	26.2	1.43	0.11	1.31	0.08	F	
intertidal	26.3	1.29	0.11	1.18	0.08	F	
intertidal	25.8	1.49	0.18	1.31	0.12	F	
intertidal	17.5	0.30	0.01	0.29	0.02	M	
intertidal	30.7	2.57	0.65	1.92	0.25	M	ripe
intertidal	15.4	0.35	0.04	0.31	0.11	M	
intertidal	30.0	2.04	0.34	1.70	0.17	M	
intertidal	17.0	0.34	0.03	0.31	0.09	M	
intertidal	27.5	1.45	0.15	1.30	0.10	F	
intertidal	26.8	1.52	0.17	1.35	0.11	M	
intertidal	22.1	0.85	0.06	0.78	0.08	M	
intertidal	25.2	1.30	0.07	1.23	0.06	M	
intertidal	22.8	0.87	0.07	0.80	0.08	F	
intertidal	24.2	1.10	0.11	0.99	0.10	F	
intertidal	29.5	1.64	0.10	1.54	0.06	F	
intertidal	29.8	2.31	0.23	2.07	0.10	F	
intertidal	33.5	2.79	0.32	2.48	0.11	F	
intertidal	19.9	0.55	0.00	0.55	0.01	F	
intertidal	17.8	0.38	0.03	0.35	0.09	M	
intertidal	20.1	0.59	0.03	0.57	0.05	F	
intertidal	19.5	0.51	0.03	0.48	0.06	M	
intertidal	21.8	0.68	0.02	0.66	0.03	M	
intertidal	17.2	0.34	0.00	0.34	0.00	I	no gonad visible
intertidal	18.3	0.54	0.01	0.53	0.01	F	
intertidal	26.3	1.15	0.21	0.94	0.18	M	
intertidal	21.7	0.65	0.00	0.65	0.00	M	
intertidal	18.5	0.49	0.03	0.45	0.07	M	
30m	24.4	1.18	0.00	1.18	0.00	I	

30m	25.5	1.28	0.03	1.25	0.02	M	
30m	22.5	0.86	0.01	0.84	0.01	M	
30m	41.2	4.79	0.54	4.26	0.11	M	
30m	41.8	5.50	0.30	5.19	0.05	M	
30m	24.4	0.96	0.44	0.52	0.46	M	
30m	31.1	1.92	0.01	1.91	0.01	F	
30m	24.2	0.85	0.00	0.85	0.00	M	
30m	24.1	0.89	0.01	0.87	0.01	M	
30m	21.9	0.68	0.01	0.68	0.01	M	
30m	17.7	0.31	0.01	0.31	0.02	M	
30m	28.3	1.72	0.06	1.67	0.03	M	
30m	18.7	0.56	0.00	0.56	0.00	I	
30m	20.1	0.44	0.00	0.44	0.00	I	
30m	18.2	0.33	0.00	0.33	0.00	I	
30m	30.1	2.25	0.05	2.20	0.02	M	
30m	21.7	0.84	0.02	0.82	0.02	M	
30m	22.5	0.89	0.01	0.89	0.01	M	
30m	25.6	1.16	0.10	1.06	0.09	M	
30m	23.1	1.15	0.04	1.11	0.03	M	
30m	24.1	1.04	0.02	1.02	0.02	M	
30m	20.3	0.61	0.00	0.61	0.00	I	
30m	18.5	0.42	0.00	0.42	0.00	I	
30m	27.7	1.24	0.06	1.18	0.05	M	
30m	30.2	1.91	0.13	1.78	0.07	M	
30m	28.0	1.31	0.03	1.28	0.02	F	
30m	23.8	0.88	0.01	0.87	0.01	F	
30m	32.8	2.06	0.00	2.05	0.00	M	
30m	27.1	1.63	0.02	1.62	0.01	F	
30m	29.0	2.07	0.02	2.04	0.01	M	
30m	26.8	1.54	0.02	1.52	0.01	F	
30m	34.7	3.54	0.04	3.50	0.01	F	
30m	40.0	4.88	0.49	4.39	0.10	M	
30m	29.5	2.18	0.14	2.04	0.07	M	

30m	29.8	1.29	0.04	1.25	0.03	F	
30m	41.0	5.28	0.56	4.72	0.11	M	
30m	30.3	2.29	0.05	2.24	0.02	F	
30m	25.7	1.39	0.00	1.39	0.00	I	
30m	39.3	4.72	0.00	4.72	0.00	I	
30m	28.3	1.44	0.08	1.36	0.05	M	
30m	26.7	1.52	0.07	1.44	0.05	M	
30m	25.5	1.05	0.01	1.03	0.01	F	
30m	32.9	2.63	0.13	2.50	0.05	M	
30m	34.7	2.95	0.12	2.83	0.04	F	
30m	32.7	2.91	0.32	2.59	0.11	M	
30m	29.0	1.44	0.02	1.43	0.01	F	
30m	24.4	0.91	0.00	0.91	0.00	I	
30m	28.6	0.71	0.11	0.60	0.15	F	
30m	23.9	0.92	0.02	0.91	0.02	F	
30m	31.2	2.53	0.01	2.53	0.00	F	
30m	26.5	1.38	0.07	1.31	0.05	M	
30m	29.0	1.60	0.09	1.52	0.05	M	
30m	26.4	1.24	0.03	1.20	0.03	M	
30m	36.7	3.11	0.17	2.94	0.05	M	

APPENDIX 1.2. Sclerochronology protocol

Based on the literature (Picken 1980, Fig. S1.2A), the distance between the apex of the shell and the different black rings are measured using a 'mesuroscope' at Biogeosciences Laboratory (Université de Bourgogne Franche-Comté, France). According to Picken (1980), black rings correspond to winter growth, and lighter bands to summer growth, which can be a proxy to characterise growth dynamics through time (Fig. S1.2A).

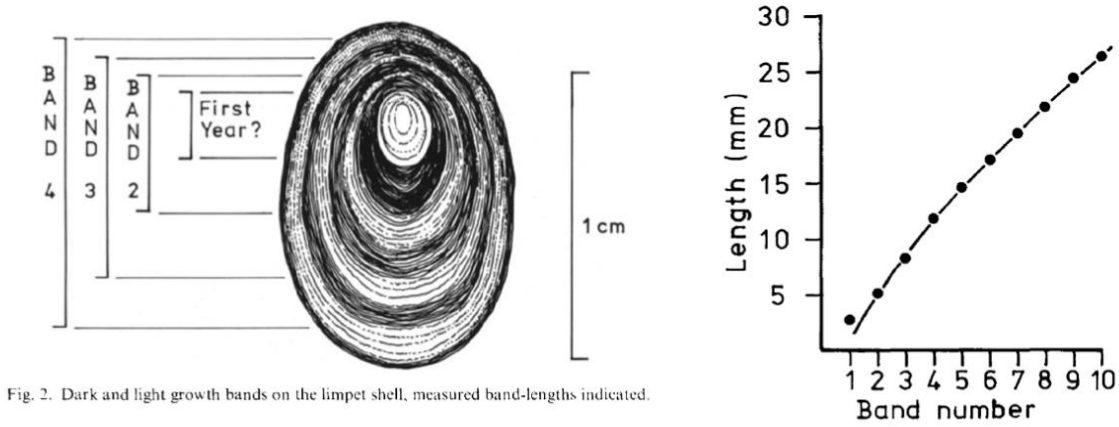


Fig. 2. Dark and light growth bands on the limpet shell, measured band-lengths indicated.

Figure S1.2A. Picken (1980)'s protocol to characterise ring growth through time. Dark rings correspond to winter growth and light rings to summer growth periods.

The 'mesuroscope' (Fig. S1.2.B) is a binocular microscope connected to a computer that enables coordinates to be marked and reported onto an excel sheet; the x,y position (movement of the horizontal plateau where the shell is fixed) and the z position (measured by the vertical movement of the plate, corresponding to zooming in or out and therefore to the height of the shell). The shell is observed with the binocular microscope, a pointer helps at positioning the focus on the screen and a button automatically saves the x, y, z positions on the excel sheet, making measurements fast, efficient and precise. Precision is 10µm. The position of each black ring, on the left side and on the right side of the apex is measured and summed to assess the total shell growth between two rings (Fig. S1.2.B). Each shell was photographed before the procedure to estimate the position of each dark ring, which is not that precise and simple for all cases (Fig. S1.2.B).

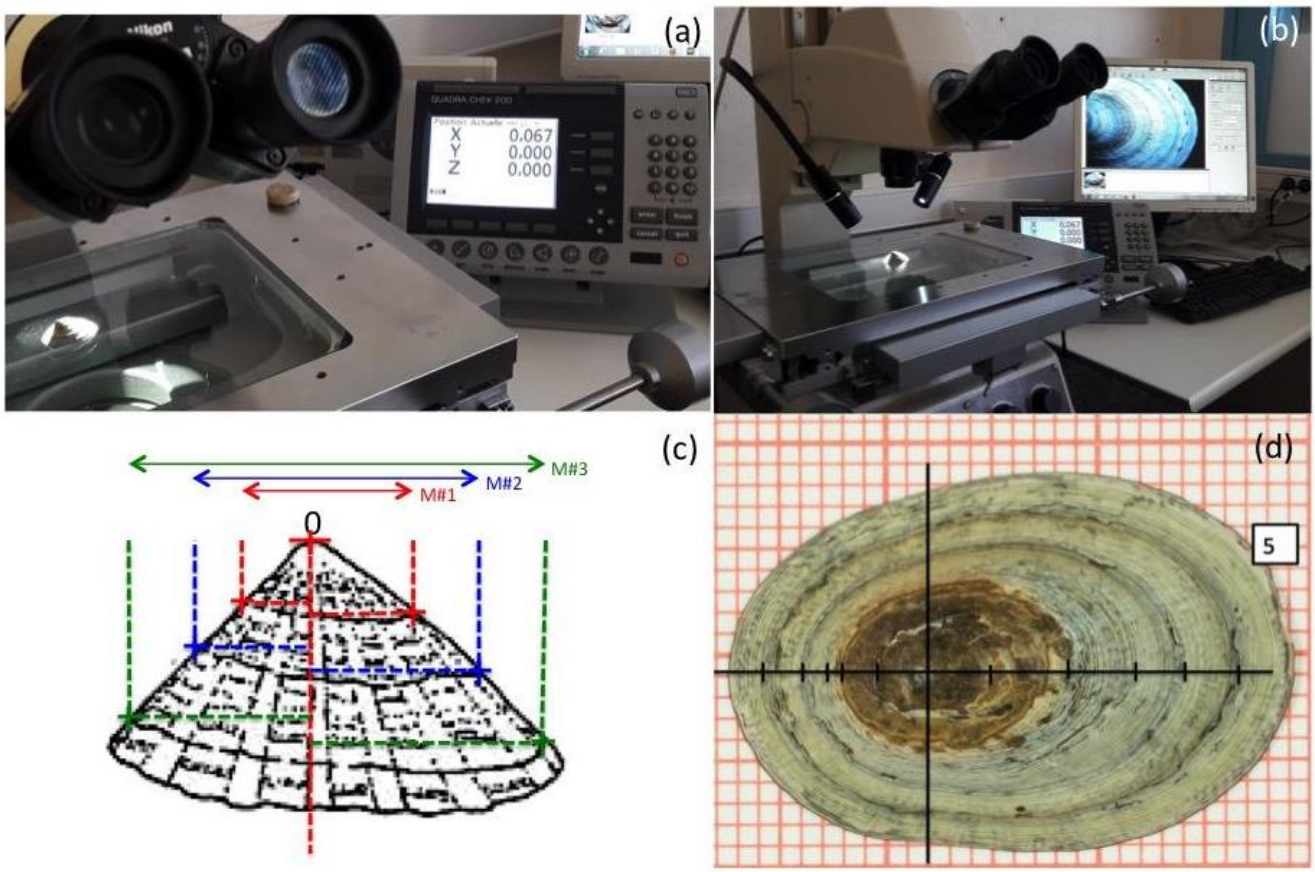


Figure S1.2.B. Details of the 'mesuroscope' (a,b) with the binocular loop connected to the computer, which automates the acquisition of the x,y,z measurements. (c) Schematic representation of the procedure adopted for the measurements of the rings. First, the apex was positioned, and the distance from the apex to the right and left part of each ring was measured (x, y) and summed to get the diameter of each dark ring (M#1, M#2, M#3). (d) Example of picture captured by P. Pernet to prepare ahead the measurement the position of the black rings.

Each of the 60 shells was measured following this protocol and 20 of them were measured twice for replicate analysis. No significant difference was observed between the replicates.
The length ~ time measurements were added to the intertidal and subtidal DEB models.

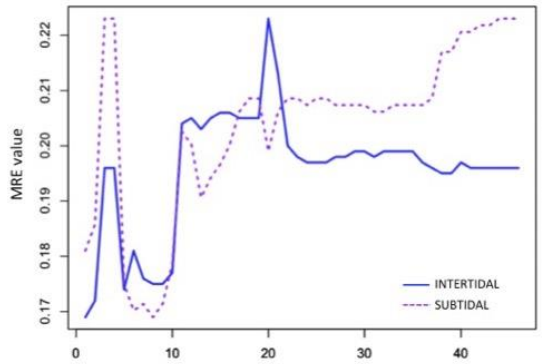
APPENDIX 1.3. Merged models, replicates

Table S1.3. Summary of goodness of fit, DEB parameter estimates at a reference temperature of T_{ref}= 20°C of the different merging trials. Predicted values for zero-variate data, relative error (RE) for the uni-variate data. MRE= Mean Relative Error. 'Trials' are defined as merging procedures where the parameters are merged in different orders, namely, Trial #1: Merge $\ddot{h}_a$, E_H^p, κ, E_H^b, z, $[\dot{p}_M]$, $[E_G]$, $\dot{v}$, E_H^j, δ_{M_larvae}, δ_M; Trial #2: merge $[E_G]$, E_H^j, E_H^b, $\dot{v}$, $\ddot{h}_a$, κ, δ_{M_larvae}, $[\dot{p}_M]$, δ_M, E_H^p, z; Trial #3: merge δ_M, E_H^b, E_H^j, E_H^p, $\ddot{h}_a$, κ, $[\dot{p}_M]$, $\dot{v}$, z, δ_{M_larvae}, $[E_G]$; Trial #4: merge z, E_H^p, δ_M, $[\dot{p}_M]$, δ_{M_larvae}, κ, $\ddot{h}_a$, $\dot{v}$, E_H^b, E_H^j, $[E_G]$; Trial #5: Merge z, $[\dot{p}_M]$, δ_{M_larvae}, κ, $\ddot{h}_a$, $\dot{v}$, E_H^b, $[E_G]$, E_H^j, E_H^p, δ_M.

	TRIAL #1		TRIAL #2		TRIAL #3		TRIAL #4		TRIAL #5	
MRE intertidal	0.196		0.197		0.196		0.196		0.196	
MRE subtidal	0.227		0.228		0.227		0.227		0.227	
Loss function	0.7936		0.7937		0.79363		0.79364		0.79363	
DEB parameters										
z	0.2586		0.259		0.2576		0.2582		0.2579	
$\{\dot{p}_{Am}\}$	6.8590		6.8688		6.86581		6.85855		6.8617	
$\dot{v}$	0.0498		0.04984		0.04987		0.04988		0.04989	
κ	0.9257		0.925		0.9254		0.9261		0.9256	
$[\dot{p}_M]$	24.56		24.54		24.66		24.6		24.62	
$[E_G]$	3953		3952		3952		3952		3952	
E_H^b	0.0014		0.0014		0.0014		0.001409		0.0014	
E_H^j	0.9179		0.9324		0.9252		0.9145		0.9206	
E_H^p	94.56		95.7		95.01		93.96		94.66	
$\ddot{h}_a$	$4.248.10^{-8}$		$4.328.10^{-8}$		$4.234.10^{-8}$		$4.242.10^{-8}$		$4.24.10^{-8}$	
s_G	10^{-4}		10^{-4}		10^{-4}		10^{-4}		10^{-4}	
δ_M	0.4249		0.4247		0.4247		0.4248		0.4247	
δ_{M_larvae}	0.7215		0.7233		0.7215		0.7215		0.7215	
s_M	8.5311		8.5484		8.54249		8.53997		8.5372	
Zero-variate										
	prediction	**// RE**	**prediction**	**// RE**	**prediction**	**// RE**	**prediction**	**// RE**	**prediction**	**// RE**
ab (d)	10.62	0.0618	10.62	0.6158	10.61	0.0611	10.61	0.0611	10.61	0.0609
ap (y)	3.66	0.0839	3.66	0.0855	3.66	0.0849	3.66	0.0850	3.66	0.0845
am (y)	14	3.10^{-4}	13.92	0.0054	14	$1.34.10^{-5}$	14	$3.75.10^{-5}$	14	$1.6.10^{-4}$
Lb (cm)	0.0228	$1.7.10^{-6}$	0.02274	0.0026	0.0228	$2.93.10^{-5}$	0.0228	$2.64.10^{-5}$	0.0228	$1.42.10^{-6}$
Lp (cm)	1.49	0.0325	1.492	0.0617	1.491	0.0624	1.49	0.0323	1.49	0.0323
Li (cm)	5.192	0.2013	5.212	0.0564	5.182	0.0618	5.19	0.201	5.184	0.2024
Ww0 (g)	$5.72.10^{-6}$	0.0137	$5.72.10^{-6}$	0.0133	$5.72.10^{-6}$	0.0136	$5.72.10^{-6}$	0.0144	$5.72.10^{-6}$	0.0138

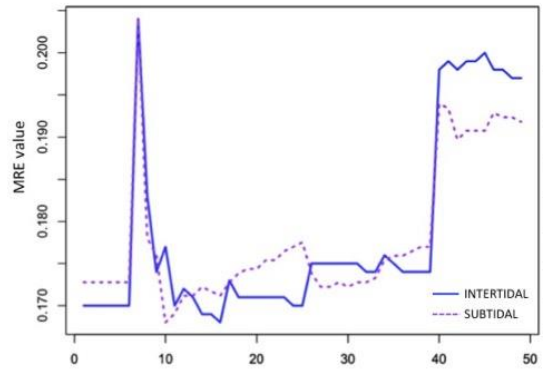
Wdp (g)	0.03956 0.6762	0.03967 0.3038	0.0396 0.3057	0.03957 0.677	0.03956 0.6762
Uni-variate					
	RE	**RE**	**RE**	**RE**	**RE**
LWd_signy (cm, g)	0.1271	0.3323	0.3322	0.1273	0.1274
LWd (cm, g)	0.2156	0.3307	0.3306	0.2158	0.216
LGSI (cm, -)	0.5828	1.061	1.054	0.5812	0.5835
LJO (cm, µmol/h)	0.2488	0.3133	0.3138	0.2487	0.2487
TJO (K, µL/h)	0.0878	0.0882	0.0874	0.08764	0.08757
tL (d, cm)	0.4096	0.4772	0.475	0.4101	0.4097

APPENDIX 1.4. Merged models, replicates: changes in MRE



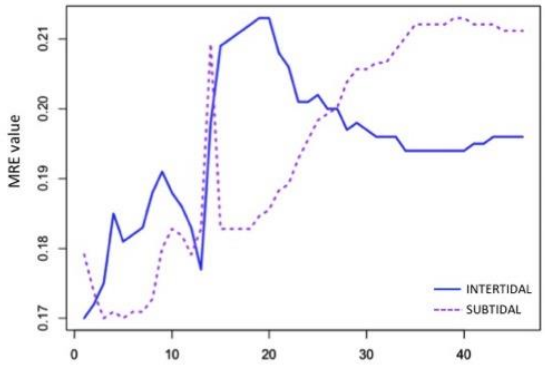
Trial #1

Merged parameter	Merging step #	Merged parameter	Merging step #
$\ddot{h}_a$	2-4	$[E_G]$	24-28
E_H^p	5-8	$\dot{v}$	29-31
κ	9-15	E_H^j	32-33
E_H^b	16-17	δ_{M_larvae}	34
z	18	δ_M	45-46
$[\dot{pM}]$	19-23		



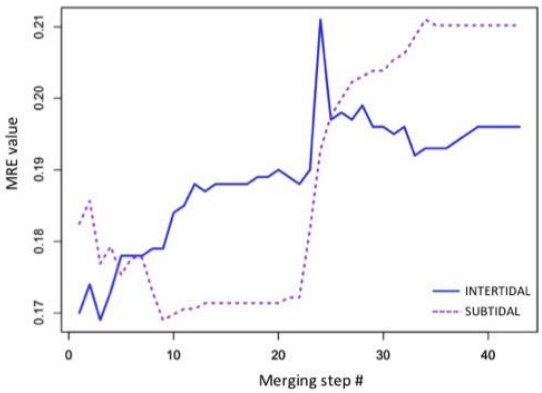
Trial #2

Merged parameter	Merging step #	Merged parameter	Merging step #
$[E_G]$	2-5	δ_{M_larvae}	24
E_H^j	6-10	$[\dot{pM}]$	25-30
E_H^b	11-12	δ_M	31-38
$\dot{v}$	13-15	E_H^p	39-44
$\ddot{h}_a$	16-17	z	45-49
κ	18-23		



Trial #3

Merged parameter	Merging step #	Merged parameter	Merging step #
δ_M	2-7	$[\dot{pM}]$	27-29
E_H^b	8-9	$\dot{v}$	30-32
E_H^j	10-12	z	33-36
E_h^p	13-17	δ_{M_larvae}	37-39
$\ddot{h}_a$	18-19	$[E_G]$	40-46
κ	20-26		



Trial #4

Merged parameter	Merging step #	Merged parameter	Merging step #
z	2-3	$\ddot{h}_a$	27
E_H^p	4-6	$\dot{v}$	28-29
δ_M	7-14	E_H^b	30-31
$[\dot{pM}]$	15-17	E_H^j	32-35
δ_{M_larvae}	18-20	$[E_G]$	36-43
κ	21-26		

Figure S1.4. Evolution of Mean Relative Error (MRE) values along the merging of the different parameters. MRE intertidal in solid blue line, MRE subtidal in dashed purple line. 'Trials' are defined as merging procedures where the parameters are merged in different orders, namely, Trial #1: Merge $\ddot{h}_a$, E_H^p, κ, E_H^b, z, $[\dot{p}_M]$, $[E_G]$, $\dot{v}$, E_H^j, δ_{M_larvae}, δ_M; Trial #2: merge $[E_G]$, E_H^j, E_H^b, $\dot{v}$, $\ddot{h}_a$, κ, δ_{M_larvae}, $[\dot{p}_M]$, δ_M, E_H^p, z; Trial #3: merge δ_M, E_H^b, E_H^j, E_H^p, $\ddot{h}_a$, κ, $[\dot{p}_M]$, $\dot{v}$, z, δ_{M_larvae}, $[E_G]$; Trial #4: merge z, E_H^p, δ_M, $[\dot{p}_M]$, δ_{M_larvae}, κ, $\ddot{h}_a$, $\dot{v}$, E_H^b, E_H^j, $[E_G]$; Trial #5: Merge z, $[\dot{p}_M]$, δ_{M_larvae}, κ, $\ddot{h}_a$, $\dot{v}$, E_H^b, $[E_G]$, E_H^j, E_H^p, δ_M. Trial 5 is presented in the main manuscript (Figure 1.4).

Individual-based model of population dynamics in a sea urchin of the Kerguelen Plateau (Southern Ocean), *Abatus cordatus*, under changing environmental conditions

Arnould-Pétré Margot[1], Guillaumot Charlène[1,2], Danis Bruno[2] , Féral Jean-Pierre[3], Saucède Thomas[1]

1 UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, EPHE, 6 bd Gabriel F-21000 Dijon, France
2 Laboratoire de Biologie Marine, Université Libre de Bruxelles, Avenue F.D.Roosevelt, 50. CP 160/15. 1050 Bruxelles, Belgium
3 Aix Marseille Université/CNRS/IRD/UAPV, IMBE-Institut Méditerranéen de Biologie et d'Ecologie marine et continentale, UMR 7263, Station Marine d'Endoume, Chemin de la Batterie des Lions, 13007 Marseille, France

Ecological Modelling, 440 (2020). Accepted October 23rd 2020

Abstract

The Kerguelen Islands are part of the French Southern Territories, located at the limit of the Indian and Southern oceans. They are highly impacted by climate change, and coastal marine areas are particularly at risk. Assessing the responses of species and populations to environmental change is challenging in such areas for which ecological modelling can constitute a helpful approach. In the present work, a DEB-IBM model (Dynamic Energy Budget – Individual-Based Model) was generated to simulate and predict population dynamics in an endemic and common benthic species of shallow marine habitats of the Kerguelen Islands, the sea urchin *Abatus cordatus*. The model relies on a dynamic energy budget model (DEB) developed at the individual level. Upscaled to an individual-based population model (IBM), it then enables to model population dynamics through time as a result of individual physiological responses to environmental variations. The model was successfully built for a reference site to simulate the response of populations to variations in food resources and temperature. Then, it was implemented to model population dynamics at other sites and for the different IPCC climate change scenarios RCP 2.6 and 8.5.

Under present-day conditions, models predict a more determinant effect of food resources on population densities, and on juvenile densities in particular, relative to temperature. In contrast, simulations predict a sharp decline in population densities under conditions of IPCC scenarios RCP 2.6 and RCP 8.5 with a determinant effect of water warming leading to the extinction of most vulnerable populations after a 30-year simulation time due to high mortality levels associated with peaks of high temperatures. Such a dynamic model is here applied for the first time to a Southern Ocean benthic and brooding species and offers interesting prospects for Antarctic and sub-Antarctic biodiversity research. It could constitute a useful tool to support conservation studies in these remote regions where access and bio-monitoring represent challenging issues.

Keywords

Ecological modelling, Kerguelen, climate change, model sensitivity, endemic echinoderm, Dynamic Energy Budget, Individual-based model

ACKNOWLEDGEMENTS
The authors would like to thank Yoann Thomas and Cédric Bacher for their valuable advice during model construction, and Pierre Magniez for its contribution for data on *A. cordatus* reproduction strategies. The present work is a contribution to the IPEV program PROTEKER (No. 1044) and the French LTER Zone ATelier Antarctique (ZATA). It is also contribution no. 39 to the vERSO project (www.versoproject.be), funded by the Belgian Science Policy Office (BELSPO, contract n°BR/132/A1/vERSO) and contribution no. 13 to the "Refugia and Ecosystem Tolerance in the Southern Ocean" project (RECTO; BR/154/A1/RECTO) funded by the Belgian Science Policy Office (BELSPO). M A-P was supported by the French Foundation for Research on Biodiversity and its partners (FRB - www.fondationbiodiversite.fr). This work was supported by a "Fonds pour la formation à la Recherche dans l'Industrie et l'Agriculture" (FRIA) and « Bourse fondation de la mer » grants to CG.

AUTHORS' CONTRIBUTION
M.A-P: Conceptualization, Methodology, Writing - Original draft; C.G.: Conceptualization, Methodology, Writing - Original draft; B.D.: Supervision, Writing - Review & Editing; J.P.F.: Writing - Review & Editing; T.S.: Supervision, Writing - Review & Editing.

1. INTRODUCTION

The Kerguelen Islands are part of the French Southern Territories (Terres australes françaises - Taf), located at the limit of the Indian and Southern oceans, in the sub-Antarctic area. The region is highly impacted by climate change, and coastal marine ecosystems and habitats are particularly at risk given that species have long adapted to cold and stable conditions (Waller et al. 2017, Gutt et al. 2018, Convey and Peck 2019). Coastal marine species of the Kerguelen Islands are threatened by temperature and seasonality shifts, which are expected to intensify in a near future (Turner et al. 2014, IPCC 5th report). Future predictions of the Intergovernmental Panel on Climate Change (IPCC 5th report) are provided as possible Representative Concentration Pathways (RCP) scenarios of climate change and can be used to infer the potential response of ecosystems to future environmental conditions. However, the insufficient spatial and time resolutions of such models constitute serious limitations for assessing the effects of future environmental changes on sub-Antarctic species (Murphy and Hofmann 2012, Constable et al. 2014b).

The echinoid *Abatus cordatus* (Verrill, 1876) is endemic to the Kerguelen oceanic plateau and common in coastal benthic habitats of the Kerguelen Islands. It is reported in the northern Kerguelen plateau, and around Heard and Kerguelen islands but most records are from shallow, coastal areas of the Kerguelen Islands where dense populations are commonly observed (Agassiz 1881, De Ridder et al. 1992, Mespoulhé 1992, Poulin 1996, David et al. 2005, Hibberd and Moore 2009, Guillaumot et al. 2016, Guillaumot et al. 2018a,b). This makes the species particularly at risk considering the synergetic effects of the multiple factors (temperature variations, significant shifts in coastal currents, sedimentation rates and phytoplanktonic blooms) affecting coastal marine communities at high latitudes (Waller et al. 2017, Stenni et al. 2017, Gutt et al. 2018). The species' endemicity can be partly related to low dispersal capabilities, which is a consequence of a particular life trait: *A. cordatus* broods its young in incubating pouches located on the aboral side of the test, and has a direct development with no larval stage and no metamorphosis. The low dispersal capacity of *A. cordatus* likely increases its vulnerability to environmental changes (Ledoux et al. 2012).

Benthic fauna of sub-Antarctic regions remains under-studied compared to pelagic species (Améziane et al. 2011, Xavier et al. 2016). Ecological niche models can represent relevant tools to study the consequences of environmental changes on the biology of these benthic organisms and on their population dynamics. Correlative niche models were used to predict the distribution of suitable areas for *A. cordatus* on the Kerguelen plateau (Guillaumot et al. 2018a,b). However, supplementary data and analyses are still needed to depict and understand the species' response to environmental changes.

In the present work, a mechanistic modelling approach using a Dynamic Energy Budget – Individual-Based Model (DEB-IBM) was used to analyse the biological response of *A. cordatus* to various environmental conditions. An individual mechanistic model (DEB) was first built using experimental and literature data (Guillaumot 2019c). A DEB model aims to represent the physiological development of an organism, from the embryo to its death based on energetic fluxes and allows considering the metabolic state of the individual at any given moment of its life cycle. It relies on biological principles and first laws of thermodynamics to recreate the metabolic development as a function of two environmental parameters, food resources and temperature (Kooijman 2010).

The DEB model was then upscaled to the population level (IBM), wherein it was implemented as iterative mathematical calculations of each organism's individual development in the population. The IBM relies on the simulation of individuals as autonomous entities forming a complex population within a dynamic system (Railsback and Grimm 2019). The DEB-IBM is used to analyse population dynamics emerging from the development and the physiological traits of individuals as a function of environmental forcing variables (i.e. food resources and temperature). The DEB-IBM can then be used to simulate population dynamics under different environmental scenarios, enabling a better quantification of the vulnerability of populations to changing environmental conditions.

Modelling population dynamics using a DEB-IBM model for a sub-Antarctic and brooding invertebrate brings a feature so far unseen in other published DEB models. The main objectives of the study were to develop a DEB-IBM model for *A. cordatus* (1) to simulate population structure

and dynamics at different sites under both current environmental conditions and future IPCC climate scenarios RCP 2.6 and RCP 8.5, and (2) to assess the feasibility of such a model for organisms in a region where low data availability and resolution may limit model building and validation. The current resolution and accuracy of future climate scenarios in sub-Antarctic areas do not allow building precise and reliable predictions for the future but they were used here as a proof of concept to test population responses to various, conceivable conditions. Sensitivity analyses were performed to test the robustness, potential and relevance of models (Grimm and Berger 2016) considering data availability. Simulations performed for various temperature conditions and food resource availabilities, if validated, may constitute a promising tool to address conservation issues.

2. MATERIAL AND METHODS

2.1. Study area

The DEB-IBM population model was generated in the geographic and environmental context of the Kerguelen Islands (Fig. 1.5) using data of the study site of Anse du Halage, a fieldwork station that has regularly been investigated through several biological studies since the 1980s (Magniez 1980, Schatt and Féral 1991, Mespoulhé 1992, Poulin 1995, Poulin and Féral 1998, Ledoux et al. 2012). The Kerguelen Islands show jagged coastlines and numerous islets and fjords that provide a large variety of habitats to the marine benthic fauna. The nature of the seafloor varies from rocky to sandy and muddy shores. The predominance of the giant kelp *Macrocystis pyrifera* is a main feature of the Kerguelen as this engineer and key species plays a decisive role in the protection and structuring of benthic shallow habitats in many places of the archipelago (Lang 1967, 1971, Arnaud 1974, Féral et al. 2019).
Located in the Morbihan Bay, a 700 km^2 semi-enclosed shallow embayment (50m depth on average) of the Kerguelen Islands, Anse du Halage is situated at the bottom of a small and shallow (2m depth) cove dominated by fine to medium sands (Magniez 1979, Poulin 1996) (Fig. 1.5). The tidal range is comprised between 0.4 and 2.1m, so that the area can exceptionally be uncovered at the lowest tides (Schatt and Féral 1991, Mespoulhé 1992). Sea surface temperature varies between 1 and 2°C in winter (September) to 7 to 8°C in summer (March), with sporadic peaks of +11°C in some places, for certain years (Schatt and Féral 1991, Féral et al. 2019). Salinity varies between 31.89 and 33.57 (Arnaud 1974).
Temperature data used in the model were collected in the framework of the Proteker program (French Polar Institute n°1044) (Féral et al. 2019) and accessed online (IPEV programme n°1044, http://www.proteker.net/-Thermorecorders-.html?lang=en accessed on 08/05/2019). They were recorded from 2012 to 2018 at three sites used in the model (Fig. 1.5): Ile Longue (for the model at Anse du Halage), Ile Haute, an island in the North-Western corner of the Morbihan Bay, and Port Couvreux, a coastal site outside the Morbihan Bay, in the Gulf of the Baleiniers on the Northern coast of the archipelago (Fig. S1.5).
The organic matter deposited on the seabed varies with seasonal phytoplankton blooms and remineralization by bacteriae (Delille et al. 1979). The sediment organic content and phytoplanktonic blooms are particularly important at Anse du Halage, with average values of 4.5% of organic carbon content. The sediment organic carbon (OC) content was monthly measured as a percentage of sediment dry weight by Delille and Bouvy (1989).

Environmental data time-series are available at a monthly timestep. The model was scaled on a single square metre patch, supposing no connectivity between neighbour locations, as no data on horizontal nor vertical water movements and matter fluxes were available.

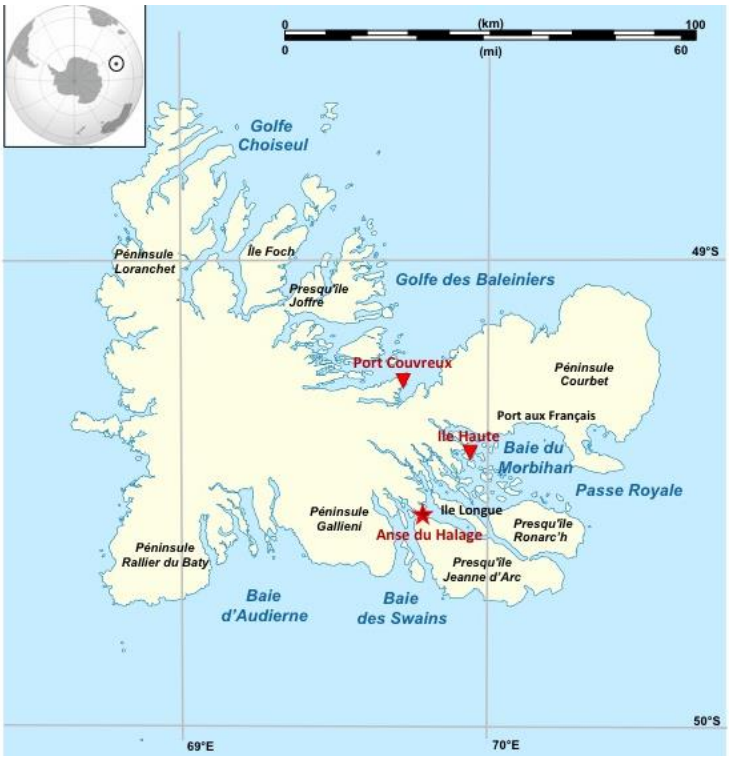


Figure 1.5. Location of the studied sites in the Kerguelen Islands, calibration site (Anse du Halage, red star) and projection sites (Ile Haute and Port Couvreux, red triangles).

2.2. Study species

Abatus cordatus (Fig. 1.6) is a shallow deposit-feeder and sediment swallower, living at 5°C average, full or half buried into soft sediments (De Ridder and Lawrence 1982). It is distributed all around the Kerguelen Islands, but population densities are highly variable depending on depth, substrate nature and exposure to the open sea. Distributed from the intertidal area to the deep shelf over 500m depth, populations highest densities are found in very shallow (0-2m depth) and sheltered areas with soft bottoms of fine to medium sand (Poulin 1996). In shallow areas, observed density vary from less than 5 individuals/m^2 (in the Fjord des Portes Noires, Poulin and Féral 1995) to 10 ind./m^2 (at Port-aux-Français, Mespoulhé 1992), 130 ind./m^2 (Ile Haute, Mespoulhé 1992, Poulin 1996), 168 ind./m^2 (Port Couvreux, Poulin 1996) and up to 280 ind./m^2 (Anse du Halage, Magniez 1980, Poulin 1996). Juveniles are commonly found sheltered in between holdfasts of the giant kelp *Macrocystis pyrifera* bordering with sandy shallow areas.
The species is relatively resistant to low salinities locally induced by freshwater run-off from the main island (Guille and Lasserre 1979). It is tolerant to temperature variations, particularly marked in shallow areas, but temperature tolerance does not exceed +12°C (personal observations). The maximum size ever observed is 4.9 cm in length (Mespoulhé 1992). Lifespan is assumed to be around six years old (Mespoulhé 1992), although it cannot be excluded that some individuals may grow older. Identified predators are gastropods, crustaceans and seagulls (Poulin and Féral 1995, Poulin 1996) from which the specimens are hidden when burrowing into the sediment (Magniez 1979, Poulin and Féral 1995).

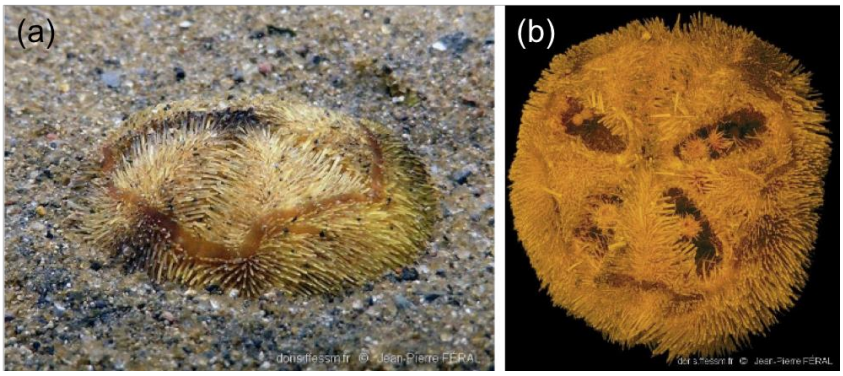


Figure 1.6. Specimens of *Abatus cordatus*. (a) Aboral view of a specimen half buried in sand, (b) Aboral view of a female showing the brood pouches with juveniles inside. © Féral J.P.

Sexual reproduction in *A. cordatus* occurs every year, with all mature females producing eggs (Magniez 1983) and incubating their young in their four brood pouches located on the aboral side of the test (Fig. 1.6). After brooding, juveniles exit the pouches and start their autonomous development on the seabed, in the vicinity of their mother (Magniez 1983, Schatt 1985). Reproduction time can greatly differ between sites: generally extending from March to May (as in Anse du Halage and Ile Haute), reproduction can also occur from June to August (Ile Suhm), from December to February (Port Matha) or from August to November (Port Couvreux) (Poulin 1996). Females usually spawn once a year (Poulin 1996). Brooding and burrowing behaviours imply a relative sedentary lifestyle and can explain a part of the species endemicity, with dense populations scattered all around the archipelago and only a few older individuals that may be found isolated from core populations (Mespoulhé 1992, Poulin and Féral 1995).

2.3. DEB modelling

Principles

DEB theory defines individuals as dynamic systems and provides a mathematical framework for modelling organisms' life cycle. It describes physiological processes using four primary state variables -reserve, structure, reproduction buffer and maturity- directly linked to mass and energy flows and influenced by two forcing variables, temperature and food availability (Fig. 1.7, Kooijman 2010, Jusup et al. 2017). Based on feeding, growth and reproduction processes, DEB models predict the metabolic and development states of organisms through time (Sousa et al. 2008, Kooijman 2010). Metabolic processes are linked to shape and size of the organism, represented by the structural volume and the structural area. Structural volume is related to maintenance processes, while structural area is closely linked to food ingestion and assimilation processes and controls the amount of energy arriving into the reserve compartment E (Fig. 1.7, van der Meer 2006).

The energy contained in the reserve compartment is allocated to organism maintenance ('somatic' and 'maturity' maintenances, priority processes that condition the organism's survival), to growth (increase of structural volume V), and to the increase of complexity (EH) or reproduction buffer (ER) (Fig. 1.7) according to the kappa-rule (Kooijman 2010). The complexity is represented as the maturity level. The amount of energy accumulated into this compartment triggers metabolic switches such as the transition (i.e. ability to feed, to reproduce) between life stages, defined in DEB theory (namely embryo, juvenile and adult life stages) (Kooijman 2010).

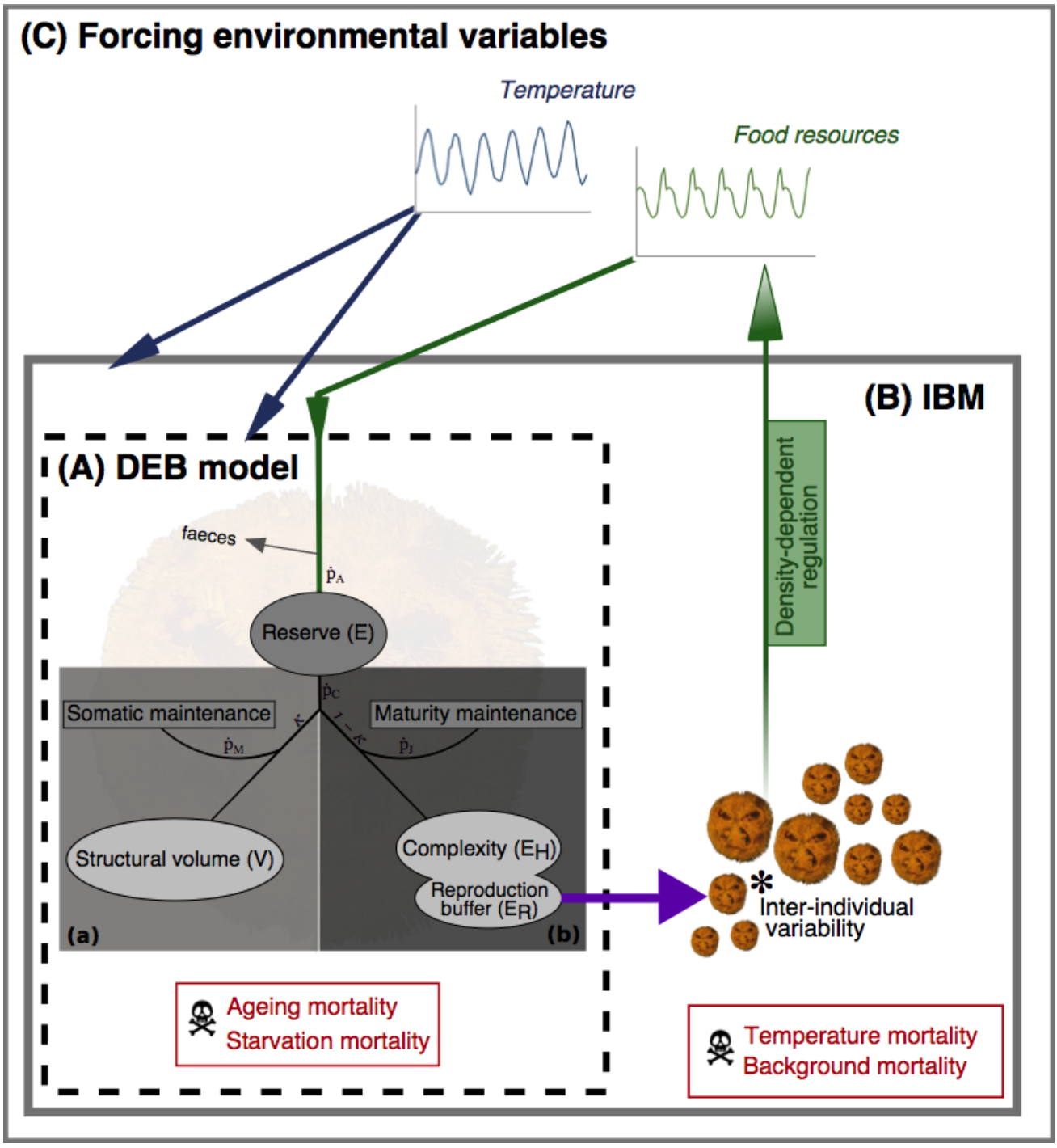


Figure 1.7. Schematic representation of the DEB-IBM (Dynamic Energy Budget – Individual-Based Model). Individuals (A) undergo development through the DEB model and reproduce (purple arrow). Altogether and with a slight inter-individual variability in DEB parameters (*), they form the population of the IBM (B) which undergoes population-specific processes (temperature and background mortalities) at the scale of a simple square metre patch at the reference site (C). The IBM population is embedded within this specific environment, whose environmental conditions (temperature and food resources) affect individual and population dynamics. Additionally, the population influences the resources availability following a density-dependence regulation.

Application of DEB model to *A. cordatus*

Parameter estimation**.** An individual mechanistic DEB model was developed for *A. cordatus* (Guillaumot 2019c). Estimated DEB parameters are reported in Table 1.4. The DEB model considers a larval growth accelerated compared to the adult stage (Schatt 1985), so-called 'abj' type model. The model was constructed using data from the literature (Table 1.5). The goodness

of fit of the DEB model to the data was evaluated by calculating the Mean Relative Error (MRE) of each dataset, which is the sum of the absolute differences between observed and expected values, divided by the expected values. MRE values are contained in the interval [0, infinity). The MRE is considered to be a reference method to assess DEB modelling performance (Lika et al. 2011b), for which the closer to 0, the better model predictions match the data.

Table 1.4. Parameters estimated for the DEB model developed for *Abatus cordatus*. Values are given for the reference temperature of 20°C. The MRE of the model is 0.121.

Parameter	Symbol	Value	Unit
Basic DEB parameters			
Volume of specific somatic maintenance[1]	[ṗM]	13.84	$J.d^{-1}.cm^{-3}$
Somatic maintenance rate coefficient[1]	k̇M	5.10^{-3}	d^{-1}
Fraction of energy allocated to somatic maintenance and growth[1]	κ	0.78	-
Volume-specific cost of structure[1]	[EG]	2395	$J.cm^{-3}$
Energy of maturity at birth[1]	$[E_H^b]$	0.5693	$J.cm^{-3}$
Energy of maturity at metamorphosis[1]	$[E_H^j]$	8.325	$J.cm^{-3}$
Energy of maturity at puberty[1]	$[E_H^p]$	1638	$J.cm^{-3}$
Arrhenius temperature[3]	TA	9000	K
DEB compound parameters			
Energy conductance[1]	v̇	0.02722	$cm.d^{-1}$
Maturity maintenance rate coefficient[3]	k̇J	0.002	d^{-1}
Shape coefficient[2]	δ	0.6718	-
Maximum structural length [1]	L_m	2.93	cm
Acceleration factor[1]	sM	2.397	-
Reproduction parameters			
Yield of structure on reserve[1]	yVE	0.865	$\#mol.mol^{-1}$
Contribution of reserve to weight[1]	w	0.647	-
Ageing parameters			
Weibull ageing acceleration[1]	ḧa	$5.02.10^{-6}$	d^{-2}
Gompertz stress coefficient[3]	sG	0.0001	-

[1] Estimated using the covariation method (Lika et al. 2011a, 2011b, Marques et al. 2018)
[2] Calculated from data for initial value and then estimated with the covariation method
[3] Fixed, guessed value

Table 1.5. Zero and uni-variate data used for the estimation of the DEB model parameters. All values are given at a measured temperature of 5°C. MRE: Mean Relative Error. Plots related to uni-variate data can be found in Appendix 1.6.

Variable	Symbol	Obs.	Prediction	Unit	MRE	Reference
Zero-variate data						
Age at metamorphosis[1]	aj	142	143	d	0.0072	Schatt (1985)
Age at puberty[2]	ap	1098	1018	d	0.0731	Mespoulhé (1992)
Life span	am	2190	2190	d	5.10^{-8}	Mespoulhé (1992)
Length at metamorphosis	Lj	0.276	0.324	cm	0.1738	Schatt (1985)
Length at puberty	Lp	1.9	1.824	cm	0.0399	Mespoulhé (1992)
Maximal observed length (at 6 years old)	L6	4.2	3.65	cm	0.1321	Mespoulhé (1992)
Ultimate maximal length	Li	8	9.507	cm	0.1883	guessed
Wet weight of the egg	Ww0	$1.78.10^{-3}$	$1.59.10^{-3}$	g	0.1081	Schatt (1985)
Wet weight at metamorphosis	Wwj	$1.03.10^{-2}$	$1.70.10^{-2}$	g	0.6482	Schatt (1985)
Wet weight at puberty	Wwp	2.9	3.03	g	0.0448	Féral and Magniez (1988)
Wet weight at 6 years old	Ww6	25	24.18	g	0.0328	Féral and Magniez (1988)
Gonado-somatic index[3]	GSI	0.07	0.078	-	0.1194	Magniez (1983)

Variable	Symbol	Obs. /Prediction	Unit	MRE	Reference
Uni-variate data					
Time since birth vs. length	tL	See Appendix 1.6	d // cm	0.2259	Schatt and Féral (1996)
Egg diameter vs. egg wet weight	LW_egg		cm // g	0.1262	Mespoulhé (1992)
Length vs. Wet weight adult	LW		cm // g	0.3248	Féral and Magniez (1988)
Length vs. O2 consumption	LJO		cm // µL/h	0.2872	Féral and Magniez (1988)

[1] moment at which the juveniles leave the brooding pouch of the mother
[2] moment at which the sea urchin is able to reproduce
[3] maximum gonad index for an animal of the maximum size, gonad index as gonad weight/total wet weight.

Maturation and development. Embryos of *A. cordatus* have a direct development in brood pouches of females (Magniez 1983, Schatt 1985). They start feeding inside pouches after 142 days of incubation (i.e. 5 months) and leave the pouches as fully developed sea urchins after 8.5 months (Schatt 1985). According to DEB theory, individuals are considered embryos until they can feed (Kooijman 2010). Before the fifth month, feeding inside the maternal pouches is not clearly attested, but feeding through epidermal uptake of Dissolved Organic Matter (DOM) is considered as the possible mechanism (Schatt and Féral 1996). At each growth step, energy is supplied to the reserve by the ingested food (Fig. 1.7, $\dot{p}A$) and then leaves the reserve compartment to be directed to growth, maturation or reproduction processes through the mobilisation flux (Fig. 1.7, $\dot{p}C$). This is performed following the kappa-rule: a κ fraction is directed towards the structure (growth compartment and somatic maintenance, Fig. 1.7, and the remaining (1-κ) fraction towards complexity (maturation, reproduction compartments and maturity maintenance, Fig. 1.7).

During the juvenile stage, the individual does not supply energy into reproduction, but accumulates energy in its maturity compartment EH until reaching the 'puberty' threshold that, according to DEB theory, defines the moment when the organism is mature enough to reproduce (Kooijman 2010). After reaching this threshold, at around 2.5 to 3 years old (Schatt 1985, Mespoulhé 1992), the organism can allocate energy into the reproduction buffer ER for gamete production (Fig. 1.7). The structural volume increases continuously along the individual's life, from birth to death, supplied in energy left from what has not been allocated to priority maintenance costs $\dot{p}M$ and $\dot{p}J$.

Starvation mortality**.** Magniez (1983) observed that the gonadal index continues to decrease slightly for around two months after reproduction. He hypothesized that it was related to the season: as the reproduction period finishes at the start of winter, food resources decrease and energy investment into reproductive organs is momentary diverted towards the maintenance of

somatic elements. This was demonstrated in the other sea urchin species *Strongylocentrotus purpuratus* (Lawrence et al. 1966) and *Arbacia lixula* (Fenaux et al. 1975) confronted with starvation.

In the model, when scaled reserve e (reserve relative to reserve capacity, no dimension) falls below the scaled structural length l (length relative to maximum length, no dimension), it is assumed that the individual is confronted to starvation: the kappa-rule is then altered as energy is entirely redirected to the somatic maintenance and all other fluxes (growth, reproduction or maturation) are set to 0. When e < 0, the organism does not have enough energy to allocate the amount necessary for survival (somatic maintenance costs) and dies. See section "7.Submodels/Starvation" in Appendix 1.12 for further details and implemented equations.

Ageing mortality. Death probability by senescence was calculated in the model using the ageing sub-model, a simulation of damages induced by lethal compounds such as free radical or other reactive oxygen species (ROS), following the DEB theory for ageing (Kooijman 2010).
The density of damage inducing compounds in the body increases as the reserve compartment is fuelled with energy that is allocated through the entire organism. It influences the hazard mortality rate $\dot{h}$, which is a function of the damage accumulated in the body and simulates the vulnerability of the individual to damages, such as the risk of dying from illness increases with age. In the model, the hazard mortality rate $\dot{h}$ is supplemented by a stochastic parameter (Martin et al. 2010) to control the ageing mortality rate. See section "7.Submodels/Ageing" in Appendix 1.12 for further details and implemented equations.

2.4 Individual-based modelling for the population

Principles
The individual DEB model is used to simulate each individual as an entity of the individual-based population model (IBM). An IBM represents the individual components (individuals of *A. cordatus*) of an environmental system (Anse du Halage) and their behaviours, enabling to feature each individual as an autonomous entity and looking at results at the scale of the whole population (DeAngelis and Mooij 2005, Grimm and Railsback 2005, Railsback and Grimm 2019). In our model, each individual does not have any direct interaction nor adaptive behaviour towards their environment nor the other members of the population. They follow a continuous development governed by metabolic fluxes (DEB model) that are influenced by environmental conditions (temperature and food resources) along their entire life. Each individual is a component of the modelled population, which is itself affected by population death rate and density-dependent processes.
The IBM was built with the software Netlogo version 6.0.4 (Wilensky 1999), using the DEB-IBM model developed by Martin et al. (2010) for the species *Daphnia magna*. The NetLogo code is available at http://modelingcommons.org/browse/one_model/6201. It contains the script to run the model, the input files of monthly food resources and temperatures for the three stations and a detailed description of the model following the ODD (« Overview, Design concepts, Details ») protocol from Grimm et al. (2010) and the associated list of variables present in the code. This detailed ODD was also included in Appendix 1.12.

Application of IBM model to *A.cordatus*
Model structure. The model includes two types of entities: the individuals and the environment. Individuals are divided into 4 types of sub-agents, depending on their life stage and sex: embryos, juveniles, adult males and adult females. The values of four primary state variables are attributed to each individual (scaled reserve UE, volumetric structural length L, scaled maturity UH and scaled reproduction buffer UR). The level of energy contained in the scaled maturity UH thresholds the life stages. These four variables are 'scaled', meaning here that the energy dimension has been removed by dividing with the surface-area-specific maximum assimilation rate $\{\dot{p}_{Am}\}$ (in $J.L^{-2}.t^{-1}$), based on DEB theory (Kooijman 2010).
Simulations were run with a monthly timestep for calculation, in regard to the slow growth of the species and the available data (an analysis of the effect of the timestep on the individual model

was conducted, Appendix 1.7). At each timestep, food and temperature conditions are first input into the model, state variable values of each individual are calculated in order to assess whether new maturity thresholds are reached or whether energy is sufficient for survival, growth or reproduction. The population state is reassessed at the end of each month. Spatially, population structure and density are simulated on a patch of one square metre at each site and individuals do not leave the patch during their entire life.

Initialisation. The initial population density value was set to 120 ind./m^2 and this figure was split into classes of equal densities of 20 individuals of different age-classes, between 0 and 5 years old, in order to stabilize the initialisation between the different replicates. An initial run is realised to capture the values of the four state variables that characterise the individual of each age class (at October 2012 temperatures and f=1), in order to initiate the model (Appendix 1.12).
The first decade of the simulation period was always considered as the initialisation phase and was removed from the analysis, the model showing important outliers (in individual metabolism and population structure) during these first ten years.

Inter-individuals variability. Each individual is characterised by similar energetic performances estimated by the DEB estimation (Table 1.4). Five DEB parameters were divided by a scatter-multiplier parameter that was generated in order to create inter-individual variability. These five DEB parameters were selected because they are associated to the four state variables that characterise the individuals and are not null at the time the individual is initiated into the model (following Martin et al. 2010): (1) maturity level at birth (U_H^b, d.cm^2), that is the amount of energy accumulated in the maturity compartment needed to reach the juvenile stage; (2) maturity level at puberty (U_H^p, d.cm^2), the amount of energy accumulated into the maturity compartment to reach the adult stage; (3) energy investment ratio (g, no dimension), the cost of the added volume relative to the maximum potentially available energy for growth and maintenance; (4) the initial energy reserve at birth (UE, d.cm^2); and (5) the initial structural length (L, cm).

The scatter-multiplier is the exponential of a random number from a normal distribution of mean 0 and standard deviation cv (0.1 by default, can be set by the user in the interface of the model). The value is therefore small enough to not affect tremendously the initial variable and generate trade-off between parameters. It is applied as soon as the individual is created in the system.

Reproduction. Sex-ratios (ratio males/females) in the studied populations are slightly contrasting between localities, from 0.94 (Ile Haute) to 0.99 (Anse du Halage) and 1.04 (Port Couvreux) (Poulin 1996). The average ratio of 0.99 was chosen in the model. By approximation, it was considered that only females undergo physiological changes during the reproduction process, males being only used as a component of the total population.

To this date, few monitoring studies have been performed on *A. cordatus* reproduction. Magniez (1983) is the only one who studied the Gonado Somatic Index (GSI), that is the proportion of ash-free gonads dry weight over the ash-free body dry weight, therefore directly linked to the accumulation of energy into the reproduction buffer. According to Magniez (1983), reproduction can occur if the GSI reaches at least 0.07%. This condition was used in our model to control the ability of the female to reproduce when time comes.

The GSI parameter was only attributed to females and was estimated for each month, with this equation (Kooijman 2010, section 4.10, eq. 4.89):

$$\mathrm{GSI} = \frac{\mathrm{time\_of\_accumulation} * \dot{k}_M * g}{f^3 * (f + \varkappa * g * y_{VE})} * \left((1-\varkappa) * f^3 - \frac{\dot{k}_J * U_H^p}{Lm^2 * s_M^3}\right),$$

where the time of accumulation is the number of days spent since the end of the reproduction period, $\dot{k}_M$ is the somatic maintenance rate coefficient (in d^{-1}), g the energy investment ratio (no dimension), f the scaled functional response (no dimension), κ the fraction of energy directed towards structure, y_{VE} the parameter for the yield of structure on reserve (mol/mol), that is the

number of moles of structure that can be produced with one mole of reserve, (1-κ) the fraction of energy directed towards complexity, $\dot{k}_J$ the maturity maintenance rate coefficient (in d^{-1}), U_H^p the scaled energy in the complexity compartment at puberty ($d.cm^2$), s_M the acceleration factor (no dimension) and L_m the maximum structural length (cm).

The reproduction period is constant from March to May for the individuals at Anse du Halage, and they only spawn once a year (Magniez 1983, Schatt and Féral 1996, Poulin 1996). After each monthly step, the model checks the GSI value for each female. If the GSI reaches the 0.07% threshold at the onset of the period (March), reproduction is triggered for this considered female.
According to the literature, when reproducing, females invest around 52% of their reproductive organs' energy into reproduction (Magniez 1983). This energy is released during the three months when spawning occurs. That is, the GSI of the female will decrease by 52% of its initial value over the 3 months period (so a decrease of one third of 52% per month, with ∂GSI = (GSIstart - 0.52 * GSIstart) / 3, where GSIstart is the level of gonadal index at the onset of reproduction). In parallel, the usual ∂UR (change in energy density in the reproduction buffer outside of the reproduction period, no unit) is set to 0 for the three months, while UR (energy density in the reproduction buffer) is forced to decrease in a similar fashion to the GSI: ∂^2UR = (UR_start - 0.52 * UR_start) / 3, with UR_start being the reproduction buffer at the start of the period.
Reproduction induces the introduction of 27 embryos in average in the system (Magniez 1983), added proportionally along the three months (9 per month).

Background mortality. No specific adult mortality rates are mentioned in the literature, as no cause have been defined precisely. Background population mortality annual rates were estimated based on size frequency distribution provided by Mespoulhé (1992), and using the formula from Ebert (2013) $N(t) = N_0 * e^{-M*t}$ with N(t) the population size at time t, N_0 the initial population size, M the mortality rate and t the time (in months). Two yearly mortality rates were defined: one for juveniles (41%) and one for adults (24%).

A percentage of embryos mortality in the pouches was calculated based on data from Poulin (1996), determining an egg survival of 65%.This mortality is associated to the fact that when the first juveniles start leaving the maternal pouches at the beginning of January, they push aside the protecting spikes of the pouch, and eggs remaining in the brood are no longer protected and die (Magniez 1980).

Mortality induced by temperatures. As no precise information is available to accurately describe *A. cordatus* temperature tolerance, three different types of sensitivity were designed to cover different hypotheses (Fig. S1.8.B). Based on experimental results obtained in the Kerguelen Islands (personal observation), mortality gradient due to temperature was applied to the population for temperatures comprised between 8 and 12°C. Over 12°C, all individuals are considered to die in the model, as none survived in the experiment. (1) A 'vulnerable' type was defined with population death rates of 25%, 35% and 45% when the sea urchins are exposed to temperatures respectively reaching 8, 9.5 and 11°C during two consecutive months. (2) The 'resistant' type was defined with a mortality rate 15% lower than the vulnerable one for the same temperature thresholds (e.g. 10% instead of 25% population mortality at +8°C), for similar exposure duration (i.e. two months). (3) The 'intermediate' type is similar to the 'resistant' type but individuals are considered to die after one month of exposure to each temperature instead of two (Fig. S1.8.B).

Density-dependent regulation. Population density autoregulates through competition for food resources. This procedure relies on the monitoring of population density in relation to the carrying capacity and allows stabilizing the model. The model calculates the current population density and quantifies the competition effect on food availability depending on how far from the carrying capacity (K) the population density (P) is, and updates food availability in accordance.
It is considered that at each timestep, a certain amount of food is available in the environment (f_{env}) but according to population size, competition for food (FC, quantified food competition) is present and influences effective food availability (f_{eff}), with $f_{eff} = f_{env} + FC$, following Goedegebuure et al. (2018). f_{eff} and f_{env} are contained between 0 and 1. FC is calculated with the following equations:

If $P < 1.9 * K$, then $FC = (1-f_{env}). (1 - \frac{P}{2.K-P})$

If $P \geq 1.9 * K$, then $FC = (1-f_{env}). (1 - \frac{P}{K/10})$

FC is positive if $P < K$, and f_{eff} tends to its maximal value 1 with a decreasing population size, as FC becomes very low and tends towards $1-f_{env}$. When $P > K$, FC turns negative and make f_{eff} decrease, with the minimal value reached at $P= 2K$.

Two equations are used because if $P = 2K$, the first formula gives an error due to a division by 0, and if $P > 2K$, then the formula gives the untrue result of less competition with a bigger population (hence the use of 1.9 as a pivot value). Competition is only effective if food availability is less than the maximum (hence the use of '$1 - f_{env}$' in the equation).

2.5. Summary of model parameterization and sensitivity analysis

The model was constructed following the ecological and physiological observations available in the literature for *A. cordatus*. These observations are summarised in the following table (Table 1.6). Once these elements were added, the ageing submodel and the carrying capacity parameters, for which no *in situ* observations are defined, were calibrated until obtaining a model stable in time, over several centuries.

The sensitivity of the model to different parameters was tested. This sensitivity analysis also served as a first form of validation in the absence of wider means of validation. Initial population number, inter-individuals variation coefficient, juvenile and adult background mortalities, number of eggs produced per female during a reproduction event, and egg survival rate were each applied variations of -30%, -20%, -10%, +10%, +20% and +30% (Table 1.6). The influence of changes in these parameter values was assessed on the average population density (ind/m^2), the average juvenile/adult ratio, the average physical length, the average reserve energy and the average structural length variation over the period of 200 years. For each analysis, models were replicated 100 times. A model was considered to 'crash' when the population is not stable and collapses entirely before the end of the simulation period. The proportion of crashes relates to the number of crashes counted for 100 simulations (i.e., for 15 crashes and 100 simulations, the proportion is 15/(100+15)~13%). Due to computing time limitations, the analysis was stopped when reaching a proportion higher than 66% of crashes (indicated by a black cross in Appendix 1.9).

The model sensitivity to the GSI threshold assumption was tested with the upper and lower values of the GSI calculated at the onset of reproduction in Magniez (1983). The minimum value did not impact the model at all, but the higher threshold value prevented most of the females from reproducing (results not presented).

Table 1.6. List of parameters integrated in the individual and population models. Descriptions and values. The source reference that justifies the choice of the parameterization is provided in the 'reference' column. The last column synthetises which parameters where modified to performed a sensitivity analysis, whose results are presented in Appendix 1.9.

Parameters	Model parameterization	Reference	Sensitivity analysis
Individuals			
Time of development until birth	8 months	254 days (Schatt 1985)	MRE DEB model
Time of development until puberty	Thresholded by U_H^b value	2.5 to 3 years old (Schatt 1985, Mespoulhé 1992)	MRE DEB model
Starvation	e < l	Kooijman (2010)	Not tested
Ageing	Probability depending on accumulated cell damages, constrained by stochasticity	Damage probability: following Martin et al. (2010) and Kooijman (2010) rules for ageing Stochasticity: calibrated at the end of model construction until reaching model stability	Not tested, calibrated parameter
Population			
Initial population density	120 ind./m^2	Rounded from literature (Guille and Lasserre 1979, Magniez 1980, Mespoulhé 1992, Poulin 1996)	-30 to +30% variation tested
Initial population structure	5 age-classes of 20 individuals	Follow average population structure observed by Mespoulhé (1992)	Not tested
Variation coefficient (cv) from the inter-individual variability	0.1	Follow IBM parameterization of Martin et al. (2010) study	-30 to +30% variation tested
Ratio females/males	50/50	Sex-ratio: 0.99 (Poulin 1996)	Not tested
Initial GSI	0.03%	Magniez (1983)	Not tested
GSI threshold for reproduction	0.07%	Magniez (1983)	Tested with the upper (0.116) and lower (0.028) values of the GSI calculated at the onset of reproduction in Magniez (1983)
Reproduction period	3 months once a year	Magniez (1983), Schatt and Féral (1996), Poulin (1996)	Not tested
Energy investment into reproduction	52% of the reproductive energy at the onset of the period	Magniez (1983)	Not tested
Number of eggs	27 eggs per adult female	Magniez (1983), Schatt (1985)	-30 to +30% variation tested
Eggs survival to juvenile stage (birth)	65%	Poulin (1996)	-30 to +30% variation tested
Yearly background mortality rates	41% of juveniles 24% of adults	Equation provided in Ebert et al. (2013), implemented with population data from Mespoulhé (1992)	-30 to +30% variation tested
Mortality induced by temperature tolerance	Three sensitivity scenarios	Designed from experimental results	Not tested
Carrying capacity	200 ind./m^2	Calibrated at the end of model construction until reaching model stability, no information available in the literature	Not tested, calibrated parameter

2.6. Forcing environmental variables

Temperature

In the frame of DEB theory, temperature influences metabolic rates following the Arrhenius function which defines the range of temperatures that affect enzyme performance, considering that metabolic rates are controlled by enzymes that are set inactive beyond an optimal temperature tolerance (Kooijman 2010, Thomas and Bacher 2018). The Arrhenius response is characterised by five parameters that describe the species tolerance range: the Arrhenius temperature TA, the temperature at the upper and lower limits of the species tolerance range T_H and T_L respectively, and the Arrhenius temperature beyond upper and lower limits of the tolerance range T_{AH} and T_{AL} respectively.

In our study, the available information is not sufficient to define the complete relationship between temperature and metabolic performances, and the temperature correction factor (TC) is only calculated using one of the five Arrhenius parameter T_A (in K), following the equation $r(T) = r * \exp(T_A/T_{ref} - T_A/T)$ with r a given metabolic rate, T_{ref} the reference temperature (293K ≈ 20°C), T the environmental temperature (in Kelvin) and $\exp(T_A/T_{ref} - T_A/T)$ being the temperature correction factor TC. The correction is applied to the metabolic rates $\dot{v}$, $\dot{k}_M$, $\dot{k}_J$, $\ddot{h}a$ (Table 1.4).

Temperatures recorded since 1993 at Port aux Français, another site in the Gulf of Morbihan, show a clear 6-year cycle of increasing and decreasing temperatures (Appendix 1.5). The [2012-2018] temperature dataset selected as input forcing variable in the model therefore constitutes an interesting proxy of temperature conditions at Anse du Halage which includes a complete overview of the environmental variability at the station. However, it is important to take this choice into consideration during the interpretation of results, as it needs to be differentiated from a cycle that would be inherent to the biology of the species.

Food resources

In DEB theory, energy is supplied to the reserve of the organism through the ingestion process which is proportional to food availability, represented in the model by a functional response f (from 0 to 1). Food assimilation ($\dot{p}_A$, Fig. 1.7) is proportional to the surface of the structure of each individual and contributes to the filling of the reserve compartment E (Fig. 1.7). The functional response f was calibrated using the values of organic carbon (OC) content in sediment as a percentage of dry weight of sediment at the station Anse du Halage at the end of each month, available in Delille and Bouvy (1989). The maximum value of 1 for f corresponds here to the maximum value of organic carbon content that was found (6.94%) and a f minimum of 0 corresponds to 0% OC.

2.7 Model projection

Present-day conditions at Anse du Halage

To assess the influence of varying environmental conditions on model outputs, after being constructed for the site Anse du Halage, the model was implemented in two other sites, Ile Haute and Port Couvreux, where *A. cordatus* is reported in high densities (Poulin 1996) (Fig. 1.4). The implementation to these two other stations was done with contrasting temperatures (from the Proteker program, as previously explained in 2.1). Food conditions at these two sites are not available and were estimated at the end of the summer to be 50% to 30% of the organic carbon values measured at Anse du Halage according to the comparative study of Delille et al. (1979). These rates were applied to year-long conditions (Fig. S1.5). Models were launched for a period of 200 years.

Future conditions

Two future scenarios predicting environmental conditions for 2100 were used, based on the IPCC scenario RCP 2.6 and 8.5 (respectively optimistic and pessimistic scenarios, IPCC 5th report), accessed at https://www.esrl.noaa.gov/psd/ipcc/ocn/ (in August 2019). Coarse IPCC predictions (1°x1° resolution) of chlorophyll a concentration were used to roughly evaluate potential changes of food availability on the east coast of the Kerguelen Island in future conditions. Scenario RCP 2.6

shows an average decrease of 10% of current food resources availability, while scenario RCP 8.5 shows an average decrease of 20% (Fig. S1.8.C). As for temperature, we defined RCP 2.6 with a linear increase of +1.1°C and +1.7°C for RCP 8.5. Models were launched for a period of 30 years.

3. RESULTS

3.1. The individual-based model

Variations in energy allocated to the reserve and the maturation buffer are the main controls of individual development. Monthly variations (∂UE and ∂UR, ∂X here stands for $\frac{dX}{dt}$) were simulated over one year under present-day environmental conditions (Fig. 1.8).
Energy in the reserve (Fig. 1.8) shows variations between -2.5 and 8 on average (no unit), with extreme range values reaching -5.8 and 13.7. This shows a relative constant energy density inside the reserve throughout the year with, however, a noticeable increase from October to December and a sharp decrease from December to January (Fig. 1.8). According to DEB theory (Kooijman 2010), the more energy is stored inside the reserve (through food assimilation), the more it can be distributed to other compartments, and the more energy can be assimilated into the reserve anew. Availability of food resources for *A. cordatus* is the highest in December (f = 1) (Fig S1.5), it is assimilated and stored as energy into the reserve compartment. Based on the energy available in the reserve at the end of December, energy is supplied in January to other compartments (such as the reproduction buffer, Fig. 1.8 and growth, Fig. S1.10), while the individual ingests the food available to replenish its reserve anew. As food availability decreases in January (f = 0.748), the reserve loses energy (Fig. 1.8) because the individual cannot assimilate as much energy as the amount transferred to other compartments.

The energy density entering the reproduction buffer (Fig. 1.8) of mature females varies between 0 and 4.9 on average in the course of the year, with a maximum of 10.7. The rate of energy input increases at an average pace of +1.1% per month from October to the onset of the reproduction period in March, when it decreases and remains null until the end of the spawning period in June. Then, energy starts accumulating again until the next reproduction period. During the three months of the spawning period, from March to May, no energy is allocated from the reserve to the reproduction buffer and the energy stored in this buffer is progressively delivered to gametes. Only females that are mature in March undergo reproduction and deliver the energy contained in the reproductive buffer to the gametes. Females that become mature during the reproduction period undergo a normal increase of the energy in the reproduction buffer, which explains the small increasing trend observed during the March-May period (Fig. 1.8).

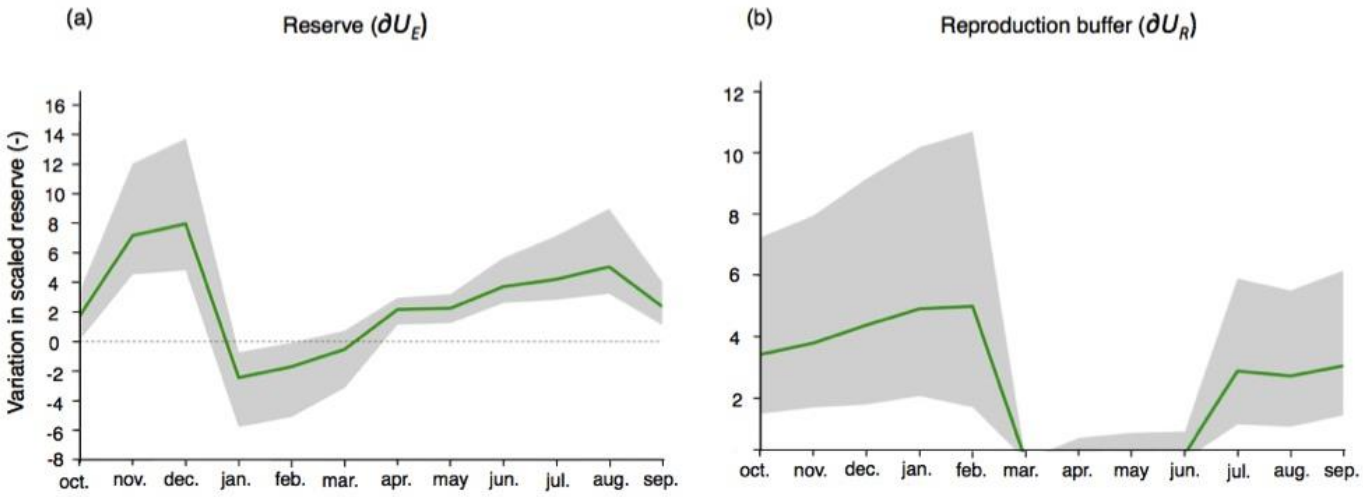


Figure 1.8. Simulation of the variation of energy allocated to the reserve (a) and the reproduction buffer (b) compartments over one year. Males were not considered in the model when simulating reproduction processes, and thus results presented here only take females into consideration. Average results for all mature females, in 100 model simulations are presented by the green line. The grey area corresponds to the variation range (variation induced by differences between individuals: age, size, energy allocation) between all females among the 120 individuals that initiate the model. The variation in energy allocated is the change in a scaled variable X: ∂X here stands for $\frac{dX}{dt}$ (for an explanation of the term "scaled" here, see section 2.4).

3.2. The population model

Modelled population dynamics at the calibration site

Based on the individual model, population dynamics were simulated over a time period of 30 years, showing a constant population density comprised between 120 and 220 individuals per square metre. Overall, the population structure remains constant through time but with well-marked yearly variations, mainly in juvenile density (Fig. 1.9). Juveniles indeed represent around 83% of the total population density and show important yearly variations due to (1) important seasonal reproductive outputs causing a surge in population density, (2) strong mortality rates causing gradual decreases in the population, (3) the transfer of the large juvenile cohort to the adult population after around 3 years, and (4) the influence of inter-individual competition for food limiting population densities and even causing its decrease. In contrast, the adult population is much more stable relative to the juvenile one, with lower density values (around 40 individuals per square metre). Both juvenile and adult population fluctuations follow a general 6-year pattern displayed over the 30 years of simulations (rectangle, Fig. 1.9). This pattern is linked to temperature cycles over the same time span and includes two sharp decline in population density over a 6-year cycle ('T' symbol, Fig. 1.9), which corresponds to high temperatures rising above +8°C during two consecutive months and causing mortality rates of 10% of the entire population.

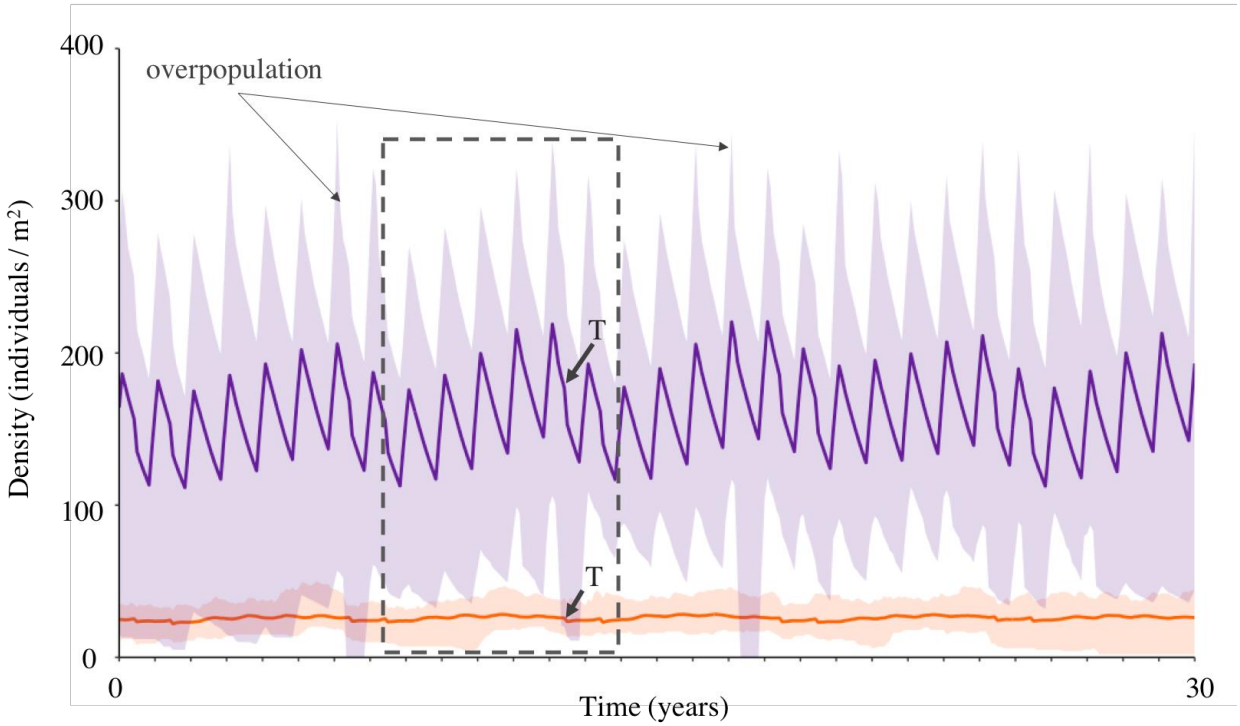


Figure 1.9. Modelled population structure and density under present-day environmental conditions: monthly values of juvenile density (purple) and adult density (orange) over 30 years (for 100 simulations). Bold lines: mean density value. Shaded areas: variation range for the 100 simulations. 'T' symbol: sharp decrease in population density due to temperature-induced mortality. Dashed-line rectangle: 6-year cycle in population dynamics (this 6-year pattern is due to the input temperature data and not to a biological cycle inherent to the population).

Sensitivity analysis

Different parameter settings for the model initiation may result in very diverging outputs (Appendix 1.9). It also influences model stability, and population collapse in particular. Overall, the initial number of individuals and the level of the inter-individual variation coefficient are parameters that have little influence on model stability and low proportion of population crashes may result. In addition, model outputs do not differ significantly between simulations. Increase in juvenile and adult mortality levels will also have little influence on model outputs but decreasing mortality levels will induce a population burst followed by a strong competition for food and a consequent population collapse.

Among all parameters set at the model initialisation, egg number and egg survival are the most important determining model stability, as they directly control juvenile density. High juvenile densities (induced by a low background juvenile mortality and a high number of eggs and egg survival) always result in fast population collapses as a result of high competition for food between individuals. As population density increases, the amount of food available for each individual decreases and individuals start starving to death. In contrast, a reduction in the number of juveniles causes a reduction in the average population density due to a strong mortality rate of juveniles. It does not imply model instability and the proportion of modelled population crashes is always lower than 15%. The reduction of population density also strongly influences the average amount of energy available for each individual: the more energy is available, the more individuals can grow in structural length.

Projections of the population dynamics model to other sites

The dynamic population model built at Anse du Halage was implemented (Appendix 1.11) for the two sites of Ile Haute (inside the Morbihan Bay) and Port Couvreux (outside the Morbihan Bay). Both models were simulated twice with initial estimates of 50% and 30% of food availability (f) compared to Anse du Halage (f_H). Temperature inputs were based on local temperature variations recorded at the two sites. Model outputs predict lower population densities at both sites compared

to Anse du Halage and interestingly, similar ratios between juveniles and adults (Table 1.7). These results are consistent with density values found in the literature, which gives between 100 and 136 individuals/m^2 at Ile Haute and 50 to 168 ind./m^2 at Port Couvreux (Mespoulhé 1992, Poulin 1996). The different observed density values reported in publications for Port Couvreux may be due to contrasting conditions that locally prevail among the three small embayments of that locality (Poulin 1996). This has recently been confirmed by our personal observations in the field (Saucède 2020). Model outputs suggest a strong influence of food availability on population densities controlled by inter-individual competition for food. Accordingly, simulations predict a drop in density values at Port Couvreux when food resources decrease at 30% of f_H, while density values are relatively stable at Ile Haute in comparison (Table 1.7). This mainly affects juvenile densities and results in a lower population ratio (Table 1.7).

Temperatures recorded at the two sites inside the Morbihan Bay (Anse du Halage and Ile Haute) are close to each other and slightly higher than outside the Bay at Port Couvreux (Fig. S1.5). Contrasting results were therefore expected between Port Couvreux and the two other sites. On the contrary, temperatures may not be contrasting enough between sites to affect population structure and density. Confidence intervals overlap between all sites for values of both population density and juveniles-adults ratio (Table 1.7).

Table 1.7. Modelled population densities (a) and juveniles over adults ratio (b) at the calibration (Anse du Halage) and projection (Ile Haute and Port Couvreux) sites. Average and standard deviation values are given for 100 model replicates and 200 years of simulation. f_H: time series of f value at Anse du Halage (Delille and Bouvy 1989).

(a)	**Anse du Halage**	**Ile Haute**	**Port Couvreux**
f_H, T°Halage	182.6 ± 49	-	-
50% of f_H, T°site	-	137.6 ± 40	137.4 ± 41
30% of f_H, T°site	-	123.2 ± 38	91.8 ± 44

(b)	**Anse du Halage**	**Ile Haute**	**Port Couvreux**
f_H, T°Halage	6.53 ± 3.12	-	-
50% of f_H, T°site	-	6.31 ± 4.40	6.32 ± 4.31
30% of f_H, T°site	-	6.04 ± 4.31	3.87 ± 2.61

Population dynamics under future predictions of climate change

Population structure and density were simulated and implemented for scenarios of temperature and food resources changes based on IPCC scenarios RCP 2.6 and 8.5, and for populations of 'resistant', 'intermediate' and 'vulnerable' organisms (Fig. 1.10). Population dynamics are all predicted to be affected by both scenarios (Fig. 1.10) with overall population densities predicted to be four to seven times lower than current population predictions. Population structures are also predicted to be affected by a lower contribution of juveniles to overall population densities. The respective effects of temperature and resource availability were simulated independently. Under temperature change only (Fig. 1.10), model predictions are close to model outputs in which both variables are combined, with a strong decrease in average population density compared to present-day conditions. The effect of changes in resources availability only is less marked, with population densities showing a close pattern to present-day models (Fig. 1.10).

Models therefore predict a stronger effect of temperature changes on populations, with population densities of ‘vulnerable' organisms predicted as very low (less than one tenth of present-day densities on average). Populations of 'vulnerable’ organisms are even predicted to go extinct in only 30 years of simulation (Fig. 1.10). Populations of organisms with 'intermediate' sensitivity are more resilient and withstand over 30 years of simulation in some cases, but they collapse at the

end of the period under IPCC scenario RCP 8.5. Overall densities are very low (around 20 or less individuals per square metre on average).

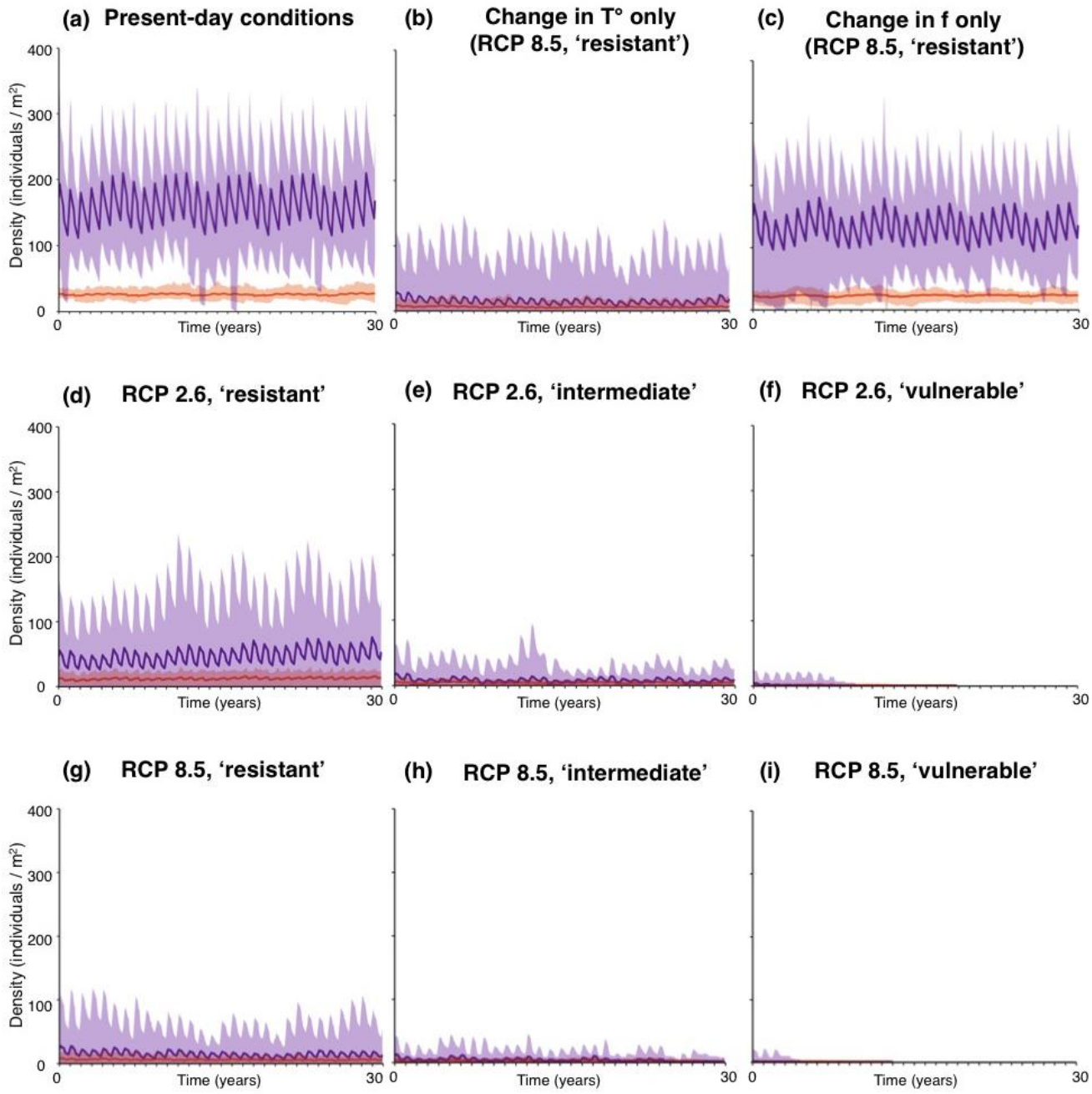


Figure 1.10. Model predictions under IPCC scenarios RCP 2.6 and RCP 8.5 (for 100 simulations). Purple: adult population, orange: juvenile population. Coloured bold lines show mean values for 100 simulations. Shaded areas are simulation variation ranges. (a) Population model under present-day conditions, (b) model under IPCC scenario RCP 8.5 of T° change only (+1.7°C compared to present) and for resistant organisms; (c) model under IPCC scenario RCP 8.5 of f change only (-20% compared to present) and for resistant organisms; (d) model under IPCC scenario RCP 2.6 of T° and f changes (-10% of f and +1.1°C compared to present) and for 'resistant' organisms; (e) model under IPCC scenario RCP 2.6 of T° and f changes and for 'intermediate' organisms; (f) model under IPCC scenario RCP 2.6 of T° and f changes and for 'vulnerable' organisms; (g) model under IPCC scenario RCP 8.5 of T° and f changes and for 'resistant' organisms; (h) model under IPCC scenario RCP 8.5 of T° and f changes and for 'intermediate' organisms; (i) model under IPCC scenario RCP 8.5 of T° and f changes and for 'vulnerable' organisms.

3.3. Population mortality under present-day and future predictions

Under present-day conditions (Fig. 1.11), background mortality and ageing are the main causes that affect population mortality each year (respectively between 65-90% and 5.8-9%). High temperatures and starvation have sporadic effects on mortality. High mortality due to high temperatures only happened in [2016-2017] and [2017-2018] and starvation contributes at the highest to 10% of overall mortality, depending on the year.
Over the course of a year (Fig. 1.10a), background mortality and ageing affect the population every month, while high temperatures (over 8°C) cause the death of half of the population in March and April. Starvation is responsible for the death of a weak proportion of the population in November and December only (austral summer), in link with the competition for food resources of the increasing population during this productive and warm period.
Under both future scenarios (Fig. 1.10e,f), mortality levels are low compared to present-day model (Fig. 1.10b), which is mostly due to small predicted population densities. Background and ageing mortalities are therefore very low. Starvation is not a cause of mortality anymore, while high temperatures cause mortality of individuals before they may starve to death. When comparing between model predictions under scenario RCP 8.5 for changes in food availability only (Fig. 1.10c), temperature change only (Fig. 1.10d), and the combined variables (Fig. 1.10f), temperature clearly appears as the main cause of mortality, at the same level as background mortality.

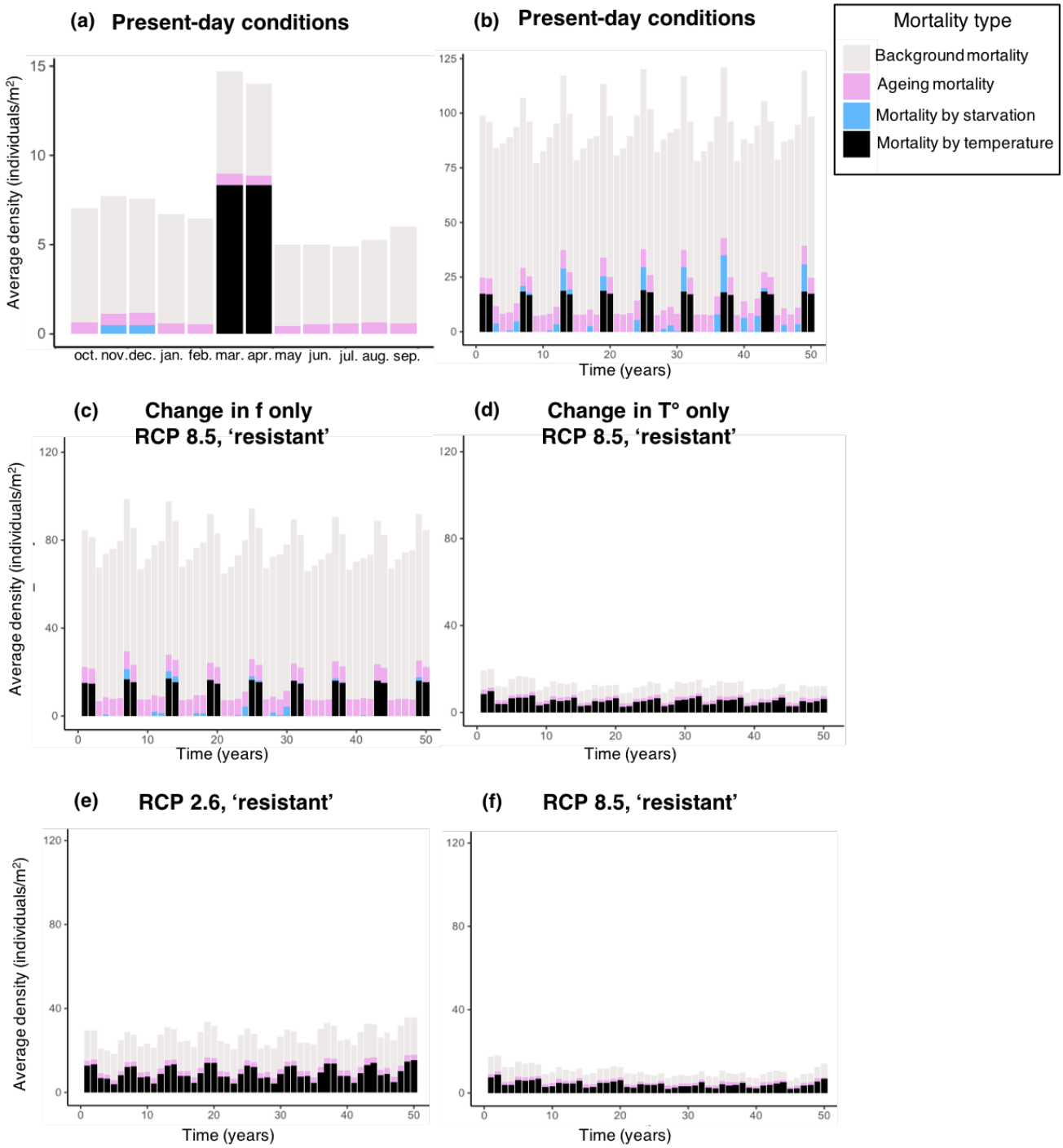


Figure 1.11. Mortality simulations (in individuals/m^2) per month (a) and year (b-f) under present-day (a-b) and future (c-f) predictions of the two IPCC scenarios (for 100 model simulations). (a) Model under present-day conditions for 12 months (year #7 [2016-2017] was chosen as an example); (b) Model under present-day conditions for 50 years of simulations; (c) Simulated mortality under scenario RCP 8.5 for resistant organisms and predicted changes in food availability only (f=-20% compared to present); (d) Simulated mortality under scenario RCP 8.5 for resistant organisms and predicted changes in temperature only (+1.7°C compared to present); (e) Simulated mortality under scenario RCP 2.6 for resistant organisms and predicted changes in both food availability and temperature (food reduction of -10%, T° increase of +1.1°C); (f) Simulated mortality under scenario RCP 8.5 for resistant organisms and predicted changes in food availability and temperature (food reduction of -20%, T° increase of +1.7°C).

4. DISCUSSION

4.1. Potential and limitations of the DEB-IBM approach

In the present work, a DEB-IBM model was built for *A. cordatus* based on our current knowledge of this vulnerable, endemic species of the Kerguelen plateau. On-site monitoring and experiments on species tolerance to changing environmental conditions remains challenging issues in the Kerguelen Islands, and in the Southern Ocean in general. Difficulties are due to the sensitivity of specimens (Magniez 1983, Schatt 1985, Mespoulhé 1992) and their inherent ecological characteristics. Models can constitute a powerful tool for Antarctic research as they can provide additional support to experimental knowledge and infer the impact of broad-scale climate change on populations. The potential of the present mechanistic modelling approach resides in its capacity to model the physiology of organisms as a response to environmental factors. Using DEB models for the representation of individual components within the IBM enables to upscale a dynamic model to an entire population (Railsback and Grimm 2019) as a function of two changing abiotic factors: temperature and food resources. Applying such a model to a sub-Antarctic, benthic, and brooding species is challenging, and had never been performed so far. The present work shows the feasibility and relevance of the DEB-IBM approach to study Southern Ocean species like *A. cordatus.*
Relevant results were obtained both at the individual and populations levels. First, simulations showed the characteristic annual evolution of energy dynamics in the organism (Fig. 1.8, Appendix 1.10) and second, population structure and density dynamics were modelled over an extended period of time decoupling juvenile and adult populations (Table 1.7, Fig. 1.10, Fig. 1.11). Projections to other sites also show the potential application of the model to other areas for which environmental data are available. Future models give an insight and add some clues to assess the potential impact of climate change and predict the biotic response of populations. Models however still need some improvements including complementary data on species ecophysiology. The model was also shown to be sensitive to mortality rates and some parameters (egg number and egg survival) settings while some population characteristics (initial population densities and inter-individual variability) have little effects (Appendix 1.9).

4.2. Limitations to the DEB-individual model

The dynamic population model built in this work uses outputs from the DEB model developed for *A. cordatus* (Guillaumot 2019c), which allows to represent as faithfully as possible the physiological dynamics of individuals during their entire life cycle. The goodness of fit of the DEB model shows that estimated parameters accurately described observed data. However, collecting additional data at the different stages of the organism's life cycle and under different conditions of temperature and food availability would contribute to improving further model accuracy and parameter predictions. In particular, data on environmental settings and species ecophysiology are still needed to improve the accuracy and relevance of the following parameters.

The Arrhenius function and the optimal temperature range
In DEB theory, the Arrhenius function determines the optimal temperature range of the organism's metabolism as a response to enzymatic tolerance (Kooijman 2010, Thomas and Bacher 2018). In the present work, calculation of the Arrhenius function relies on fragmented datasets. The ascending part of the Arrhenius curve that is, the temperature range in which faster metabolic rates are determined by higher temperatures was estimated, but values are still missing for the descending slope (i.e. the temperature range beyond the optimal temperatures in which the metabolic rates slow down with higher temperatures) (Kooijman 2010). The present model assumes that higher temperatures favour more suitable conditions with no limit (Appendix 1.10), which has to be corrected arbitrarily using our personal field and experimental observations on the echinoid ecology (Appendix 1.8). Further experiments should help improve the calculation of the Arrhenius function. They would consist in measurements of respiratory rates as a function of temperature variations (e.g. Uthicke et al. 2014) and will enable more accurate simulations of *A. cordatus*' ecophysiology and the direct effect of temperature on the organism's metabolism, a prerequisite to better model population mortality.

Age, size and growth estimates

Most parameters used in the DEB model were taken from the literature and experimental studies, except for some of them that were assumed based on physiological traits of counterparts. In particular, organisms' maximum age, growth rate and size are not sufficiently known due to difficulties in setting up long-term experiments in the Kerguelen Islands. The relationship between echinoid growth, size and shape cannot be assessed based on growth lines measurement because there is no linear relationship between echinoid size and age (Ebert 1975) and because resorption may occur during periods of starvation (Brockington et al. 2001, Ingels et al. 2012). The most reliable method would consist in monitoring organisms' growth through time using tagging methods (Ebert 2013). However, such an approach is time-consuming and challenging as even small measurement errors may have a significant effect on results (Ebert 2013) and no experimental data are available so far.

Former studies (Mespoulhé 1992) showed that after 4 to 5 years, specimens of *A. cordatus* only slowly increase in size and echinoids' test tend to become distorted, a common feature in large spatangoid echinoids in which test plates tend to overlap while body size does not increase anymore (Mespoulhé 1992). However, this slow growth rate in aged specimens could also result from other causes affecting optimal food intake for instance. At the calibration site of Anse du Halage, a study of echinoid cohorts suggests that few individuals grow older than six years old (Poulin and Féral 1994). Overall, the absence or nearly absence of growth in old invertebrate organisms makes age estimates delicate to assess. In the present model, based on the combination of the ageing sub-model and other mortality processes, most individuals are calibrated to die within the assumed maximum age (before 6 years old), although some individuals may reach over ten years old due to the chosen stochasticity introduced in the sub-model.

Juveniles inside brood pouches were assumed to grow at a constant and same rate as adults but it has sometimes been assumed that the brooded young may already feed and develop at a faster rate (Schatt 1985, Schatt and Féral 1996). At this stage, offsprings are particularly fragile and need protection in the brood pouches to survive, which prevents any monitoring of growth rates and feeding behaviours (Magniez 1983, Schatt 1985, Mespoulhé 1992).

4.3. Ecological relevance of the IBM population model

Upscaling the DEB-individual model to the population level in the IBM enables to simulate population structure and dynamics as a response to temperature and food resource availability. In particular, the IBM enables to predict the targetted effect of environmental changes on the population at the different life stages of individuals. Additional environmental data would help enhance IBM reliability and improve our knowledge of populations and environmental conditions in remote areas.

Field works are also subject to uncertainties due to the species burrowing habit which renders the assessment of population structure difficult, the brittleness of specimens also limiting counting replicates (Magniez 1980, Mespoulhé 1992). Important variations in population densities were noted across studies (Guille and Lasserre 1979, Mespoulhé 1992, Poulin 1996, personal observations) for a same site, which may suggest either important variations in population density and structure through time, which was however refuted by Poulin and Féral (1994), or important biases in sampling due to the aggregative behaviour of individuals and the patchiness of distribution patterns (Poulin 1996).

The sensitivity analysis (Appendix 1.9) showed that the model is not very much dependent on assumptions made on initial population densities because the model density-dependent regulation operates through intra-specific competition for food resources only. There is no agonistic behaviour among conspecific individuals as it was reported in other echinoid species (e.g. *Echinometra sp*. Shulman 1990) and there is no evidence of competition for space in *A. cordatus* based on field observation. Intra-specific competition in shallow-water echinoids is a common phenomenon under food-limited conditions (Stevenson et al. 2015). McClanahan and Kurtis (1991) stated that in *Echinometra mathaei*, when predation pressure and intra-specific competition are low, populations

increase without limitation and regulation operates through a decrease in food availability for individuals. The same could hold true for *A. cordatus* as well.

Intra-specific competition for oxygen could also have a regulatory effect (Ferguson et al. 2013) since *A. cordatus* shows a high oxygen consumption rate (Guille and Lasserre 1979, Magniez and Féral 1988). In muddy substrates, specimens are usually observed unburied, positioned onto the sediment (instead of underneath), which was often interpreted as a result of difficulties to breath inside fine sediments.

The sensitivity analysis also showed that the number of eggs produced by females is a controlling parameter of model stability as well. There is a high variability in the number of eggs produced among females (from 9 to 106 eggs per female, personal communication from P. Magniez). Taking into account such a variability would introduce an enhanced stochasticity in the population dynamics model if implemented and linked to each female's reproductive buffer UR and GSI values (Martin et al. 2010, e.g. for zebrafish in Beaudouin et al. 2015).
Finally, the model was also shown to be sensitive to background mortality (Appendix 1.9). Although monitoring mortality rates in the field is challenging, such data would greatly enhance the reliability of the IBM.

In general, the sensitivity tests showed that the model works with the current quality and quantity of data available for this species in this habitat. However, on the matter of the temporal resolution, our model needs to be expanded and further consolidated, and we consider this element as a limitation to our work in its current state.

Modelled food resources
The organic content of sediments is one of the main food resources for detritus feeders and sediment ingestors like *A. cordatus* (Snelgrove and Butman 1994) and Antarctic echinoids (Michel et al. 2016). In the present model, the organic carbon content of sediments was used as a proxy for food availability for *A. cordatus*. Intra-specific competition for food has a stronger effect on resources availability than seasonal variations in resource availability. This is in line with ecological evidences that populations of *A. cordatus* survive periods of low food resources that prevail during the austral winter. High seasonality in food resources is a common feature of polar ecosystems and species have long adapted their diet accordingly (De Ridder and Lawrence 1982, Michel et al. 2016). This has been shown in Antarctic benthic invertebrates such as shallow-water brachiopods (Peck et al 2005), cnidarians (Orejas et al. 2001), and echinoids (Brockington et al. 2001, Ingels et al. 2012). For instance, the Antarctic sea urchin *Sterechinus neumayeri* is believed to be capable of mobilizing energy from gut tissues, gonads and the body wall during the austral winter (Brockington et al. 2001), a strategy that may have been evolved in *A. cordatus* as well (Magniez 1983). Shrinking and resorption, which are sometimes hypothesized as a survival mechanism in other echinoids facing long periods of starvation, are phenomena which are still understudied (David and Néraudeau 1989, Ebert 2013) and have not been verified in *A. cordatus*. In the present model, starvation results in the redirection of the energy flow exclusively towards maintenance of structure, at the expense of other compartments. Although Magniez (1983) observed a decrease in gonadal material after the reproduction period, it is very small in females (- 0.3%) and slightly bigger in males (-1.6%), and the exact cause has not been studied. It is not known whether this decrease in gonadal material can be directly attributed to a reabsorption for survival purposes or some other mechanism. The use of previously stored energy in the different compartments to sustain the maintenance of structure is assumed to be non-existent in our model. Such starvation processes could be tested in future implementation, provided sufficient data is obtained through experimental setups observing the phenomenon.

The two scenarios of future food availability were based on coarse IPCC and NOAA projection models for the region. These simulations and associated outputs are here considered as conceivable scenarios of the influence of food and temperature changes on population dynamics. They are used as a proof of concept and are by no means considered as definite and reliable scenarios of population dynamics in the future. Future accurate predictions should imply the integration of complex mechanisms influencing the production, transport and deposition of organic

matter in the ocean, the possibility of species to adapt to changing environmental conditions, and more experimental data are needed to integrate the detailed influence of temperature on physiological processes. First observations suggest a low response towards the applied changes in food availability, in comparison with the influence of temperature. However, it cannot be concluded that the species would not be affected by future conditions in food resources in the area.

Temperature resilience
Important differences were obtained between population structures and densities depending on future scenarios and model projections made for contrasting food resources and temperature. Most importantly, the 'resistant' population model of *A. cordatus* at Anse du Halage is predicted to sustain the expected changes in temperature and food resources under both future scenarios although population density is also predicted to be strongly reduced. In contrast, the 'vulnerable' population model predicts population extinction after a few decades of simulation. This implies that a precise evaluation of the species resilience to temperatures is needed for more robust and decisive models. Moreover, Antarctic echinoids were shown to present varied responses to ocean warming depending on species and life stages, with higher vulnerability to warm temperatures in juveniles than in adults (Ingels et al. 2012). Such a contrast suggests that more data could help fine-tune the present model.

4.4. Relevance of the DEB-IBM approach for Southern Ocean studies

DEB-IBM models are being developed for various applications and research fields. They are considered a powerful tool for environmental risk assessment, such as the effect of toxicity (e.g. Beaudouin et al. 2015, David et al 2019, Vlaeminck et al. 2019) and the impact of environmental changes on population dynamics (e.g. Saraiva et al. 2014, Malishev et al. 2018, Thomas and Bacher 2018, Goedegebuure et al. 2018). They can also be used to predict the behaviour of microbial systems (Jayathilake et al. 2017) or bring to light underlying mechanisms of life history strategies (e.g. Gatti et al. 2017). The DEB model brings ontogenetic and phenotypic variations to the population model while the IBM brings stochasticity, population dynamics (e.g. competition for food), as well as learning and interaction mechanisms (DeAngelis et al. 1991, Martin et al. 2012) to complement the model. The potential of the DEB-IBM approach resides in the combination of both models to predict population dynamics as a response to changing environmental conditions (i.e. at the individual level in the DEB model and at population level in the IBM).

In the present work, the DEB-IBM was used to improve our understanding of the dynamics of *A. cordatus*' populations. Applications could be further developed to address conservation issues such as the designation of priority areas and the definition of management plan strategies. Vast areas of the French Southern Territories have recently been placed under enhanced protection of a national nature reserve based on experts' knowledge and ecoregionalisation approaches (Koubbi et al. 2010, Fabri-Ruiz et al. 2020). Most areas however could not have benefited from thorough benthic field studies, and ecological models can represent interesting tools to assess the relevance of defined protection areas for target species and ecosystems. Such models can be useful when drafting management plan strategies for determining favored ship traffic routes or areas where human activities can be implemented in coastal areas of the national nature reserve of the French Southern Territories. Dynamic population models allow testing different ecological scenarios in a quite straightforward way to illustrate research designs and proposals. They can provide some clues to investigate the potential effect of environmental changes on key species for which conservation efforts should be directed in a short to long-term strategy (Fulton et al. 2015). Dynamic models can also prove useful for adaptable conservation strategies like the designation of dynamic protected areas as a consequence of changing environments and ecosystems. Finally, dynamic models could be further implemented into studies of ecosystem functioning and the impact of environmental changes on the alteration of sub-Antarctic ecosystems.

APPENDIX 1.5. Forcing environmental conditions

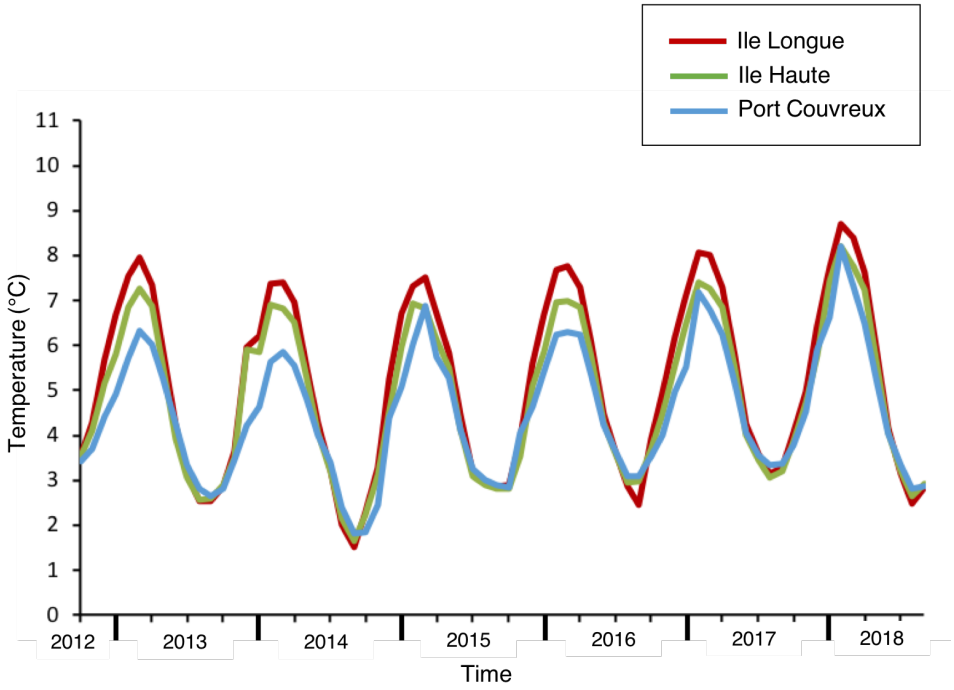


Figure S1.5.A. Onsite temperature records (monthly mean values) at the three sites used in the model. Red: Ile Longue (reference site for Anse du Halage), Green: Ile Haute, Blue: Port Couvreux. Data provided by IPEV program Proteker (n°1044); Ile Longue: http://www.proteker.net/Ile-Longue-5m-depth.html?lang=en; Ile Haute: http://www.proteker.net/Ile-Haute-5m-depth.html?lang=en; Port Couvreux: http://www.proteker.net/Ilot-des-Trois-Bergers-5m-depth.html?lang=en (accessed on 08/05/2019).

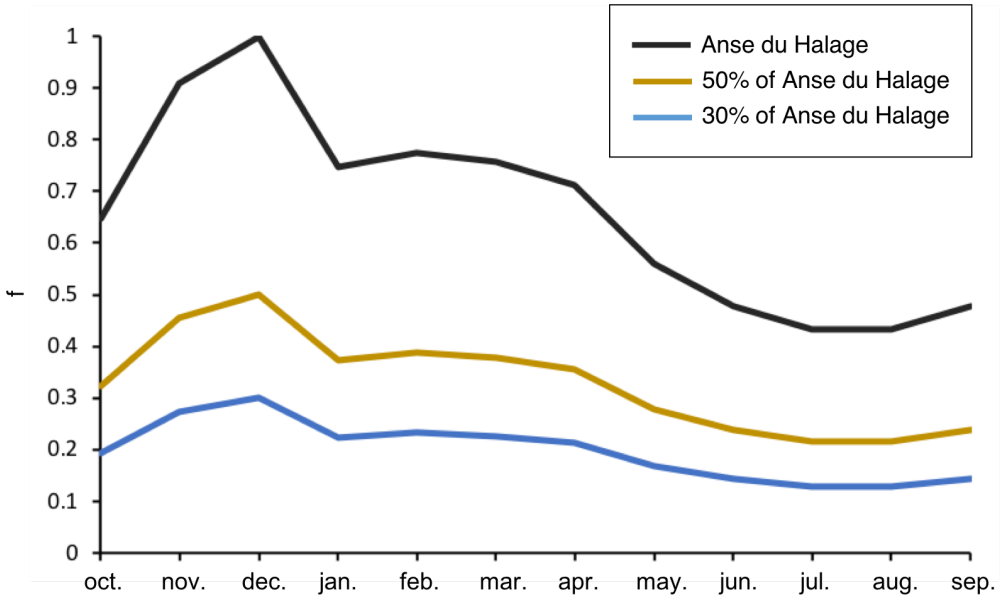


Figure S1.5.B. f values (food resources) used as input in the model. Black: Anse du Halage (data from Delille and Bouvy 1989). Yellow: 50% values of Anse du Halage. Blue: 30% values of Anse du Halage.

APPENDIX 1.6. Plots of uni-variate data from the DEB model

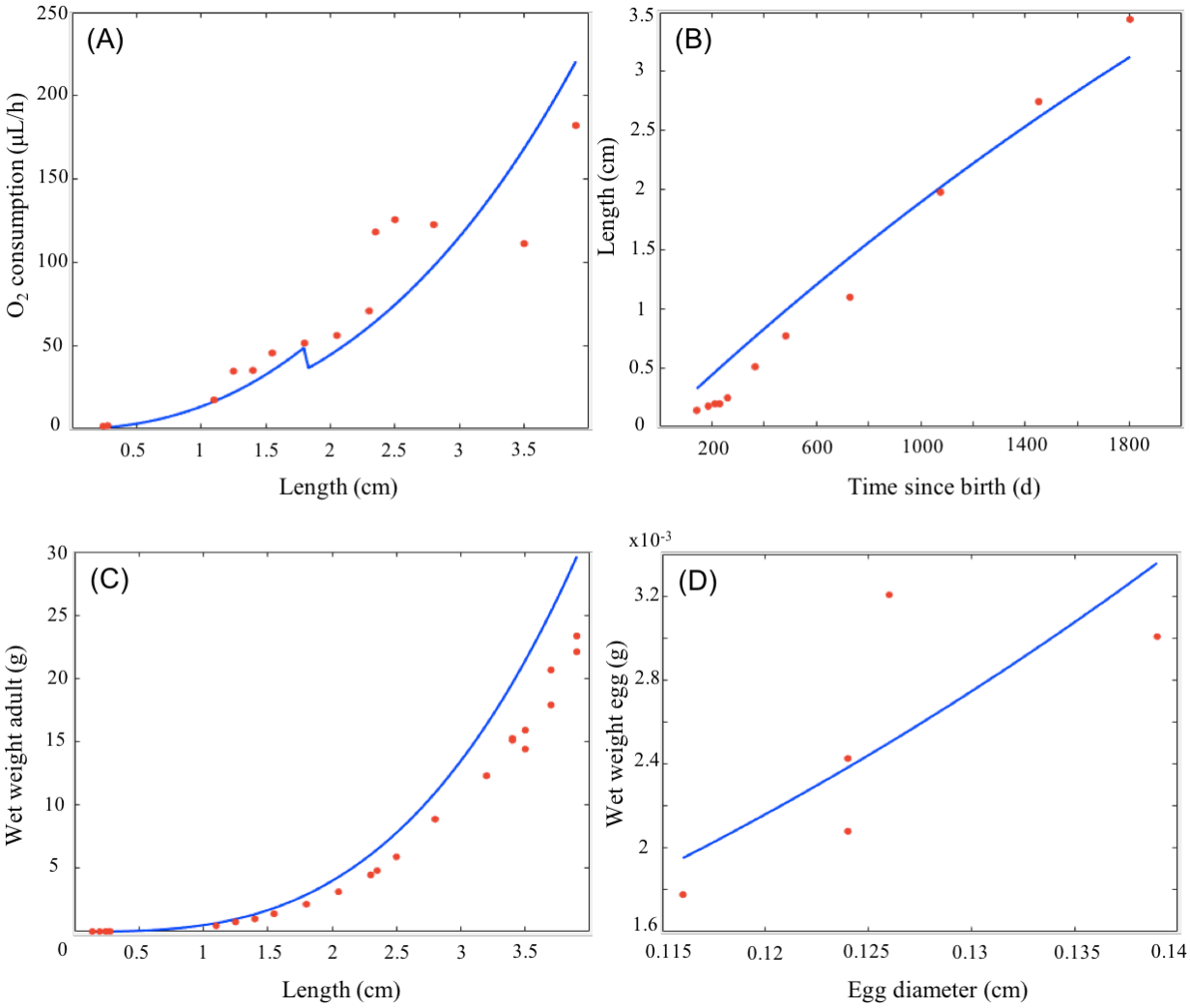


Figure S1.6. Uni-variate observations (red dots) used to calibrate the DEB model of *Abatus cordatus*. Blue lines correspond to model predictions, with Relative Error provided in Table 1.5. (A) Relationship between oxygen consumption (µl/h) and length (cm) (Féral and Magniez 1988), (B) growth rate (Schatt 1985, Mespoulhé 1992), (C) Length (cm)-Weight (g) relationship of adults (Féral and Magniez 1988), (D) Length (cm)-Weight (g) relationship of eggs (Schatt 1985).

APPENDIX 1.7. Influence of timestep changes

The sensitivity of the model to the chosen timestep was tested by comparing model outputs of the initial monthly calibrated model with a daily implementation (simple repetition of the monthly value for each day of the corresponding month).

In order to assess whether models can similarly predict individual performances when calibrated with different timesteps, models generated with a monthly and a daily timestep were compared. In these models, all processes outside of the individual development and the production of new offspring into the model were deleted (no competition, no mortality of any cause). Only individual development processes were kept, including starvation and reproduction, which directly influence individual energy fluxes without any stochastic effect. This facilitates model comparison. The population was assumed to be composed of female individuals only that reproduce following the procedure explained in the main manuscript. The model was initialized with 120 female entities, and individual metabolic performances (dUR, dUE, dL) over five years were compared between the two models.

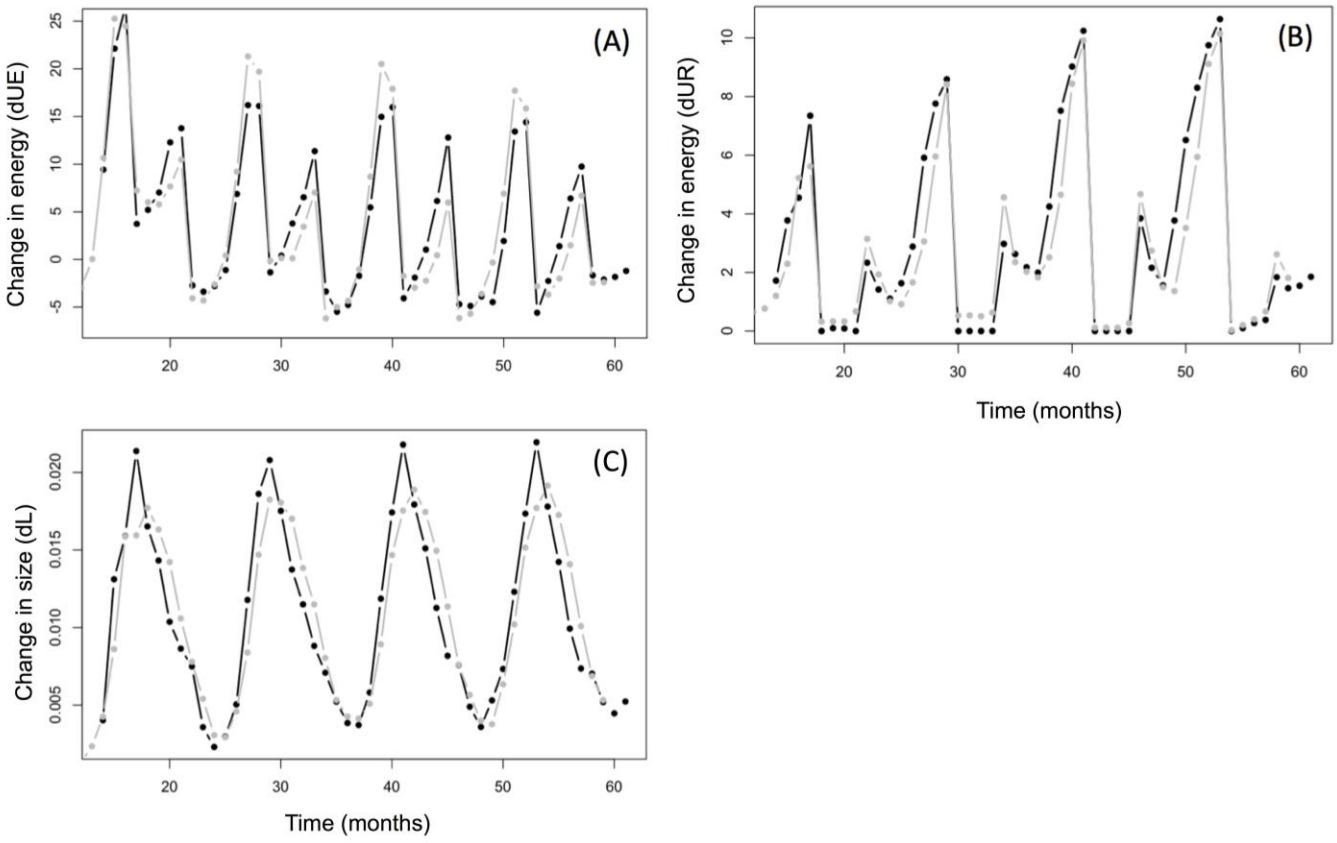


Figure S1.7. Comparisons of individual metabolic performances between models calibrated with a monthly (black line) or daily (grey line) timestep: (A) variation in scaled reserve (dUE), (B) in the reproduction buffer (dUR) or (C) variation in structural length (dL) over time. Models are generated without any process structuring populations except for reproduction events (i.e. recruitment of new juveniles). Average values of all individuals of the population are represented.

These results strongly highlight that models calibrated with different timesteps present very close patterns, and suggest that changing model timestep does not influence the shape and order of magnitude of individual metabolic performances predictions.

In our model, time is continuous for individual processes (individual development is modelled using ordinary differential equations), but not for all population processes which were taken from literature and experimental sources. These population processes are based on a monthly scale and used as a baseline to model population dynamics over time. Running the model at a smaller timestep implies altering all population processes to fit a narrower time increment, which is also not always relevant ecology-wise when studying population dynamics, since *A. cordatus* is a slow growing individual that lives in stable environmental conditions. Environmental changes do not occur often enough to significantly influence individual metabolism on a day-to-day basis and to consider mortality due to temperature changes at a daily step.

Similarly, information used for population background mortality rates were only available in the literature on a yearly range (Mespoulhé 1992, Poulin 1996, Ebert et al. 2013), and applying mortality for each day-step seems inappropriate for our study. Recruitment of newborn juveniles is also a yearly event, rendered possible when the reproduction buffer contains enough energy at a certain period of the year to enable females to release gametes. Reproduction development (GSI, reproduction buffer) is a continuous process in the model but specific reproduction events (releasing gametes, brooding and releasing offspring) are more fitting to monthly rather than daily triggers. In consideration of the ecological basis of these population processes and the very low sensitivity of the individual model to changes in timestep, it was therefore decided in this study to implement the model on a monthly timestep, although Fig. S1.7 proves that methodologically speaking, a different timestep could be applied to the core of our model.

APPENDIX 1.8. Future environmental scenarios and population resistance scenarios

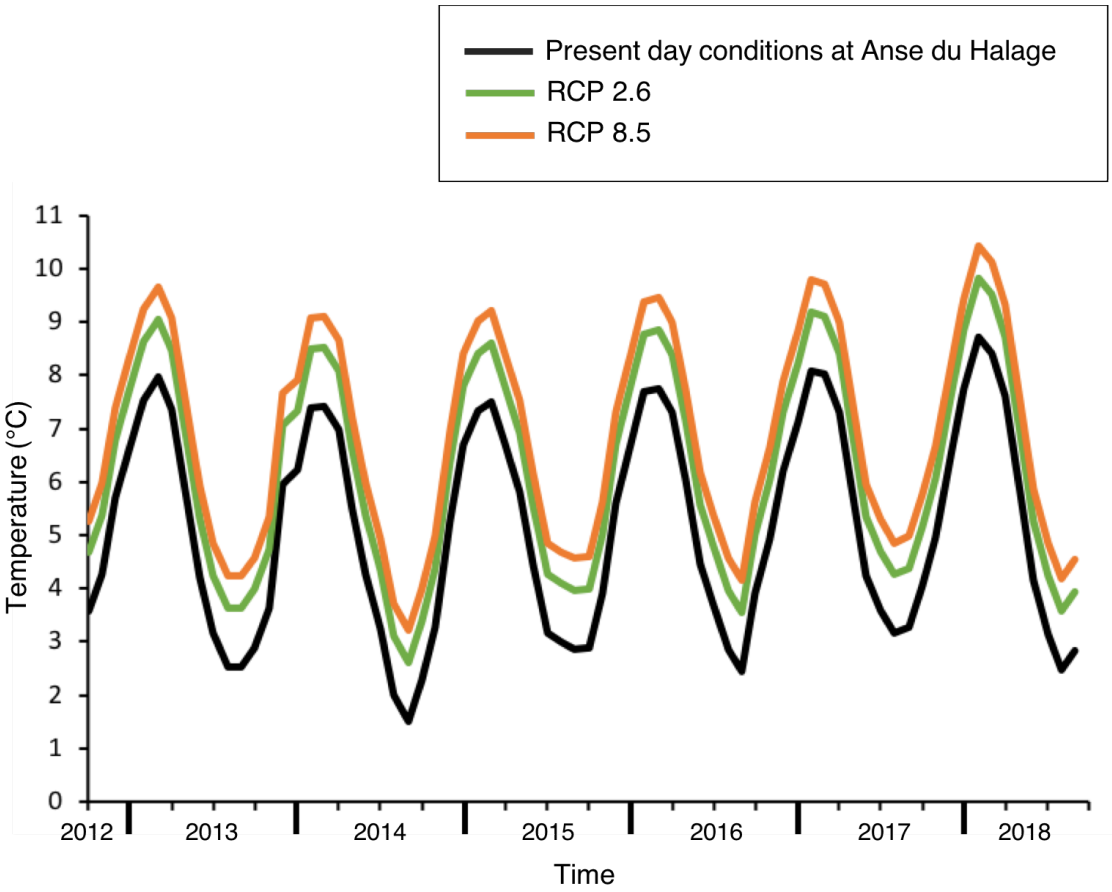


Figure S1.8.A. Temperatures for the different future projections based on the 2012-2018 dataset. Black: reference temperatures at Anse du Halage under present-day conditions. Green: Projection for scenario RCP 2.6 (+1.1°C warming). Orange: Projection for scenario RCP 8.5 (+1.7°C warming).

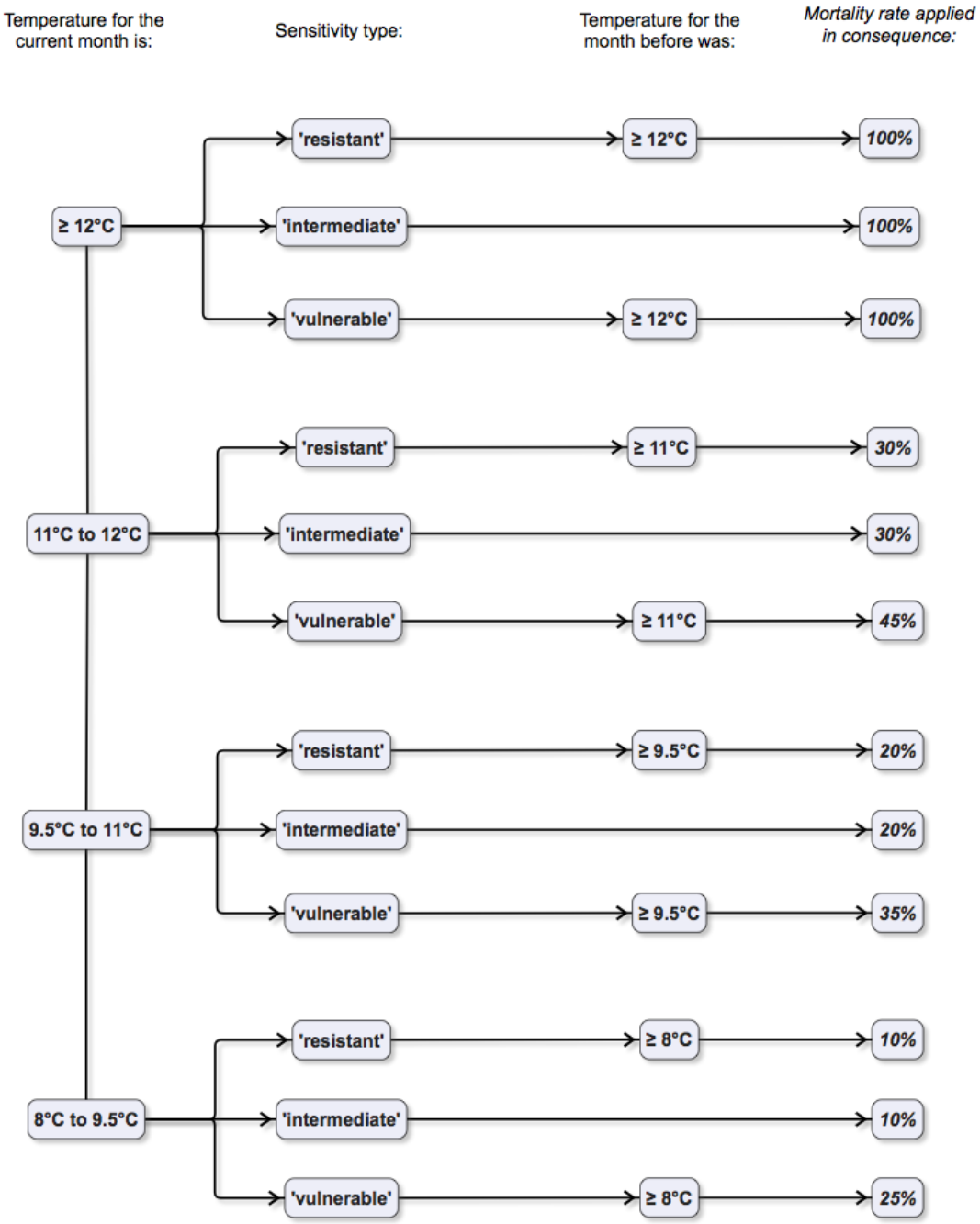


Figure S1.8.B. Decision tree explaining the three types of sea urchin sensitivity available in the model for the population temperature mortality rates. A triple population resistance scenario was used: 'resistant', 'intermediate, 'vulnerable'. For each threshold of the gradient of population mortality due to temperature, the 'resistant' population endures a mortality rate 15% lower than the 'sensitive population' (i.e. at 8°C for two months, the 'resistant' population suffers a 10% mortality rate, the 'sensitive' population suffers a 25% mortality rate). For the 'short resistance' population, rates are the same as for the 'resistant' population, but mortality takes effect only after one month of temperatures reaching over the threshold (rather than two months in the other cases).

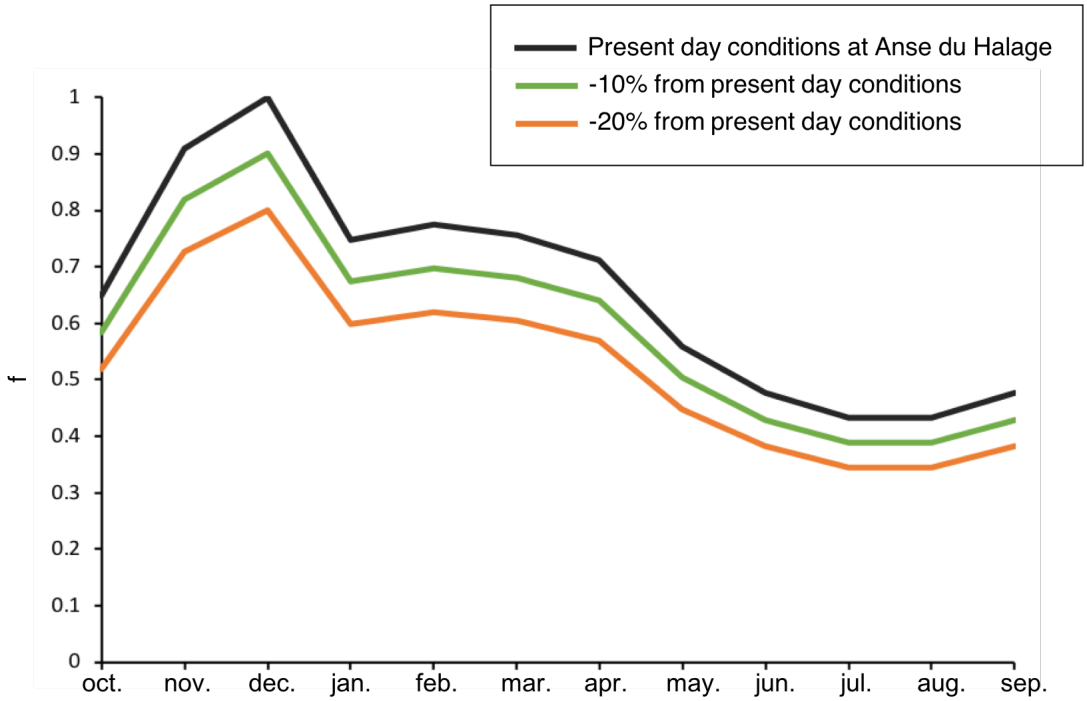
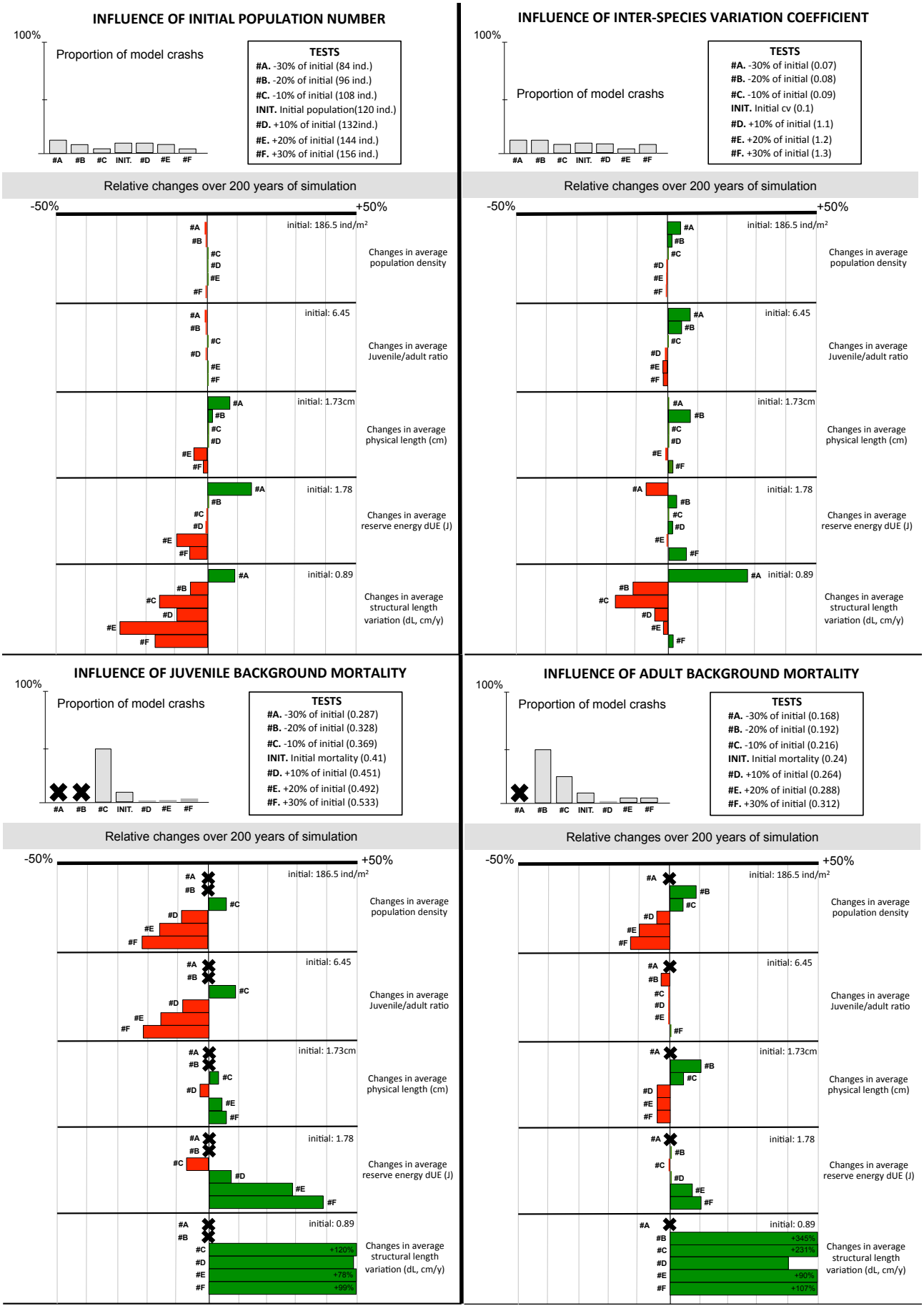


Figure S1.8.C. f values (food resources availability) estimated over one year for the different future projections. Black: reference f at Anse du Halage under present-day conditions from Delille and Bouvy (1989). Green: Projection for scenario RCP 2.6 (linear decline of 10% of food availability compared to present-day conditions). Orange: Projection for scenario RCP 8.5 (decline of 20% of food availability compared to present-day conditions).

APPENDIX 1.9. Sensitivity analysis



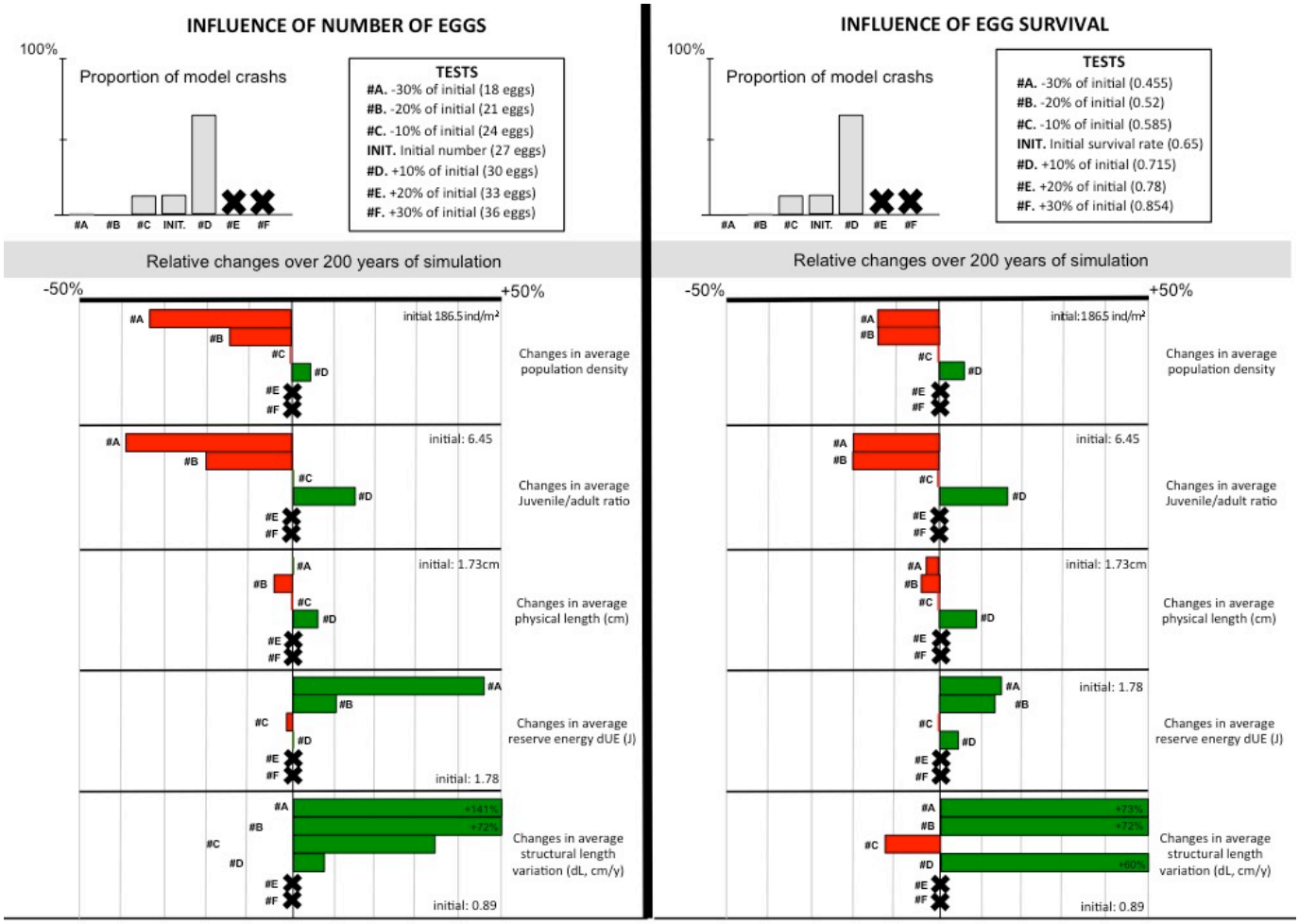


Figure S1.9. Model sensitivity to (a) the initial population number, (b) inter-species variation coefficient, (c, d) juvenile and adult background mortalities, (e) egg number produced per female during a reproduction event, and (f) the egg survival rate. Variations of -30%, -20%, -10%, +10%, +20% and +30% were tested for these parameters (#A to #F). For each analysis, the model was run until 100 simulations of 210 years of simulation were obtained. The model is considered as a 'crash' when the population is not stable and collapses before the end of the simulation period. The proportion of crashes relates to the number of crashes counted for 100 simulations (i.e., for 15 crashes and 100 simulations, the proportion is 15/(100+15)). Due to computing time limitations, the analysis was stopped when reaching a proportion higher than 66% of crashes (indicated by a black cross).

The percentage of changes obtained between the initial and the #A to #F scenarios values was calculated for average population density (ind/m^2), average juvenile/adult ratio, average physical length (cm), average reserve energy dUE and average structural length variation (dL) over the period of 200 years (210 years minus the first 10 years needed for model calibration). Decreasing values are indicated in red, increasing values in green.

APPENDIX 1.10. Individual growth under different scenarios

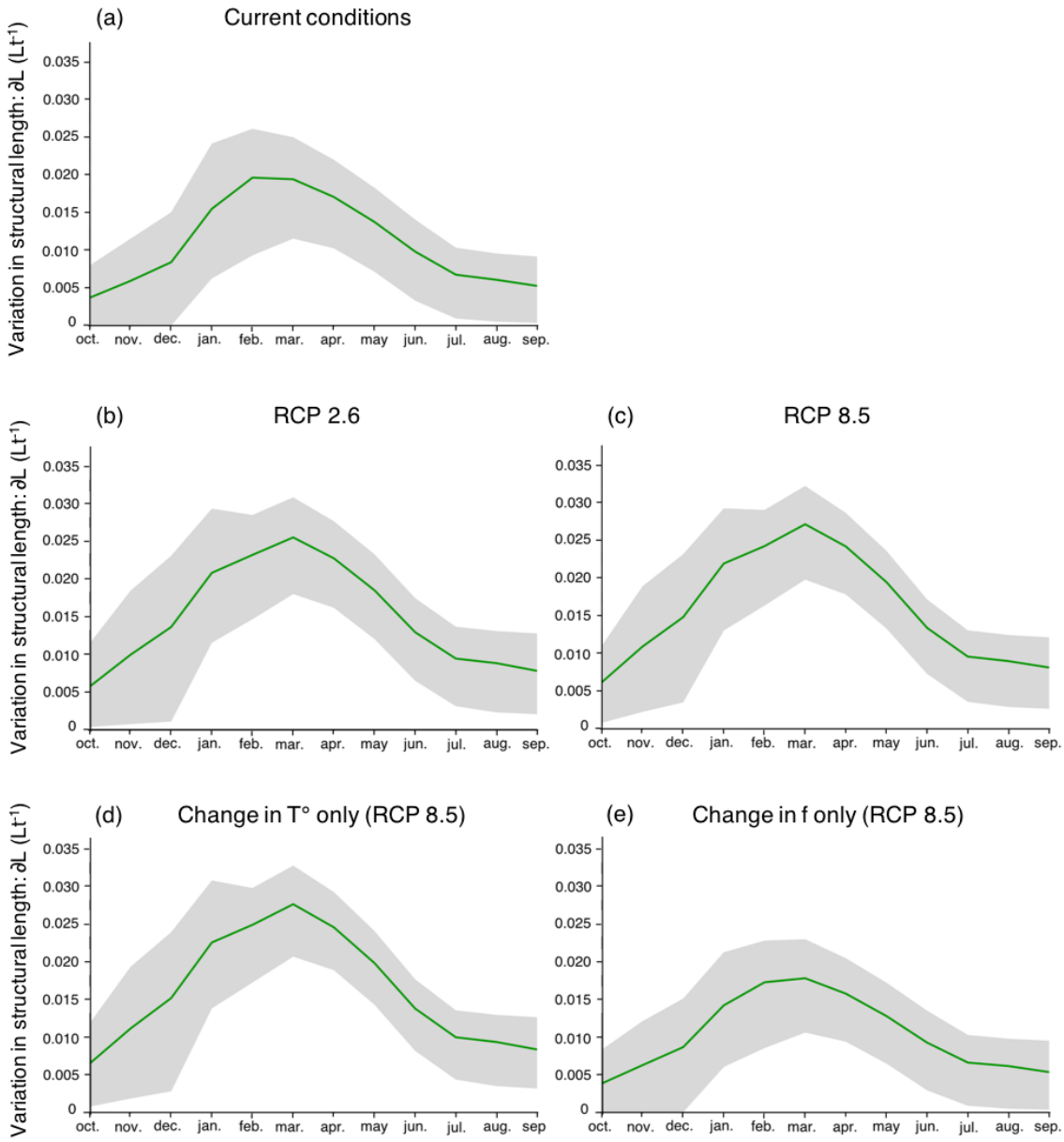


Figure S1.10. Simulation of the monthly variation of structural length (∂L, stands for $\frac{dL}{dt}$, dimension L.t^{-1}) over one year. Average results for all individuals (all sex and age), in fifty simulations, are presented by the green line. The grey area corresponds to the variation range (variation induced by differences between individuals: age, size, energy allocation) between all individuals, with the same number of individuals at model initiation (population then varies over time). (a) Model under present-day conditions; (b) model under IPCC scenario RCP 2.6 of T° and f changes (-10% of f and +1.1°C compared to present); (c) model under IPCC scenario RCP 8.5 of T° and f changes (-20% of f and +1.7°C compared to present); (d) model under IPCC scenario RCP 8.5 of T° change only (+1.7°C compared to present); (e) model under IPCC scenario RCP 8.5 of f change only (-20% compared to present).

These results give an illustration of what is discussed in section 4.2.1: In the DEB model of this work, we do not have the data to infer the descending slope of the Arrhenius curve, that is the temperature range beyond the optimal temperatures in which the metabolic rates slow down with higher temperatures. Thus, in its current implementation, the model gives better results at the individual level when confronted to higher temperatures, which is not in accordance with field and experimental observations. This is corrected at the population level with the use of the rate of mortality induced by temperature.

APPENDIX 1.11. Projection of the population model

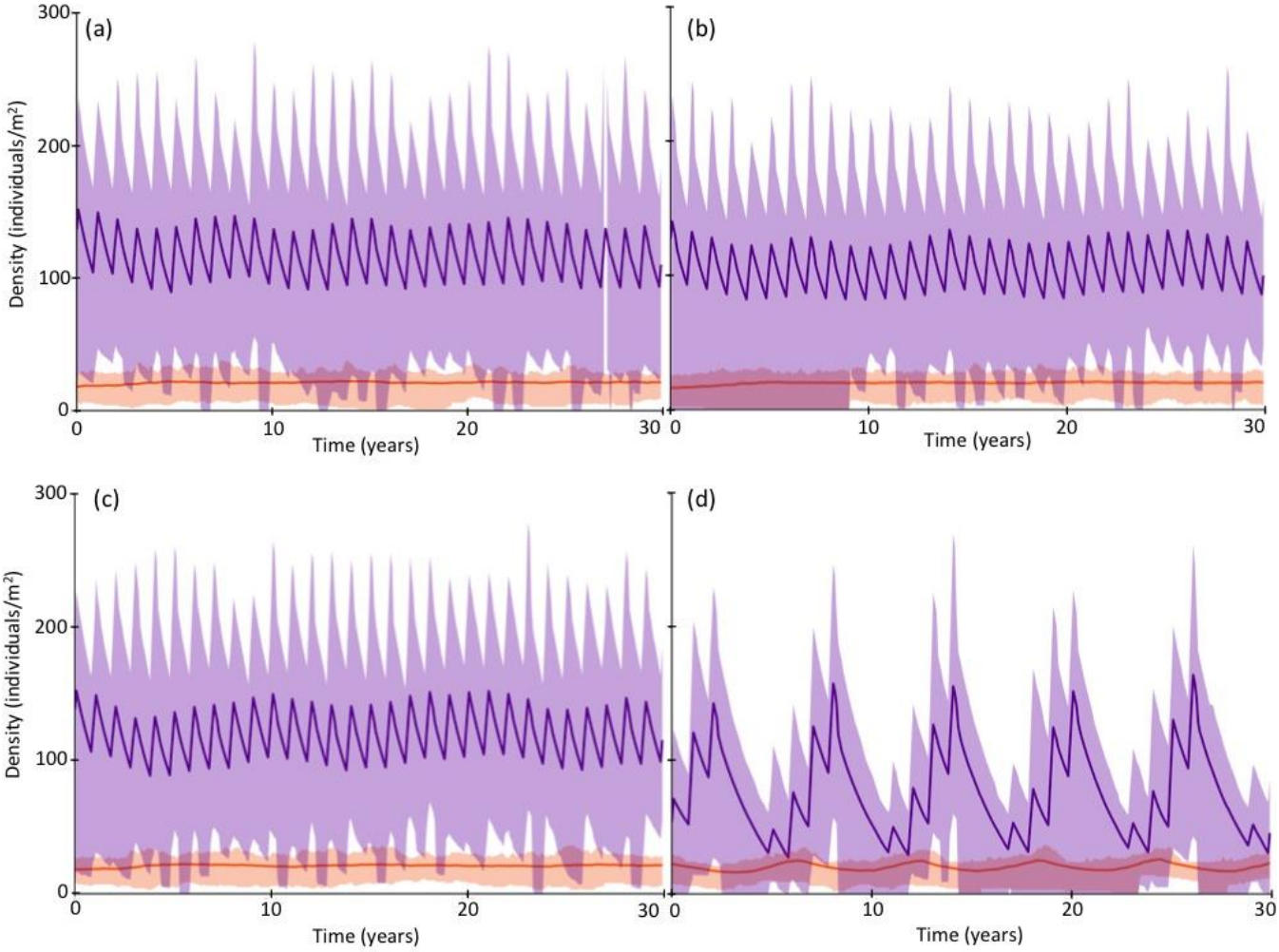


Figure S1.11. Modelled population structure and density under current environmental conditions calibrated at Anse du Halage and projected for two sites: Ile Haute (a, b) and Port Couvreux (c, d). Monthly density of juveniles (purple) and adults (orange) over 30 years (100 simulations). Bold lines: mean density value. Shaded areas: variation range. (a) Projection at Ile Haute based on local temperature records and 50% *f* values (Fig. S1.5.B, yellow), (b) Projection at Ile Haute based on local temperature records and 30% *f* values (Fig. S1.5.B, blue), (c) Projection at Port Couvreux based on local temperature records and 50% *f* values (Fig. S1.5.B, yellow), (d) Projection at Port Couvreux based on local temperature records and 30% *f* values (Fig. S1.5.B, blue).

The pattern observed at Port Couvreux with *f*=30% of *f* values estimated at Anse du Halage (Fig. S1.7) is due to local conditions of low food availability and temperatures (2013 and 2014 temperature data for Port Couvreux, Fig. S1.5.A, blue) impeding the conception of new individuals when females are few and do not reach GSI values high enough to reproduce. Some years, background mortality is not compensated by new cohorts and the population continuously decreases. The cyclic pattern is controlled by temperature data input in the model (six years period transposed for the entire simulation time).

APPENDIX S1.12. ODD of the DEB-IBM model

The ODD is available in the "Info" section of the NetLogo code, found at http://modelingcommons.org/browse/one_model/6201.

MODEL PRESENTATION

The present DEB-IBM (Dynamic Energy Budget - Individual-Based Model) was built to simulate and predict population dynamics in an endemic common benthic species of the Kerguelen Plateau (sub-Antarctic region), the sea urchin *Abatus cordatus*. It upscales the individual mechanistic DEB model to the population level, enabling to model the population dynamics through time as a product of individual physiological responses, and to predict the species response to a changing environment through comparisons between sites and between predicted future scenarios. The main objective of this work was to develop the model using the available data for this species living in a remote environment that is impacted by climate change and where logistical challenges strongly hinder the scientific research.

How to run this model ?

A few simple steps are necessary to run this model under its basic implementation:

1/ Download the model and the environmental files from the ''Files'' tab in the NetLogo modeling commons (http://modelingcommons.org/browse/one_model/6201#model_tabs_browse_files).
For the basic implementation, the two files needed are *"temp_time_monthavg_Halage.txt"* for temperatures and *"inputRSces_M_f_Delille.txt"* for resources. Make sure the model (.nlogo) and the data (.txt) files are stored in the same folder on your computer.

2/ Once you have opened the model (.nlogo), the interface is generally the first thing that is visible.
Navigation between the interface, the information page, and the code of the model is done through the three tabs at the top of the software ('Interface', 'Info', 'Code').
Make sure that the following elements are selected in the interface:

- *Sites*: 'Anse du Halage'
- *projection*: 'present'
- *future*: 'mixed temp & food'
- *sensitivity*: 'resistant'
- *competition*: 'On'
- *run_time*: '210'
- *cv*: '0.1'
- *add-my-pet?*: 'Off'

Also ensure that none of the green boxes ('input paramaters') is empty. If any of them is empty, switch the *add-my-pet?* button 'On' and fill the relevant boxes with the basic DEB parameters for *A. cordatus* as taken from the Add-my-Pet database (https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/entries_web/Abatus_cordatus/Abatus_cordatus_res.html): [$\dot{p}_M$], E_H^b, E_H^p, [E_G] and L_m, which correspond to these boxes respectively: p_M, E_H^b, E_H^p, E_G, zoom.

Except in the aforementioned case, do not modify any of the parameters in the green boxes placed under the line « Input parameters » on the interface.

3/ Click on the purple *setup* button. This initializes the model, and should barely take a second on an average computer. A sure way of knowing the model has finished setup, is that color shapes appear in the small black square that is on the bottom-right of the purple buttons.

4/ Once setup is finished, click on the purple *go* button. This will run the model for the simulated duration input in the *run_time* box (number of years). Clicking on the *go* button again before the end of the simulation will pause the model, clicking on it after the end of the simulation will continue the simulation without a temporal limit. The *go once* button will only run the model for a single loop, that is a simulation of one month.

5/ The model can be run for future projections with different combinations of food and/or temperature scenarios. For this, select in the interface the desired RCP scenario under *projection* and the wanted combination under *future*. Three types of sensitivity to high temperature are also available under *sensitivity* (see « Check temperature » submodel in the ODD below for a short explanation of the difference between the three types).

For now, the model only fully works if the site Anse du Halage is selected. Sites Port Couvreux and Ile Haute can also be selected, however the site-specific data is only available for temperature and not for resources. Thus, if one of those two sites is selected, the file for the resources data will be that of Anse du Halage. This is merely to test the model for different temperature data using real time-series records rather than future projections.

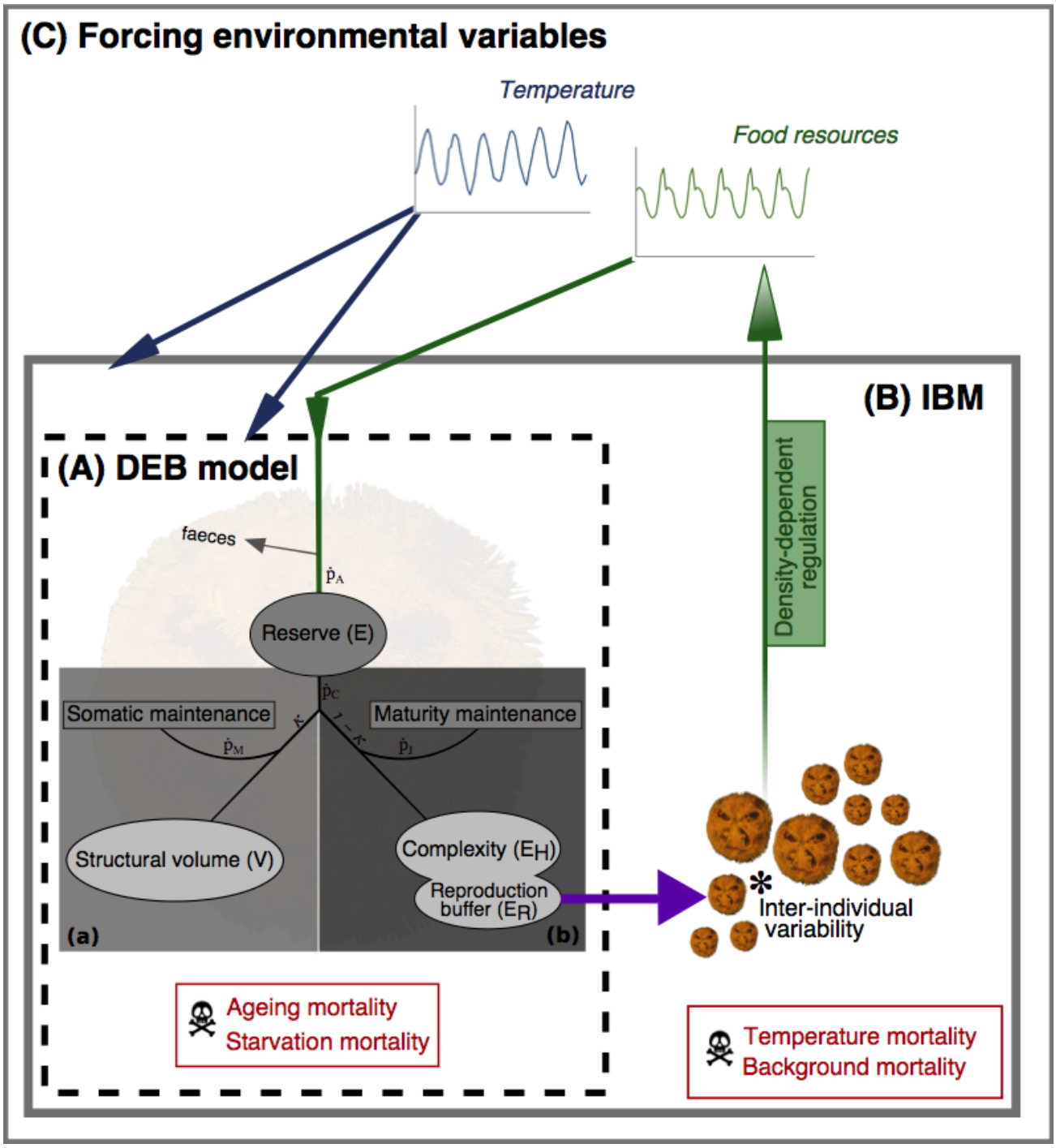
NB: Here again, make sure that the relevant temperature files (*"temp_time_monthavg_Couvreux.txt"* or *"temp_time_monthavg_Haute.txt »*, also available in the ''Files'' tab in the NetLogo modeling commons) are stored in the same folder as the model.

Other things to keep in mind in this implementation:

- The temporal resolution is a monthly interval.
- Temperature and functional response f are in the form of time-series. (To test with constant and DEB standard values, uncomment the corresponding lines at the beginning of the *go* procedure in the code).
- This species lives in waters of 5°C in average, however the upper limit of the Arrhenius relationship is not available to model *Abatus cordatus* physiological response to higher temperatures. Because of this, the mortality due to temperature was manually forced in the model using survival data obtained during an experiment led in Kerguelen in November 2018.

MODEL DESCRIPTION

Here is a schematic representation of the DEB-IBM model and a short summary description (the letters in brackets refer to the figure):



Individuals (A) undergo development through the DEB model and reproduce (purple arrow). Altogether and with a slight inter-individual variability in DEB parameters (*), they form the population of the IBM (B) which undergoes population-specific processes (temperature and background mortalities) at the scale of a simple square metre patch at the reference site (C). The IBM population is embedded within this specific environment, whose environmental conditions (temperature and food resources) affect individual and population dynamics. Additionally, the population influences the resources availability following a density-dependence regulation.

Below is the description of the model following the ODD (Overview, Design concepts and Details) protocol from Grimm et al. (2010), where information on the following characteristics can be found:

1. **Purpose**
   Short presentation of the objective of the model
2. **Entities, state variables & scales**
   Parameters with their DEB notation, code notation, dimension and signification
   Types of entities in the model, their state variables
   Temporal resolution and temporal extent of the model, spatial resolution
3. **Process overview & scheduling**
   Pseudo-code of a simulation (what the model does in one 'go once' simulation)
   Order in which agents execute commands, when variables get updated
   How time is represented in the model, presentation of calendars and timers
4. **Design concepts**
   - •Basic principles — Underlying DEB theory and principles
   - •Emergence — Emergent and imposed results
   - •Adaptation — Adaptative traits of individuals
   - •Interaction — Individuals interactive behaviours
   - •Stochasticity — Randomization in processes
   - •Collectives — How individuals are grouped under types of entities, specific characteristics
   - •Observations — Outputs of the model for results or tests
5. **Initialisation**
   Elements to select and parameters to input for initialisation
   Initial state of the model at setup
   Origin of initial parameters, Initialisation of individuals
   What differs from one initialisation to another
6. **Input data**
   Data taken from external files, How the files are compiled and read
   Specific data for the standard basic model (Anse du Halage)
7. **Submodels**
   Detailed description of each submodel, with equations and processes:
   - •Update calendar
   - •Update environmental variables
   - •Competition and f
   - •Convert parameters with TC (temperature correction factor)
   - •Change in reserve
   - •Change in maturity or reproduction buffer
   - •Change in structural length
   - •Starvation
   - •Ageing
   - •Update individuals
   - •Update reproduction and birth timers
   - •Reproduction
   - •Calculate GSI
   - •Background mortality
   - •Check temperature
   - • Population monitoring
   - • Update time

1. Purpose

The model was developed from the individual mechanistic Dynamic Energy Budget (DEB) model to study the response of the populations of the sea urchin *Abatus cordatus*, endemic to the

Kerguelen Plateau (sub-Antarctic region), to changes in environmental conditions (temperature and resources), through comparisons between sites and between predicted future scenarios.

2. Entities, state variables and scales

DEB notation	Notation in the code	Dimension and description
e	e_scaled	(-) scaled reserve density per unit of structure, e = [E]/[E_m]
E_H^b	E_H^b	(e) maturity at birth
E_H^p	E_H^p	(e) maturity at puberty
[E_G]	E_G	(e·L^{-3}) volume-specific costs of structure
g	g	(-) energy investment ratio
$\dot{h}$	h_rate	(t^{-1}) specific death probability rate
$\partial\dot{h}$	dh_rate	(-) change of hazard rate in time
$\ddot{h}_a$	h_a	(t^{-2}) Weibull ageing acceleration
κ	kap	(-) fraction of mobilised reserve allocated to soma
κ_R	kap_R	(-) reproduction efficiency
$\dot{k}_M$	k_M_rate	(t^{-1}) somatic maintenance rate coefficient
$\dot{k}_J$	k_J_rate	(t^{-1}) maturity maintenance rate coefficient
L	L	(L) structural length
∂L	dL	(-) change of structural length in time
L_w	Lphy	(L), physical length (L/∂_M), with ∂_M being the shape coefficient (-)
L_b	L_b	(L) structural length at birth
L_m	zoom	(L) maximum volumetric length
[$\dot{p}_M$]	p_M	(e·L^{-3}·t^{-1}) specific volume-linked somatic maintenance rate
$\ddot{q}$	q_acceleration	(t^{-2}) ageing acceleration
$\partial\dot{q}$	dq_acceleration	(-) change of ageing acceleration in time
S_A	S_A	(L^2) assimilation flux (scaled)
S_C	S_C	(L^2) mobilisation flux (scaled)
s_G	sG	(-) Gompertz stress coefficient
s_M	s_M	(-) acceleration factor
U_E	U_E	(t·L^2) scaled reserves, U_E = E / {$\dot{p}_{Am}$}
∂U_E	dU_E	(-) change of scaled reserves in time
U_H^b	U_H^b	(t·L^2) scaled maturity at birth
U_H^p	U_H^p	(t·L^2) scaled maturity at puberty
U_H	U_H	(t·L^2) scaled maturity
∂U_H	dU_H	(-) change of scaled maturity in time
U_R	U_R	(t·L^2) scaled energy in reproduction buffer
∂U_R	dU_R	(-) change of energy in reproduction buffer
$\dot{v}$	v_rate	(L·t^{-1}) energy conductance
y_{VE}	y_VE	(mol/mol) yield of structure on reserve
Other variables	**Notation in the code**	**Dimension and description**
f	f	(-) scaled functional response
K	car-cap	(#) carrying capacity
density	pop_density	(#/m^{-2}) current population density
competition	food_compet	quantification of the scale of competition
TC	TC	(-) temperature correction factor
GSI	GSI	(-) gonado-somatic index
eggs	eggs	(#) number of eggs per female
Ri	Ri	(#) reproductive output (number of juveniles born per female)
Scatter multiplier	scatter-multiplier	(-) parameter used to put a random variation in individual parameters

The model includes two types of entities: individuals and the environment. Individuals are divided into 4 types of sub-agents, depending on life-stage and sex: embryos, juveniles, adult males and adult females. Life-stages are considered using DEB definitions. Individuals are characterised by four primary state variables, based on DEB theory (Kooijman 2010): scaled

reserve (U_E), structural length (L), scaled maturity (U_H) and scaled reproduction buffer (U_R). Additional state variables for individuals are age (in months and in years), ageing acceleration ($\ddot{q}$) and death probability rate ($\dot{h}$). State variables in DEB are originally dependent on energy (unit of Joules), but in order to simplify the calculations of differential equations so that they do not require measurements of energy, each state variable was discarded of their energy unit by dividing with the maximum surface-area specific assimilation rate {$\dot{p}_{Am}$} (dimension $e.L^{-2}.t^{-1}$) following Martin et al. (2010) and Kooijman (2010). Females have supplementary attributes linked to reproduction processes for which the state variable is the gonado-somatic index (GSI) that is the ratio of gonad weight over the weigth of the entire body. Finally, each individual has a variable called scatter-multiplier, used to implement a slight variation in three standard DEB parameters (U_H^b, U_H^p, g) and initial energy reserve at birth (U_E_0yo). This scatter-multiplier is the exponential of a number taken randomly on a normal distribution of mean 0 and standard deviation cv (set by the user in interface).

The environment in the model is characterised by two state variables, monthly average temperature (T, unit: °C) and monthly resources availability represented by the proportion of food an individual can intake on a scale of 0 to 1 (f, no dimension). Values for these variables are input into the model from external files, as time-series of monthly values. Temperature data was collected from existing thermo-recorders on the corresponding sites (implemented by the PROTEKER program, http://www.proteker.net/?lang=en), while resources data comes from the publication by Delille and Bouvy (1989) for the site Anse du Halage.

Models provide results at a monthly resolution, over a temporal extent that can be modified by the user in the interface (example set at 210 years). The first ten years are assumed to be the initialisation phase and should be removed for the analysis of results. Changes are applied to individuals on a monthly basis and thus each update corresponds to the state of the system at the end of the displayed month. In this implementation, the model runs on one single patch of environment representing one square meter, and thus density of population is equal to the number of individuals present in the model. Movements of individuals are not taken into account, and each individual born on the patch grows and dies on that same patch. There is no information about water movements in the area, and the species is known to mostly feed on sediment matter. This model is non-spatial, and connectivity between the patches (e.g. for food or individual movements) is assumed to have little enough significance to be absent from this model.

3. Process overview and scheduling

At each timestep, the model runs the following commands in that order:

```
Reset the death counts
 Update calendar

 (For each patch:
        Update environmental variables
        If competition ON [
                Calculate competition ]
        Calculate f)

 (For each individual:
        Remove if marked as deceased
        Convert relevant parameters with temperature correction factor
        Calculate change in reserve
        If not mature [
                Calculate change in maturity]
        If mature [
                Calculate change in reproduction buffer ]
        Calculate change in structural length
        If scaled reserve < scaled length [
                Starve]
        Calculate ageing)

 Update individuals

 (For each individual:
        Update reproduction timers)

 (For each females:
        Update birth timer
        If first month of reproduction period [
                If reproduction ON [
                        Mark GSI down
                        Prepare eggs]]
        Calculate GSI
        If GSI >= 0.07 [
                Turn ON reproduction
                If month before reproduction period [
                        Mark U_R down
                        Launch reproduction (with birth_time)]]
        If GSI < 0.07 [
                Turn OFF reproduction]
        If reproduction ON [
                If within reproduction period [
                        reproduce]
                If within birth-giving period [
                        release offsprings]])

 Background mortality
 Check temperature
 Monitoring of population
 Update time
```

Individuals execute the same command one by one in a fixed order before going to the next command, but all have access to the same state of the environment since it updates once at the beginning of the timestep only.
In this model, time is represented continuously using ordinary differential equations (ODE) for the individual state variables, and all other variables are calculated in a discreet manner (every month, with one month rounded to 30.5 days).
Below is an illustration of the different calendars and timers used in the model, beginning at the point where the simulation starts (Gregorian calendar for reference):

repro_time	7	8	9	10	11	0	1	2	3	4	5	6	7	8	…
GSI_time	5	6	7	8	9	10	11	12	1	2	3	4	5	6	…
birth_time	1	0	0	0	0	8	7	6	5	4	3	2	1	0	…
month_time	10	11	12	1	2	3	4	5	6	7	8	9	10	11	…
Gregorian calendar	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	…

'repro_time' follows the reproduction year, runs on a twelve steps loop,
'GSI_time' follows the GSI cycle year, runs on a twelve steps loop,
'birth_time' is a countdown tracker for the time between the start of reproduction and the following release of offspring in a year. It is triggered by the launching of reproduction and counts backward, staying at 0 if not triggered. 'month_time' follows the Gregorian calendar

4. Design concepts

BASIC PRINCIPLES

The IBM was built using the DEB-IBM model developed by B. Martin for *Daphnia magna*, along with its DEB-IBM user manual and model description (Martin et al. 2010). The underlying theory for the individual development in the model follows the Dynamic Energy Budget theory (Kooijman 2010). The population is studied following the IBM principles (Railsback and Grimm 2019), as a dynamic system composed of autonomous individuals affected by the environmental conditions throughout their life-cycle. Each individual undergoes a continuous development from birth till death, following the DEB principles with a slight variation between individuals at their initialisation, and represents a component of the IBM population, which is itself affected as a whole by variables such as population death rates and density-dependent processes. The emerging state of the population is then observed, and compared between different scenarios of environmental variations.

EMERGENCE

The model illustrates the evolution of the population structure following the response of the individuals to the environmental conditions input. Metabolic responses, life-stages, ability to reproduce, starvation and ageing processes of the individuals, emerging from the mechanistic representation of their development, affect the population structure and average characteristics. A background mortality rate and a mortality caused by above normal temperatures are forced into the model, and the same reproductive output is imposed to all females that are able to reproduce.

ADAPTATION

Agents do not have an adaptive behavior. Individual traits vary among individuals in a population, but each individual carry the same traits along their entire lifespan and do not change nor learn from the events they experience or from each other. Consequently, the design concepts "objectives", "learning", "prediction", and "sensing" do not apply.

INTERACTION

Individuals do not have any direct interaction. They only affect each other indirectly, as the size of the population influences the resources availability and thus the capacity of each individual to access food.

STOCHASTICITY

In the model, stochasticity is used in the ageing submodel: there is a 50% chance that the ageing process is activated and observed for the individual. This stochastic element can be modified in the code by changing the numbers x and y in the 'update individual' procedure. Stochasticity is also implemented in four of the initial variables for each individual (scaled maturity at birth, scaled maturity at puberty, energy investment ratio, energy reserve at birth), using the scatter-multiplier, the exponential of a number taken randomly on a normal distribution of mean 0 and standard deviation cv (set by the user in the interface of the model, at 0.1 for the standard model). (taken from Martin et al. 2010, Kooijman 2010).

COLLECTIVES

The individuals are grouped under a particular type of entity depending on their life-stage and sex, and update their life stage along time: beginning at the "juveniles" type (from around 0 to 2 years old), they are

then belonging to the “males” or “females” type after reaching puberty (around 3 years old). The age at which a juvenile reaches puberty is an emergent property of its development. The sex is arbitrarily and randomly imposed on the individual that becomes an adult so that the sex-ratio (males/females) of the population is around 0.99. Depending on which group they belong to, some variables are different: juveniles do not modify their reproduction buffers, males and female do not modify their maturity compartment, females possess some proper variables such as GSI (Gonado-somatic index), eggs (number of eggs produced) and Ri (reproductive output). These collectives do not emerge from individual behaviour, but instead are implemented by the modeller in order to distinguish the life stages and sex of the individuals.

OBSERVATIONS

The main output of the model are plots of population structure with densities of population at the different life-stages, plots of the cumulative counts of individual deaths (and proportions of associated causes), plots of mean values in state variables U_R, U_E and L and change in these state variables (∂U_R, ∂U_E and ∂L) for the different individual types. These plots allow observing the response of the population to contrasting environmental conditions and individual metabolic responses in the population in relation to these environmental conditions. Additionally, plots of the mean age at death of individuals dying due to the ageing submodel were used to calibrate the ageing submodel itself.

5. Initialisation

For the standard model, the following elements must be selected in the interface:

- *Sites*: ‘Anse du Halage’
- *projection*: ‘present’
- *future*: ‘mixed temp & food’
- *sensitivity*: ‘resistant’
- *competition*: ’On’
- *run_time*: ‘210’
- *cv*: ‘0.1’

The initial DEB parameters can be calculated by the model if the *‘add-my-pet?’* switch is set to ON in the interface and the basic DEB parameters [$\dot{p}_M$], E_H^b, E_H^p, [E_G] and L_m for the species as taken from the Add-my-Pet database are input into the relevant boxes (respectively: p_M, E_H^b, E_H^p, E_G, zoom).
The standard model is run for 210 years in total for the site Anse du Halage under present-day conditions, with a population sensitivity set to ‘resistant’ and competition affecting resources availability. At setup, the values for the temperature and f at the site are taken from the time-series data found in the input files and compiled into lists usable by the model. The model is initialised with environmental conditions of October (month_time 10). If the model is set for future projections, the values are modified according to the chosen scenario (i.e. either one of RCP 2.6 and RCP 8.5 with food only, temperature only or food and temperature combined).
The carrying capacity is set at 200 ind./m^2 and the proportion of females at 0.5.

Initial parameters are based on *A. cordatus* DEB model parameters, developed in Guillaumot (2019c). Two simulation procedures are run for the initialisation of individuals: (1) a simulation of embryonic development to determine the initial reserve at birth, and (2) a simulation of the development of one individual from 0 till 5 years old at constant f and temperature values.
The first simulation uses a bisection method to determine the initial reserve at birth. The loop simulates the embryonic development from conception till birth, while testing for different scaled reserve *e* at conception. When the scaled reserve reaches the aimed value after a few loops of development, there are two possible situations: either the development is before or after the birth stage, and thus the loop is reset with new values of initial scaled reserve *e* set accordingly and the simulation relaunched, or the development is at birth stage and the value of scaled reserve *e* that was obtained is saved.
The second simulation starts off where the first one finishes, using the resulting reserve density at birth. It runs a loop for the development of the individual from birth till five years old, with standard parameters and a constant functional response f = 1. The simulation keeps track of the age of the individual, and for each year the values for the state variables are set aside and the simulation continues until the following year. These values are stored in variables and will be used to initialise models.

When running a model, an initial population of 120 individuals is created and contains a similar proportion of the individuals belonging to six age classes from 0 to 5 years old. Each of these individuals receives the set of parameters corresponding to its age class (variables stored in the ‘second simulation’ mentioned in the above paragraph), with stochasticity applied on some of these parameters (see section ‘Stochasticity’).

6. Input data

Each individual sets its calendars with GSI_time at 5 and repro_time at 7, and females are given initial values of 0.03 for their GSI and set their birth_time timer at 0.

The model reads environmental variables from input .txt files containing monthly time-series of food resources (from Delille and Bouvy 1989) and temperatures (PROTEKER program IPEV n°1044). The text files contain an ordered list of values (see below for Anse du Halage data). The temperature file contains 72 values of monthly average temperatures corresponding to temperature records from October 2012 to September 2018. The food resources file contains 12 values, taken from the measurements of organic carbon content in sediments published in Delille and Bouvy (1989) and scaled by the maximal value to create a proxy of f, contained between 0 and 1.

Data for resources and temperatures at Anse du Halage:

Resources:

month	f value
01	0.748
02	0.775
03	0.756
04	0.712
05	0.559
06	0.477
07	0.432
08	0.432
09	0.477
10	0.648
11	0.909
12	1.000

Temperatures:

time	mean temperature
0	3.576178523
1	4.272769444
2	5.693903226
3	6.680989247
4	7.540075893
5	7.95569852
…	
68	4.157559722
69	3.148745968
70	2.482346774
71	2.836618056

7. Submodels

Update calendar
The model timestep is a month.

Update environmental variables
The model takes the temperature and f values of the corresponding month from the input files.

Competition and f
The model calculates the current population density and quantifies the competition effect on food availability (FC, food competition) depending on how far from the carrying capacity (K) the population density (P) is:
If P < 1.9 * K, then FC = $(1- f_{env}).(1 - \frac{P}{2.K-P})$
If P ≥ 1.9 * K, then FC = $(1- f_{env}).(1 - \frac{P}{K/10})$

Two equations are used because if P = 2K, the first formula gives an error due to a division by 0, and if P > 2K, then the formula gives the untrue result of less competition with a bigger population (hence the use of 1.9 as a pivot value).
Competition is only effective if food availability is less than the maximum (hence the use of *'(1 – f)'* in the equation).

Then the model updates the *f* value in accordance with the quantified food competition:
$f_{eff} = f_{env} + FC$, where f_{env} is the food available in the environment as input into the model from the external files, and f_{eff} the effective food availability.

Proportionally to how much *f* is lessened compared to the maximum, the size of the current population has an influence on how important the competition is: If the population is below the carrying capacity (K), then food is more available for the present individuals, but if the population is over the carrying capacity, the availability of food is lessened.

Therefore,
if P > K, FC < 0 <=> f decreases
if P = K, FC = 0 <=> f constant
if P < K, FC > 0 <=> f increases

Meaning that the competition is actually calculated depending on how far from the carrying capacity is the population density, and how far from maximum *f* is the food availability (f_{env}). The less food available and the bigger the population, the higher the competition.
The regulatory effect of this competition lies in the starvation of individuals at lower food availability, which leads to a reduction of the population size with higher competition (combination of low food availability and big population size). *f* is always contained between 0 and 1.
If the competition is turned off, *f* is the direct value taken from the input list.

Convert parameters with TC
A temperature correction factor (TC) is calculated using the Arrhenius temperature (T_A) and applied to conductance $\dot{v}$, somatic maintenance rate $\dot{k}_M$, maturity maintenance rate $\dot{k}_J$ and Weibull ageing acceleration ḧa, which are all affected in the same way by the correction factor.
A given metabolic rate X at temperature T is thus modified with:
$X(T) = X * \exp(T_A/T_{ref} - T_A/T)$ with $\exp(T_A/T_{ref} - T_A/T)$ being the correction factor TC, T_{ref} the reference temperature (293.15K), T_A = 9000K, and T the actual temperature (in Kelvin) of the organism's life environment.

Change in reserve
The reserve is supplied from ingested food, that is represented in the model by the functional response *f* (from 0 to 1). Scaled assimilation rate S_A is found with $\dot{p}_A$ /{$\dot{p}_{Am}$}, where $\dot{p}_A$ is the assimilation flux (in energy per time) and {$\dot{p}_{Am}$} the maximal assimilation flux per surface area of structure (in energy per time per surface). Since $\dot{p}_A$ = {$\dot{p}_{Am}$} * f * L^2, with L the structural length, then S_A = f * L^2.
A flux of mobilized energy goes outside of the reserve compartment: the scaled mobilisation flux S_C is the scaled equivalent of $\dot{p}_C$ / {$\dot{p}_{Am}$} therefore equal to:
$L^2 * (g * e / (g + e)) * (1 + L * \dot{k}_M / \dot{v}) = S_C$, where g is the energy investment ratio (the cost of the added volume for this timestep relative to the maximum potentially available energy for growth and maintenance), *e* is the scaled reserve density (reserve density relative to maximum reserve density) and L the structural length, and with $\dot{k}_M$ the rate of mobilisation of the κ fraction of S_C for somatic maintenance, proportional to structural length, and $\dot{v}$ the energy conductance. The reserve dynamics calculated at each time step correspond to $\partial U_E = S_A - S_C$.

Change in maturity or reproduction buffer
Before puberty (ability to reproduce), changes in maturity level are calculated as the flux of energy going into the maturity compartment, that is the fraction 1-κ of the mobilisation flux after paying for the maintenance costs of the maturity compartment U_H:
$\dot{p}_R = (1-\kappa) * \dot{p}_C - \dot{p}_J$.
The maturity level of the compartment U_H changes each month through the scaled formula for ∂U_H:
$S_R = (1-\kappa) * S_C - \dot{k}_J * U_H = \partial U_H$, when the reproduction buffer U_R does not receive, ∂U_R is set to 0.
Juveniles keep growing until they reach puberty, when the maturity level U_H is equivalent to U_H^p. At this point, they are able to reproduce, thus the energy flux S_R is redirected entirely to the reproduction buffer U_R and the maturity compartment does not increase anymore: U_H is constant and equal to U_H^p.
Therefore, after puberty, and except for females undergoing reproduction:
$S_R = (1-\kappa) * S_C - \dot{k}_M * U_H^p = \partial U_R$

Change in structural length
The structural length L is updated thanks to remaining energy of the fraction κ of the mobilisation flux S_C after that somatic maintenance has been paid. The structural length change is equal to $\partial L = (1 / 3) * ((\dot{v} / g * L^2) * S_C - \dot{k}_M * L)$.

Starvation
When scaled reserve value is below the scaled structural length *l* value (length relative to maximum length), that is when e < l, it is assumed that the individual is confronted to starving conditions. The kappa rule is

altered and as energy is entirely redirected to somatic maintenance and all other fluxes (growth, reproduction or maturation) are set to 0. The model follows these conditions:
Mobilisation flux $S_C = ([\dot{p}_M] / L^3) / \{\dot{p}_{Am}\}$.
Since $[\dot{p}_M] = [E_G] * \dot{k}_M$ and $[E_G] = g * \kappa * \{\dot{p}_{Am}\} / \dot{v}$
we can rewrite $S_C = (L^3 * \dot{k}_M * g * \kappa * \{\dot{p}_{Am}\} / \dot{v}) / \{\dot{p}_{Am}\} = (L^3 * \dot{k}_M * g * kap) / \dot{v}$
then recalculate $\partial U_E = S_A - S_C$.
When $e < 0$, the organism doesn't have enough energy to pay somatic maintenance and dies.
The starvation strategy used in the population model was chosen among the ones presented in Kooijman (2010) based on Magniez (1983) research on *A. cordatus* reproduction and development.

Ageing
Two ordinary differential equations are calculated: changes in ageing acceleration $\ddot{q}$ (also called the scaled density of damage inducing compounds) and changes in hazard mortality rate $\dot{h}$:
$\partial\ddot{q} = (\ddot{q} * (L / L_m)^3 * s_G + \ddot{h}_a) * e * (\dot{v} / L - \dot{r}) - \dot{r} * \ddot{q}$
$\partial\dot{h} = \ddot{q} - \dot{r} * \dot{h}$, with $\dot{r} = (3 / L) * \partial L$

These equations are used for the simulation of the accumulation of damage inducing compounds and their effect, following the DEB theory for ageing (Kooijman 2010). Damage inducing compounds density is proportional to reserve mobilisation S_C and influences the hazard mortality rate $\dot{h}$, which is a function of the damage accumulated in the body. Damage inducing compounds are diluted via growth $\dot{r}$, and additionally ageing is calculated with two other parameters, the Weibull ageing acceleration $\ddot{h}_a$ and the Gompertz stress coefficient s_G. In other words, the hazard mortality rate is the simulation of the vulnerability of the individual towards damage, such as the risk of dying from an illness increasing as the individual ages. Additionally, in our model, the ageing submodel relies on a stochastic element, where the individual has a 50% chance of looking into its death probability rate $\dot{h}$.

Update individuals
The calculated changes are applied to each state variables of the individual:
The temporal resolution is a monthly interval: each ODE is calculated then the resulting ∂ is applied * 30.5. For a state variable X, $X = X + \partial X * 30.5$.
If the individual has reached a maturity level corresponding to a threshold, it updates its life stage (i.e. its breed in NetLogo language) accordingly. The individual also updates its age.

Update reproduction and birth timers
The reproduction calendar (repro_time) and the GSI calendar (GSI_time) advance by one month each timestep, and fall back to the start in a twelve months cycle. The starting date of the two calendars is not the same (March for repro_time and June for GSI_time, see table previously).
The birth timer (birth_time) is only owned by females and is not always running. It is set off if the female has launched reproduction, and it counts down instead of up (e.g. if it was at 7 the month before, the timer will be set to 6 this month). As long as the female has not launched reproduction, the birth timer will stay set at 0. Once the reproduction period starts, the birth timer is what allows to verify if the individual is undergoing reproduction and to adapt its state variables accordingly: at birth_time 8, 7 and 6, females are reproducing (i.e. conceiving offspring by decreasing the energy in their reproduction buffer, see below); at birth_time 3, 2 and 1, females release offspring (i.e. a number of new juveniles proportional to the number of females having reproduced is initiated into the model, see below).

Reproduction
Only females are considered in the reproduction processes. The value of the Gonado-somatic index (GSI = 100 * ((ash-free gonads dry weight) / (ash-free body dry weight)) is increasing monthly until the reproduction period, when the amount of energy accumulated will be checked by the model to allow, or not, the female to participate to reproduction. Whenever the level of GSI reaches at least 0.07%, the female can reproduce, if not, she will continue updating the energy into the reproduction buffer until the next reproduction period.

When females are reproducing, conception of offsprings causes a decrease in energy in their reproduction buffer: their usual ∂U_R is set to 0 for the three months, while U_R is forced to decrease: for each month of the reproduction period the female decreases its buffer by a third of 52% of the energy stored: $\partial_2 U_R = (U_{R\_start} - 0.52 * U_{R\_start}) / 3$,

with U_{R_start} the reproduction buffer at the start of the period (Magniez 1983). The GSI follows a similar pattern (see submodel 'Calculate GSI').

When females release offsprings, five months after conception, 65% of the eggs are assumed to have survived until birth. The reproductive output (Ri) is therefore equal to Ri=0.65*eggs number For each of the three months of offspring release, the number of juveniles (Ri / 3) are initiated into the model.

Calculate GSI
The GSI is estimated for each month according to the time of accumulation of energy into the reproduction buffer from the end of the reproduction, following the equation:

$$GSI = \frac{time\_of\_accumulation * \dot{k}_M * g}{f^3 * (f + \varkappa * g * y_{VE})} * ((1 - \varkappa) * f^3 - \frac{\dot{k}_J * U_H^p}{Lm^2 * s_M^3}),$$

where the time of accumulation is the number of days since the end of the reproduction period, $\dot{k}_M$ the somatic maintenance rate coefficient and $\dot{k}_J$ the maturity maintenance rate coefficient, g the energy investment ratio, *f* the scaled functional response, y_{VE} is the parameter for the yield of structure on reserve, that is the number of moles of structure that can be produced with one mole of reserve, s_M the acceleration factor, U_H^p the scaled energy in the complexity compartment at puberty, κ the fraction of energy directed towards structure and (1- κ) the fraction of energy directed towards complexity and L_m the maximum volumetric length (see "Entities, State variables and scales" for the dimensions and in-code notations).
The GSI of a reproducing female will decrease by 52% of its initial value over the 3 months period of reproduction (i.e. a decrease of one third of 52% per month):
$\partial GSI = (GSI_{start} - 0.52 * GSI_{start}) / 3$, where GSI_{start} is the level of gonadal index at the onset of reproduction.

Background mortality
The background mortality rate is applied to the overall population: 3.42% of juveniles and 2% of adults (males and females) die each month (calculated from size frequency distribution provided by Mespoulhé 1992). Depending on the cause of death, the individuals set on a certain flag (deceased_bg or deceased_old) and a 'deceased' flag and are removed from the system.

Check temperature
Depending on the temperature for the current and prior month and on the type of sensitivity to temperatures chosen for the model, a mortality rate is applied to the population for temperatures from 8 to 12°C. For a "vulnerable" setting, temperatures exceeding thresholds of 8, 9.5, 11 and 12°C for two consecutive months cause a mortality rate of 25%, 35%, 45% and 100% respectively. For an "intermediate" setting, temperatures exceeding thresholds of 8, 9.5, 11 and 12°C for only one month cause a mortality rate of 10%, 20%, 30% and 100% respectively.
For a "resistant" setting, temperatures exceeding thresholds of 8, 9.5, 11 and 12°C for two consecutive months cause a mortality rate of 10%, 20%, 30% and 100% respectively.

Population monitoring
At the end of each timestep, population density is calculated and data collected for monitoring and plotting mean values of state variables. Plots are built on the lists compiled out of all individual state variable values.

CHAPTER 2

CHAPTER 2 CORRELATIVE MODELS

Chapter 2 is a synthesis on the potential, limits and methodological issues of Species Distribution Models (SDMs) applied to Southern Ocean benthic case studies. SDMs have been used for a long time in ecology to assess species realised niche. However, methods that have been developed for SDMs in other regions of the world need to be adapted to Southern Ocean dataset peculiarities.

•The first study reviewed these Southern Ocean dataset peculiarities, highlighted the main methodological limits to SDMs applied to Southern Ocean case studies and provided some new methods (from the studies below) to generate more accurate models.

•In the second study, the focus was on model evaluation. Cross-validation procedures aim at splitting occurrence datasets into training and test subsets. However, Southern Ocean occurrence datasets are often spatially aggregated, which violates the independency criterion between training and test subsets and biases model evaluation accuracy. In this study, we compared several cross-validation procedures (random vs. spatial partition of training-test subsets) for the case study of the sea star *Odontaster validus* Koehler 1906.

•In the third study, six sea star species with a circumpolar distribution were used as case studies to generate SDMs with contrasting numbers of environmental descriptors. The influence of the number of these environmental predictors and of the collinearity between them was assessed.

•The fourth study focussed on extrapolation uncertainty in SDM predictions. Considering the reference dataset of environmental conditions for which species presence records are modelled, extrapolation corresponds to the part of the projection area for which one environmental value at least falls outside of the reference dataset. Due to the broad extent of the Southern Ocean and data gaps in occurrence datasets, extrapolation represents an important part of model predictions. Using the case study of six sea stars species, extrapolation was highlighted and methods were provided to improve model predictions.

•The last study of this chapter, presented in the appendix section, analysed the influence of spatial and temporal aggregation of occurrence datasets on modelling performances. The case study of four sea urchin species of the Kerguelen Plateau was analysed. Methods to correct for the effect of spatial sampling bias were applied and their efficiency was proved to generate more accurate predictions.

•**Guillaumot C,** Danis B, and Saucède T (2021). Species Distribution Modelling of the Southern Ocean benthos : methods, main limits and some solutions. *Antarctic Science*, 1-24.

•**Guillaumot C**, Artois J, Saucède T, Demoustier L, Moreau C, Eléaume M, Agüera A and Danis B (2019). Broad-scale species distribution models applied to data-poor areas. *Progress in Oceanography,* 175, 198-207.

•**Guillaumot C**, Danis B and Saucède T (2020). Selecting environmental descriptors is critical to modelling the distribution of Antarctic benthic species. *Polar Biology*, 1-19.

•**Guillaumot C,** Moreau C, Danis B and Saucède T (2020). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*. 188, 102438.

•[*Appendix section*] **Guillaumot C**, Martin A, Saucède T and Eléaume M (2018) Methods for improving species distribution models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau. *Marine Ecology Progress Series*. *594*, 149-164.

Species Distribution Modelling of the Southern Ocean benthos: a review on methods, cautions and solutions

Guillaumot Charlène[1,2], Danis Bruno[1], Saucède Thomas[2]

[1] Université Libre de Bruxelles, Marine Biology Lab. Avenue F.D. Roosevelt, 50. CP 160/15 1050 Bruxelles, Belgique
[3] UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, 6 bd Gabriel F-21000 Dijon, France

Antarctic Science, accepted April 8th 2021.

Abstract

Species Distribution Modelling (SDM) studies the relationship between species occurrence records and their environmental setting, providing a valuable approach to predicting species distribution in the Southern Ocean (SO), a challenging region to investigate due to its remoteness and extreme weather and sea conditions. The specificity of SO studies, including restricted field access and sampling, the paucity of observations and difficulties in conducting biological experiments, limit the performance of SDMs.

In this review, we discuss some issues that may influence model performance when preparing datasets and calibrating models, namely the selection and quality of environmental descriptors, the spatial and temporal biases that may affect the quality of occurrence data, the choice of modelling algorithms and the spatial scale and limits of the projection area.

We stress the importance of evaluating and communicating model uncertainties, and the most common evaluation metrics are reviewed and discussed accordingly. Based on a selection of case studies on SO benthic invertebrates, we highlight important cautions to take and pitfalls to avoid when modelling the distribution of SO species, and we provide some guidelines along with potential methods and original solutions that can be used for improving model performance.

Key-words

Antarctica, marine benthic invertebrates, modelling performance, biases, limits

ACKNOWLEDGEMENTS

This work was supported by a “Fonds pour la formation à la Recherche dans l’Industrie et l’Agriculture” (FRIA) and “Bourse Fondation de la Mer” grants to C. Guillaumot.

This is contribution no. 47 to the vERSO project (www.versoproject.be), funded by the Belgian Science Policy Office (BELSPO, contract n°BR/132/A1/vERSO). Research was also financed by the “Refugia and Ecosystem Tolerance in the Southern Ocean” project (RECTO; BR/154/A1/RECTO) funded by the Belgian Science Policy Office (BELSPO), this study being contribution number 24.

1. INTRODUCTION

Due to its remoteness and extreme weather and sea conditions, the Southern Ocean (SO) is a challenging region in which to carry out biological studies (Kaiser et al. 2013, Gutt et al. 2018). It is also one of Earth's regions where we observe the most rapid and dramatic environmental changes in marine ecosystems, motivating the study of these marine communities (Turner et al. 2014, Ashton et al. 2017, Clark et al. 2019). Ecological modelling approaches are now well established and can be used to predict spatial patterns of organisms', populations' and species' distributions and assess their environmental drivers (Peterson et al. 2011). Based on field observations and experimental datasets, ecological modelling encompasses valuable approaches to helping to analyse biological data and interpolating our knowledge of species distributions in relation to environmental descriptors (Kennicutt et al. 2014).
Species distribution models (SDMs) are ecological models that study the statistical relationship between species occurrence records and environmental factors, determining the set of environmental conditions that is suitable to a species distribution (Elith et al. 2006, Elith and Leathwick 2009, Peterson et al. 2011). They represent the species realised niche (Pearson 2007, Sillero 2011), being the ensemble of abiotic conditions in which the species survives and reproduces, adding into consideration the influence of biotic interactions (competition, predation, parasitism, symbiosis, *etc*.) (Hutchinson 1957). SDMs have been widely used in various fields of ecology, such as conservation biology, biogeography, palaeoecology and global change biology (Pearson 2007). In recent years, a growing number of ecological studies have used SDMs to analyse the distribution of marine pelagic and benthic species in the SO (e.g. marine invertebrates, fish, seabirds and marine mammals) and to determine species environmental preferences (Loots et al. 2007, Pierrat et al. 2012, Xavier et al. 2015, Nachtsheim et al. 2017), to compare ecological niche predictions in response to changing environments (Basher and Costello 2016, Gallego et al. 2017, Guillaumot et al. 2018b, Jerosch et al. 2019) or to identify diversity hotspots for conservation purposes (Pinkerton et al. 2010, Hibberd 2016, Thiers et al. 2017).

However, the quality of ocean-wide models is often limited by the heterogeneity, amount and spatial distribution of data, along with limited temporal and spatial resolutions. For all of these reasons, both modelling methods and model construction should be tested for accuracy and robustness prior to interpretation, and these indicators should be transparently communicated to ensure that model outputs are relevant given the specificities of the datasets used for modelling.
In the present paper, we review the most common methodological issues encountered in species distribution modelling applied to the SO, following the flowchart in Fig. 2.1. Challenges regarding occurrence and environmental dataset peculiarities are described. The choice of SDM algorithm, and procedures to implement and evaluate models are addressed. Based on benthic invertebrate case studies, we stress important precautions to take and pitfalls to avoid during common steps of SDM implementation. Finally, we aim to provide some guidelines with a set of potential methods and original solutions that can be used for improving model performance.

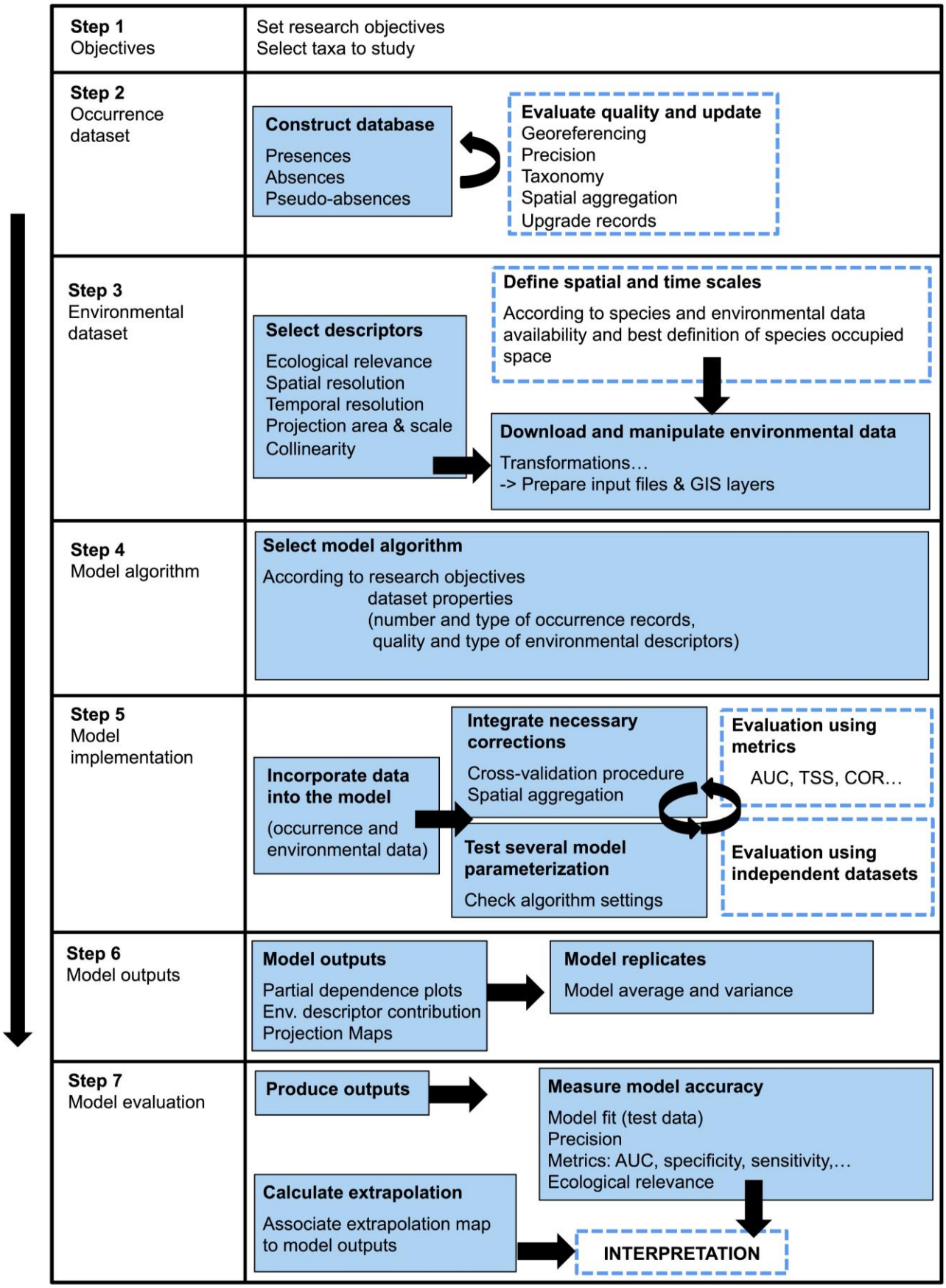


Figure 2.1. Flow chart of the species distribution model construction process. Steps 1 to 4 concern data collection and treatment. Steps 5 to 7 integrate procedures for model implementation and evaluation. Dashed rectangles allow for a possible step backwards when assessing model uncertainties or evaluating model performance. GIS: Georeferencing Information System. AUC = area under the receiver operating curve. TSS: True Skill Statistics. COR: Pearson correlation.

2. Quality of datasets

2.1.1. Environmental datasets: field data.

Preparing environmental datasets is the first encountered challenge when generating models (Gutt et al. 2012, De Broyer et al. 2014). The SO, here defined as waters south of 45°S latitude, covers an extensive area of > 20 million km² (Breitzke 2014). Having access to environmental data with good temporal and spatial resolutions at such a broad scale is challenging, an issue common to all broad-scale oceanographic studies (Robinson et al. 2017). 'Broad scale' is defined here as the entire SO, 'regional scale' as smaller areas of a few hundred square kilometres and 'local scale' as a few square kilometres to square metres (Gage 2004).
Oceanographic data acquisition in the field is strongly conditioned by weather and sea conditions along with the seasonality of polar regions (polar night and dense sea-ice coverage in winter) that prompt recurring gaps in the acquisition of environmental data in the SO. Data are also much more frequently sampled close to research stations and along main sailing routes (Guillaumot et al. 2019 - Chapter 2). This is particularly striking in regions such as the southwestern Weddell Sea, along the shores of the western Antarctic Peninsula and in the Bellingshausen and Amundsen seas (Clarke et al. 2007, Griffiths et al. 2014).

2.1.2. Environmental datasets: satellite-derived data.

Satellite-derived data form a significant source of information for SO oceanographic studies. Providing valuable environmental indicators at broad spatial scale, they can give details about continuous and long-term measurements of water masses including sea-ice coverage, extent and duration, sea-surface temperatures and salinities, biogeochemical parameters, sea level, primary production and typical meteorological parameters (El Mahrad et al. 2020).
The accuracy of satellite data however should be considered with care, given detection limits, interpolations that reduce the influence of atmospheric particulate scatter and the use of interpolation and gap-filling methods that smooth raw data at broad spatial and temporal scales (Pope et al. 2017, Stock et al. 2020).
Whenever possible, it is recommended to validate environmental data derived from satellite products at regional and local scales by comparing pixels on a satellite image with 'real' field observation data (Henson et al. 2015, Trull et al. 2018). Simple correlation analyses or more complex ground-truth processes are available to compare satellite and *in situ* data and to secure the interpretation of satellite-derived products (White-Newsome et al. 2013, Allan 2014). This, however, constitutes a huge task, even if such *in situ* data are available, and is not performed generally before implementing SDMs.

2.1.3. Environmental datasets: access to datasets.

Environmental data generated at the scale of the entire SO can be accessed for free through different web portals such as the NASA's OceanColor Web (https://oceancolor.gsfc.nasa.gov/), where satellite-derived data, averaging different temporal measurements down to 4 km resolution are available at the scale of the entire SO dating from 2000. These images are post-processed to characterize sea-surface temperature or ocean colour as proxies of surface productivity.
The National Oceanic and Atmospheric Administration's (NOAA) data centre (WOCE2013, https://www.nodc.noaa.gov/OC5/woa13/woa13data.html) also makes available post-processed data of ocean temperature, salinity, oxygen concentration and nutrients at different grid formats, down to 0.25° resolution, averaging over six decades (from 1955 to 2012). Bio-ORACLE (https://www.bio-oracle.org/) compiles a large panel of marine data layers at 1° spatial resolution for different depth layers and time periods, for the present (2005-2012) and the future (2040-2050; 2090-2100) (Assis et al. 2018). Finally, GEBCO (https://www.gebco.net/) is the reference platform for very-high-resolution bathymetry data (~500 m resolution) of the world's oceans.
Several works also make available compilation of these SO datasets dedicated to ecological modelling in the SO; they represent a valuable source of information for starting with data preparation and modelling (Raymond 2012, Fabri-Ruiz et al. 2017b, Guillaumot et al. 2018c).

An increasing amount of environmental data collected during SO oceanographic campaigns have been made accessible for regional-scale studies. Several web portals aggregate all of these field measurements and provide them open access (e.g. https://www.marine-geo.org/collections/#!/collection/USAP#summary; https://www.pangaea.de/).

2.1.4. Environmental datasets: spatial and temporal resolutions.

Most environmental data are accessible through broad-scale maps from the aforementioned data portals and are available with a finest spatial resolution of ~ 4 km, if not coarser (Raymond 2012, De Broyer et al. 2014, Fabri-Ruiz et al. 2017b, Guillaumot et al. 2018c). This low resolution strongly hampers the precise assessment of relationships between species occurrences and environmental descriptors (Pittman 2017, Staveley et al. 2017) and consequently the accuracy of model predictions (Connor et al. 2018), because the relevance of environmental descriptors represents a trade-off between their resolution and their spatial and temporal coverage (Guisan et al. 2007, Seo et al. 2009, Lauzeral et al. 2013, Vale et al. 2014). It is recommended that the resolution of environmental descriptors used in SDM should be in line with the scale of ecological processes at play and for which species ecophysiological responses show the highest variations, if models are expected to capture most species-environment relationships (Austin and van Niel 2011).

The published environmental datasets are often averaged over relatively long periods of time (from years to decades for WOCE2013 or Bio-ORACLE). The analysis of inter-annual variations can complement the interpretation of model predictions: the absence of such information does not preclude running models but this should be kept in mind when it comes to interpreting model outputs (Guillaumot et al. 2018a - Appendix). Important environmental variations within a reference time period may not satisfy the equilibrium criterion between species distribution and environmental conditions, which is a strong prerequisite of SDM (Elith et al. 2006) and may affect the relevance and accuracy of model predictions (Guillaumot et al. 2018a - Appendix). In this respect, an alternative for improving modelling performance would be using seasonal averages or extreme values as environmental descriptors rather than pluri-annual to annual averages (Franklin 2010a, Bradie and Leung 2017).

2.1.5. Environmental datasets: cartographic projections.

Considering the poles in numerical analyses has long been a source of difficulty in spatial modelling as the convergence of meridians distorts shapes, surfaces, angles or distances towards high latitudes when using standard cylindrical representations such as the Mercator projection (Deleersnijder et al. 1993, Eby and Holloway 1994, Murray 1996). Working with conical or azimuthal projections (e.g. polar stereographic system) helps maintain the consistency of angles and shapes and therefore better meets the requirements of SO studies, although areas and distances are progressively distorted when moving away from the pole (Mulcahy and Clarke 2001). Mapping environmental descriptors and projecting model predictions can be carried out with either square or hexagonal pixels. Each option does not alter image quality and hexagonal shapes may even offer some advantages (Kamgar-Parsi and Sander 1989, Tirunelveli et al. 2002). However, some contrasts may be present between images using square or hexagonal pixels, as each pixel measures the average environmental conditions in the considered surface (Vanden Berghe et al. 2013).
Subdividing the study area into sub-regions and using different pixel shapes can be a good solution for improving the relevance of representations (Vanden Berghe et al. 2013, Cryer 2015). Evaluating the accuracy of environmental values captured both in square and hexagonal pixels using baseline *in situ* field measurements can also be suggested. This is yet to be tested for ecological modelling studies for the SO.

2.1.6. Environmental datasets: future forecasts.

Since 1992, future climate models have been constantly updated through the efforts of the Coupled Model Intercomparison Projects (CMIP) featured by the Intergovernmental Panel on

Climate Change (IPCC) Assessment Reports (ARs) with the aim of providing a plausible representation of future climate linked to potential anthropogenic impacts (IPCC 2000, Mearns et al. 2001). Recent updates (CMIP5 and CMIP6) of climate models are driven by different possible future greenhouse gas emission scenarios (Representative Concentration Pathways RCP2.6, RCP4.5, RCP6.0 and RCP8.5, from the least to the most pessimistic scenario for CMIP5 and Shared Socioeconomic Pathways, SSP1 to SSP5 for CMIP6) and are built upon the average of an ensemble of simulations (Hayhoe et al. 2017). Future climate models for the SO are available through two main online platforms, Bio-ORACLE (https://www.bio-oracle.org/, Assis et al. 2018) and the NOAA's portal (https://psl.noaa.gov/ipcc/ocn/).

The relevance of using future predictions based on global assessment scenarios for marine studies has been widely questioned (Flato et al. 2014, Frölicher et al. 2016, de la Hoz et al. 2018), including their use in SDMs, given that, climate models mainly rely on untestable assumptions (Beaumont et al. 2008, Gotelli and Stanton-Geddes 2015, Cavanagh et al. 2017, Freer et al. 2018), future layers are not always available for oceanographic studies (Fabri-Ruiz 2018, Guillaumot et al. 2018a - Appendix, 2018b), discrepancies between present observations and future predictions can be problematic (Jiménez-Valverde et al. 2020), and models are based on a representation of the climate system that has a complex cascading effect on ecological processes (Cavanagh et al. 2017). Cavanagh et al. (2017) examined how well IPCC-class models reproduced sea-ice conditions. By subsetting CMIP5 models that best describe spatial extent and temporal ice cover, they improved the precision of the projected future sea-ice distribution, which was better suited to ecological analyses. Extending this method to other key oceanographic parameters should contribute to improving the accuracy of future climate models for the SO and their relevance to ecological studies.

2.2.1. Occurrence datasets: historical compilation.

Biological sampling in the SO began with the first expeditions of the HMS *Challenger* (1873-1876). Sampling effort has considerably increased over the second part of the twentieth century and during these last decades in particular, following technological advances that have enabled the access to remote regions and sample processing (Fig. 2.2).

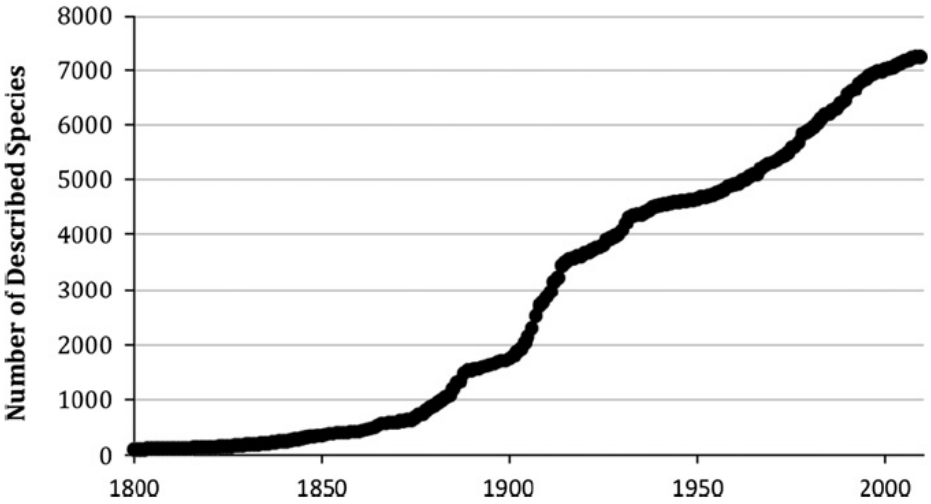


Figure 2.2. Cumulative number of Antarctic species described over time, according to data available in the Register of Antarctic Marine species (until March 2010). From De Broyer and Danis (2011).

This long-lasting and irregular effort in biogeographical (occurrence) data collection has had an impact on data compilation and has resulted in heterogeneous datasets, as observed in several data papers and associated Integrated Publishing Toolkit (IPT) databases such as Guillaumot et al. (2016), Fabri-Ruiz et al. (2017a) or Moreau et al. (2018), or in the general platform *biodiversity.aq* web portal.
The historical compilation of biological data includes (1) taxon misidentifications and taxonomic inconsistencies due to the various taxonomic revisions published through time, (2) errors in the

georeferencing of occurrence records due to contrasting nomenclatures used to report latitude and longitude, (3) the accumulation of errors in metadata through the different generations of curation and (4) errors due to the use of different coordinate projection systems. Finally, in cases where species distributions may have shifted with time, species environmental preferences may have changed or non-contemporaneous environmental or occurrence datasets are used, discrepancies between occurrence records and environmental conditions can be present and violate the environment-occurrence equilibrium assumption necessary to generate SDMs. All of these side effects were reviewed in detail by Newbold (2010). The impacts on species niche definition and SDM predictions have been reported in many works (Ensing et al. 2012, Lahoz-Monfort et al. 2014, Monk 2014, Aguiar et al. 2015, Tessarolo et al. 2017, Guillaumot et al. 2018a - Appendix) that all advise us to thoroughly check datasets for quality management prior to running models.

2.2.2. Occurrence datasets: spatial aggregation.

Most species occurrence data were collected in the vicinity of research bases or their surroundings or along recurrent maritime routes, leading to clear spatial aggregation patterns in biological datasets (Fig. 2.3) (Griffiths et al. 2014, Guillaumot et al. 2019 - Chapter 2).

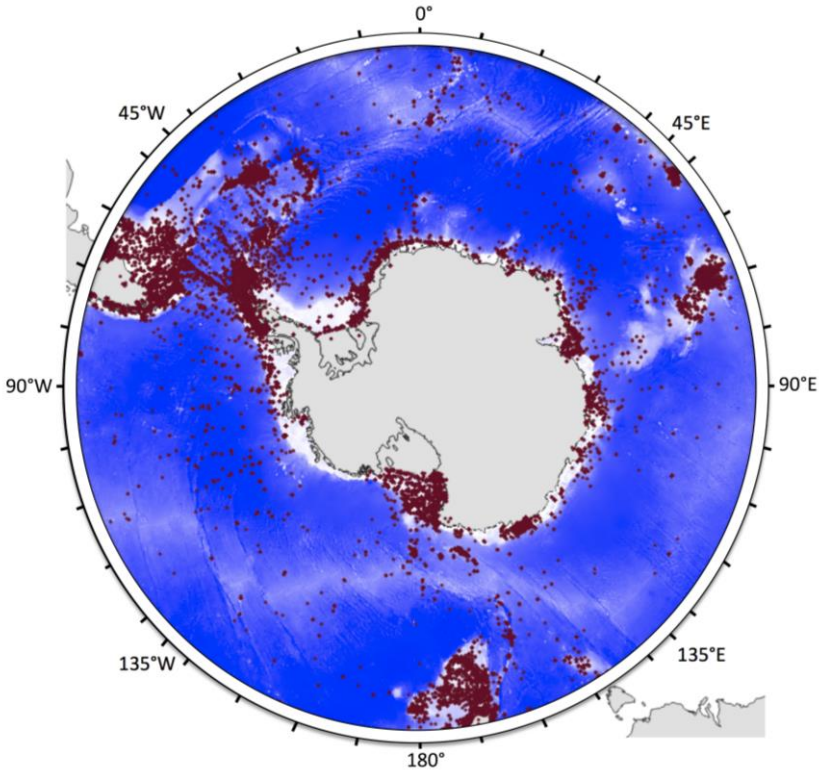


Figure 2.3. Distribution of benthos sampling sites (red dots) in the Southern Ocean (SO, < 45°S). Sampling sites are not evenly distributed in the SO, showing important spatial aggregation in the Scotia Arc region and Western Antarctic Peninsula with several clusters along the Antarctic shelf, and over the Kerguelen and Campbell plateaus. In contrast, deep-sea regions and remote areas of the Antarctic shelf are under-sampled. From Guillaumot et al. (2019 - Chapter 2), updated from Griffiths et al. (2014).

Spatial aggregation can affect model accuracy, as aggregated presence records do not fully and homogeneously represent the entire environment that is occupied by given species. This aggregation also violates an initial assumption of SDMs that requires independence between records (Araújo and Guisan 2006, Hijman 2012). This may bias model predictions (Luoto et al. 2005, Segurado et al. 2006, Dormann 2007, Kühn 2007, Crase et al. 2012), leading to statistical artefacts and generating inaccurate patterns (Bahn and McGill 2007, Currie 2007).
Spatial aggregation of data and the effect of this spatial aggregation on model outputs can be quantified using the Moran's *I* index, which estimates the spatial autocorrelation between the presence records used to build the model and predicted presence probabilities (Luoto et al. 2005). This spatial autocorrelation implies that close pixels are expected to present more similar predicted

probabilities than distant ones due to the short geographical distance between records rather than environmental similarities alone. Testing and correcting for this bias should help to reduce its impact on model predictions (see section 3.3) (Diniz-Filho et al. 2003, Kühn 2007).

2.2.3. Occurrence datasets: presence-only records.

SDMs based on presence/absence data are recognized as having better predictive performance than models using presence-only data (Zaniewski et al. 2002, Brotons et al. 2004, Lobo et al. 2010, Wisz and Guisan 2009, Smith 2013, Carvalho et al. 2015, Peel et al. 2019). However, except for some local-scale studies (e.g. Robinson et al. 2011), in most oceanographic studies species absence records are usually not available for SDMs, and working with presence-only records is the only alternative (Lobo et al. 2010). SDMs are then built by associating presence-only records with a random selection of background records that will be used to characterize the full environmental conditions (Franklin 2010b, Barbet-Massin et al. 2012). Background records should not be mistaken for pseudo-absence records that are artificial absence data, where the species is supposed (but not confirmed) to be absent. Pseudo-absence records do not represent the overall conditions of the study area. Presence/pseudo-absence models represent another modelling approach, predicting occupied and unoccupied habitats rather than suitable and less suitable habitats for presence/background modelling (Sillero and Barbosa 2020).

Presence-only datasets may contain several uncertainties that can bias model predictions. (1) Working on rare or cryptic species is generally prone to taxonomic misidentifications that may either contract or, alternatively, expand the extent of predicted species distributions (Costa et al. 2015, Aubry et al. 2017). Such biases due to taxonomic errors were shown to be highly variable and to depend on experts identifying specimens, as suggested by Beale and Lennon (2012) who worked on a compilation of several collections. (2) Sampling gear may have an impact on species detectability, which varies inconsistently across the model domain and is generally not taken into account by presence-only methods. Inaccurate species observations may generate false-positive results (species predicted as being present when they were not sampled or observed in the field) and false-negative results (species predicted as being absent when they were sampled or observed in the field) during model initialization (Guillera-Arroita 2016). Species presence records should be carefully scrutinized prior to modelling (Lozier et al. 2009), or at least records should be categorized into different subsets of data verifiability (Aubry et al. 2017). (3) Georeferencing errors are a frequent issue in databases (Murphey et al. 2004, Maldonaldo et al. 2015). This is especially the case in large databases compiling independent datasets using species presences recorded with varying levels of precision (Graham et al. 2008, Bloom et al. 2018). Several studies have simulated virtual random georeferencing errors and have shown that these errors lead to significant drops in model performance and inconsistencies in the respective contributions of environmental descriptor contributions, influencing model interpretation (Graham et al. 2008, Osborne and Leitão 2009, Naimi et al. 2011). These side effects seem to be minimized in local-scale models, here again advocating for the use of local-scale models whenever possible (Mitchell et al. 2017).

2.2.4. Occurrence datasets: dealing with small datasets.

Usually, the number of species presence records available for modelling is relatively limited considering the wide geographical extent of the SO (De Broyer et al. 2014). Generating SDMs with small datasets may include many pitfalls: (1) it reduces the potential of SDMs to transfer in space and time (Hernandez et al. 2006, Raes 2012), (2) it truncates predicted distribution and niche definition (Hortal et al. 2007, 2008, Rocchini et al. 2011, Sánchez-Fernández et al. 2011, Titeux et al. 2017, El-Gabbas and Dormann 2018), (3) it reduces modelling goodness-of-fit as the model may wrongly represent reality (Stockwell and Peterson 2002, McPherson et al. 2004, Pearson et al. 2007, Wisz et al. 2008, Liu et al. 2019), (4) it increases instability between model replicates (Guillaumot et al. 2018a - Appendix), (5) it gives rise to metholodogical constraints on threshold selection (Jiménez-Valverde and Lobo 2007, Bean et al. 2012), (6) it gives rise to methodological constraints on the application of evaluation metrics (Pearson et al. 2007), (7) it complicates the identification of model optimal complexity (Galante et al. 2018) and (8) it leads to a reduction in

model accuracy because presence and background datasets would not differ markedly (Luoto et al. 2005).
Alternatives are being developed to produce more accurate models based on a limited amount of presence records. One solution is generating several models performed on restricted areas and datasets with more detailed information and then averaging them with a weighted ensemble approach. This 'ensemble of small models' approach showed improved performance compared to single models (Lomba et al. 2010, Breiner et al. 2015, 2018).
Another alternative is to restrict the prediction area according to where occurrence records are found and ensuring upstream that the number of records is sufficient to precisely characterize the species environmental preferences: trivial advice that is surprisingly neglected, as recently pointed out by Morales et al. (2017) and Araújo et al. (2019).

2.2.5. Occurrence datasets: definition of species-occupied environmental space.

Spatial aggregation, along with heterogeneity, limited size and uncertainties in datasets can strongly bias the quantification of the species-occupied environmental space (Hortal et al. 2008, Newbold 2010, Tessarolo et al. 2017). However, accurately defining species-occupied space is the cornerstone of SDM initialization (Elith et al. 2006, Boulanger et al. 2018).
Moreover, SDMs suppose that species are in equilibrium with the environmental conditions that they inhabit. SDMs do not take into consideration potential vagrants that have dispersed out of their usual environmental range or populations that could momentarily survive in unsuitable habitats because doing so violates the equilibrium assumption between species distribution and environmental conditions (Beale and Lennon 2012). These elements should be cautiously considered when preparing datasets prior to generating models by removing any atypical records.

Over the last two decades, field data acquisition has expanded through the use of biologging technology with electronic devices attached to seabirds and marine mammals in order to access the positions of species all year long (Raymond et al. 2015, Ropert-Coudert et al. 2020). These data uncover the hidden behaviours of marine animals and constitute a powerful way of better estimating species-occupied space; they can also be used to validate and refine our understanding of the environmental conditions prevailing in those species distribution areas (Arthur et al. 2017, Nachtsheim et al. 2017, Hindell et al. 2020).

3. Adapting model implementation to datasets

3.1. The choice of modelling algorithms.

To run performant SDMs, several assumptions must be tested and computing methods adapted to each case study (Austin 2002, de la Hoz et al. 2019). Among them, the choice of the modelling algorithm should be of major concern, since no algorithm works best for all species, in all areas, at all spatial scales and in all time periods (Jarnevich et al. 2015, Qiao et al. 2015). The selection and parameterization of modelling algorithms proved to be major causes of variation between SDM predictions (Diniz-Filho et al. 2009, Dormann et al. 2008, Buisson et al. 2010, Watling et al. 2015, Boulanger et al. 2018). Each algorithm is particularly suited for dealing with a specific type and quality of data (Guisan and Zimmermann 2000, Austin 2002, Elith et al. 2006, Peterson 2011, Guisan et al. 2017), which will determine the final model outputs (Aguirre-Gutiérrez et al. 2013, Beaumont et al. 2016).

When modelling species distribution, it is necessary to select appropriate algorithms that have good transferability performances (i.e. have good abilities to correctly transfer predictions to other geographic space and time periods; Randin et al. 2006) and that they limit overfitting (i.e. mitigate model complexity) while being flexible in integrating complex environmental relationships. Machine-learning algorithms (e.g. maximum entropy MaxEnt, boosted regression trees BRT, random forests RF, support vector machines SVMs, Vapnik 1998, Breiman 2001, Elith et al. 2008, 2011) give access to important aspects of computing performance (Zhou 2012), and are relevant approaches for handling complex relationships between species occurrences and the environment

(Olden et al. 2008, Elith and Leathwick 2009). The BRT and RF algorithms are particularly suited to complex and heterogeneous datasets (Fig. 2.4, Guillaumot et al. 2020b - Chapter 2). They were proven to be efficient in generating performant models with limited overfitting (Elith et al. 2006, Wisz et al. 2008, Wenger and Olden 2012). They can automatically select the most informative features among a large set (Merow et al. 2014, García-Callejas and Araújo 2016, Guillaumot et al. 2020b - Chapter 2) and perform well at generalizing predictions in the absence of information or, conversely, at dealing with redundant information provided by correlated factors (Breiman 1984, De'ath and Fabricius 2000, Friedman 2001).

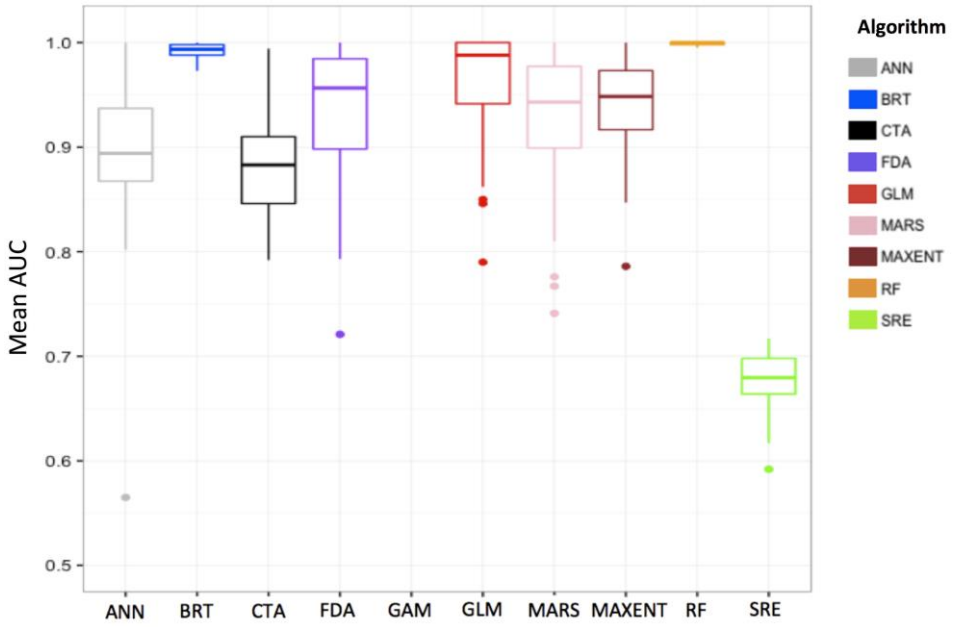


Figure 2.4. Compared Area Under the Curve (AUC) performances of species distribution models generated with different algorithms (ANN=Artificial Neural Network, BRT=Boosted Regression Trees, CTA=Classification Tree Analysis, FDA=Flexible Discriminant Analysis, GAM=Generalized Additive Model, GLM=Generalized Linear Model, MARS=Multivariate Adaptive Regression Splines, MAXENT=Maximum Entropy, RF=Random Forest, SRE=Surface Range Envelope) to predict the distribution of the sea urchin *Sterechinus diadema* in the Southern Ocean. Results show a good performance for BRT and RF, adapted to small, historically compiled datasets (temporally heterogeneous) and spatially aggregated presence-only data. Models were calibrated with presence-only data and 200 background data randomly sampled in the study area. Average scores of 100 model replicates. See Guillaumot et al. (2018b) for details.

The different fields of application and the respective performance of existing algorithms have been extensively compared in various works based on (1) a single species (Pearson et al. 2006: plants in South Africa, Elith and Graham 2009: plant distribution in South Australian landscapes, Marmion et al. 2009: European butterflies, Lorena et al. 2011: plants in South America, Beaumont et al. 2016: mammals in Australia), or (2) an ensemble of worldwide distributed terrestrial (Elith et al. 2006) or marine species (Ready et al. 2010), for (3) certain regions only (Guisan et al. 2007: trees in Switzerland, Tsoar et al. 2007: snails, birds and bats in Israel, Reiss et al. 2011: benthic marine species in the North Sea, Bucklin et al. 2015: vertebrates of Florida) or (4) using virtual species (Meynard and Quinn 2007, García-Callejas and Araujo 2016, Qiao et al. 2015).
However, in order to generate such comparisons (Fig. 2.4), it is important to specifically adjust each algorithm to the case study. Algorithms all perform differently with regards to overfitting, spatial aggregation and transferability, and comparing model performances using different parameter settings is challenging (Merow et al. 2014) given that model parameterization has strong effects on the quality of model outputs (Anderson and Gonzalez 2011, Rodda et al. 2011, Warren and Seifert 2011, Yackulic et al. 2013, Radosavljevic and Anderson 2014, Moreno-Amat et al. 2015, Halvorsen et al. 2016, Galante et al. 2018, Lieske et al. 2018).

Initially developped in the 1990s, ensemble modelling has been increasingly used since then (Hansen and Salamon 1990, Schapire 1990). Ensemble modelling consists of combining several

algorithms (Zhou 2012), input datasets (occurrence or environmental descriptors datasets) or parameterizations (Araújo and New 2007, and see Hao et al. 2019 for a review of applications). The approach is interesting as it can provide predictions that take into account the variability of several models (Araújo and New 2007, Hao et al. 2019).
Ensemble modelling has been used for various studies with SDMs (Araújo and New 2007, Marmion et al. 2009, Thuiller et al. 2009, Buisson et al. 2010, Luedeling et al. 2014, Trolle et al. 2014, Carvalho et al. 2015, Scales et al. 2016, Jerosch et al. 2019) and has benefitted from the development of R packages to implement them (*Biomod*: Thuiller et al. 2009, *BiodiversityR*: Kindt et al. 2017, *biomod2*: Thuiller et al. 2018, *sdm*: Naimi et al. 2019).
The main benefits of using ensemble models lie in the fact that the different algorithms will perform differently for various input cases (regardless of their overall performance). The models thus complement each other, avoiding some biases that might have resulted from using a single algorithm (Marmion et al. 2009, Knutti 2010, Zhou 2012). However, model interpretation is much more difficult when mixing algorithms implemented differently, with contrasting ways of presenting outputs (Sillero 2011) and different definitions of thresholds for identifying habitat suitability (Perrault-Hébert 2019), requiring the normalization of predictions, which is rarely applied (Zhang and Mahadevan 2019). This is the main limitation to the approach and could offset the gains in model performance (Crimmins et al. 2013, Zhu and Peterson 2017, Hao et al. 2020). Such gains were contested, especially since model evaluation was often performed without using an independent evaluation dataset (Hao et al. 2019). Combining predictions of different models generated with contrasting assumptions is therefore tricky when interpreting the results (Perrault-Hébert 2019). Optimizing the parameterization of a single algorithm (which could be correctly evaluated) may therefore constitute a more valuable approach (Perrault-Hébert 2019). Comparing the performance of different algorithms can be helpful in the first stage of the modelling process in order to select the most suitable algorithm and to calibrate the models (Massada et al. 2013).

3.2. The choice of environmental descriptors.

The selection of environmental descriptors is also a crucial step in the modelling process (Franklin 2010b, Austin and Van Niel 2011, Petitpierre et al. 2017). Ideally, environmental descriptors should be selected for their ecological relevance to the studied organisms (Austin and van Niel 2011, Dormann et al. 2012b, Bradie and Leung 2017), they must capture environmental discontinuities and constraints in the distribution area (Jarnevich et al. 2015), and they should also be detailed enough to represent the habitat complexity and variability in order to allow for good SDM accuracy and performance (Elith and Leathwick 2009, Barbet-Massin et al. 2012, Bucklin et al. 2015, Petitpierre et al. 2017).
In most studies, the final number of descriptors selected to depict the species environment is generally close to 10 (Pierrat et al. 2012, Mormède et al. 2014c, Guillaumot et al. 2018a - Appendix, Fabri-Ruiz et al. 2019). Overall, a small number of descriptors will allow for the generation of less complex models and facilitate interpretation (Austin and van Niel 2011, Braunisch et al. 2013, Bucklin et al. 2015, Petitpierre et al. 2017). In contrast, increasing the number of descriptors potentially increases the effect of any collinearity between them (i.e. correlation between values of descriptors), which may lead to statistical artefacts in model predictions if the algorithms cannot handle information redundancy (Dormann et al. 2012b, Merow et al. 2014). Therefore, collinearity is usually tested for beforehand and collinear descriptors are adjusted (in practice, one descriptor of a pair is removed) before running the model (Dormann et al. 2012b, Merow et al. 2013, Fois et al. 2018). However, Guillaumot et al. (2020b - Chapter 2) showed that model complexity, transferability and accuracy do not significantly change between models generated with different sets, including from 4 to 58 collinear descriptors when using the BRT algorithm. BRTs automatically keep the most relevant descriptors to describe species distribution and can deal with redundant information (De'ath and Fabricius 2000, Whittingham et al. 2006, Elith et al. 2008), which is not the case for all algorithms (Merow et al. 2014).

Selecting environmental descriptors therefore implies that several tests should be performed upstream in order to determine the best set to be used depending on research objectives. Fois et al. (2018) recommended first calibrating models with a large set of descriptors of various natures (proximal *vs*. distal descriptors) that will be pruned stepwise, after analysing their ability to

accurately describe the habitat and after testing for collinearity (El-Gabbas and Dormann 2018). Generating, testing and comparing several sets of descriptors is a widespread strategy to target in a stepwise manner the set that gives the best predictive accuracy (Snickars et al. 2014, Bucklin et al. 2015, Bradie and Leung 2017, Petitpierre et al. 2017). Replacing environmental descriptors by principal components of a factorial analysis also proved to be efficient because complex environmental gradients of the study area are simplified in fewer, orthogonalized components (Kühn 2007, Petitpierre et al. 2017). So far, this latter method has never been applied to SO case studies, and it should be tested in order to evaluate the interpretability of model results.

3.3. Correcting spatial sampling biases.

Generating a model based on spatially aggregated presence-only records may bias predictions with a higher probability of occurrence predicted in highly sampled areas (Dormann 2007, Guillaumot et al. 2018a - Appendix). To compensate for such a bias, a first approach is to sample background records according to the spatial bias introduced by the aggregated presence records themselves (Phillips et al. 2009). The background dataset is used to define the environmental background: its boundaries and variability constitute essential information for building and projecting model outputs (Wisz and Guisan 2009, Barbet-Massin et al. 2012). The choice of the number of background records to be sampled and the extent of their distribution should be considered carefully when calibrating a model because it can strongly influence model predictions (Chefaoui and Lobo 2008, Lobo et al. 2010, Barbet-Massin et al. 2012, Jarnevich et al. 2017). This number should be with respect to the prevalence score, being the ratio between the species-occupied space (represented by presence record locations) and the total surface of the study area (represented by background locations: McPherson et al. 2004). Some advice is provided in Barbet-Massin et al. (2012) for selecting the correct number of background records according to prevalence scores.

Targeting background records has been extensively tested, and several procedures have been developed to significantly improve the relevance of models (Fig. 2.5). Background records can be sampled within predefined areas (i.e. ‘discs’ or ‘buffers’) close to presence records (Hengl et al. 2009, Phillips et al. 2009, Fourcade et al. 2014, Bertrand et al. 2016), following the presence or absence of other species (Phillips et al. 2009, Syfert et al. 2013, Iturbide et al. 2015, Molloy et al. 2017, Phillips et al. 2017, Ranc et al. 2017), according to probabilities given by a kernel density estimator (KDE) of the sampling frequency (Fourcade et al. 2014, Jarnevich et al. 2017, Guillaumot et al. 2018a - Appendix, Fabri-Ruiz et al. 2019) or according to additive descriptors of accessibility and sampling effort (El-Gabbas and Dormann 2018). Once again, the selected method should be adapted to each case study and its efficiency tested prior to model interpretation (Støa et al. 2018).

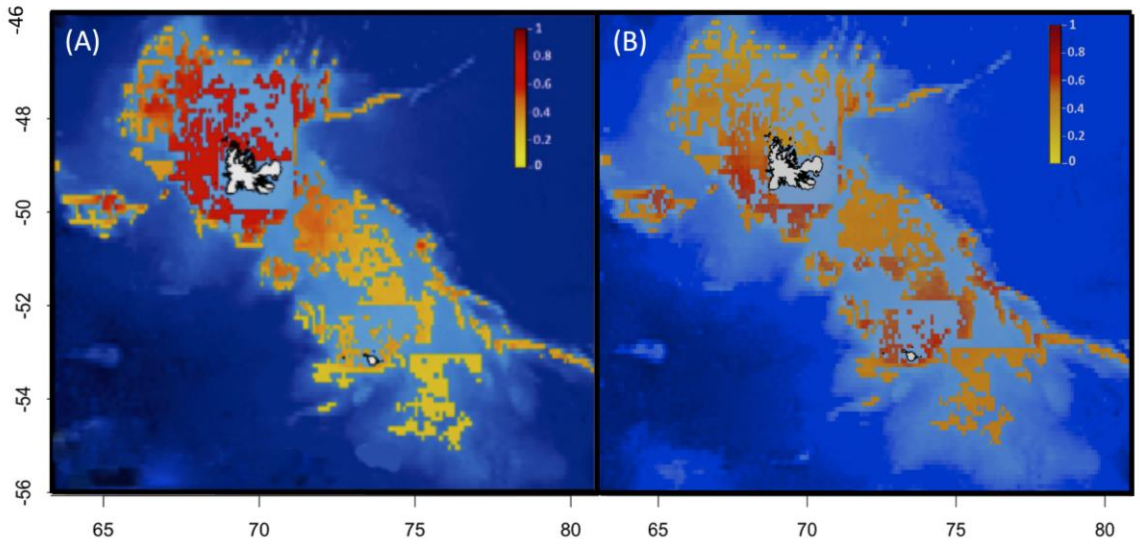


Figure 2.5. Comparison of predicted distribution probabilities (between 0 and 1) of the sea urchin *Ctenocidaris nutrix* on the Kerguelen Plateau: (A) without compensating for sampling bias; (B) with a kernel density estimator (KDE) correction: more background data are sampled in highly sampled areas. The spatial aggregation of presence-only records near the shoreline of the Kerguelen Islands strongly biases model predictions. The KDE correction was proven to be efficient at correcting for such a bias and provides more relevant predictions. From Guillaumot et al. (2018a - Appendix).

A second method consists of filtering the available presence data to reduce the influence of the clustering of species records (Segurado et al. 2006, Kramer-Schadt et al. 2013, Boria et al. 2014). This is an efficient method compared to the background targeted sampling approach detailed above, but the remaining number of presence records after filtering should be sufficient to correctly determine species-occupied space (Kramer-Schadt et al. 2013). Reliable information should also be available to characterize the bias in species occurrence data (Aiello-Lammens et al. 2015, Sillero and Barbosa 2020). The filtering protocol requires meeting many prerequisites, but priority is given to keeping presence data independent and minimizing records clustering (Alagador D. personal communication 2019).

Overall, if several methods are developed to correct for the effect of spatial aggregation on model outputs, it is recommended that one should interpret model projections performed for poorly sampled areas with great caution (Phillips et al. 2009, Iturbide et al. 2018).

4. Model outputs

4.1. Taxonomic bias and population variability.

SDMs are usually parameterized using all presence records available for a species and all environmental conditions prevailing in the species records (Elith and Leathwick 2009). When modelling species distribution at a broad spatial scale, it is often assumed that all populations of a species have the same relationship to environmental conditions over the entire distribution area (Pierrat et al. 2012, Xavier et al. 2015, Guillaumot et al. 2018b, Fabri-Ruiz et al. 2019). However, occurrence datasets may include a set of populations with different phenotypic plasticities (Chevin et al. 2010), transgenerational adaptations (Dixon et al. 2015) or simply different habitat selection in the case of vagile species. Therefore, the modelled species can actually present different abilities to respond to environmental changes. In particular, physiological performances of populations are likely to vary in marine species with wide distribution ranges and high dispersal capabilities over long distances (Thatje 2012). This is particularly relevant with regards to future predictions that do not integrate inter-population variability in the potential acclimation of species, and this may lead models to alternatively over- or under-estimate the distribution of species-suitable environments (Cacciapaglia and van Woesik 2017, Thyrring et al. 2017).

Phylogeographical studies have also regularly revealed the existence of cryptic species in the SO benthos, which show similar morphologies for distinct genotypes and potentially, distinct ecological requirements and geographical distributions (Lozier et al. 2009). Such studies often stress the need for taxonomic revisions (González-Wevar et al. 2019, Ocaranza-Barrera et al. 2019, Moreau et al. 2020). SDMs can be generated based on a spatial subdivision of presence records according to the genetic structure of taxa, and in a second step, the different predictions can be merged together to the broader scale (Knowles et al. 2007, Marcer et al. 2016, Cacciapaglia and van Woesik 2017, Ikeda et al. 2017, Roberts et al. 2017, Pardo-Gandarillas et al. 2018). However, defining the genetic structure of benthic species in the SO is a long-term endeavour that requires a constantly renewed sampling effort, considering the extent and complexity of the study area (Moreau et al. 2017, Fraser et al. 2018, Moore et al. 2018). Waiting for taxonomic revisions and enhanced sampling efforts to best depict relationships between genetic units and environmental conditions (Vandersteen 2011) and combining SDM with experimental data or mechanistic approaches can be alternatives for taking into account the possible physiological contrasts between populations (Kearney and Porter 2009, Kearney et al. 2010, Buckley et al. 2010, Fordham et al. 2013, Briscoe et al. 2016, Feng and Papes 2017, López-Farrán/Guillaumot et al. in press - Chapter 3).

4.2. Definition of region of interest ("projection area").

The limitations in the current knowledge of species distribution also affect the quality of information available for estimating their potential distribution (Thuiller et al. 2003). When the limits of species environmental ranges are not fully captured, this uncertainty can significantly impact the accuracy of SDM predictions (Hortal et al. 2007, 2008, Rocchini et al. 2011, Sánchez-Fernández et al. 2011, Titeux et al. 2017, El-Gabbas and Dormann 2018). It reduces the applicability of models for predictive purposes (Thuiller et al. 2004), induces model overfitting (Tsoar et al. 2007, Barve et al. 2011, Guillaumot et al. 2018b) and can lead to overestimating the extent of suitable areas (Anderson and Raza 2010).
This bias can be partly overcome by reducing the extent of the projection area to the known distribution of the available occurrence records (Anderson and Raza 2010), and by increasing knowledge regarding species ecology and physiology in order to identify the environmental conditions that are unsuitable for their survival or development (Byrne et al. 2016).

4.3. Model extrapolation.

Models are said to extrapolate when a portion of the predicted area includes environmental conditions that are outside the range of values for which the model was calibrated. Model extrapolation may occur when model predictions are transferred, either in space or time. When extrapolated, model predictions are in non-analogue conditions compared to the initial calibration conditions because calibration data may not encompass the entire environmental range of each of the predictors (Guillaumot et al. 2020c - Chapter 2). The set of projected environmental conditions can otherwise still be within the range of conditions, but specific combinations of environmental descriptors may be new, also leading to extrapolation (Mesgaran et al. 2014). In such conditions, predictions might be ecologically and statistically invalid and model interpretations inaccurate (Randin et al. 2006, Williams and Jackson 2007, Williams et al. 2007, Fitzpatrick and Hargrove 2009, Owens et al. 2013).
Among the different approaches, Elith et al. (2010) propose estimating and quantifying model extrapolation using the Multivariate Environmental Similarity Surface (MESS) index to identify the most influential descriptors that lead to extrapolation. Grid-cell pixels for which at least one environmental descriptor has a value outside the range of environmental values defined by presence-only records (calibration range) are considered to be extrapolations. In these cases, the MESS index assigns negative values and the ensemble of pixels containing negative values defines the extrapolation area (Elith et al. 2010, Guillaumot et al. 2020c - Chapter 2). Most often, for SDMs performed at the scale of the SO, the number of records available to define the environmental space occupied by species is limiting and the resolution of environmental descriptors relatively low (see section 2). As a consequence, SDM projections sometimes include

wide extrapolation areas that may cover > 75% of the predicted regions (Fig. 2.6) (Guillaumot et al. 2020c - Chapter 2).

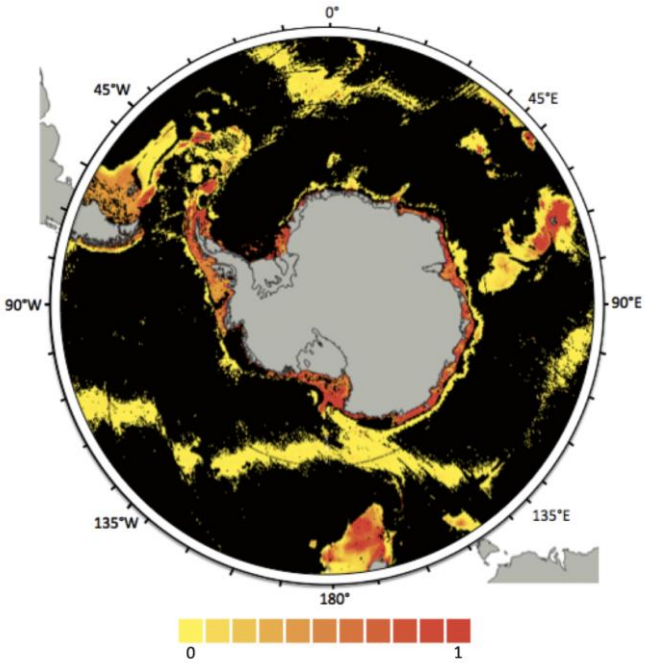


Figure 2.6. Extrapolation map of the species distribution model generated for the sea star *Acodontaster hodgsoni*, with all presence-only records available. Extrapolation corresponds here to the ensemble of environmental conditions that are outside of the boundaries of the calibration range. The extrapolation area is displayed in black and covers 78.6% of the entire projection area; coloured pixels (yellow-red colour palette) show distribution probabilities (included between 0 and 1). Extracted from Guillaumot et al. (2020c - Chapter 2).

In addition to quantifying the overall extrapolation area (Fig. 2.6), it is possible to fine-tune the analysis and define which environmental descriptors and areas are concerned with extrapolation (Owens et al. 2013, Guillaumot et al. 2021 – Thesis material) (Fig. 2.7). Such information could be used to resample the environmental descriptors implemented in the model.
In any case, it has been recommended to provide information on model extrapolation and more generally to other concepts of uncertainties (species detection, errors, *etc*.), along with model predictions, because they are essential to accurate interpretation (Beale and Lennon 2012, Addison et al. 2013, Guisan et al. 2013).

Limiting model projections to 'realistic' depth ranges or some other environmental limiting factor based on a robust knowledge of species ecology (i.e. some expert-driven decision) was proven to be efficient at reducing extrapolation (Kearney and Porter 2009, Hare et al. 2012, De Villiers et al. 2013, Guillaumot et al. 2020b – Chapter 2). Such a strategy is transitional until complementary samples and more comprehensive occurrence datasets are made available to better define the species-occupied space (Guillaumot et al. 2020b – Chapter 2).

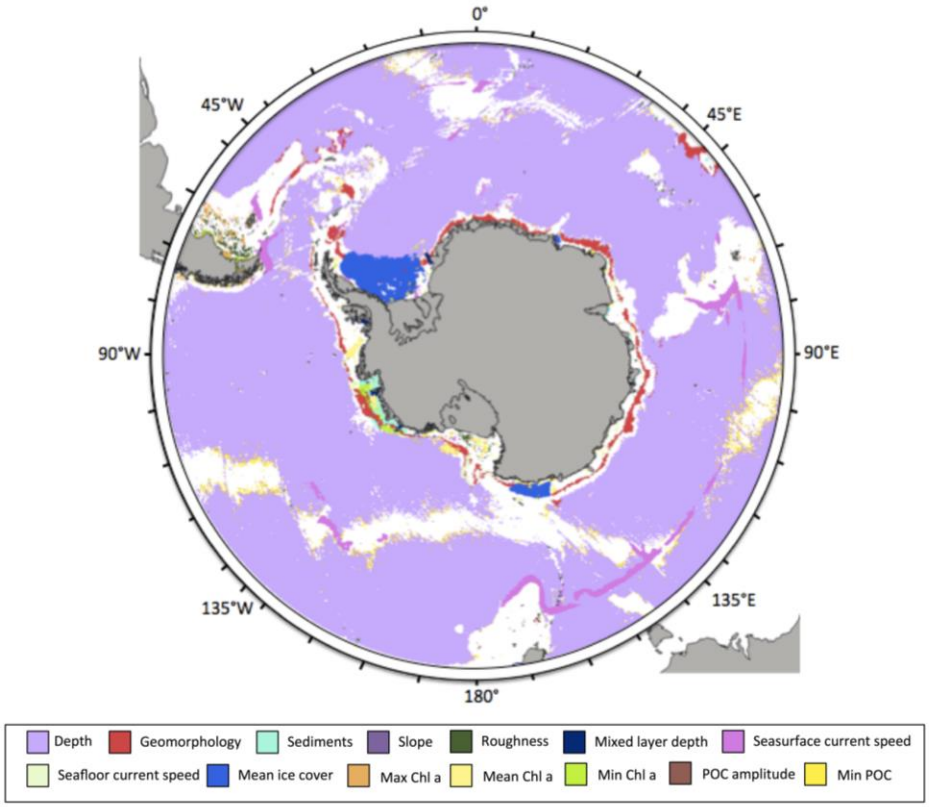


Figure 2.7. Extrapolation map of the species distribution model generated for the sea star *Acodontaster hodgsoni* indicating environmental descriptors responsible for extrapolation (black pixels of Fig. 2.6 are here coloured according to the descriptor responsible for extrapolation; i.e. for each pixel, the predictor in question lies outside the calibration range). In this case study, 14 environmental descriptors are responsible for extrapolation, with depth being the main contributor. White pixels correspond to areas where the model does not extrapolate (the corresponding model predictions are shown in Fig. 2.6). POC stands for 'Particulate Organic Carbon' and Chla is the concentration in chlorophyll a on the sea-surface. Generated from Guillaumot et al. (2021).

5. Model validation and accuracy of model predictions

5.1. Some common metrics for the evaluation of model predictions.

Once models are generated, the accuracy of their predictions must be assessed in order to evaluate the validity of the models with regards to scientific issues to address, to compare different model outputs and to allow for the formulation of reliable interpretations (Zurell et al. 2020). Several metrics were developed in order to evaluate the performance of models (Fielding and Bell 1997, Allouche et al. 2006). Most of them are based on the calculation of an error matrix (or confusion matrix) that displays the proportion of presence and absence records that are correctly predicted by the model (Allouche et al. 2006).

In most biological studies focused on the SO benthos, absence records are usually unavailable and SDMs are generated based on a set of presence/background records (see section 2.2). As a consequence, the statistics that are commonly used for presence/absence datasets may not be appropriate for model evaluation (Wiley et al. 2003, Phillips et al. 2006, Braunish et al. 2013), such as the Kappa statistic (Allouche et al. 2006). In contrast, the Area Under the Curve, or Area Under the Receiver Operating Curve (AUC), is one of the most used and appropriate metrics for measuring the performance of model predictions based on presence/background data (Hand 2009). The AUC is an objective measure that remains stable with low-prevalence datasets (i.e. low frequency of occurrences with regards to the projection space) and is not sensitive to threshold effects (Manel et al. 2001, Hand 2009, van Proosdij et al. 2016). However, for presence/background models, specificity (the fraction of correctly predicted absences) might be

overestimated when the number of background records is much higher than the number of presence-only records or when background and presences are associated with very different environmental values. This incidentally inflates AUC scores (Phillips et al. 2006, Raes and ter Steege 2007, Lobo 2008, Jiménez-Valverde 2012) and invalidates the relevance of the AUC metrics (van Proosdij et al. 2016).
Even when properly employed, the AUC cannot be used to compare models when SDMs are generated for different species based on different environmental descriptors or projected on distinct regions because the values depend on the relative size of suitable areas and prevalence scores may contrast (see section 3.3) (Wisz et al. 2008, Anderson and Gonzalez 2011). The AUC metrics must be used as a simple measure of the relative ranking of model predictions associated with a specific dataset (El-Gabbas and Dormann 2018). Overall, each statistic is characterized by specific advantages and potential biases, so that it is recommended that one uses several statistics for evaluating model predictions (Allouche et al. 2006).
The accuracy of model predictions can also be evaluated by testing the classification of independent test data, where the available occurrence dataset can be split into independent subsets to train or test the model (for a review, see Fielding and Bell 1997).

5.2. Cross-validation procedures.

Cross-validation procedures are aimed at evaluating model predictions using a subset of presence or absence records retrieved from the initial dataset used for modelling in order to assess how well the test data match with the modelled predictions (Bahn and McGill 2013). When working with presence-only datasets, two subsets of presence records are used: one subset is used to train the model (the training group) and the second subset is used to test the model (the test group). Test data and training data must be spatially independent from each other (Hijmans 2012, Bahn and McGill 2013). In most modelling exercises, standard cross-validation procedures are commonly used, in which the initial presence dataset is randomly split into a training and test subset. Frequently, as previously discussed, presence data are spatially aggregated in SO datasets and the necessary condition of independence between training and test data is seldom met, making the model accuracy evaluation overly optimistic (Telford and Birks 2009, Hijmans 2012, Radosavljevic and Anderson 2014). In contrast to random procedures, spatial cross-validation procedures improve the performance of the validation step by spatially segregating the training and test subsets, ensuring the spatial independence between data even when they are spatially aggregated in the initial datasets (Dhingra et al. 2016, Roberts et al. 2017, Guillaumot et al. 2019 - Chapter 2, see also
http://cran.rapporter.net/web/packages/blockCV/vignettes/BlockCV_for_SDM.html).

Several spatial cross-validation procedures have been proposed (Fig. 2.8), and the most appropriate one can be determined by comparing the different procedures in order to define the one that is the most suitable for the study (Muscarella et al. 2014, Radosavljevic and Anderson 2014, Valavi et al. 2018, Guillaumot et al. 2019 - Chapter 2) depending on the spatial scale of the analysis, the number and spatial distribution of the presence data and the selected algorithm (and its associated complexity) used for modelling (El-Gabbas and Dormann 2018, Hao et al. 2020).

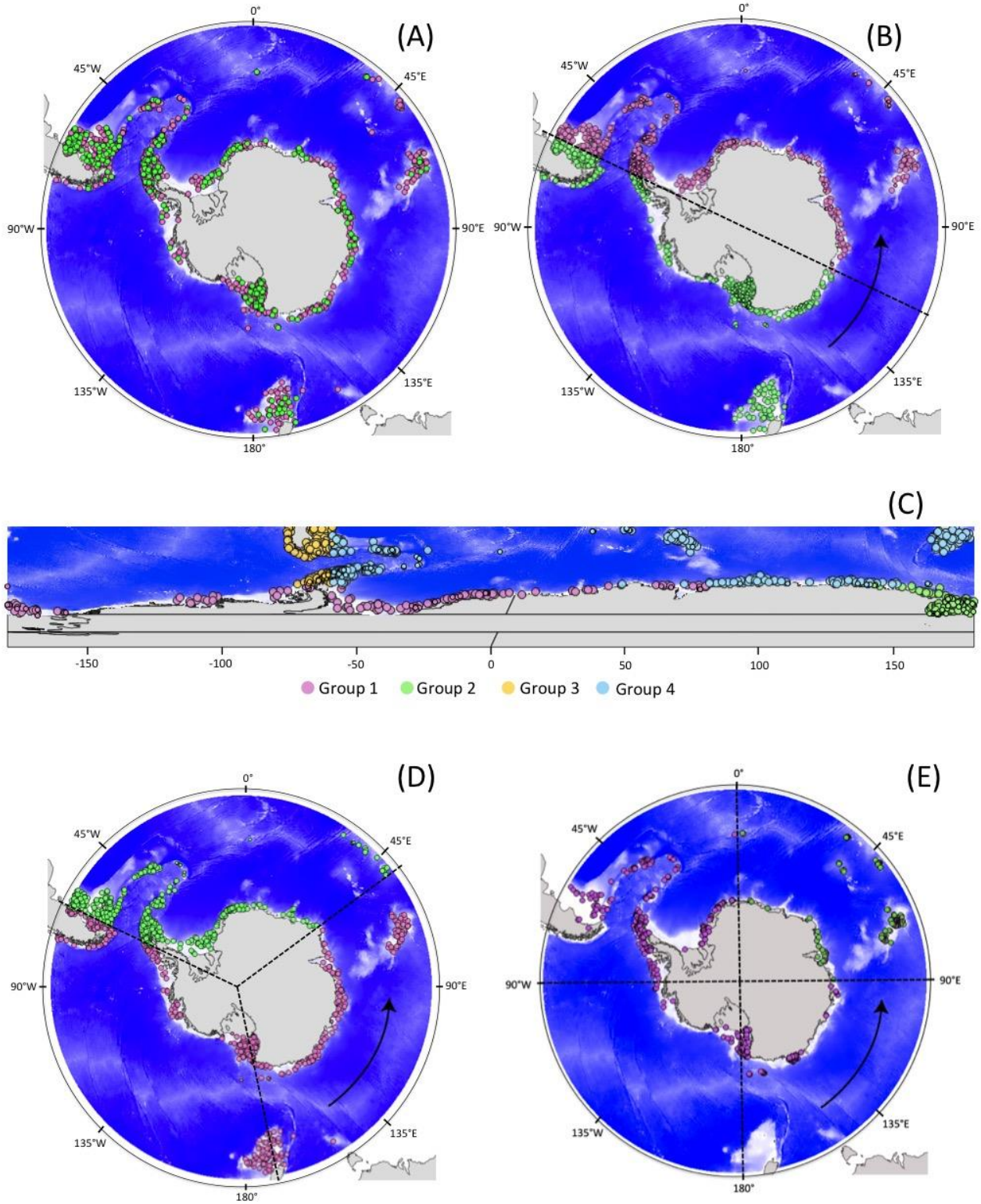


Figure 2.8. Different cross-validation procedures based on the study of the sea star *Odontaster validus*, showing presence-only records and a random set of 1,000 background data selected according to a Kernel Density Estimation (KDE) weighting scheme from the dataset of Griffiths et al. (2014) on sampling effort of the Southern Ocean benthos. Data are split into training (pink) and test (green) subsets. The blue background corresponds to bathymetry and grey areas to emerged lands. (A) Random cross-validation procedure, with a random split into 75% training and 25% test data. (B) '2-fold CLOCK' clustering by random spatial partition of the dataset into two groups (one training and one test). (C) 'BLOCK' splitting, generated according to median latitudinal and longitudinal values (Muscarella et al. 2014). After the generation of four groups (corresponding to the four colours), one group is randomly defined as the test subset and the other three groups as the training subsets. A different system of projection was used to represent this map in order

to highlight the latitudinal and longitudinal definition of the transects. (D) '3-fold CLOCK' clustering by random spatial partition of the dataset into three groups (two training and one test). (E) '4-fold CLOCK' clustering by random spatial partition of the dataset into four groups (three training and one test). Figure extracted from Guillaumot et al. (2019 - Chapter 2).

6. Conclusions and future prospects

This review summarizes some points and issues to be considered during SDM construction for modelling the distribution of SO species (Fig. 2.1). It shows that accurate and efficient SDMs can be produced for SO species when considering potential common biases and issues and correcting for their side effects. Proposed corrections must be adjusted to each case study: no consensus method nor implementation procedure always perform best, each case study requires proper analyses in order to generate the most relevant and accurate predictions. This means that, for each model, several procedures to implement the model should be tested in order to select the most suitable one, ideally giving priority to the availability of independent datasets for evaluating the models. We discuss how SDMs perform best when the species-occupied space is accurately described, using extensive occurrence datasets with both presence and absence records, and when data are checked for positioning and georeferencing errors. A good knowledge of species' ecology, life history traits and populational variations within the overall species distribution and environmental range help to improve model quality (Fois et al. 2018). The compilation, examination and preparation of datasets prior to modelling are essential steps in generating efficient models. Estimating and communicating the uncertainties associated with model predictions are also important tasks to be highlighted. This process may include a 'simple' interpretation of the ecological relevance of SDM outputs by experts (Merow et al. 2017) for the mapping of model extrapolations, as illustrated here. Model uncertainties are part of model outputs and should not be omitted (Guisan et al. 2013, Grimm et al. 2014, Grimm and Berger 2016).

Remaining challenges for constructing relevant SDMs for SO studies include more efforts regarding data collection outside of the main sampling hotspots and filling in knowledge gaps in SO species taxonomy. Some methodological perspectives, developed in other regions, address the integration of physiological information into SDMs. This facilitates the understanding of species environmental preferences and helps one to better estimate the ecological niches of species (Kearney and Porter 2009, Talluto et al. 2016, Mathewson et al. 2017, Rodríguez et al. 2019, Gamliel et al. 2020). Such studies have recently been developed for SO benthic species: in López-Farrán/Guillaumot et al. (in press - Chapter 3), the combination of physiological experimental results and SDM projections allowed for the assessment of the invasive potential of the Patagonian crab *Halicarcinus planatus* (Fabricius, 1775) on Antarctic coasts, as was similarly done in Byrne et al. (2016) for the Arctic sea star *Asterias amurensis* Lutken, 1871. Hybrid modelling approaches constitute another exciting approach, where information from both SDMs and physiological models are fully integrated, using the physiological information as a prior to inform the SDM (Gamliel et al. 2020). Recently applied to an endemic sea urchin of the Kerguelen Plateau (Guillaumot et al. submitted - Chapter 3), the method allows for more precise prediction of the effects of seasonal variations on species habitat suitability.
Other interesting methodological approaches include the consideration of biotic interaction information, dispersal capacity estimates or population dynamics in complement to SDM predictions in order to generalize the understanding of the main drivers of species distribution (Pellissier et al. 2010, Meier et al. 2011, Pagel and Schurr 2012, Conlisk et al. 2013, Pellissier et al. 2013, Leach et al. 2016, Anderson 2017). These, however, necessitate a deep knowledge of the species ecology and of the surrounding environment, suggesting that their first applications should be expected in local- or regional-scale studies.

A final take-home message is that model outputs should be interpreted carefully and model predictions always considered with a critical eye. Models are simple representations of complex systems and should be used to complement other approaches in order to support conservation strategies or to address fundamental research objectives (Porfirio et al. 2014, Kampichler and Sierdsema et al. 2018).

Species distribution models in a data-poor and broad scale context

Guillaumot Charlène[1,3], Artois Jean[2], Saucède Thomas[3], Demoustier Laura[1], Moreau Camille[1,3], Eléaume Marc[4], Agüera Antonio[1,5], Danis Bruno[1]

[1] Université Libre de Bruxelles, Marine Biology Lab. Avenue F.D. Roosevelt, 50. CP 160/15 1050 Bruxelles, Belgique
[2] Université Libre de Bruxelles, Spatial Epidemiology Lab. (SpELL). Avenue F.D. Roosevelt, 50. CP 160/15 1050 Bruxelles, Belgique
[3] UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, 6 bd Gabriel F-21000 Dijon, France
[4] Muséum national d'Histoire naturelle, Dpt Systématique et Évolution, UMR ISYEB 7205, 57 rue Cuvier, F-75231 Paris, France
[5] Danish Shellfish Center, DTU-aqua, Øroddevej 80, 7900 Nykøbing Mors, Denmark.

Progress in Oceanography, 175 (2019), Accepted April 20th, 2019

Abstract

Species distribution models (SDMs) have been increasingly used over the past decades to characterise the spatial distribution and the ecological niche of various taxa. Validating predicted species distribution is important, especially when producing broad-scale models (i.e. at continental or oceanic scale) based on limited and spatially aggregated presence-only records.

In the present study, several model calibration methods are compared and guidelines are provided to perform relevant SDMs using a Southern Ocean marine species, the starfish *Odontaster validus* Koehler, 1906, as a case study. The effect of the spatial aggregation of presence-only records on modelling performance is evaluated and the relevance of a target-background sampling procedure to correct for this effect is assessed. The accuracy of model validation is estimated using k-fold random and spatial cross-validation procedures. Finally, we evaluate the relevance of the Multivariate Environmental Similarity Surface (MESS) index to identify areas in which SDMs correctly interpolate and conversely, areas in which models extrapolate outside the environmental range of occurrence records.

Results show that the random cross-validation procedure (i.e. a widely applied method, for which training and test records are randomly selected in space) tends to over-estimate model performance when applied to spatially aggregated datasets. Spatial cross-validation procedures can compensate for this over-estimation effect but different spatial cross-validation procedures must be tested for their ability to reduce over-fitting while providing relevant validation scores. Model predictions show that SDM generalisation is limited when working with aggregated datasets at broad spatial scale. The MESS index calculated in our case study shows that over half of the predicted area is highly uncertain due to extrapolation.

Our work provides methodological guidelines to generate accurate model assessments at broad spatial scale when using limited and aggregated presence-only datasets. We highlight the importance of taking into account the presence of spatial aggregation in species records and using non-random cross-validation procedures. Evaluating the best calibration procedures and correcting for spatial biases should be considered ahead the modelling exercise to improve modelling relevance.

Key-words

Boosted regression trees (BRTs), presence-only, cross-validation, extrapolation, modelling evaluation

ACKNOWLEDGEMENTS
This work was supported by a “Fonds pour la formation à la Recherche dans l’Industrie et l’Agriculture” (FRIA) grant to C. Guillaumot. This is contribution no. 26 to the vERSO project (http://www.versoproject.be), funded by the Belgian Science Policy Office (BELSPO, contract n°BR/132/A1/vERSO). We are also thankful to the anonymous reviewers that help improve this work with relevant remarks and advices.

AUTHORS’ CONTRIBUTIONS
CG, JA, TS conceived the ideas and designed the methodology; LD provided a part of the data; CM, ME, AA and BD contributed to the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

1. INTRODUCTION

Species Distribution Models (SDMs) have been increasingly used during the past decades. The diversity of applications has widened to include a vast panel of topics from studies of invasive species distribution range shifts to assessment of species responses to environmental drivers and conservation issues from local to global scales (Guisan and Thuiller 2005, Ficetola et al. 2007, Guisan et al. 2013, Beaumont et al. 2016, Phillips et al. 2017). In vast and remote areas such as the Southern Ocean, modelling species distributions is challenged by (1) the paucity of biotic data available (a serious constraint when describing species realised niche), (2) by the heterogeneous quality of environmental data describing environmental conditions (e.g. missing data in coastal areas, low resolution of environmental layers, limited number of environmental descriptors available), and (3) by the sampling bias (spatial and temporal aggregation of data collection) (Barry and Elith 2006, Robinson et al. 2011, Hortal et al. 2012a, Tessarolo et al. 2014, Guillaumot et al. 2018a - Appendix). Sampling effort has mostly been carried out offshore or in the vicinity of research stations during the austral summer while remote shallow areas are seldom accessed and dense winter sea ice conditions limit oceanographic studies (Gutt et al. 2012).

Several studies have proposed model corrections or alternatives to separately mitigate the induced impacts of spatial and temporal biases on modelling performance (Phillips et al. 2009, Newbold 2010, Barbet-Massin et al. 2012, Hijmans 2012, Tessarolo et al. 2014, Guillera-Arroita et al. 2015, Guillaumot et al. 2018a - Appendix, Valavi et al. 2018). However, to our knowledge, no study has yet proposed methodological guidelines to address such issues when dealing with data-poor and broad spatial areas (i.e. at continental or oceanic scales).
Several statistical tools such as the Area Under the Curve of the Receiver Operating characteristic (AUC), the True Skill Statistic, or the Point Biserial Correlation are commonly used to evaluate the relevance of SDM predictions (Fielding and Bell 1997, Allouche et al. 2006). Using these indices for models performed with presence-only data has been widely discussed because background-data are usually considered as absences, leading to confusion in model interpretation and violating most test assumptions (i.e. computing AUC and TSS statistics requires the use of true absences) (Jiménez-Valverde 2012, Li and Guo 2013). These methods can also be biased when applied to limited and broadly distributed data. Machine-learning algorithms are widely used in SDMs to fit complex relationships between species occurrences and environmental data (Elith et al. 2006). The resulting models may be highly complex and poorly efficient under changing environmental conditions as they may fit a response to any variation including the random noise (=model overfitting), (Wenger and Olden 2012). Models' ability to predict in new environmental conditions is described as the generalization performance by Friedman et al. (2001).
Producing reliable SDMs implies finding a good trade-off between model complexity and predictive and generalisation performances (Anderson and Gonzalez 2011, Radosavljevic and Anderson 2014). The relevance of modelling and generalisation performance, and the optimal level of model complexity can be tested using independent data. The method has been commonly applied and referred to as the cross-validation procedure (Araujo and Guisan 2006, Valavi et al. 2018). The cross-validation procedure uses a training subset of occurrence data to fit the model and a separate test subset to validate the predictions and the statistical relationships between the studied variables (Fielding and Bell 1997). 'Random cross-validation' procedures are widely used and randomly split the occurrence dataset into training and test subsets. However, the spatial aggregation of occurrence data can lead to the violation of the independence assumption between training and test data randomly sampled, and in turn to false confidence in modelling validation performances (Hijmans 2012). The violation of the independence assumption can also lead to generate highly complex and overfitted models (Boria et al. 2014, Merow et al. 2013, Radosavljevic and Anderson 2014). Therefore, the cross-validation procedure should be adapted to each given dataset and case study, so that, different 'spatial cross-validation' procedures have been developed and compared in this study. The spatial cross-validation procedures aim at spatially splitting the occurrence dataset into a training and a test subset by increasing the geographical distance between the two subsets (Veloz 2009, Brenning 2012, Muscarella et al. 2014, Radosavljevic and Anderson 2014, Brown et al. 2017, Valavi et al. 2018). The spatial cross-validation reduces spatial correlation between training and test data in situations where spatial autocorrelation is significant in the occurrence dataset, a common issue in ecology (Roberts et al.

2017).
Uncertainties in SDMs represent another limitation to model usage that should be quantified and the effects must be specifically assessed or taken into account during model interpretation (Barry and Elith 2006, Carvalho et al. 2011, Beale and Lennon 2012, Guisan et al. 2013). Model extrapolation outside the range of the known species environmental conditions leads to misinterpretation of SDM outputs and can be a real issue when using SDM predictions as a support tool for conservation decisions. Therefore, areas of optimal predictions and limited uncertainties must be identified. This can be achieved using indicators such as the Multivariate Environmental Similarity Surface (MESS). Developed for SDMs, the MESS index highlights areas where environmental conditions are outside the range of conditions observed in data (Elith et al. 2010).
In the present study, model uncertainties and the performance of several spatial cross-validation procedures were analysed using the case study of the sea star *Odontaster validus* Koehler, 1906. Distributed over the entire Southern Ocean (< 45°S), *O. validus* is a common and abundant species in shallow-water benthic habitats (McClintock et al. 2008a, Lawrence 2013), characterised by an opportunistic feeding behaviour (from suspension-feeding to algivory, deposit-feeding and predation). It has been shown to play a significant role in structuring benthic communities and regulating populations of other benthic taxa (McClintock et al. 2008a). The species physiology was recently modelled using the Dynamic Energy Budget approach (Agüera et al. 2015) which allows for the assessment of the metabolic performance of the species under different environmental conditions. Here, SDMs were produced to interpolate the known distribution of *O. validus* over its entire geographic range using an available dataset of environmental descriptors. The influence of spatial data aggregation on model outputs was analysed and the performance of correction procedures evaluated. In a second step, several cross-validation procedures were assessed and compared to test for modelling accuracy, optimal level of complexity and predictive performance. A final 'optimum' model is proposed, which takes into account uncertainty estimates. Results are generalised and formalised as guidelines for further SDM works, showing the relevance of the approach when working at broad spatial scale with a limited number of spatially aggregated presence-only records.

2. MATERIAL AND METHODS

2.1.Model selection and calibration procedures

SDMs were generated using the Boosted Regression Trees (BRTs) algorithm. BRTs were selected for their ability to fit complex relationships between species records and the related environment, while guarding against over-fitting (Elith et al. 2008, Reiss et al. 2011). BRTs are also adapted to deal with incomplete datasets (Elith et al. 2008), can perform well with low prevalence datasets (Barbet-Massin et al. 2012), are weakly sensitive to species niche width (Qiao et al. 2015) and were recognised to transfer well in space and time (Elith et al. 2006, Elith and Graham 2009, Heikkinen et al. 2012). BRTs were calibrated using the method proposed by Elith et al. (2008) to select the optimal number of trees in the final model (Appendix 2.1). The combination of parameters that minimises the optimal number of trees to build the model (reduction of complexity) while reaching a minimum predictive deviance to the test data (reduction of error) was selected. The following parameters were used to calibrate the models: tree complexity=4, bag fraction=0.75 and learning rate=0.007 (Fig. S2.1B). The number of background data sampled in the area was set at 1000 sampled points after evaluating the optimal number of data points to be sampled (see Appendix 2.1 for details). This number constitutes the best trade-off between describing environmental conditions and being as close as possible to the number of species presence records available (Barbet-Massin et al. 2012). All background sampling was restricted in space to areas shallower than 1500m depth, which corresponds to the species deepest record, in order to avoid model extrapolation at depths known as unsuitable for the species survival based on knowledge of the species ecology (McClintock et al. 2008a, Lawrence 2013). Sampling was restricted to a single background data per pixel. Similarly, presence records falling on a same 0.1° grid-cell pixel were filtered before model calibration in order to reduce spatial over-weighting (Segurado et al. 2006, Boria et al. 2014).

2.2. Occurrence dataset

SDMs were generated using presence-only data made available for the sea star *O. validus* by Moreau et al. (2018). Presence-only records of *O. validus* are strongly aggregated in space (i.e. concentrated in "easily" accessible and frequently visited areas characterised by relatively low sea ice concentrations), a condition also prevailing in the total dataset available for Southern Ocean benthic taxa (updated from Griffiths et al. (2014), Fig. S2.2), making *O. validus* a representative case study for Southern Ocean benthic studies. Models were generated using the environmental descriptors published as raster layers by Fabri-Ruiz et al. (2017b). They were collected from different sources and modified to fit modelling requirements at the scale of the Southern Ocean (from 45°S latitude to Antarctica coasts). Collinearity between environmental descriptors was tested using the Variance Inflation Factor (VIF) stepwise procedure of the 'usdm' R package (Naimi et al. 2014) and Spearman correlations (rs). Surface temperature and roughness, a depth-derived variable, were respectively correlated to ice cover and depth. They were omitted according to the commonly used thresholds of VIF > 5 and rs > 0.85 (Pierrat et al. 2012, Dormann et al. 2012b, Duque-Lazo et al. 2016). A final set of 16 environmental descriptors at 0.1° resolution was compiled to build the models (Table S2.3).

2.3. Evaluation and correction spatial aggregation

The significance of spatial aggregation of occurrence data was tested by measuring spatial autocorrelation (Legendre and Fortin 1989) on model residuals using the Moran's I index (Segurado et al. 2006, Dormann 2007, Crase et al. 2012). A positive Moran's I value (between 0 and 1) indicates that spatially close residuals will share similar values. A negative (close to −1) or null value respectively indicates a maximal dispersion or a random dispersion of residuals in space (Cliff and Ord 1981). Detecting significant spatial autocorrelation in presence-only records will assess the degree of aggregation of species records in the studied area.
Two null models were generated and their respective outputs compared to each other in order to evaluate the importance of spatial aggregation in the total Southern Ocean benthic dataset (Fig. S2.2). Null model #1 was produced to evaluate the overall spatial aggregation of benthic records in the Southern Ocean due to sampling effort. It was generated by randomly sampling n=309 occurrence records (corresponding to the number of non-duplicate presence-only data available for *O. validus*) in the total Southern Ocean benthic dataset (Fig. S2.2). 1000 background records were randomly sampled in the entire Southern Ocean. The Moran's I score was calculated by comparing model #1 predictions to the distribution of the total Southern Ocean benthic dataset (Fig. S2.2). Null model #2 was built to compute a reference Moran's I score for a model generated with randomly distributed records. 309 presence data and 1000 background data were randomly sampled in the entire Southern Ocean. Null model #2 will provide a reference value for spatial autocorrelation scores due to the intrinsic structure of environmental data. It will serve as a reference model for comparison with Moran's I scores of model null #1 and to assess the degree of spatial aggregation due to sampling effort. To correct for the effect of spatial aggregation on modelling performance, a target-background correction method was applied (Phillips et al. 2009). The total Southern Ocean benthic dataset (Fig. S2.2) was used to create a Kernel Density Estimation layer that provides an estimate of the probability to find a benthic presence data for each pixel. The Kernel Density Estimation was calculated with the 'kde2d' function of the MASS R package (Ripley 2015) on the extent of the Southern Ocean (n and lims parameters defined to fit a raster layer of extent (−180, 180, −80, −45) and 0.1° resolution). Null model #1 was corrected by randomly sampling 1000 background records according to the weighting scheme of the Kernel Density Estimation layer. After evaluating spatial aggregation in the total Southern Ocean benthic dataset, spatial autocorrelation was specifically assessed for *O. validus*. Spatial autocorrelation was measured for two models generated without (model A) and with (model B) Kernel Density Estimation correction. Comparison between the two models aimed at assessing the efficiency of the Kernel Density Estimation correction for *O. validus*. Model A (without correction) was built using all presence-only data available for *O. validus* and 1000 background records randomly sampled in the Southern Ocean. Model B (with correction) was built using all presence-only data available for *O. validus* and 1000 background records that were sampled following the weighting

scheme of the Kernel Density Estimation layer. Each model was generated 100 times and the two averaged models (average models A and B) were compared to each other. Differences between models A and B quantify the importance of spatial aggregation on model outputs. Finally, model relevance was assessed using three statistics: the Area Under the Receiver Operating Curve (AUC) (Fielding and Bell 1997), the Point Biserial Correlation between predicted and observed values (COR, Elith et al. 2006) and the True Skill Statistic (TSS, Allouche et al. 2006).

2.4. Testing different cross-validation procedures

SDMs validation was performed using different cross-validation procedures. Background data were first sampled in the entire area following the Kernel Density Estimation scheme and the compilation of presence-only and background data was then split into a training and a test subset to build the cross-validation procedure. Two splitting procedures were followed; they differ between each other in the spatial independence between the training and the test subset. (1) The random cross-validation procedure, commonly used in SDMs, aims at randomly splitting the dataset into training and test subsets (Fielding and Bell 1997, Hijmans 2012) which may lead to close spatial vicinity between the two datasets (Hijmans 2012), and, (2) the spatial cross-validation procedure that aims at spatially spitting the dataset in order to reduce spatial correlation and may improve independence between the two subsets (Hijmans 2012, Muscarella et al. 2014). The random procedure was therefore compared to four different spatial cross-validation procedures. (1) In the 'BLOCK' method developed by Muscarella et al. (2014), different subsets of equal occurrence numbers are created. For each replicate, this k-fold procedure divides the dataset into four equal subsets according to the mean latitude and mean longitude positions of occurrence data (Fig. 2.9c), then three of these four subsets are randomly selected to train the model (75%) and the last one is used to test the model (25%). (2) In the 'CLOCK' methods, the dataset was divided according to random longitudinal transects, splitting the Antarctic Circle into two parts (2-fold 'CLOCK' method, Fig. 2.9b), (3) three parts (3-fold 'CLOCK' method, Fig. 2.9d) or (4) four parts (4-fold 'CLOCK" method, Fig. 2.9e). In the 2-fold 'CLOCK' method, one subset was considered as the training subset, the second one as the test subset; in the 3-fold 'CLOCK' method, two subsets were defined for training and the third one for testing; in the 4-fold 'CLOCK' method, three subsets were considered for training and one for testing (Fig. 2.9). Different cross-validation procedures were tested using the 'gbm.step' procedure available in the *dismo* R package (Elith et al. 2008, Hijmans et al. 2017). Once the dataset is split in different folds, Elith et al. (2008) apply an iterative procedure that enable to find the minimum deviance to the test data, and relates it to the optimal number of trees (optimal model complexity) to generate the model. If test and training data are spatially correlated, the number of trees required to build BRTs will be overestimated. Therefore, the use of Elith et al. (2008) procedure will enable to accurately interpret and compare optimal complexity and performance scores of models calibrated with either randomly or spatially segregated folds (i.e. with contrasting distances between training and test subsets), and thus will help explain the influence of occurrence spatial aggregation on model complexity and performance. R scripts written to generate the models and the different cross validation procedures are provided online at: https://github.com/charleneguillaumot/THESIS/.

Independence between training and test subsets was evaluated using the Spatial Sorting Bias index (SSB) (Hijmans 2012). SSB compares the distance between training-presence and testing-presence data with the distance between training-presence and training-background. SSB=0 (non independence) means that the ''distance between training-presence and test-presence sites will tend to be smaller than the distance between training-presence and test-background sites'' (Hijmans 2012). SSB=1 indicates that the two distances are comparable (independent enough) (Hijmans 2012). SSB was calculated with the *dismo* R package (Hijmans et al. 2017).

SDMs evaluation was generated by computing the percentage of test data that fall on grid-cell pixels predicted as suitable. Suitable pixels were defined using the Maximum sensitivity plus specificity threshold (MaxSSS) that splits models into suitable (> MaxSSS value) and unsuitable areas (< MaxSSS value). MaxSSS is accepted as a relevant threshold for presence-only SDMs (Liu et al., 2013). The averaged optimal number of trees required to generate BRTs was compared between models and used as a proxy of model complexity.

Statistical differences between models generated with the different cross-validation procedures (AUC, TSS, COR, percentage of correctly classified test data, number of trees) were tested using

the non-parametric Mann-Whitney Wilcoxon pairwise comparison.

2.5. Assessment of model extrapolation

The Multivariate Environmental Similarity Surface (MESS) index was estimated following the procedure described by Elith et al. (2010) using the *dismo* R package (Hijmans et al. 2017). The MESS calculation consists in extracting the environmental conditions where presence-only data were recorded and determining for each pixel of the model projection layer if environmental conditions are covered by presence-only records. Negative MESS values indicate areas of model extrapolation in which the value of at least one environmental descriptor is beyond the environmental range covered by available presence-only records. Conversely, positive MESS values indicate areas of model projection in which values of environmental descriptors are within the environmental range covered by presence-only records. According to the number of environmental descriptors that are not included inside the range of presence records values, MESS outcome can strongly vary. The MESS evaluation deals with each environmental descriptor equally (unweighted analysis) and in this study, a pixel was considered as unsuitable as soon as a single descriptor value does not match the environmental range of presence-only records. On a projection map, SDM predictions were darkened according to the MESS extrapolation range to visualise the uncertain area due to extrapolation. Extrapolation performance of SDMs was assessed by comparing the proportion of the environment predicted as suitable by the model with the total set of environmental conditions.

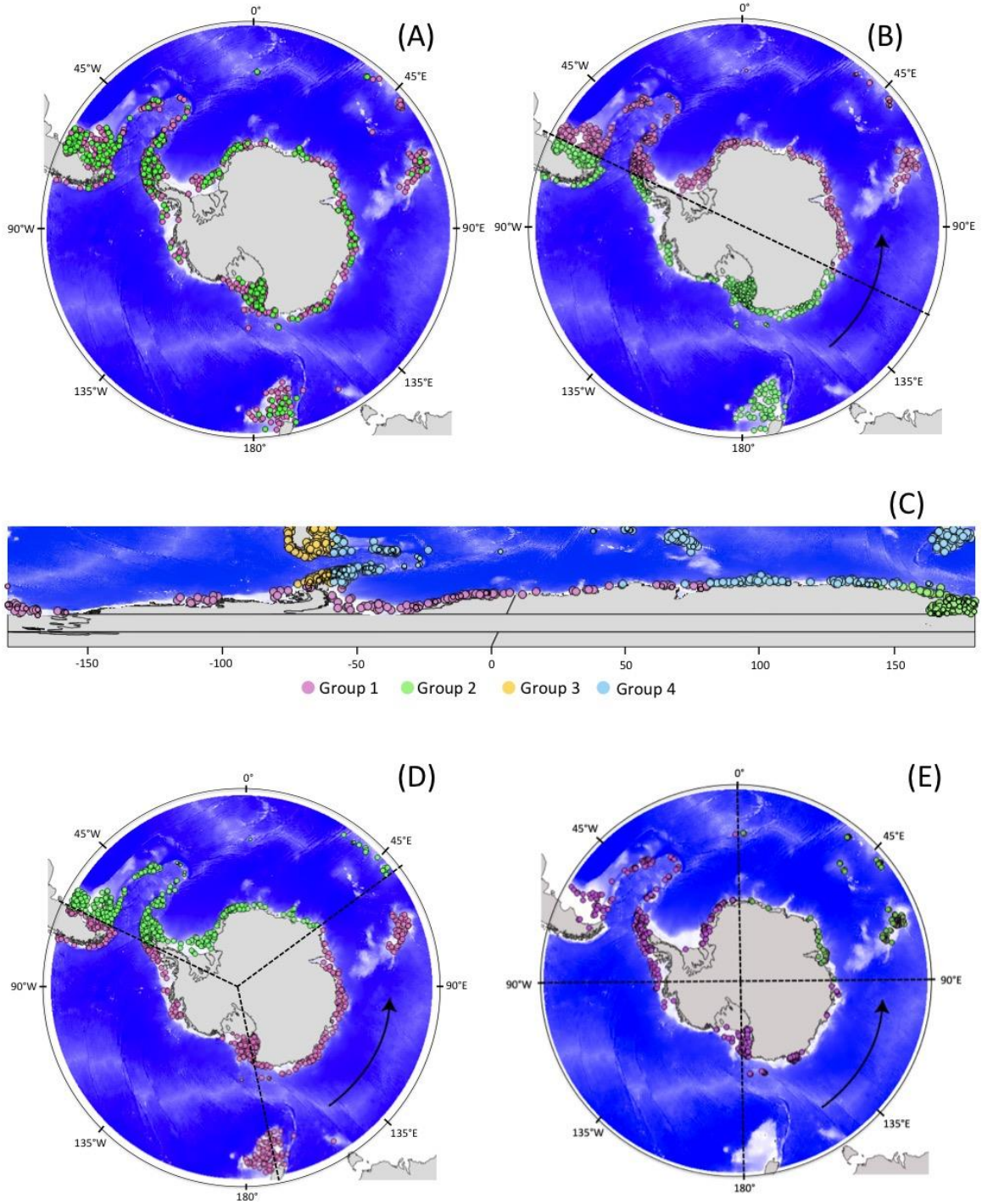


Figure 2.9. Comparison of the different cross-validation procedures. Dots represent *Odontaster validus* presence-only records and a random set of 1000 background data, sampled according to the Kernel Density Estimation weighting scheme. Colors indicate data splitting into training (pink) and test (green) subsets. Blue background corresponds to bathymetry and grey areas to emerged lands. For each case, 100 replicates of random background-data sampling and transects partitioning are performed, symbolised by the arrows on the figure. (A) Random cross-validation procedure, with a random splitting into 75% training and 25% test data. (B) '2-fold CLOCK' clustering by random spatial partition of the dataset into two groups (one training, one test). (C) 'BLOCK' splitting, generated according to the median latitudinal and longitudinal values (Muscarella et al., 2014). After generation of four groups (corresponding to the four colors), one group is randomly defined as the test subset, the other three groups as the training subset. A different system of projection was used to represent this map to highlight the latitudinal and longitudinal definition of the transects. (D) '3-fold CLOCK' clustering by random spatial partition of the dataset into three groups (2 training, 1 test). (E) '4-fold CLOCK' clustering by random spatial partition of the dataset into four groups (3 training, 1 test).

3. RESULTS

3.1. Available data and spatial autocorrelation

Distribution records available for *Odontaster validus* display a circumpolar and patchy spatial pattern (Fig. 2.10a). The niche occupied by *O. validus* does not cover the entire range of environmental conditions prevailing in the projection area (Fig. 2.10b). *O. validus* is recorded in conditions close to zero and sub-zero seafloor temperatures (Fig. 2.10b) and is mainly distributed in shallow and coastal areas. Most of *O. validus* presence records are aggregated in regions where scientific benthic surveys are most often led and where sampling effort was privileged due to access facilities (e.g. the Ross Sea and the Antarctic Peninsula). Overall, this holds true for presence records of all benthic Southern Ocean taxa as well (Fig. S2.2), although, in this case, most environmental conditions are covered by the total benthic samples (Fig. 2.10b).

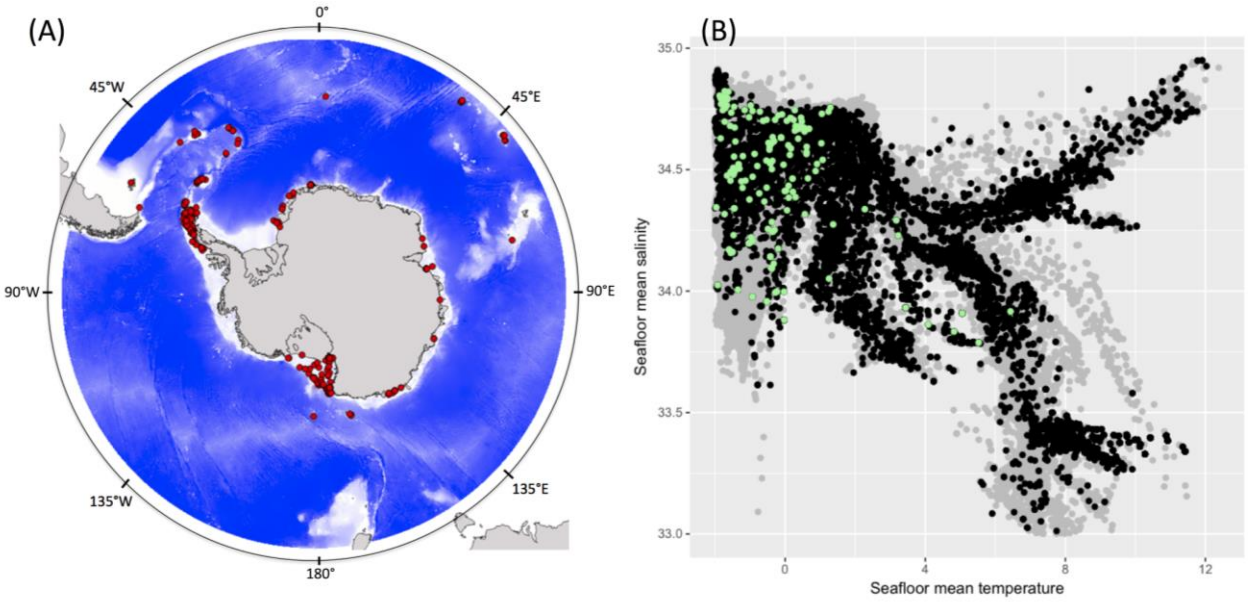


Figure 2.10. (A) Presence-only records of the sea star *Odontaster validus* in the Southern Ocean. Duplicates (occurrences falling on a same 0.1° resolution pixel) were removed from the display. (B) Values of the environmental range covered by the entire benthos sampling dataset presented in Fig. S2.2 (black dots), by presence-only records of *O. validus* (green dots) in comparison with a set of 1000 background dots randomly sampled according to the Kernel Density Estimation scheme (grey dots) for two environmental descriptors: mean seafloor temperature (°C) and mean seafloor salinity (PSU). A part of the environment (grey dots) does not contain benthic occurrence samples (black dots), illustrating that sampling effort is not geographically exhaustive.

Spatial autocorrelation was measured for both the total Southern Ocean benthic dataset (null models) and or *O. validus* alone (models A and B) (Table 1). Moran's I scores were tested significant for all models, null model #2 excepted. The absence of spatial autocorrelation (I=0.005 ± 0.004; p=0.19) in null model #2 shows that environmental data are not strongly aggregated in space. In contrast, presence-only records of the total Southern Ocean benthic dataset are spatially aggregated. The degree of spatial aggregation due to sampling effort is evidenced by the comparison between null model #1 and #2, scores of model #1 being 10 times higher than those of null model #2 (Moran's I=0.050 ± 0.011 and 0.005 ± 0.004, respectively).
Values of Moran's I computed for models of *O. validus* (models A and B) are higher than those computed for the total Southern Ocean benthic dataset (null model #1 and #1 with Kernel Density Estimation). The sampling bias is therefore more pronounced for *O. validus* than for the majority of other benthic species.
Model correction by the Kernel Density Estimation procedure was shown to reduce spatial autocorrelation with Moran's I values decreasing from 0.050 to 0.034 for null model #1, and from 0.085 to 0.069 for *O. validus* models A and B (Table 2.1). However, although lower, Moran's I values remain significant after correction, indicating that the applied corrections do not entirely remove the presence of spatial autocorrelation.

Table 2.1. Comparison between models of spatial autocorrelation values measured on model residuals (average and standard deviation of Moran's I values computed for 100 model replicates). Moran's I significance is indicated by p-values; for $p < 0.05$, the absence of spatial autocorrelation (null hypothesis) is rejected. Null model #1: 309 presence records were randomly sampled among occurrences of the total Southern Ocean benthic dataset (Fig. S2.2) and background data are composed of 1000 points randomly sampled in the entire Southern Ocean; model #2: 309 records (to define presence records) and 1000 background data both randomly sampled in the entire Southern Ocean; model #1 with Kernel Density Estimation: similar to model null #1 but with 1000 background data randomly sampled following the Kernel Density Estimation weighting scheme; model A: 309 presence records of *Odontaster validus* and 1000 background data were randomly sampled in the entire Southern Ocean; model B: similar to model A but with the 1000 background data sampled following the Kernel Density Estimation weighting scheme. AUC: Area Under the Receiver Operating Curve, TSS: True Skill Statistic, COR: Point Biserial Correlation.

	Null model #1	**Null model #2**	**Null model #1** with Kernel Density Estimation	**Model A**	**Model B**
Spatial autocorrelation (Moran's I)	0.050 ± 0.011 *p<0.001*	**0.005 ± 0.004** ***p=0.19***	0.034 ± 0.011 ***p<0.001***	0.085 ± 0.009 *p<0.001*	0.069 ± 0.006 *p<0.001*
AUC	0.976 ± 0.010	0.710 ± 0.014	0.964 ± 0.015	0.997 ± 0.001	0.948 ± 0.003
TSS	0.674 ± 0.013	0.331 ± 0.020	0.660 ± 0.019	0.698 ± 0.002	0.696 ± 0.003
COR	0.850 ± 0.028 *p<0.001*	0.336 ± 0.018 *p<0.001*	0.801 ± 0.037 *p<0.001*	0.944 ± 0.011 *p<0.001*	0.923 ± 0.015 *p<0.001*

3.2. Comparison of cross-validation procedures

For the BRTs fitted with the random cross-validation procedure, all overall goodness-of-fit metrics (AUC, TSS, COR) were good with predictive accuracy Area Under the Curve (AUC) values higher than 0.9 (Table 2.2). However, when evaluated through spatial cross-validation procedure, the AUC scores decreased in all BRTs. These results show that BRTs tend to overfit the data if the independence between training and test data is not ensured. Indeed, the random cross-validation procedure presents SSB values close to zero, indicating that training and test subsets may be highly correlated (Fig. 2.9a). In contrast, all spatial cross-validation procedures have SSB values close to 1, indicating a better spatial independence between training and test data (Table 2.2).
The generalisation performance (AUC and correctly classified test data) are very high for the random cross-validation procedure, with more than 89.4% of test-presence records falling correctly in areas predicted as suitable by the model (Table 2.2). The random cross-validation procedure generates more complex BRTs compared to the spatial methods (significantly higher number of trees for the random cross-validation procedure compared to the spatial cross-validation procedures). As the model closely fits the dataset used for its construction, high AUC, TSS and COR scores were obtained but these results may be misleading and overestimated. In contrast, spatial cross-validation procedures generate less complex models (more general), which could account for lower AUC, TSS and COR scores.

Table 2.2. Average Spatial Sorting Bias (SSB) and standard deviation values for the 100 model replicates (background sampling+test/training clustering). AUC: Area Under the Receiver Operating Curve; Correctly classified test data (%): percentage of presence-test and background-test records falling on predicted suitable areas (prediction > maximum sensitivity plus specificity (maxSSS) threshold); TSS: True Skill Statistic; COR: Point Biserial Correlation; ntrees: averaged optimal number of trees required to generate BRTs. Stars are indicated for spatial cross-validation groups significantly different from the random cross-validation procedure (nonparametric pairwise Mann-Whitney Wilcoxon test, p-value < 0.01).

	Random cross-validation Random splitting	Spatial cross-validation Block method	Spatial cross-validation 2-fold Clock method	Spatial cross-validation 3-fold Clock method	Spatial cross-validation 4-fold Clock method
Mean SSB	0.101 ± 0.04	0.802 ± 0.37	0.832 ± 0.09	0.803 ± 0.23	0.848 ± 0.32
AUC	0.947 ± 0.013	0.854* ± 0.06	0.811* ± 0.053	0.818* ± 0.078	0.824* ± 0.089
Correctly classified test data (%)	89.452 ± 1.523	80.946* ± 7.504	80.039* ± 3.489	80.713* ± 5.421	79.471* ± 8.538
Test data (% of total dataset)	25%	[13-38]%	[19-81%]	[1-68%]	[1-66%]
TSS	0.715 ± 0.041	0.542* ± 0.188	0.465* ± 0.088	0.490* ± 0.136	0.576* ± 0.165
COR	0.792 ± 0.029	0.632* ± 0.126	0.584* ± 0.089	0.591* ± 0.12	0.483* ± 0.197
ntrees	1580 ± 251.058	543.5* ± 88.9	375* ± 91.9	424.5* ± 131.1	379* ± 98.5

3.3. Proposed model and uncertainty map

We decided to maximise the spatial independence between training and test subsets, minimise model complexity and optimise generalization performances in *O. validus* model. Using these criteria, we found that the '2-fold CLOCK' modelling method was well adapted to *O. validus* dataset (second highest TSS and COR scores; high proportion of test data being correctly classified, with the lowest standard deviation score (80.04 ± 3.49%); an important proportion of the total dataset used a test subset [19–81%] and the lowest model complexity (ntrees = 375 ± 91.9)).
The MESS index was calculated in order to define the part of this extrapolated area, that is, the part of the geography for which at least one environmental descriptor is outside the environmental conditions of the sampled presence records. The MESS index was compiled as a raster layer and projected on the probability distribution map by darkening uncertain areas (Fig. 2.11). Uncertain areas due to extrapolation represent 64.2% of the entire projected surface, the major part being also predicted by the model as unsuitable (Table 2.3). Almost 9.5% of the area was however predicted as suitable by the model although considered as an extrapolated area.

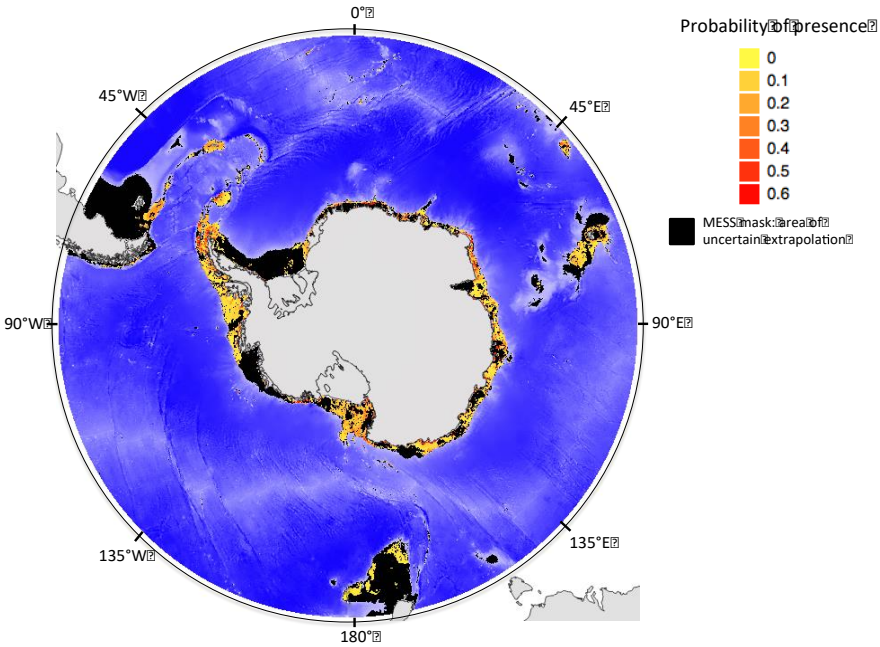


Figure 2.11. SDM performed with the spatial cross-validation '2-fold CLOCK' method. Average of 100 model replicates. Distribution probabilities are darkened according to the Multivariate Environmental Similarity Surface (MESS) layer, with dark pixels corresponding to regions where the model extrapolates outside of the environmental conditions in which the species was sampled. Dark pixels represent 64.2% of the entire projected area. Probabilities of presence are contained between 0 and 1 but the colorbar was scaled until 0.6 to enhance visual contrast.

Table 2.3. Proportion of interpolated and extrapolated pixels according to the averaged SDM predictions. Interpolation (or uncertain extrapolation respectively) refers to areas where environmental conditions within the pixel are inside (or outside, respectively) of the species ecological range, as defined by the Multivariate Environmental Similarity Surface (MESS). Suitable pixels were defined using the MaxSSS threshold that splits model predictions into suitable (> maxSSS mean score) or unsuitable areas (< maxSSS mean score).

MESS classification	**Model prediction**	
	Suitable pixels	Unsuitable pixels
Interpolation	10.24%	25.57%
Uncertain extrapolation	9.42%	54.77%

4. DISCUSSION

4.1. Evaluating SDM performance

Using independent datasets to test SDM performance is a prerequisite for relevant validation analyses (Peterson et al. 2011). At broad spatial scale and in data-poor areas, the number of available data is limited and data distribution often patchy, which really challenges the success of validation procedures. Estimating the performance of SDM predictions and the level of extrapolation in such areas is a necessity.

The cross-validation procedure has been proposed as a reliable approach to evaluate modelling performances (Fielding and Bell 1997, Hijmans 2012, Dhingra et al. 2016, Roberts et al. 2017). Cross-validation procedures must however be adapted to spatially aggregated data because training and test subsets may be sampled in close areas, violating the independence assumption (Segurado et al. 2006, Hijmans 2012). Such a potential bias is rarely taken into account. In the present work, we compared SDM performance using five different cross-validation procedures for

modelling, at broad spatial scale, the distribution of a species for which available data are limited in number and are spatially aggregated. Results show strong differences between procedures, which highlights the importance of testing and selecting the most appropriate method when evaluating model performance.

4.2. Correction for SAC and spatial bias

Strong significant Moran's I scores were measured on model residuals, revealing the presence of spatial autocorrelation in the total Southern Ocean benthic dataset (Fig. S2.2). The difference between null models #1 and #2 evidences the influence of sampling aggregation on spatial autocorrelation values (Table 2.1) as discussed by Guillaumot et al. (2018a - Appendix). *O. validus* presence-only dataset follows the same pattern, with records aggregated in coastal areas where sampling effort has been mostly concentrated (Table 2.1, Fig. 2.10). A target-group background sampling was applied and proved to be efficient to reduce spatial autocorrelation (as assessed using Moran's I statistic), although it still remains at a significant level. Spatial autocorrelation scores are strongly dependent on the resolution of environmental raster layers. The coarse resolution of environmental data used in the present study may be responsible for the over-estimation of spatial autocorrelation scores. This could account for spatial autocorrelation remaining significant even after the Kernel Density Estimation correction.

4.3. Selection of cross-validation procedures

The random cross-validation procedure has been widely used in ecological modelling to evaluate model predictions (Fielding and Bell 1997, Merow et al. 2013, Mainali et al. 2015, Torres et al. 2015, Phillips et al. 2017) but the method has been rarely compared to alternative procedures. The present study shows that contrasting model assessments are obtained when using different cross-validation procedures (Radosavljevic and Anderson 2014, Roberts et al. 2017). Applying a random cross-validation to an aggregated dataset at a broad spatial scale can result in training and test subsets being sampled in the same area, and leads to an inflation of modelling performances (Veloz 2009, Hijmans 2012, Radosavljevic and Anderson 2014, Wenger and Olden 2012). In the context of this study, SDMs produced with a broad-scale and spatially aggregated occurrence dataset and a random cross-validation procedure are more complex and likely over-fit the training dataset. This also may account for the high evaluation scores obtained (AUC, TSS, COR) and may also explain the apparent high generalization performance of BRTs fitted with random cross-validation. The lack of model generality can *a posteriori* lead to strong caveats and unreliable models with poor transferability performance when projected on a new environmental space (Wenger and Olden 2012, Crimmins et al. 2013). Methods that select the most parsimonious BRT, combine low model complexity and high modelling performance should therefore be preferred. The spatial cross-validation procedures tested in this study were shown to produce less complex models than the random cross-validation procedure. Increased model generality (i.e. decrease in model overfitting) and forced spatial segregation between training and test subsets result in decreasing SDM validation scores. These results show that applying a random cross-validation procedure for a patchy dataset can lead to over-estimation of SDM predictive performance if training and test subsets are not independent. This is in line with several works (Brenning 2005, Elith et al. 2010, Anderson 2013, Muscarella et al. 2014) in which a decrease of AUC scores can be reported when using a spatial cross-validation procedure instead of a random procedure. Machine-learning algorithms have been reported to be the best approaches to generate SDMs but the influence of over-fitting on model evaluation are under-estimated (Reiss et al. 2011, Duan et al. 2014, Beaumont et al. 2016, Thuiller et al. 2016) although its effect has been pointed out in several works (Elith et al. 2008, Jiménez-Valverde 2008, Wenger and Olden 2012). Our results show that the evaluation of SDM performance can be strongly influenced by the choice of the evaluation procedure. In this work, several spatial cross-validation procedures were compared with each other but no single and best procedure emerged, a common case in ecological modelling (Qiao et al. 2015). The appropriate method to be used is highly dependent on the species and dataset under study. For instance, the 'BLOCK' method introduced by Muscarella et al. (2014) should not be used at broad spatial scale, where too important latitudinal contrasts in environmental conditions are present. In this study, such contrasting environmental conditions (due to the

presence of an environmental latitudinal gradient between sub-Antarctic and Antarctic regions, with occurrence aggregation in the two regions) lead to higher variability in generalisation performance during model projection, depending on the data subsets selected to train and test the model (Roberts et al. 2017). The 'BLOCK' method favors the independence between training and test subsets but models are slightly more complex because they are calibrated on contrasting environmental conditions (sub-Antarctic vs. Antarctic areas) and over-fit the training dataset that could also present a patchy distribution. The 'BLOCK' method is therefore more adapted to case studies without strong patchy and contrasting environmental conditions. The 'CLOCK' procedures developed in this study helped reduce the effect of latitudinal patchy occurrences distribution by mixing presence records sampled in Antarctic and sub-Antarctic regions to define training and test subsets. The 'CLOCK' methods generate less complex models and were proved more efficient to define spatially independent training and test subsets. However, the number of training and test records sampled between model replicates is not constant, which contributes to an important variability in validation performance scores. The selection of the different 'CLOCK' methods also depends on the importance of data aggregation and patchy patterns within environmental conditions. For strong data aggregation, the '2-fold CLOCK' approach will help reduce the influence of patchy patterns during model calibration and will help generalise the model and decrease its complexity. '3 or 4-fold CLOCK' methods present close modelling performances but the proportion of occurrence records used to test the model can be very low. Alternative SDM evaluation procedures can be found in the literature: for instance, calibrated cross-validation procedures aim at removing occurrences from the test subset when considered too close to the training subset (and considered as non-informative according to a statistical threshold) (Hijmans 2012). For limited presence-only datasets, removing a part of the available occurrence data may lead to the removal of a proportion of informative records, which does not constitute a reasonable option (Bean et al. 2012, van Proosdij et al. 2016). The leave-one-out method can also provide a relevant estimate of model goodness-of-fit, even for spatially aggregated datasets (Olden et al. 2002, Wenger and Olden 2012). The method aims at randomly excluding a single record from the total dataset. The model is trained on the remaining data and predicts the model response on the single removed point to test for model prediction. The procedure is replicated several times, providing a powerful evaluation of model accuracy. However, assessment of generalisation performances is not permitted with this approach (Wenger and Olden 2012). In addition to cross-validation procedures, the relevance of model validation performance is also strongly dependent on the quality of environmental descriptors available. The number of no-data pixels as well as grid-cell resolution can critically affect model evaluation. This is especially true in the present study because environmental variables, measured or interpolated, rarely extend to coastal areas, and resolution in the Southern Ocean can rarely be better than 10 km^2. Good quality datasets are needed and such limitations must be taken into account when interpreting model outputs.

4.4. Uncertainty assessment in SDMs predictions

SDM uncertainty assessment has been a widely discussed topic (Barry and Elith 2006, Carvalho et al. 2011, Beale and Lennon 2012, Guisan et al. 2013). Uncertainty in model predictions has been often assessed as the variation among the predicted distribution probabilities (Buisson et al. 2010) but this approach does not provide precise information on the origin of uncertainty (Tessarolo et al. 2014). The MESS metric is a relevant indicator of SDM extrapolation performance (Elith et al. 2010, Dhingra et al. 2016). The Mobility Oriented Parity (MOP) introduced by Owens et al. (2013) was recently proposed as an alternative to the MESS index. MESS considers extrapolation on a pixel as uncertain when at least one environmental value falls outside the environmental range of presence records. In contrast, MOP offers more flexibility by defining an extrapolated area when all environmental values fall outside the sampled environmental range. Therefore, MESS is more conservative than MOP to define species ecological envelope. Here, MESS was used to assess the proportion of the projected area for which models extrapolate. Our results show that more than half of the area corresponds to environmental conditions for which presence records have not been sampled. 9.42% of this extrapolated area is even predicted as a suitable environment. This highlights the weakness of SDMs for spatial generalisation and the risk of providing inaccurate SDMs for conservation purposes, especially if the communication between modellers and environmental managers is neglected (Guisan et al. 2013). Our results show the importance of

providing uncertainty maps along with SDM outputs in order to help interpret models with the necessary caution.

5. CONCLUSION

This work highlights the importance of assessing the relevance of SDM evaluation procedures. When applied to occurrence datasets, spatially autocorrelated and broad-scale presence-only datasets, the random cross-validation procedure may over-estimate model validation scores due to the violation of independence between training and test subsets. Applying a spatial cross-validation procedure that spatially segregates training and test data was shown to be effective to provide a reliable analysis of model performance. Spatial cross-validation methods also help reduce model complexity and therefore improve generalisation performances. The 'CLOCK' methods developed in this paper were proved to be appropriate to our Southern Ocean case study and could be applied to other non-polar case studies. This study proves the importance of testing and comparing several spatial cross-validation procedures to identify the procedure most adapted to each case study. The MESS index was used to visualise areas where SDMs extrapolate outside the range of the environmental conditions where presence records were sampled. Such results show the importance of providing information on model uncertainty to correctly interpret SDM outputs.

APPENDIX 2.1. Model calibration

Models were calibrated using all presence-only records available and a random selection of background points sampled within the species environmental range (< 1,500m depth). Different numbers of background data were sampled and compared to the total environmental range using convex hulls (Fig. S2.1.A). The best background data number to be used to calibrate SDM was the one describing well environmental conditions (e.g. mean seafloor salinity, depth, mean seafloor temperature, seafloor temperature amplitude, Fig. S2.1.A) while being as close as possible to the number of species presence records (Barbet-Massin et al. 2012). 1,000 background data were finally sampled to perform the model.

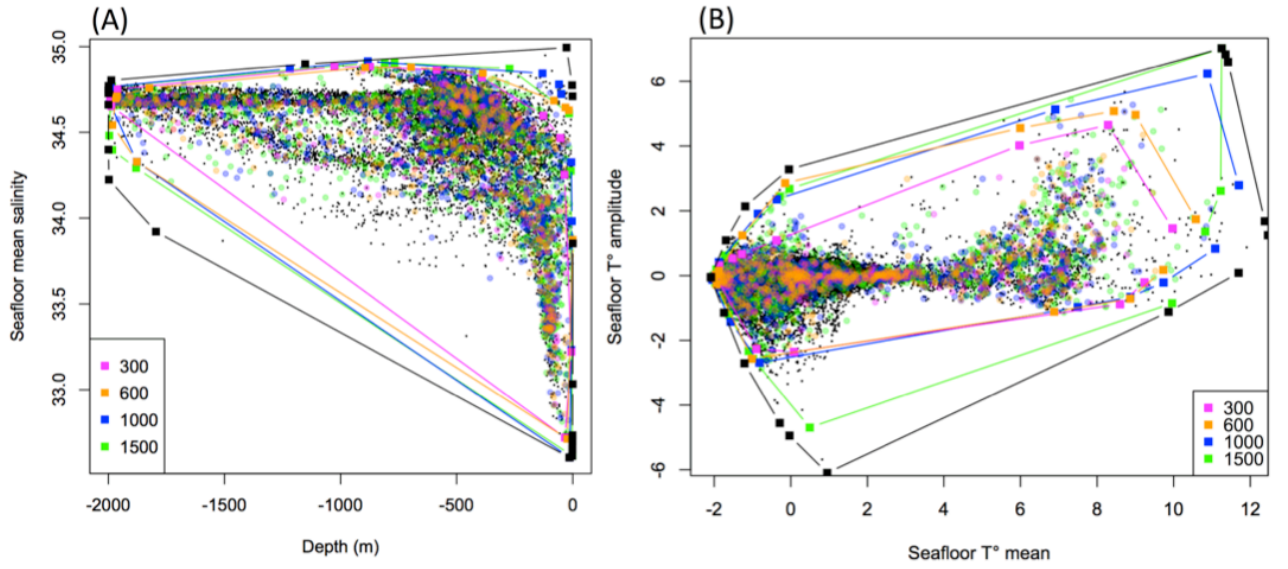


Figure S2.1.A. Values of the environment available (black dots) and of the background sample environment randomly sampled on the environment limited at 1,500m depth (coloured dots). 300, 600, 1000 and 1500 background data were sampled. Convex hulls were calculated with the *chull* function of the *grDevices* R package. They delimit the environment described by the background data sample.

BRT models were generated using the cross validation procedure of Elith et al. (2008) and the *gbm* R package (Ridgeway et al. 2006) with codes provided in the publication's supplementary material. We forced a maximum number of 10,000 trees and models were calibrated with the combination of parameters that minimizes the predictive deviance while producing the lowest number of trees (Fig. S2.1.B). The parameters values finally selected to generate the models are: tree complexity= 4, learning rate= 0.007, and bag fraction= 0.75.

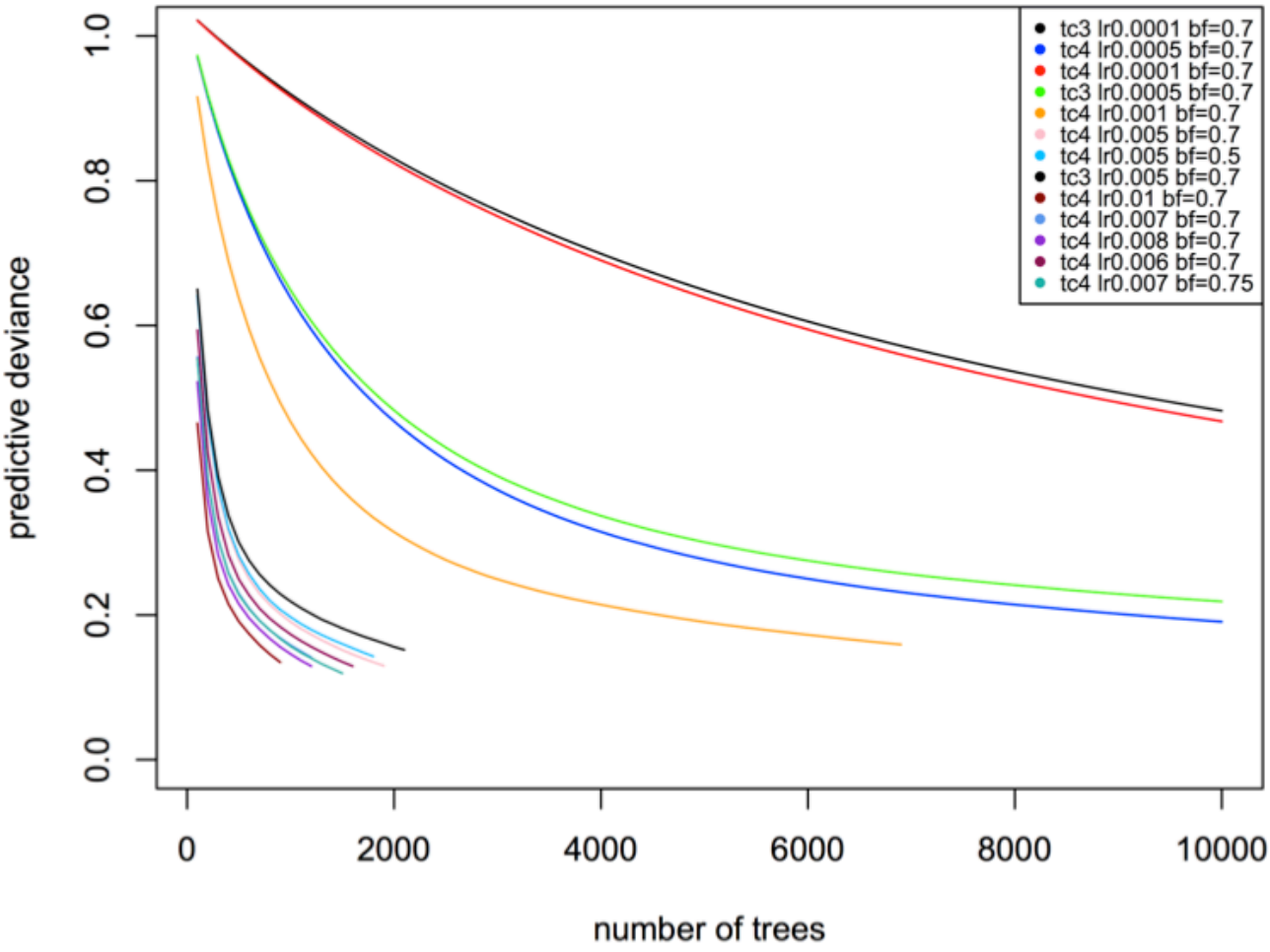


Figure S2.1.B. Comparison of the predictive deviance of models generated with different combination of parameters. Tc: tree complexity, lr: learning rate; bf: bag fraction (see Elith et al. 2008 for details).

APPENDIX 2.2. Benthic occurrence records in the Southern Ocean

Benthos occurrence records available for the Southern Ocean (Fig. S2.2) were obtained by completing the dataset of benthos sampling sites published in the Biogeographic Atlas of the Southern Ocean (Chapter 2, Griffiths et al. 2014) with recent datasets published after 2014 (Table S2.2).

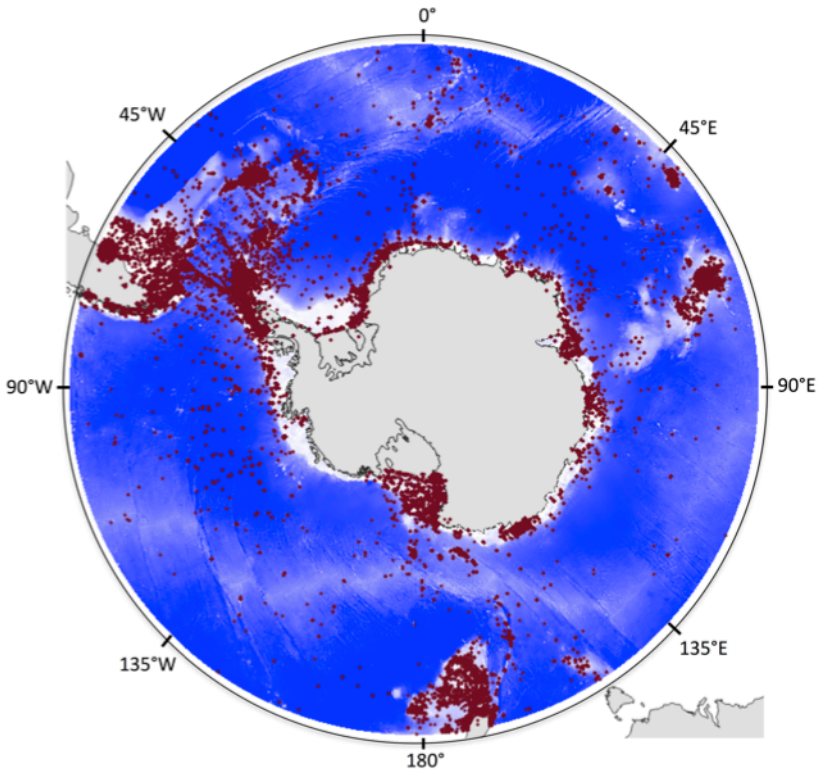


Figure S2.2. Map of the benthic Southern Ocean sampling sites updated, from the Atlas of the Southern Ocean (<45°S)(Griffiths et al. 2014).

Table S2.2. List of IPT (Integrated Publishing Toolkit) data (collected and published after 2014) added to the map of the Southern Ocean benthic sites.

Taxon	Author	Public release	URL
Tanaidacea	Italian National Antarctic Museum, Italy	2014-08-08	http://ipt.biodiversity.aq/resource?r=mna_database_tanaidacea
Intertidal taxa	Bristish Antarctic Survey, United Kingdom	2015-11-18	http://ipt.biodiversity.aq/resource?r=bas_intertidal
Macroalgae	Alfred Wegener Institute, Germany	2016-10-17	http://ipt.biodiversity.aq/resource?r=baso_macroalgae
Asteroidea	Université Libre de Bruxelles, Belgium	2017-06-30	http://ipt.biodiversity.aq/resource?r=asteroidea_southern_ocean
Ophiuridae	Italian National Antarctic Museum, Italy	2017-08-30	http://ipt.biodiversity.aq/resource?r=mna_antarctic_ophiuroidea
Echinoidea	Université de Bourgogne Franche Comté, France	2017-09-22	http://ipt.biodiversity.aq/resource?r=echinoids_occurrences_southern_ocean

APPENDIX 2.3. Datacatalog

Table S2.3. List of environmental descriptors selected for the species distribution models available for [2005-2012]. Spatial extent of the data: 78°S; 45°S/-180; 180°W. Spatial resolution: 0.1°

Environmental descriptor	Unit	Description	Source
Depth	Meters	Bathymetric grid around the Kerguelen Plateau	This study. Derived from [6]
Sea surface temperature amplitude*	°Celsius degrees	Difference between austral summer (mean January-March) and winter (mean July-September) sea surface temperature	This study. Derived from World Ocean Circulation Experiment 2013 [1] sea surface temperature layers
Seafloor mean temperature*	°Celsius degrees	Mean seafloor temperature	This study. Derived from World Ocean Circulation Experiment 2013 [1] sea surface temperature layers
Seafloor temperature amplitude*	°Celsius degrees	Difference between austral summer (mean January-March) and winter (mean July-September) seafloor temperature	This study. Derived from World Ocean Circulation Experiment 2013 [1] sea surface temperature layers
Sea surface mean salinity*	PSS	Mean sea surface salinity	This study. Derived from World Ocean Circulation Experiment 2013 [1] sea surface salinity layers
Sea surface salinity amplitude*	PSS	Difference between austral summer (mean January-March) and winter (mean July-September) sea surface salinity	This study. Derived from World Ocean Circulation Experiment 2013 [1] sea surface salinity layers
Seafloor mean salinity*	PSS	Mean seafloor salinity	This study. Derived from World Ocean Circulation Experiment 2013 [1] seafloor salinity layers
Seafloor salinity amplitude*	PSS	Difference between austral summer (mean January-March) and winter (mean July-September) seafloor salinity	This study. Derived from World Ocean Circulation Experiment 2013 [1] sea surface salinity layers
Mean surface chlorophyll a	mg/m^3	Surface chlorophyll a concentration. Summer mean over 2002-2009	MODIS AQUA (NASA) 2010 [2]
Sediments	Categorical	Sediment features	[7], updated by Griffiths 2014 (unpublished)
Geomorphology	Categorical	Geomorphologic features	ATLAS ETOPO2 2014 [8]
Slope	Unitless	Bathymetric slope	[6]
Mean seafloor oxygen concentration	mL/L	Mean seafloor oxygen concentration over 1955-2012	This study. Derived from World Ocean Circulation Experiment 2013 [1] sea surface oxygen concentration layers
Ice cover	-	Proportion of time during which ocean is covered by sea ice of concentration 85% of higher. Projection 2003-2010	This study. Derived from Australian Antarctic Data Centre [3]
POC export	$gC/m^2/day$	Particulate organic carbon 2002-2015 averages	This study. Published on Australian Antarctic Data Center [4]

References

[1] WOCE 2013, link: https://www.nodc.noaa.gov/OC5/woa13/woa13data.html accessed 2016
[2] MODIS Aqua, link: https://oceancolor.gsfc.nasa.gov/cgi accessed 2016
[3] AADC, link: http://webdav.data.aad.gov.au/data/environmental/derived/antarctic/ accessed 2017
[4] AADC POC export data, link: https://data.aad.gov.au/metadata/records/Particulate_carbon_export_flux_layers, created 2017
[5] NOAA, link: https://www.esrl.noaa.gov/psd/ipcc/ocn/ accessed 2017
[6] Smith WH and Sandwell DT (1997) Global sea floor topography from satellite altimetry and ship depth soundings. Science. 277: 1956–1962.
[7] McCoy, F.W. 1991. Southern Ocean sediments: circum-Antarctic to 30°S. Marine Geological and Geophysical Atlas of the circum-Antarctic to 30°S. (ed. by D.E. Hayes) – Antarctic Research Series.
[8] Douglass LL, Turner J, Grantham HS, Kaiser S, Constable A and others (2014) A hierarchical classification of benthic biodiversity and assessment of protected areas in the Southern Ocean. PloS one. 9: e100551.

Selecting environmental descriptors is critical for modelling the distribution of Antarctic benthic species

Guillaumot Charlène[1, 2], Danis Bruno[1], Saucède Thomas[2]

[1] Université Libre de Bruxelles, Marine Biology Lab. Avenue F.D. Roosevelt, 50. CP 160/15 1050 Bruxelles, Belgium
[2] UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, 6 bd Gabriel F-21000 Dijon, France

Polar Biology, 43 (2020). Accepted June 25th 2020

Abstract

Species Distribution Models (SDMs) are increasingly used in ecological and biogeographic studies by Antarctic biologists, including for conservation and management purposes. During the modelling process, model calibration is a critical step to ensure model reliability and robustness, especially in the case of SDMs, for which the number of selected environmental descriptors and their collinearity is a recurring issue. Boosted Regression Trees (BRT) was previously considered as one of the best modelling approach to correct for this type of bias. In the present study, we test the performance of BRT in modelling the distribution of Southern Ocean species using different numbers of environmental descriptors, either collinear or not. Models are generated for six sea star species with contrasting ecological niches and wide distribution ranges over the entire Southern Ocean. For the six studied species, overall modelling performance is not affected by the number of environmental descriptors used to generate models, BRT using the most informative descriptors and minimizing model overfitting. However, removing collinear descriptors also helps reduce model overfitting. Our results confirm that BRTs may perform well and are relevant to deal with complex and redundant environmental information for Antarctic biodiversity distribution studies. Selecting a limited number of non-collinear descriptors before modelling may generate simpler models and facilitate their interpretation. The modelled distributions do not differ noticeably between the different species despite contrasting species ecological niches. This unexpected result stresses important limitations in using SDMs for broad scale spatial studies, based on limited, spatially aggregated data, and low-resolution descriptors.

Keywords: Species Distribution Models (SDMs), Boosted regression trees (BRT), Southern Ocean, Collinearity, Asteroidea, Conservation, Environmental descriptors

ACKNOWLEDGEMENTS

This work was supported by a “Fonds pour la formation à la Recherche dans l’Industrie et l’Agriculture” (FRIA) and “Bourse Fondation de la mer” grants to C. Guillaumot. This is contribution no. 31 to the vERSO project and no. 10 to the RECTO project (www.rectoversoprojects.be), funded by the Belgian Science Policy Office (BELSPO, contracts no. BR/132/A1/vERSO and no. BR/154/A1/RECTO). This is contribution to the IPEV programs n°1124 REVOLTA and n°1044 PROTEKER. We are grateful to the crew and participants of all the cruises and research programs involved in the capture of the samples included in this dataset (see Moreau et al. 2018): POKER 2, REVOLTA 1 & 2, CEAMARC, JR144, JR179, JR230, JR262, JR275, JR287, JR15005.

1. INTRODUCTION

The Southern Ocean is one of the regions on Earth that is undergoing climate change at the fastest pace (Convey et al. 2009, Turner et al. 2014, Henley et al. 2019). Predicting the response of Antarctic species and communities to environmental changes is challenging but it has become a pressing need to address conservation issues and support guidance for the management of living resources in a dynamic context (Gutt et al. 2012, Ingels et al. 2012, Constable et al. 2014a, De Broyer et al. 2014, Convey and Peck 2019). The Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) actively works for the sustainable management of Antarctic marine ecosystems and marine life (see *https://www.ccamlr.org/en/organisation*, access August 2019). Recent proposals from CCAMLR and existing marine protected areas (MPAs), such as those newly designated around the South Orkney Islands or in the Ross Sea (CCAMLR 2009, 2016), partly rely on species distribution modelling (SDM) (Ballard et al. 2012, Anderson et al. 2016, Davis et al. 2017, Arthur et al. 2018).

SDM is a correlative approach that depicts the relationship between the distribution of species occurrence records and a set of environmental descriptors, to interpolate and predict the potential distribution of species over their entire distribution range (Elith et al. 2006, Peterson et al. 2011). Over the last decades, SDMs have been increasingly used to address conservation issues (Guisan et al. 2013, Ross and Howell 2013, Marshall et al. 2014, Reiss et al. 2014, Arthur et al. 2018), predict species suitable areas (Meier et al. 2011, Reiss et al. 2011, Nachtsheim et al. 2017, Phillips et al. 2017), including potential distribution shifts (Ficetola et al. 2007, Václavík and Meentemeyer 2009, Jiménez-Valverde et al. 2011, Tingley et al. 2014), and guide sustainable management plans for commercial purposes (Valavanis et al. 2008, Maxwell et al. 2009). They have particularly proved useful to improve our understanding of species distribution in poorly sampled and seldom accessed areas (Elith et al. 2006, Peterson et al. 2011) and for the conservation of Southern Ocean marine life (De Broyer et al. 2014, Basher and Costello 2016, Hogg et al. 2018, Jansen et al. 2018, Jerosh et al. 2019).

Calibration is a critical step in SDM procedures, influencing their relevance, robustness and accuracy (Barbet-Massin et al. 2012, Guisan et al. 2013, Anderson et al. 2016). The selection of environmental descriptors is also important, as it shapes model accuracy and performance (Elith and Leathwick 2009, Austin and van Niel 2011, Dormann et al. 2012b, Braunisch et al. 2013, Bucklin et al. 2015, Bradie and Leung 2017, Petitpierre et al. 2017). The inappropriate selection of descriptors has been shown to cause overfitting in SDMs, especially when the number of descriptors is high compared to the number of occurrences available (Anderson and Gonzalez 2011, Braunisch et al. 2013, Kramer-Schadt et al. 2013, Synes and Osborne 2011, Petitpierre et al. 2017), leading to over-complex models, reduced transferability performances and underestimation of predicted suitable areas (Beaumont et al. 2005).

Collinearity between descriptors is another major concern when addressing the quality of SDMs (Dormann et al. 2012b). Collinearity occurs when at least two descriptors are linearly related in a statistical model (Dormann et al. 2012b). In regression models, multicollinearity increases variance values between independent descriptors. It can cause incorrect estimations of beta regression coefficients and bias interpretation, making it difficult to disentangle the respective contributions of independent variables to explaining the dependent variable (Hair et al. 2014). Collinear descriptors are traditionally removed from datasets to calibrate SDMs (Dormann et al. 2012b, Pierrat et al. 2012, Merow et al. 2013, Fabri-Ruiz 2018, Guillaumot et al. 2018b), while a recent study showed that collinear descriptors could also improve the model's fit (Freer et al. 2019).

Machine-learning algorithms can effectively model complex relationships between environmental conditions and occurrence records (Olden et al. 2008, Elith and Leathwick 2009). They can harness incomplete datasets and missing data, as well as contrasting and extreme values, and generate predictive models with high transferability performances and low sensitivity to species niche width (Elith et al. 2006, Elith et al. 2008, Elith and Graham 2009, Reiss et al. 2011, Barbet-Massin et al. 2012, Heikkinen et al. 2012, Qiao et al. 2015). In machine-learning algorithms, the

Boosted Regression Trees approach (BRT) has been shown to be particularly efficient when dealing with non-informative environmental descriptors or conversely, with redundant information provided by correlated factors (Breiman 1984, De'ath and Fabricius 2000, Elith et al. 2008).

In the present work, we test the robustness of SDMs generated with BRT for various numbers of environmental descriptors and different collinearity values. Models are generated for six common and abundant asteroid (sea star) species that have been extensively sampled and studied; here used as representative case studies for the Antarctic benthos: *Acodontaster hodgsoni* (Bell, 1908), *Bathybiaster loripes* (Sladen, 1889), *Glabraster antarctica* (Smith, 1876), *Labidiaster annulatus* Sladen, 1889, *Odontaster validus* Koehler, 1906 and *Psilaster charcoti* (Koehler, 1906) (McClintock et al. 2008a, Mah and Blake 2012, Lawrence 2013, Brandt et al. 2014, Danis et al. 2014, Moles et al. 2015, Moreau et al. 2018).
Because the Southern Ocean is scarcely accessed and sampled, spatial analyses of species distribution are usually based on aggregated and relatively small presence-only datasets, often compiled from historical records (De Broyer et al. 2014, Guillaumot et al. 2016, Fabri-Ruiz et al. 2017a, Guillaumot et al. 2018a - Appendix, Moreau et al. 2018), which strongly hampers SDM performances (Hortal et al. 2008, Loiselle et al. 2008, Phillips et al. 2009, Costa et al. 2010, Newbold 2010, Guillera-Arroita et al. 2015, Guillaumot et al. 2018b). The objectives of this study are to assess the limits and potential of BRT to generate robust models for Southern Ocean benthic species and to provide some recommendations on the selection of environmental descriptors.

2. MATERIAL AND METHODS

2.1. Selection of environmental descriptors

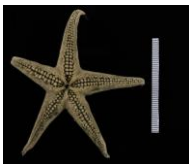
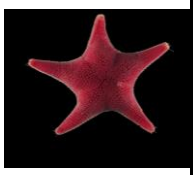
A set of 58 environmental descriptors was compiled from different sources (Appendix 2.4). This set can be downloaded from the *blueant* R package (https://github.com/AustralianAntarcticDivision/blueant), following the procedure given in the "data_for_SDM_vignette" at https://australianantarcticdivision.github.io/blueant/articles/SO_SDM_data.html.
Most descriptors are average abiotic conditions taken from the WOCE database (Appendix 2.4) and describe the average abiotic conditions for the [2005-2012] time period (*i.e*. temperature, salinity, chlorophyll-a, particulate organic carbon flux). Some descriptors are available for longer time periods only ([1957-2017] and [1955-2012] for sea ice cover and seafloor oxygen concentration respectively). More recent or precise datasets are not available at the scale of the Southern Ocean. Raster layers were compiled with a 0.1 x 0.1° pixel resolution (11km approximately), each 0.1 x 0.1° pixel being used as a single grid-cell pixel, and cropped to the extent of the Southern Ocean (herein defined as waters south of 45°S latitude) for a total of 1.26 million pixels. Missing values are not interpolated to avoid potential biases. Available descriptors are selected according to their ecological relevance to benthic studies and following previous recommendations provided for species distribution modelling (Franklin 2010b, Anderson 2013) and Antarctic studies (Saucède et al. 2014). The selected descriptors best document the main characteristics of the species physical habitat (depth, sea water temperature, geomorphology, sediment nature, slope, roughness), geography (distance to the Antarctic continent, to canyons, to continental shelves, to the maximal sea ice extent in winter), seasonality (sea ice concentration and thickness), food resources (chlorophyll-a concentration and Particulate Organic Carbon [POC] exported on the sea bottom) and chemical environment (oxygen concentration and seafloor salinity). Minimal, maximal, and range values (min-max difference) of some descriptors are computed to complement the dataset (Franklin 2010a, Bradie and Leung 2017, Guillaumot et al. 2018a - Appendix, 2018b). Extreme weather conditions and climate events were shown to strongly impact natural environments, notably species survival and distribution (Easterling et al. 2000, Wernberg et al. 2013). Here, supplementary descriptors are specially developed for the intensity and frequency of monthly changes in seafloor temperature, salinity, oxygen and chlorophyll-a concentrations. For each pixel and one year, these layers document how many times monthly average values are respectively higher ('maximal extreme event') or lower ('minimal extreme

event') than the yearly median value (Appendix 2.5, codes available at https://github.com/charleneguillaumot/THESIS).

2.2. Biological records

Antarctic sea stars play an important role in the structuring of benthic communities (McClintock et al. 2008a, Mah and Blake 2012, Lawrence 2013), they have contrasting ecological niches and life history traits (e.g. feeding diets, reproduction and development modes) that condition habitat preferences and dispersal abilities (Moreau et al. 2017, Table 2.4). Here, SDMs are generated for six sea star species using presence-only records obtained from the "Antarctic and sub-Antarctic asteroid database" published by Moreau et al. (2018): *Acodontaster hodgsoni* (Bell, 1908), *Bathybiaster loripes* (Sladen, 1889), *Glabraster antarctica* (Smith, 1876), *Labidiaster annulatus* Sladen, 1889, *Odontaster validus* Koehler, 1906 and *Psilaster charcoti* (Koehler, 1906). The studied species are abundant and have been regularly sampled during benthic expeditions to the Southern Ocean, making them some of the best-documented occurrence records on database available for Southern Ocean benthic species (Moreau et al. 2018). The working database (Moreau et al. 2018) includes presence-only records obtained by trawling and scuba diving during numerous expeditions to the Southern Ocean ranging from 1872 to 2016 (Appendix 2.6). Occurrence data collected during the last 50 years are the most abundant with an intense sampling effort carried out in the framework of the International Polar Year (IPY: 2007-2009) and the Census of Antarctic Marine Life (CAML: 2005-2010). All occurrence data are selected to ensure that a sufficient number of records are available to run the models (Stockwell and Peterson 2002, van Proosdij et al. 2016) and exhaustively cover the geographical space occupied by the considered species. Presence-only records are spatially aggregated near coastal areas and scientific stations (Appendix 2.7 and see De Broyer et al. 2014, Guillaumot et al. 2019 - Chapter 2). Presence record duplicates found in the same grid-cell pixel are removed to reduce spatial replication as described by Segurado et al. (2006) and Boria et al. (2014). Because the considered species have different depth ranges (Moreau et al. 2018), model projection is performed for each species independently and bounded by maximal depth value defined by the species deepest record (see Table 2.4 for details).

Table 2.4. The six studied species and their respective ecological traits. Presence-only records duplicates present on a same grid-cell pixel are removed from the count of occurrences. The model maximum depth is defined for each species independently according to the density distribution of recorded depth values. Images sources: Brueggeman 1998, BIOMAR ULB database (P. Pernet), proteker.net, B121 expedition (Q. Jossart).

	Acodontaster hodgsoni (Bell 1908)	*Bathybiaster loripes* (Sladen 1889)	*Glabraster antarctica* (Smith 1876)	*Labidiaster annulatus* Sladen 1889	*Odontaster validus* Koehler, 1906	*Psilaster charcoti* (Koehler, 1906)
						
Feeding diet	Predator (mainly sponges) (Brueggeman 1998)	Detritivorous (Dearborn 1977)	Deposit feeder, predator, or scavenger (Brueggeman 1998)	Predator (Dearborn et al. 1991)	Opportunistic feeder (suspensivorous, deposit feeder,predator, scavenger) (Brueggeman 1998)	Deposit feeder, predator (Brueggeman 1998)
Reproduction and development modes	Broadcaster with non-feeding planktonic larvae (Bosch and Pearse 1990)	Broadcaster with non-feeding planktonic larvae (Bosch and Pearse 1990)	Broadcaster with feeding planktonic larvae (Bosch 1989)	Broadcaster and probably feeding planktonic larvae (Janosik et al. 2008)	Broadcaster with feeding planktonic larvae (Bosch and Pearse 1990)	Broadcaster with non-feeding planktonic larvae (Bosch and Pearse 1990)
Occurrence number	297	585	844	373	309	350
Model maximum depth	1,500 m	4,000 m	4,000 m	1,500 m	1,500 m	4,000 m

2.3. Model calibration

Boosted regression tree (BRT) is chosen as a robust method to test the influence of descriptor selection on model performance. This machine-learning algorithm has been shown to be well suited to accommodate presence-only data and incomplete datasets, to fit complex relationships between species records and environmental descriptors, to limit model overfitting and to have high transferability performances (Elith et al. 2006, Elith et al. 2008, Hastie et al. 2009, Ward et al. 2009, Reiss et al. 2011, Heikkinen et al. 2012, Mainali et al. 2015, Guillaumot et al. 2019 - Chapter 2), transferability being defined as the ability of models to predict in new environmental conditions (Friedman et al. 2001).

BRT models are calibrated following the procedure detailed in Guillaumot et al. (2019 - Chapter 2) and using the *gbm* R package (Elith et al. 2008, Ridgeway 2015). BRT parameters are set to minimize both the optimal number of trees used to build the model and the minimal predictive deviance (learning rate, bag fraction and tree complexity are provided for each species in Appendix 2.8). A set of 1,000 background records are randomly sampled in the environmental space (maximal depth limit depending on the studied species, Table 2.4). This number is tested sufficient enough to represent the whole spectrum of environmental conditions existing in the geographic area of interest (Guillaumot et al. 2019 - Chapter 2: supplementary material) while being as close as possible to the number of records used to generate the model (Barbet-Massin et al. 2012). One hundred background data samples are generated as model replicates. Spatial aggregation of occurrence records is a recurrent bias in Antarctic benthic species databases (Fabri-Ruiz 2018, Guillaumot et al. 2018a - Appendix, Guillaumot et al. 2019 - Chapter 2). To reduce the effect of spatial aggregation on model outputs, background records are sampled

following a target-group approach (Philipps et al. 2009). In this approach, background data are randomly sampled in the area of interest, following a weighting scheme defined by a Kernel Density Estimation (KDE) of sampling effort in the Southern Ocean (Guillaumot et al. 2018a - Appendix, supplementary material in Guillaumot et al. 2019 - Chapter 2).
When using spatially aggregated records, standard cross-validation procedures used to evaluate modelling performances can be strongly biased (Hijmans 2012, Roberts et al. 2017, Guillaumot et al. 2019 - Chapter 2). The random selection of training and test data leads to the violation of independence between training and test subsets, which can induce an over-estimation of correctly predicted test data by the model (Hijmans 2012). Using cross-validation procedures that spatially segregate training and test data (defined based on presence and background subsets) is a good alternative to accurately evaluate the performance of SDMs based on aggregated datasets. In the present study, a "6-fold CLOCK" cross-validation approach adapted from Guillaumot et al. (2019 - Chapter 2) was applied. This procedure randomly defines six sectors around Antarctica according to longitude, three for training data and three for test data.

2.4. Collinearity and the selected number of environmental descriptors

Collinearity between the 58 selected descriptors is analysed following a stepwise approach that eliminates layers with a Variance Inflation Factor (VIF) > 10, using the 'vif.step' function of the *usdm* R package (Naimi et al. 2014). VIF > 10 is defined as the threshold above which the effect of multicollinearity on model predictions is considered significant (Hair et al. 2014) and too strong to be automatically corrected by machine-learning algorithms (Dormann et al. 2012b). Multicollinearity is measured on projection areas, that is the portion of the environment for which SDMs do not extrapolate. Extrapolation areas are defined for each species independently using the Multivariate Environmental Similarity Surface index (MESS, Elith et al. 2010). They correspond to all grid-cell pixels where descriptor values are not contained within the range of environmental conditions on which presence-only data are recorded. Models generated with the 58 environmental descriptors are compared to models for which collinear descriptors are removed.
A stepwise procedure is used to test the effect of the selected number of environmental descriptors on model performance. SDMs are first generated for the six species using the total set of 58 environmental descriptors. Then, the six descriptors that contribute the least to the average model are iteratively pruned at each step of a series of SDMs successively generated with 58, 52, 46, 40, 34, 28, 22, 16, 10, and four environmental descriptors.

2.5. Model evaluation and comparisons

The percentage of presence data correctly predicted (i.e. correctly classified test data) is computed to assess the performance of SDMs in terms of transferability. Model performances are also assessed using the Area Under the Receiver Operating Curve (AUC, Fielding and Bell 1997), the Point Biserial Correlation between predicted and observed values (COR, Elith et al. 2006) and the True Skill Statistic (TSS, Allouche et al. 2006). Suitable areas are classified using the Maximum Sensitivity plus Specificity threshold (MaxSSS), which is the most adapted index for SDMs using presence-only data (Liu et al. 2013). MaxSSS enables to split model projections into suitable (>MaxSSS value) and unsuitable areas (<MaxSSS value). The average number of regression trees produced by BRT to generate models (*gbm* R package, Elith et al. 2008) is calculated to evaluate model complexity. Scores of SDM series generated with a decreasing number of environmental descriptors are compared between each other using the Mann-Kendall non-parametric trend test to assess the presence of a monotonic trend (Hipel and McLeod 1994). Differences between model performances (AUC, TSS, COR, percentage of correctly classified test data), model properties (number of trees) and outputs (percentage of predicted suitable area) are tested using a Wilcoxon-Mann-Whitney pairwise test.

2.6. Final SDM outputs

Six final SDMs are proposed for the six considered species for the [2005-2012] time period after selection of the optimal number of descriptors and after removing collinear descriptors. The

contribution of descriptors and their marginal effects (partial dependence plots) are provided and compared between each other. Environmental conditions predicted as suitable for species distribution are plotted through a principal component analysis (PCA) to display the predicted species occupied environmental space. PCA is compared between species having the same projection depth threshold, either 1,500 m or 4,000 m depth (Table 2.4).

3. RESULTS

3.1. Contribution of environmental descriptors

All models generated for the six species and with the total set of 58 descriptors perform well with an average AUC score value of 0.853 (min. 0.827; max. 0.883) and an average of 67.2% of correctly predicted test data (59.5-75.1%). 'Extreme events' descriptors specifically computed for this study (Appendix 2.5) never contribute more than 1% to SDMs, some extreme chlorophyll-a layers excepted (Table 2.5). Overall, parameters that contribute the most to all SDMs are depth, currents, ice thickness and seafloor properties (Table 2.5, Fig. 2.12). Few contrasts are obtained in contributions between species models except for the contribution of seafloor current speed and POC concentrations that respectively vary from 1.95 to 10.84% and 0.49 to 7.05% between SDMs (Fig. 2.12).

Table 2.5. Average contribution of each environmental descriptor (based on 100 model replicates) generated for the six studied species using the total set of 58 descriptors. In dark blue, descriptors always contribute more than 1% to all models. In light blue, descriptors contributing more than 1% to some species models only (A: *Acodontaster hodgsoni*, B: *Bathybiaster loripes*, G: *Glabraster antarctica*, L: *Labidiaster annulatus*, O: *Odontaster validus*, P: *Psilaster charcoti*). In red, descriptors never contributing more than 1% to all species models. The description of the different environmental descriptors is provided in Appendix 2.4.

Descriptor	Contribution	Descriptor	Contribution	Descriptor	Contribution
depth		ice_thickness_range		seafloor_sali_2005_2012_min	
geomorphology	A,B,G,L	chla_ampli_alltime_2005_2012		seafloor_sali_2005_2012_sd	
sediments	A,B,G,O,P	chla_max_alltime_2005_2012	A	seafloor_temp_2005_2012_ampli	
slope		chla_mean_alltime_2005_2012	A,B,L,P	seafloor_temp_2005_2012_max	
roughness		chla_min_alltime_2005_2012	A,B,G,L,P	seafloor_temp_2005_2012_mean	
mixed_layer_depth		chla_sd_alltime_2005_2012	A,B,L,P	seafloor_temp_2005_2012_min	B,G,L,P
seasurface_current_speed		POC_2005_2012_ampli	A,B,G,O,P	seafloor_temp_2005_2012_sd	
seafloor_current_speed		POC_2005_2012_max	A,B,G,O,P	extreme_event_max_chl_2005_2012_ampli	
distance_antarctica		POC_2005_2012_mean	A,B,G,O,P	extreme_event_max_chl_2005_2012_max	
distance_canyon		POC_2005_2012_min		extreme_event_max_chl_2005_2012_mean	
distance_max_ice_edge		POC_2005_2012_sd	A,B,G,O,P	extreme_event_max_chl_2005_2012_min	
distance_shelf	A,B,G,O,P	seafloor_oxy_19552012_ampli		extreme_event_min_chl_2005_2012_ampli	
ice_cover_max		seafloor_oxy_19552012_max		extreme_event_min_chl_2005_2012_max	
ice_cover_mean		seafloor_oxy_19552012_mean		extreme_event_min_chl_2005_2012_mean	B,G,O,P
ice_cover_min		seafloor_oxy_19552012_min		extreme_event_min_chl_2005_2012_min	P
ice_cover_range		seafloor_oxy_19552012_sd		extreme_event_min_oxy_1955_2012_nb	
ice_thickness_max		seafloor_sali_2005_2012_ampli		extreme_event_max_sali_2005_2012_nb	
ice_thickness_mean	B,G,L,O,P	seafloor_sali_2005_2012_max		extreme_event_min_sali_2005_2012_nb	
ice_thickness_min		seafloor_sali_2005_2012_mean		extreme_event_max_temp_2005_2012_nb	
				extreme_event_min_temp_2005_2012_nb	

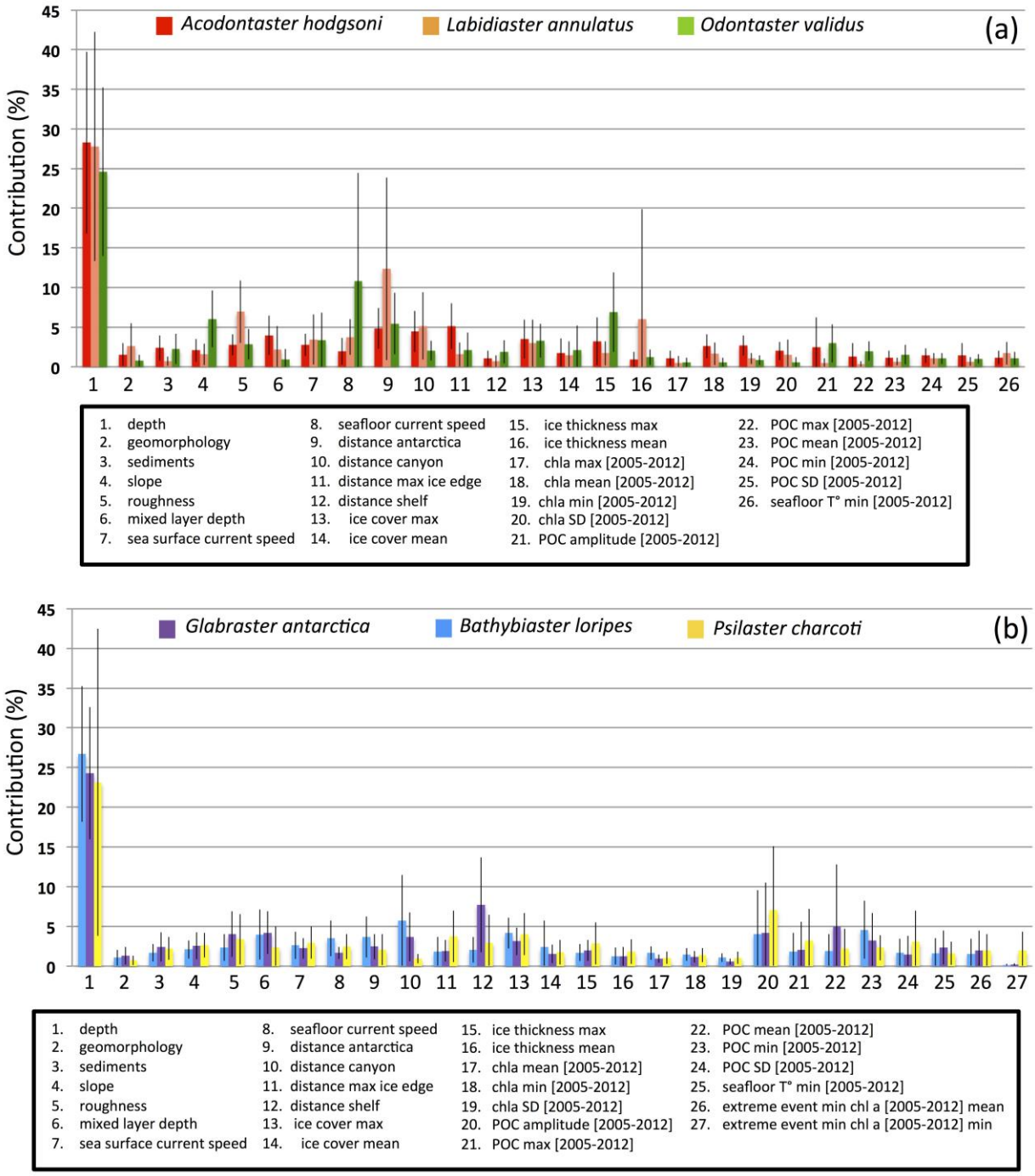


Figure 2.12. Contribution of environmental descriptors to SDMs projected until (a) 1,500 m and (b) 4,000 m depth for the six species. Environmental descriptors contributing less than 1% to all models are not shown. Error bars correspond to standard deviation values of the contribution percentages (100 replicates of background sampling and spatial cross-validation splitting).

3.2. Number of environmental descriptors

Overall, models generated with different numbers of environmental descriptors do not show significant changes in model performance (Mann-Kendall trend tests, Table 2.6). Models generated with four environmental descriptors only show a significant decrease in AUC, COR, and TSS values, and in the percentage of correctly classified test data for all species but *G. antarctica* (Fig. 2.13, Appendix 2.9). Significant differences in model performance are model-specific, whatever the number of descriptors used (Fig. 2.13, Appendix 2.9). Differences in the number of trees used to generate models and in the size of suitable areas are never tested significant (Table 2.6).

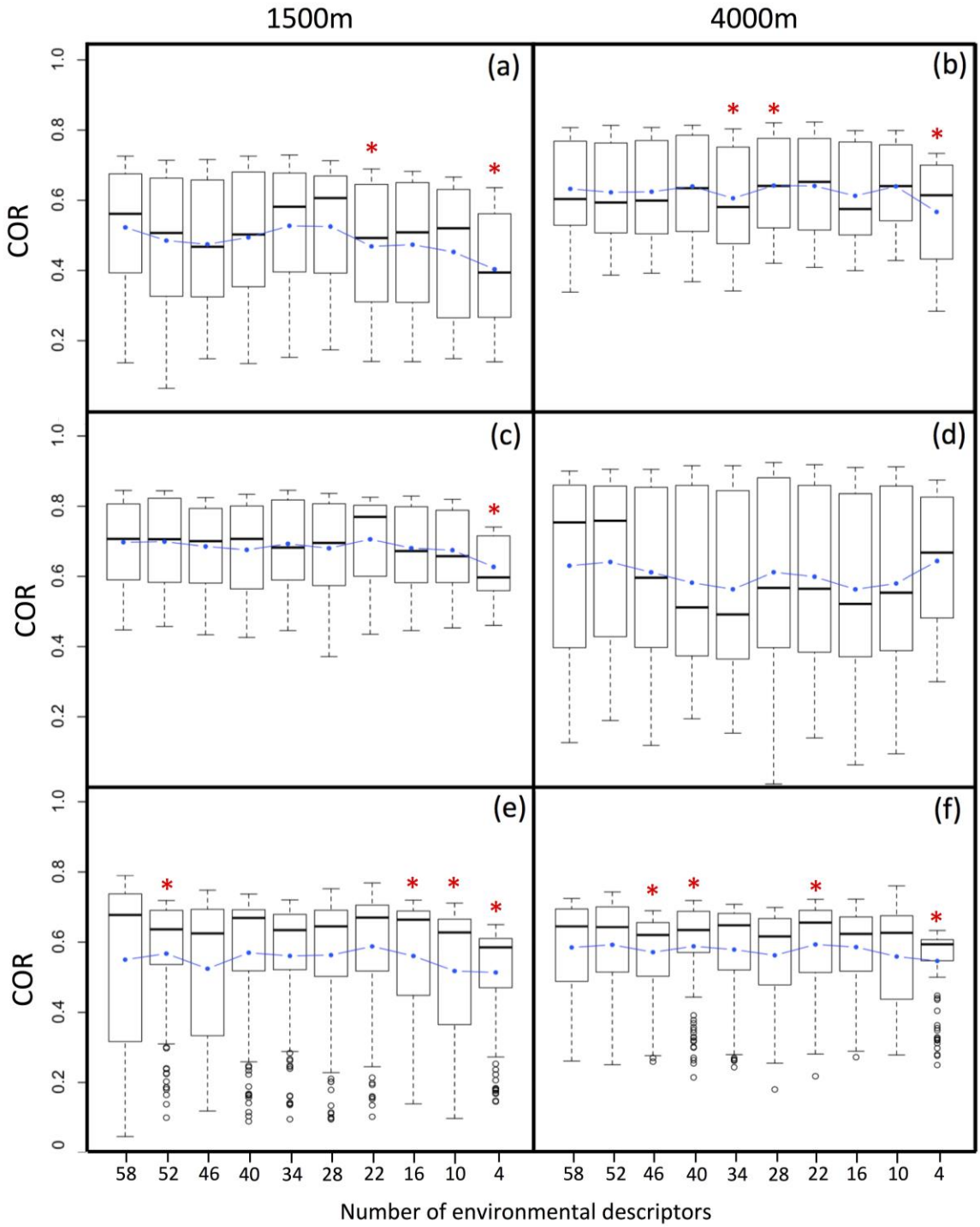


Figure 2.13 Influence of the number of environmental descriptors on SDM performance. Boxplot of 100 model replicate scores. Changes in biserial correlation (COR) values for (a) *Acodontaster hodgsoni*, (b) *Bathybiaster loripes*, (c) *Labidiaster annulatus*, (d) *Glabraster antarctica*, (e) *Odontaster validus* and (f) *Psilaster charcoti*. Average values are indicated in blue. Red asterisks indicate significant changes in median values between the series and preceeding value (Wilcoxon rank paired test, p-value < 0.05). The left-side and right-side columns correspond to species for which models are respectively projected until 1,500 m and 4,000 m depth.

Table 2.6. Mann-Kendall statistic scores (τ). Models are built with 58, 52, 46, 40, 34, 28, 22, 16, 10 and 4 environmental descriptors respectively. 100 replicates are generated in each case. The Mann-Kendall trend test is realised on the median value of the 100 replicates. All tests are not significant. The direction of the monotonic trend is given by the sign of the τ values. AUC: Area Under the Curve, COR: biserial Pearson correlation, TSS: True Skill Statistic. The percentage of correctly classified test data is defined by the proportion of presence test data correctly predicted by the model.

	Acodontaster hodgsoni	*Bathybiaster loripes*	*Glabraster antarctica*	*Labidiaster annulatus*	*Odontaster validus*	*Psilaster charcoti*
AUC	-0.111	0.022	-0.644	-0.067	-0.378	-0.289
COR	-0.111	0.156	-0.556	-0.244	-0.289	-0.289
TSS	-0.244	-0.067	-0.600	-0.067	-0.289	-0.422
Number of trees	0.205	0.675	0.303	-0.322	0.023	-0.210
% correctly classified test data	-0.067	0.511	0.156	0.511	-0.156	0.511
Average number of suitable pixels	0.167	0.111	0.200	0.333	0.333	0.289

3.3. Collinearity

Most SDMs generated with and without collinear descriptors show similar performance statistics (AUC, TSS, COR, and percentage of correctly classified test data) and a comparable number of trees is used to build models (Table 2.7). However, for *A. hodgsoni* and *G. antarctica*, lower AUC, TSS and COR values are obtained for models generated without collinear descriptors. The percentage of correctly classified test data remains unchanged except in models generated without collinear descriptors for *A. hodgsoni* (-9.9%) and *O. validus* (-19.5%) in which it significantly decreases. For all species but *G. antarctica*, the proportion of predicted suitable area increases in models generated without collinear descriptors (Table 2.7).

Table 2.7. Mann-Whitney Wilcoxon pairwise test (*W*) comparing statistics of models generated without collinear descriptors and models run with the total set of 58 environmental descriptors. Associated p-values are summarized by asterisks (no star p >0.05, * p < 0.05, ** p < 0.01 and *** p < 0.001). AUC: Area Under the Curve, COR: biserial Pearson correlation, TSS: True Skill Statistic. The percentage of correctly classified test data is defined by the proportion of presence test data correctly predicted by the model.

	1,500m			4,000m		
	Acodontaster hodgsoni	*Labidiaster annulatus*	*Odontaster validus*	*Bathybiaster loripes*	*Glabraster antarctica*	*Psilaster charcoti*
AUC	6041*	4754	5738	5578	5931*	5280
COR	5842*	4783	5867*	5596	5964*	5247
TSS	6138**	4792	5748	5596	5840*	5425
% correctly classified test data	6234**	5546	6247.5**	5590	5145	4512.5
Number of trees	5359	4352	4811	5031.5	4641.5	4312
% suitable area	3526***	6272**	8695***	9759***	4796	8571***

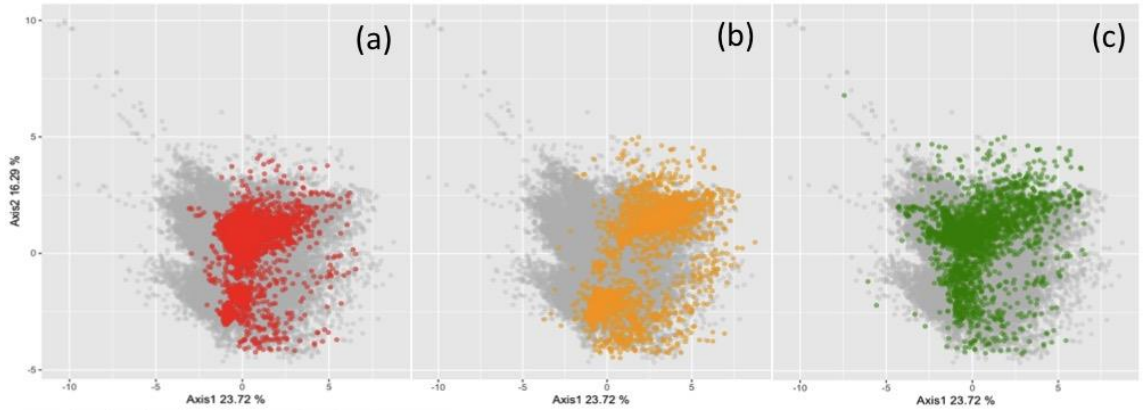
3.4. Comparison between final SDMs

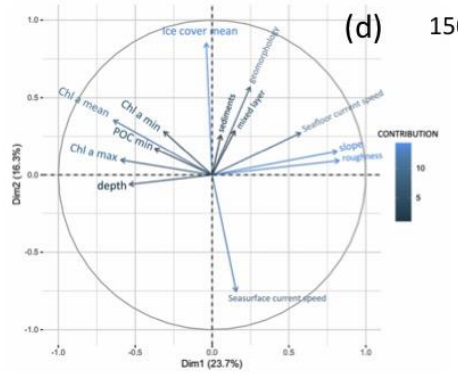
Distance layers (i.e. distance to Antarctic coasts, to shelves, to the nearest canyons, to the maximum ice edge in winter, see Appendix 2.4) are used as descriptors in a first phase of the analysis to test for the effect of collinearity and the number of descriptors on model performance because they are commonly used in SDMs performed for Southern Ocean species (Mormède et al. 2014c). However, although relevant when interpolating species distribution patterns (Table 2.5), interpreting the contribution of such descriptors is not straightforward when it comes to describe species ecological niche. Therefore, these descriptors are excluded from analyses in the final set of SDMs. In addition, descriptors that never contribute more than 1% to SDMs (Table 2.5) as well as collinear descriptors (depending on species) are removed from the initial set of descriptors. Depending on the species under study, a set of 14 to 16 descriptors is used to calibrate final

models: 13 of these descriptors are common to the six studied species and for three species, additional descriptors on extreme events on chlorophyll-a concentration are used (Appendix 2.10). The performance of final models is good for all species, with AUC values ranging from 0.810 ± 0.09 to 0.872 ± 0.07 (mean and standard deviation values), TSS values from 0.461 ± 0.121 to 0.546 ± 0.08, COR values from 0.503 ± 0.136 to 0.656 ± 0.121 and correctly classified test data from 51.6 ± 23.7% to 80.7 ± 10.1% (Appendix 2.11).

The PCA (Fig. 2.14) shows an important contribution of both the physical environment (slope, roughness) and food resources (chlorophyll-a concentrations) to SDMs projected down to 1,500 m depth (strong correlation with PC1) and a weaker and independent contribution of mean sea-ice cover and seafloor current speed (strong correlation with PC2, Fig. 2.14d). In contrast, food resources (chlorophyll-a and POC concentrations), sea ice cover and depth are the main contributors to SDMs projected down to 4,000 m depth (high correlation with PC1) with weaker contributions of the physical environment (slope and roughness) (correlation with PC2, Fig. 2.14h). Major differences are obtained between "shallow" and "deep" models (Fig. 2.12, Fig. 2.14) whatever the other species ecological traits (Table 2.4).

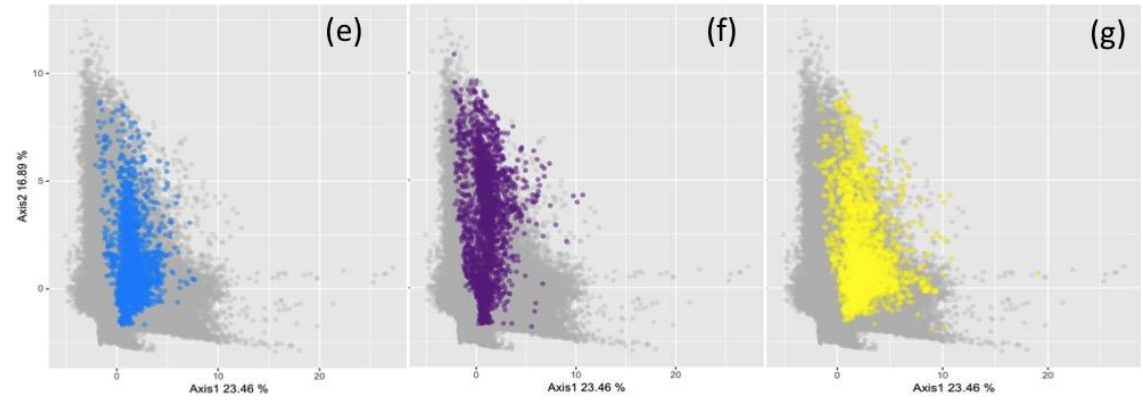
Spatial projections of SDMs also show important contrasts in distribution patterns between “shallow” (1,500 m) and “deep” (4,000 m) models (Fig. 2.15). Shallow models present low probability values along the Antarctic coasts and higher probabilities in the sub-Antarctic Islands, in the Kerguelen or South Georgia archipelagos, except for *O.validus*. The three SDMs projected down to 4,000 m depth show common patterns, with high probabilities predicted close to the Antarctic coasts where most occurrences are recorded (Appendix 2.7). High probabilities are also predicted on the Kerguelen Plateau for *B. loripes* and *G. antarctica*, while low probabilities are predicted for *P. charcoti* in the sub-Antarctic Islands (Fig. 2.15).

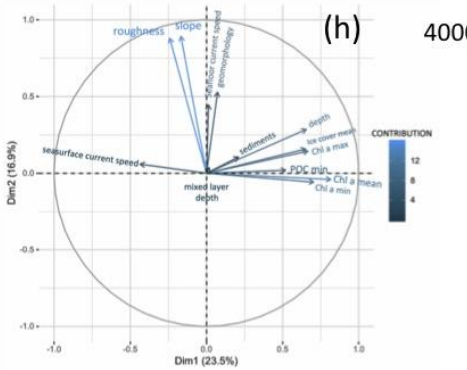




1500m models

Axis 1	Correlation	Axis 2	Correlation
roughness	0.83	Ice cover mean	0.85
slope	0.81	geomorphology	0.56
seafloor current speed	0.58	sediments	0.29
Chl a min	-0.31	mixed layer	0.28
POC min	-0.37	seasurface current speed	-0.75
depth	-0.54		
Chl a max	-0.59		
Chl a mean	-0.64		





4000m models

Axis 1	Correlation	Axis 2	Correlation
Chla mean	0.82	slope	0.89
Chla max	0.70	roughness	0.88
Ice cover mean	0.67	seafloor current speed	0.52
Depth	0.65	geomorphology	0.45
POC min	0.65		
ChL a min	0.52		
sediments	0.21		
Seasurface current speed	-0.44		

Figure 2.14. PCA of environmental values (grey dots) from descriptors used in final species distribution models, and that are common between the six species (Appendix 2.10: depth, geomorphology, sediments, slope, roughness, mixed layer depth, seasurface and seafloor current speed, ice cover mean, chlorophyll-a min, max and mean concentrations for 2005-2012, POC minimum concentrations for 2005-2012), limited to 1,500 m (a-c) and 4,000 m depth (e-g) respectively. Colour dots: species suitable area (probabilities > average maxSSS scores) for (a) *Acodontaster hodgsoni*, (b) *Bathybiaster loripes*, (c) *Labidiaster annulatus*, (e) *Glabraster antarctica*, (f) *Odontaster validus*, (g) *Psilaster charcoti.* PCA plot of environmental descriptors (d,h) and appended tables with the associated correlations to PC1 and PC2. All correlation values are significant ($p < 0.05$).

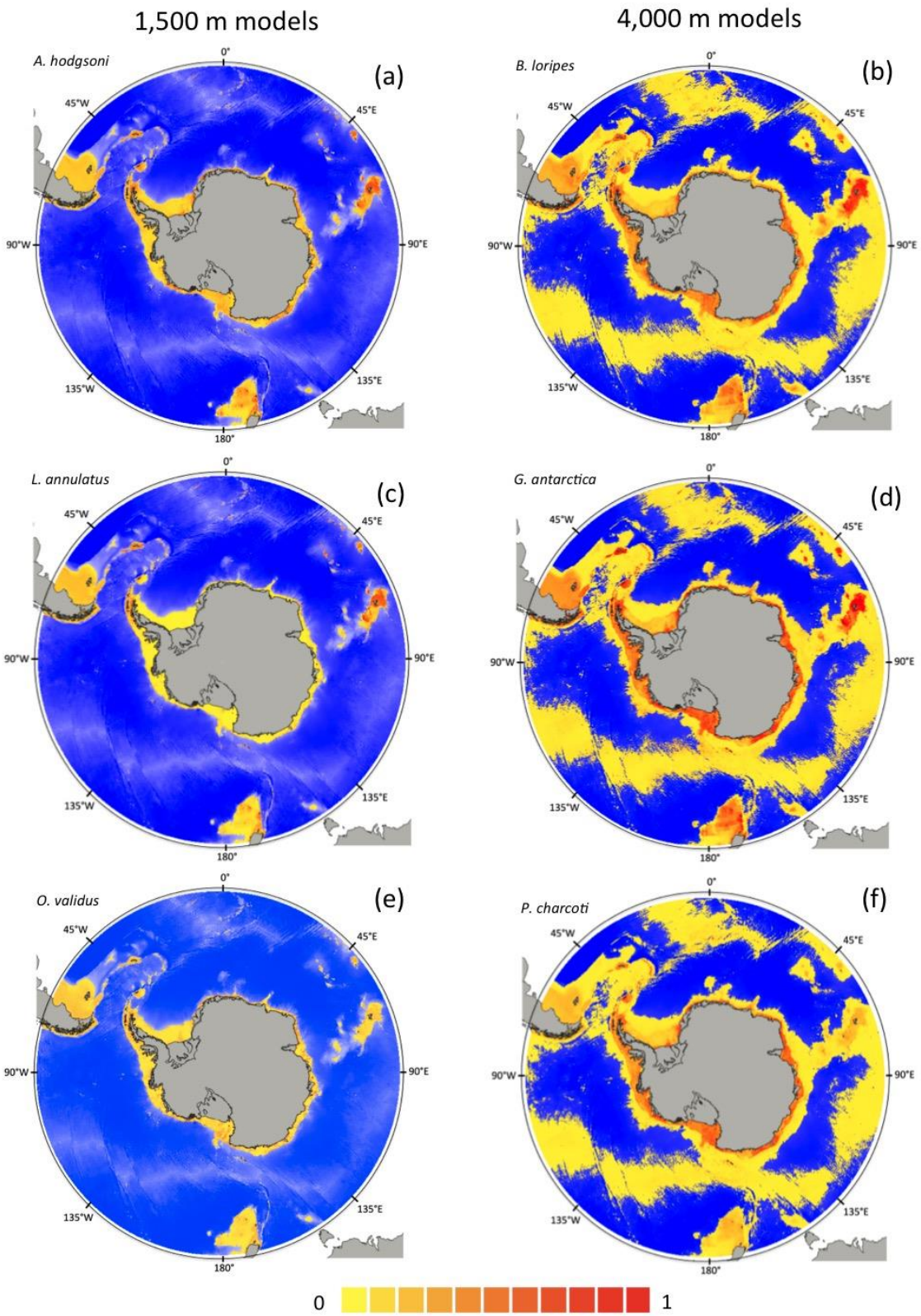


Figure 2.15. SDMs generated based on the final selection of environmental descriptors for the six studied species (Appendix 2.10). Projection areas are limited to 1,500 m depth (left-hand column) or 4,000 m depth (right-hand column) for (a) *Acodontaster hodgsoni*, (b) *Bathybiaster loripes*, (c) *Labidiaster annulatus*, (d) *Glabraster antarctica*, (e) *Odontaster validus*, (f) *Psilaster charcoti.* Blue colours correspond to depth gradient. The colour chart indicates species presence probability comprised between 0 and 1. Polar stereographic projection.

4. DISCUSSION

4.1. Influence of the number of descriptors on modelling performance

SDMs performed at the scale of the Southern Ocean are usually based on a limited mass of occurrence data, patchy datasets and using low-resolution environmental descriptors. Recent studies have questioned the relevance of using such SDMs considering the spatial and temporal heterogeneities of datasets and the importance of sampling biases (Fabri-Ruiz 2018, Guillaumot et al. 2018a - Appendix). In the present work, we focus on the selection of environmental descriptors as a critical step for model calibration (Bucklin et al. 2015, Petitpierre et al. 2017). Machine-learning algorithms such as BRT were proved efficient to deal with non-informative descriptors (De'ath and Fabricius 2000, Elith et al. 2008) and to correct for the influence of collinearity between descriptors (Dormann et al. 2012b). The performance of BRT to model the distribution of Antarctic benthic species at large spatial scale is herein evaluated.
Successive models were generated from four to 58 environmental descriptors. All models have similar accuracy (AUC, TSS, COR) and transferability (percentage of correctly classified test data) performances. Models generated with four environmental descriptors only (depending on each species) show significant differences in performance values and low capacities to describe and predict species distribution. SDMs generated for the species *G. antarctica* depart from this general result with no significant differences in modelling performances between models generated with four to 58 descriptors. This may be due to the large number of occurrence data available to describe the species distribution and conversely, the limited number of environmental descriptors contributing to the models (Appendix 2.7, Table 2.4).

Many studies have stressed the risk of model overfitting when using too many descriptors (Anderson and Gonzalez 2011, Synes and Osborne 2011, Braunisch et al. 2013, Kramer-Schadt et al. 2013, Petitpierre et al. 2017) or the risk of underestimating the extent of suitable areas due to reduced transferability performances (Beaumont et al. 2005). In contrast, our results show that models generated with a different number of predictors are characterised by similar performance levels. This is congruent with results obtained by Bucklin et al. (2015) who highlighted that the random addition of descriptors has a minor influence on modelling performances when using machine-learning algorithms. The absence of significant changes in the number of trees used to build BRT models, using a different number of environmental descriptors show that BRT is not sensitive to model overfitting, and only selects the relevant information needed for model calibration, a property formulated as the stagewise selection by Elith et al. (2008). Non-informative environmental data that might complexify SDMs are automatically pruned when generating BRT trees, and the most relevant descriptors only are retained to model species distribution (De'ath and Fabricius 2000, Whittingham et al. 2006, Elith et al. 2008). However, selecting a reduced number of environmental descriptors allows the production of simpler models for which descriptor contributions can be easily interpreted (Bucklin et al. 2015).

4.2. Influence of collinearity on modelling performance

Removing collinear descriptors from datasets has remained an usual approach in species distribution modelling (Dormann et al. 2012b, Merow et al. 2013, Fabri-Ruiz 2018, Guillaumot et al. 2018b). However, this strategy has recently been questioned when SDMs are not used for extrapolation (Braunisch et al. 2013, Bucklin et al. 2015, Li et al. 2016, Petitpierre et al. 2017). In the present study, results show that modelling performances (AUC, TSS, COR and percentage of correctly classified test data) of some SDMs significantly decrease when collinear descriptors are removed (i.e. *A. hodgsoni*, *O. validus* and *G. antarctica*). Removing collinear variables that significantly contribute to SDMs may induce model instability and reduce modelling performance. The observed decrease in AUC scores may be due to the reduction of model overfitting when removing collinear descriptors (Dhingra et al. 2016).
Machine-learning algorithms are efficient modelling tools that take into account the multiple interactions among descriptors (Segurado and Araújo 2004, Araújo and Guisan 2006, Dormann et al. 2008, Elith et al. 2008, Braunisch et al. 2013) and can correct for collinearity between

environmental descriptors if not too strong (Dormann et al. 2012b). In the present work, this is shown by the fact that performance of SDMs produced for the species *B. loripes*, *L. annulatus* and *P. charcoti* remains unaffected when collinear descriptors are removed from the analysis.

However, using collinear descriptors in SDMs can make model outputs difficult to interpret when temporal and spatial relationships between descriptors are unknown (Dormann et al. 2012b), because collinearity induces complex relationships between environmental drivers and the underlying processes (Guisan and Thuiller 2005, Elith and Leathwick 2009, Merow et al. 2013). Several methods have been documented to correct for strong collinear effects. The sequential regression approach is one of them and aims at replacing correlated variables by a linear or non-linear model (Leathwick et al. 2006, Dormann et al. 2008). A second method consists in using descriptor score values on PCA principal components rather than descriptor raw values themselves (Kühn 2007, Dormann et al. 2008). However, in this latter approach, SDMs and species ecological preferences are difficult to interpret.

4.3. Selection of environmental descriptors

'Distance layers' (Appendix 2.4) have been commonly used as descriptors in previous SDMs performed for Southern Ocean studies (Cheung et al. 2008, Murase et al. 2013, Mormède et al. 2014c, Nachtsheim et al. 2017). In the present work, 'distance layers' were used in the first set of SDMs and they all showed strong contributions to model outputs. 'Distance layers' may be strongly correlated to environmental gradients, and especially to latitudinal gradients, or may integrate the multiple effects of diverse environmental variations (Bradie and Leung 2017, Ferrari et al. 2018). Interpreting the contribution of such descriptors to SDMs can remain problematic and depends on research objectives, especially depending on whether ecological significance or statistical contributions only are sought. The statistical contribution of a descriptor to the model is the independent contribution of the descriptor deduced from what other descriptors already bring (Dormann et al. 2012b), it may not necessarily imply a direct ecological significance. Consequently, 'distance layers' were removed from the initial set of environmental descriptors along with collinear descriptors and descriptors that contributed the least to models (28 descriptors out of the 58 available, Table 2.5). This reduces the set to 14 or 16 descriptors only depending on the species under study (Appendix 2.10).

4.4. Final model outputs

In the present study, SDMs performed for *A. hodgsoni*, *L. annulatus* and *O. validus* showed lower performances (lower AUC, TSS, COR and correctly classified test data) compared to SDMs performed for *B. loripes*, *G. antarctica* and *P. charcoti*. For these last three species, a higher number of records were available and contributed to the high model performances as species niches were better described during model calibration (Qiao et al. 2015, van Proosdij et al. 2016, Guillaumot et al. 2018a - Appendix). Despite these differences in model performance, descriptor contributions and species predicted distributions are mostly similar between models (Fig. 2.12, 2.14-15). This is an unexpected result as the six studied species were initially selected for their contrasting ecological niches and life traits, which should have determined distinct occupied environments and biogeographic patterns. This unexpected result stresses the limits of SDMs performed at broad spatial scale. The low resolution (in space and time) of environmental descriptors, the heterogeneous sampling and the relative low number of occurrence records available are cumulative limitations to model accuracy and species ecological requirements were not precisely captured by models. In contrast, models are all structured by large-scale and common environmental drivers relating to broad-scale latitudinal gradients that prevail between Antarctic and sub-Antarctic regions (Clarke and Johnston 2003, Linse et al. 2006, De Broyer et al. 2014, Moreau et al. 2017).

5. CONCLUSIONS AND RECOMMENDATIONS

This work aimed at testing the influence of the number of selected environmental descriptors and their collinearity on model performance. Models were generated at the scale of the entire Southern Ocean using BRT. The BRT algorithm is a machine-learning approach that automatically selects descriptors that best characterise species niches (Elith et al. 2008). This matches our results that highlight that all models generated with different number of environmental descriptors showed similar performances. In contrast, in most SDMs generated without collinear descriptors, model overfitting tends to be minimized in comparison with models generated with the whole set of 58 descriptors. In three species only, no difference in model performance was observed between models using either collinear or non-collinear descriptors.

Final models were generated using a subset of 14 to 16 environmental descriptors that best explain species distributions. The selected descriptors are not collinear to limit interpretation errors, reduce model complexity and favour the ecological relevance of models (Austin and van Niel 2011, Braunisch et al. 2013, Bucklin et al. 2015, Petitpierre et al. 2017). However, final SDMs are not very contrasted between species despite significant differences in species ecological niches (McClintock et al. 2008a, Mah and Blake 2012, Lawrence 2013, Brandt et al. 2014, Danis et al. 2014, Moles et al. 2015). The performed SDMs are more sensitive to the number of occurrence records available and to the extent of the projection area. This final result questions the ecological relevance of using modelling approaches at broad spatial scale when based a limited number of occurrence data, spatially aggregated and using descriptors with coarse spatial and temporal resolutions.

These results match those obtained in previous studies and suggest that the validation of model predictions should use independent data, appropriate statistics and expert-based interpretations (Guisan et al. 2013, Fois et al. 2018, Fourcade et al. 2018, Leroy et al. 2018). Combining model outputs performed at narrow spatial scale and complementary data on biotic interactions (Wisz et al. 2013, Leach et al. 2016, Van der Putten et al. 2017), habitat features (Ferrari et al. 2018) and physiological traits (Kearney and Porter 2009, Fordham et al. 2013, Wittmann et al. 2016, Feng and Papes 2017, Mathewson et al. 2017, Pertierra et al. 2017) constitutes a good alternative. This can enhance the relevance of explanatory models and their use for ecological studies and conservation purposes. Downscaling SDM studies also has the advantage of improving model accuracy relating to particular, local to regional phenotypic or physiological traits of populations, which may differ at broader scale (Thatje 2012). Waiting for more data and ensuring the taxonomic quality of datasets, we recommend the use of SDMs for narrow-scale studies using scrutinized and comprehensive occurrence datasets, as much as possible, while selecting non-collinear and ecologically relevant descriptors to minimize model overfitting (El Gabbas and Dormann 2018, Fois et al. 2018).

APPENDIX 2.4 List of environmental descriptors and sources

Table S2.4. List of environmental descriptors selected for species distribution models. Downloadable on the 'blueant' R package (https://github.com/AustralianAntarcticDivision/blueant). The procedure to download the data is explained in the "data_for_SDM_vignette" at https://github.com/AustralianAntarcticDivision/blueant/tree/data_Charlene/vignettes. Spatial extent of the data: latitude: 45°S_80°S / longitude: -180°_180°W. Spatial resolution: 0.1° x 0.1° (approximately 11km). Complementary information about "extreme events" layers can be found in Appendix 2.5.

Environmental descriptor	Unit	Description	Source
Depth	meters	Bathymetry. Downloaded from GEBCO 2014 (0.0083°= 30sec arcmin resolution) and set at 0.1° resolution. Completed with the bathymetry layer manually corrected and provided in Fabri-Ruiz et al. (2017b) [1]	This study. Derived from GEBCO [2]
Geomorphology	categorical	Derived from the seafloor geomorphic feature dataset of O'Brien et al. (2009) [3]. 27 categories	This study. Derived from Australian Antarctic Data Centre [4]
Sediments	categorical	Sediment features (14 categories)	Griffiths 2014 (unpublished)
Slope	degrees	Derived from bathymetry with the *terrain* function of the 'raster' R package (Hijmans 2019) [6]. Computation according to Horn (1981) [5], i.e. option neighbor=8. The computation was done on the GEBCO bathymetry layer (0.0083° resolution) and the resolution was then changed to 0.1°.	This study. Derived from GEBCO [2]
Roughness	unitless	Derived from bathymetry with the *terrain* function of the 'raster' R package (Hijmans 2019) [6]. Roughness is the difference between the maximum and the minimum value of a cell and its 8 surrounding cells. The computation was done on the GEBCO bathymetry layer (0.0083° resolution) and the resolution was then changed to 0.1°.	This study. Derived from GEBCO [2]
Mixed layer depth	m	Summer mixed layer depth climatology from ARGOS data. Re-gridded at 0.1° resolution from a 2-degree grid using a nearest neighbor interpolation	This study. Derived from Australian Antarctic Data Centre [4]
Sea surface current speed	$m.sec^{-1}$	Current speed near the surface (2.5m depth); derived from the CAISOM model (Galton-Fenzi et al. 2012 [7], based on ROMS)	This study. Derived from Australian Antarctic Data Centre [4]
Sea floor current speed	$m.sec^{-1}$	Current speed near the sea floor; derived from the CAISOM model (Galton-Fenzi et al. 2012 [7], based on ROMS)	This study. Derived from Australian Antarctic Data Centre [4]
Distance antarctica	km	Distance to the nearest part of Antarctic continent	This study. Derived from Australian Antarctic Data Centre [4]
Distance canyon	km	Distance to the axis of the nearest canyon	This study. Derived from Australian Antarctic Data Centre [4]
Distance max ice edge	km	Mean maximum winter sea ice extent derived from daily estimates of sea ice concentration. Distance of each grid point to this extent.	This study. Derived from Australian Antarctic Data Centre [4]
Distance shelf	km	Distance to the nearest area of sea floor of depth 500m or less	This study. Derived from Australian Antarctic Data Centre [4]
Ice cover max	-	Ice concentration fraction, maximum on 1957-2017 time period	BioOracle accessed 24/04/2018, see Assis et al. (2018) [8]
Ice cover mean	-	Ice concentration fraction, mean on 1957-2017 time period	BioOracle accessed 24/04/2018, see Assis et al. (2018) [8]
Ice cover min	-	Ice concentration fraction, minimum on 1957-2017 time period	BioOracle accessed 24/04/2018, see Assis et al. (2018) [8]
Ice cover range	-	Ice concentration fraction, difference maximum-minimum on 1957-2017 time period	BioOracle accessed 24/04/2018, see Assis et al. (2018) [8]
Ice thickness max	m	Ice thickness, maximum on 1957-2017 time period	BioOracle accessed 24/04/2018, see Assis et al. (2018) [8]

Ice thickness mean	m	Ice thickness, mean on 1957-2017 time period	BioOracle accessed 24/04/2018, see Assis et al. (2018) [8]
Ice thickness min	m	Ice thickness, minimum on 1957-2017 time period	BioOracle accessed 24/04/2018, see Assis et al. (2018) [8]
Ice thickness range	m	Ice thickness, difference maximum-minimum on 1957-2017 time period	BioOracle accessed 24/04/2018, see Assis et al. (2018) [8]
chla_ampli_alltime_2005_2012	$mg.m^{-3}$	Chlorophyll-a concentrations obtained from MODIS satellite images. Amplitude of pixel values (difference between maximal and minimal values encountered by each pixel during all months of the period 2005-2012)	MODIS Aqua [9]
chla_max_alltime_2005_2012	$mg.m^{-3}$	Chlorophyll-a concentrations obtained from MODIS satellite images. Maximal value encountered by each pixel during all months of the period 2005-2012	MODIS Aqua [9]
chla_mean_alltime_2005_2012	$mg.m^{-3}$	Chlorophyll-a concentrations obtained from MODIS satellite images. Mean value of each pixel during all months of the period 2005-2012	MODIS Aqua [9]
chla_min_alltime_2005_2012	$mg.m^{-3}$	Chlorophyll-a concentrations obtained from MODIS satellite images. Minimal value encountered by each pixel during all months of the period 2005-2012	MODIS Aqua [9]
chla_sd_alltime_2005_2012	$mg.m^{-3}$	Chlorophyll-a concentrations obtained from MODIS satellite images. Standard deviation value of each pixel during all months of the period 2005-2012	MODIS Aqua [9]
POC_2005_2012_ampli	$gC.m^{-2}.d^{-1}$	Particulate organic carbon; model Lutz et al. (2007) [10]. Amplitude value (difference maximal and minimal values, see previous layers) of all average seasonal layers of 2005-2012	This study. Following Lutz et al. (2007) [10], data available on Australian Antarctic Data Centre [11]
POC_2005_2012_max	$gC.m^{-2}.d^{-1}$	Particulate organic carbon; model Lutz et al. (2007) [10]. Maximal value encountered on each pixel among all seasonal layers of 2005-2012	This study. Following Lutz et al. (2007) [10], data available on Australian Antarctic Data Centre [11]
POC_2005_2012_mean	$gC.m^{-2}.d^{-1}$	Particulate organic carbon; model Lutz et al. (2007) [10]. Mean of all seasonal layers of 2005-2012	This study. Following Lutz et al. (2007) [10], data available on Australian Antarctic Data Centre [11]
POC_2005_2012_min	$gC.m^{-2}.d^{-1}$	Particulate organic carbon; model Lutz et al. (2007) [10]. Minimal value encountered on each pixel among all seasonal layers of 2005-2012	This study. Following Lutz et al. (2007) [10], data available on Australian Antarctic Data Centre [11]
POC_2005_2012_sd	$gC.m^{-2}.d^{-1}$	Particulate organic carbon; model Lutz et al. (2007) [10]. Standard deviation all seasonal layers of 2005-2012	This study. Following Lutz et al. (2007) [10], data available on Australian Antarctic Data Centre [11]
seafloor_oxy_19552012_ampli	$mL.L^{-1}$	Amplitude (difference max/min) value encountered for each pixel on all month layers of seafloor oxygen concentrations over 2005-2012, modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] oxygen concentration layers
seafloor_oxy_19552012_max	$mL.L^{-1}$	Maximum value encountered for each pixel on all month layers of seafloor oxygen concentrations over 2005-2012, modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] oxygen concentration layers
seafloor_oxy_19552012_mean	$mL.L^{-1}$	Mean seafloor oxygen concentrations over 2005-2012 (average of all monthly layers), modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] oxygen concentration layers
seafloor_oxy_19552012_min	$mL.L^{-1}$	Minimum value encountered for each pixel on all month layers of seafloor oxygen concentration over 2005-2012, modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] oxygen concentration layers
seafloor_oxy_19552012_sd	$mL.L^{-1}$	Standard deviation seafloor oxygen concentration over 2005-2012 (of all monthly layers), modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] oxygen concentration layers
seafloor_sali_2005_2012_ampli	PSS	Amplitude (difference max/min) value encountered for each pixel on all month layers of seafloor salinity over 2005-2012, modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] salinity layers
seafloor_sali_2005_2012_max	PSS	Maximum value encountered for each pixel on all month layers of seafloor salinity over 2005-2012, modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] salinity layers
seafloor_sali_2005_2012_mean	PSS	Mean seafloor salinity over 2005-2012 (average of all monthly layers), modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] salinity layers
seafloor_sali_2005_2012_min	PSS	Minimum value encountered for each pixel on all month layers of seafloor salinity over 2005-2012, modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] salinity layers
seafloor_sali_2005_2012_sd	PSS	Standard deviation seafloor salinity over 2005-2012 (of all monthly layers), modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] salinity layers

seafloor_temp_2005_2012_ampli	°C	Amplitude (difference max/min) value encountered for each pixel on all month layers of seafloor temperature over 2005-2012, modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] temperature layers
seafloor_temp_2005_2012_max	°C	Maximum value encountered for each pixel on all monthly layers of seafloor temperature over 2005-2012, modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] temperature layers
seafloor_temp_2005_2012_mean	°C	Mean seafloor temperature over 2005-2012 (average of all monthly layers), modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] temperature layers
seafloor_temp_2005_2012_min	°C	Minimum value encountered for each pixel on all monthly layers of seafloor temperature over 2005-2012, modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] temperature layers
seafloor_temp_2005_2012_sd	°C	Standard deviation seafloor temperature over 2005-2012 (of all monthly layers), modified from WOCE	Derived from World Ocean Circulation Experiment 2013 [12] temperature layers
extreme_event_max_chl_2005_2012_ampli	integer	Amplitude value (Maximum-Minimum) of the number of extreme events (maximal chlorophyll-a concentrations) recorded between 2005 and 2012	Derived from chla_max_alltime_2005_2012 layer
extreme_event_max_chl_2005_2012_max	integer	Maximum number of extreme events (maximal chlorophyll-a concentrations) recorded between 2005 and 2012	Derived from chla_max_alltime_2005_2012 layer
extreme_event_max_chl_2005_2012_mean	integer	Mean of the number of extreme events (maximal chlorophyll-a concentrations) recorded between 2005 and 2012	Derived from chla_max_alltime_2005_2012 layer
extreme_event_max_chl_2005_2012_min	integer	Minimum number of extreme events (maximal chlorophyll-a concentrations) recorded between 2005 and 2012	Derived from chla_max_alltime_2005_2012 layer
extreme_event_min_chl_2005_2012_ampli	integer	Amplitude value (Maximum-Minimum) of the number of extreme events (minimal chlorophyll-a concentrations) recorded between 2005 and 2012	Derived from chla_min_alltime_2005_2012 layer
extreme_event_min_chl_2005_2012_max	integer	Maximum number of extreme events (minimal chlorophyll-a concentrations) recorded between 2005 and 2012	Derived from chla_min_alltime_2005_2012 layer
extreme_event_min_chl_2005_2012_mean	integer	Mean of the number of extreme events (minimal chlorophyll-a concentrations) recorded between 2005 and 2012	Derived from chla_min_alltime_2005_2012 layer
extreme_event_min_chl_2005_2012_min	integer	Minimum number of extreme events (minimal chlorophyll-a concentrations) recorded between 2005 and 2012	Derived from chla_min_alltime_2005_2012 layer
extreme_event_min_oxy_1955_2012_nb	integer	Number of extreme events (minimal seafloor oxygen concentration records) that happened between January and December of the year	Derived from seafloor_oxy_19552012_min layer
nb_extreme_event_max_sali_2005_2012	integer	Number of extreme events (maximal seafloor salinity records) that happened between January and December of the year	Derived from seafloor_sali_2005_2012_max layer
nb_extreme_event_min_sali_2005_2012	integer	Number of extreme events (minimal seafloor salinity records) that happened between January and December of the year	Derived from seafloor_sali_2005_2012_min layer
extreme_event_max_temp_2005_2012_nb	integer	Number of extreme events (maximal seafloor temperature records) that happened between January and December of the year	Derived from seafloor_temp_2005_2012_max layer
extreme_event_min_temp_2005_2012_nb	integer	Number of extreme events (minimal seafloor temperature records) that happened between January and December of the year	Derived from seafloor_temp_2005_2012_min layer

References

[1] Fabri-Ruiz S, Saucède T, Danis B & David B (2017b). Southern Ocean Echinoids database–An updated version of Antarctic, Sub-Antarctic and cold temperate echinoid database. ZooKeys, (697), 1.

[2] GEBCO. https://www.gebco.net/data_and_products/gridded_bathymetry_data/

[3] O'Brien PE, Post AL & Romeyn R (2009) Antarctic-wide geomorphology as an aid to habitat mapping and locating vulnerable marine ecosystems. CCAMLR VME Workshop 2009. Document WS-VME-09/10

[4] AADC, link: http://webdav.data.aad.gov.au/data/environmental/derived/antarctic/ accessed 2018; complementary metadata at https://data.aad.gov.au/metadata/records/Polar_Environmental_Data

[5] Horn BKP (1981). Hill shading and the reflectance map. Proceedings of the IEEE 69:14-47
[6] Hijmans, R.J., Van Etten, J., Sumner, M. *et al.* (2019). R package "raster". https://CRAN.R-project.org/package=raster
[7] Galton-Fenzi BK, Hunter JR, Coleman R, Marsland SJ, Warner RC (2012) Modeling the basal melting and marine ice accretion of the Amery Ice Shelf. Journal of Geophysical Research: Oceans, 117, C09031. http://dx.doi.org/10.1029/2012jc008214
[8] Assis J, Tyberghein L, Bosch S, Verbruggen H, Serrão EA & De Clerck O (2018). Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. Global Ecology and Biogeography, 27(3), 277-284.
[9] MODIS Aqua, link: https://oceandata.sci.gsfc.nasa.gov/MODIS-Aqua/Mapped/Monthly/9km/chlor_a/
[10] Lutz MJ, Caldeira K, Dunbar RB & Behrenfeld MJ (2007). Seasonal rhythms of net primary production and particulate organic carbon flux to depth describe the efficiency of biological pump in the global ocean. Journal of Geophysical Research: Oceans, 112(C10).
[11] AADC POC export data, link: https://data.aad.gov.au/metadata/records/Particulate_carbon_export_flux_layers, created 2017
[12] WOCE 2013, link: https://www.nodc.noaa.gov/OC5/woa13/woa13data.html accessed 2018

APPENDIX 2.5. Extreme events layers

In this study, raster layers are produced to depict extreme events occuring in a year for each pixel. This is done for chlorophyll-a and oxygen concentrations, seafloor temperatures and salinities. The aim is to describe the frequency of environmental changes occuring in the area and in a second step, model its contribution to explaining species occurrence distribution through the SDM analysis.
In a annual series, an “extreme event” is defined as a value that is higher than the median value of the series. This analysis is pixel-specific.

Using monthly raster layers, the code extracts the series of values ($Y_{i,1}$, $Y_{i,j,}$…$Y_{i,12}$) from each pixel i, for the corresponding month (j=1,..,12). A vector of 12 values is obtained and used to calculate the median value of the annual series and the associated MAD value (Median Absolute Deviation), an equivalent to the standard deviation computed for the median (MAD= median ($|Y_{i,j}$-median($Y_{i,j}$))))
A maximal extreme event is counted when $Y_{i,j}$+$MAD_{i,j}$ and $Y_{i,j}$-$MAD_{i,j}$ values are higher than the median value of the series, and similarly, a minimal extreme event is counted when $Y_{i,j}$+$MAD_{i,j}$ and $Y_{i,j}$-$MAD_{i,j}$ values are lower than the median value of the series (Fig. S2.5.A).

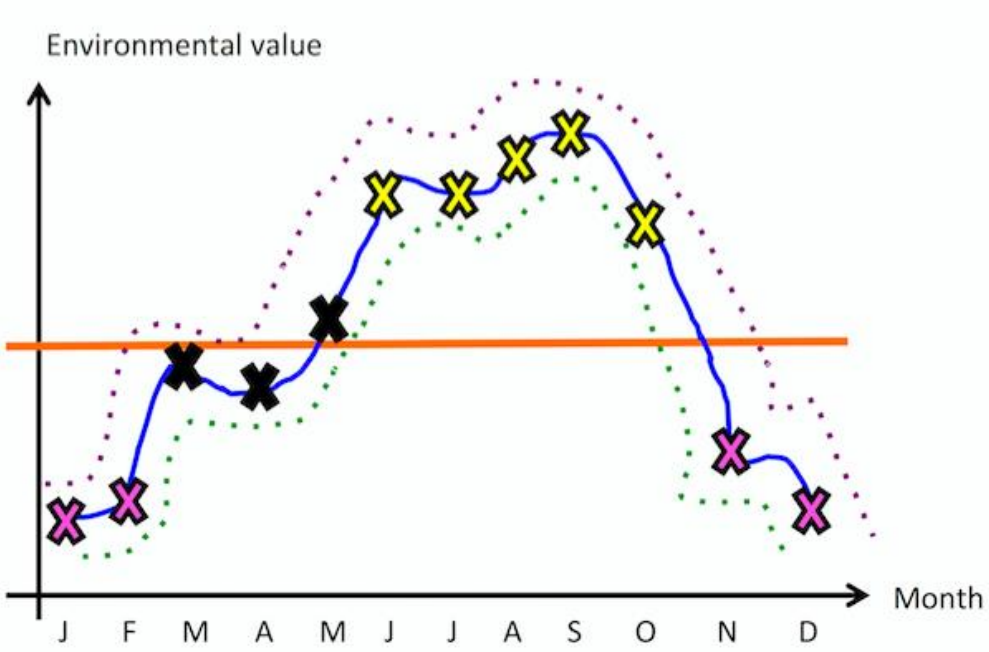


Figure S2.5.A. Theoretical plot showing the determination of extreme events. Crosses: $Y_{i,j}$ values of the raster layer for a pixel i and a month j ; orange continuous line: median value of the series ; purple dotted line: $Y_{i,j}$+$MAD_{i,j}$; green dotted line: $Y_{i,j}$-$MAD_{i,j}$. In this example, yellow crosses are maximal extreme events because $Y_{i,j}$, $Y_{i,j}$+$MAD_{i,j}$ and $Y_{i,j}$-$MAD_{i,j}$ are higher than the median value of the series (orange line); pink crosses are minimal extreme events because $Y_{i,j}$, $Y_{i,j}$+$MAD_{i,j}$ and $Y_{i,j}$-$MAD_{i,j}$ are lower than the median value of the series (orange line); the black crosses are not considered as extreme events because $Y_{i,j}$, $Y_{i,j}$+$MAD_{i,j}$ or $Y_{i,j}$-$MAD_{i,j}$ are cutting the MAD line.

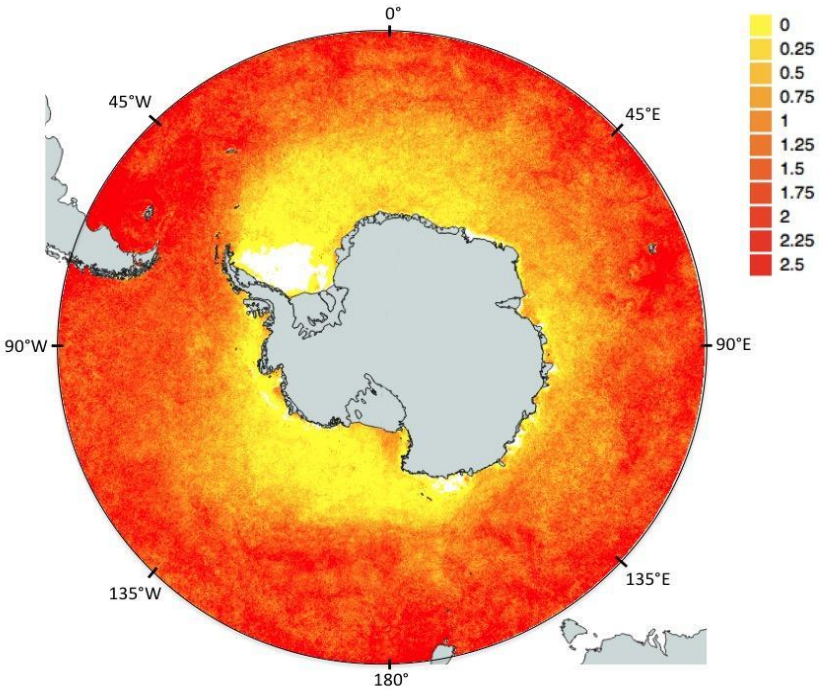


Figure S2.5.B. Example an extreme event raster layer. Average number of maximum chlorophyll-a concentrations extreme events per pixel compiled between 2005 and 2012.

APPENDIX 2.6. Cumulative occurrence collection curved through time

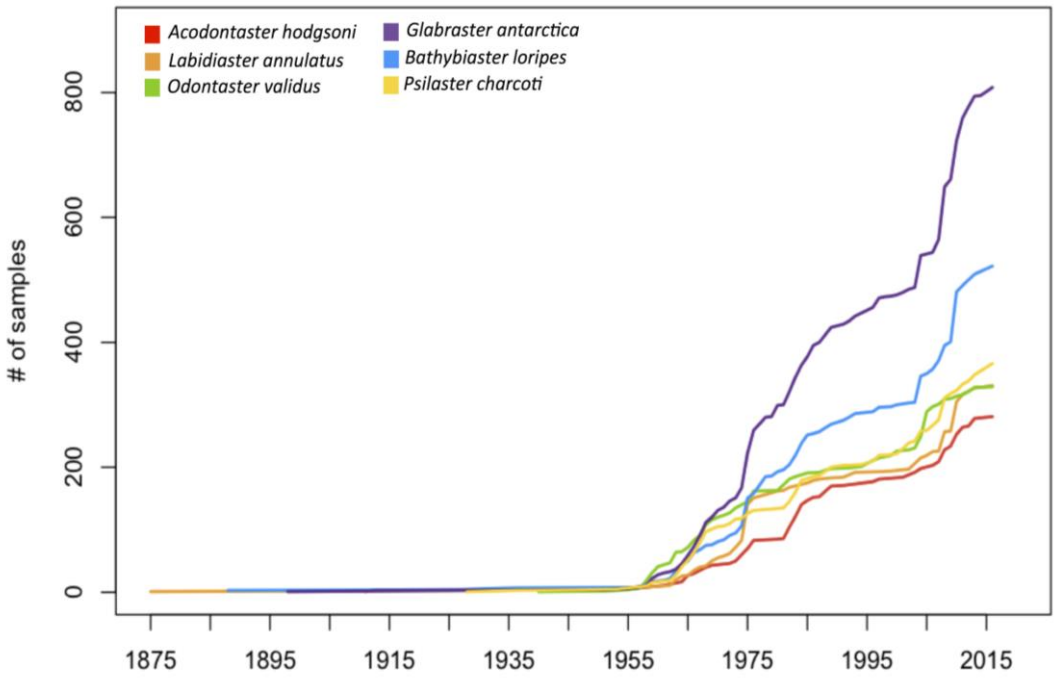


Figure S2.6. Cumulative occurrence collective curves through time and per species.

APPENDIX 2.7. Available presence-only data of the modelled species

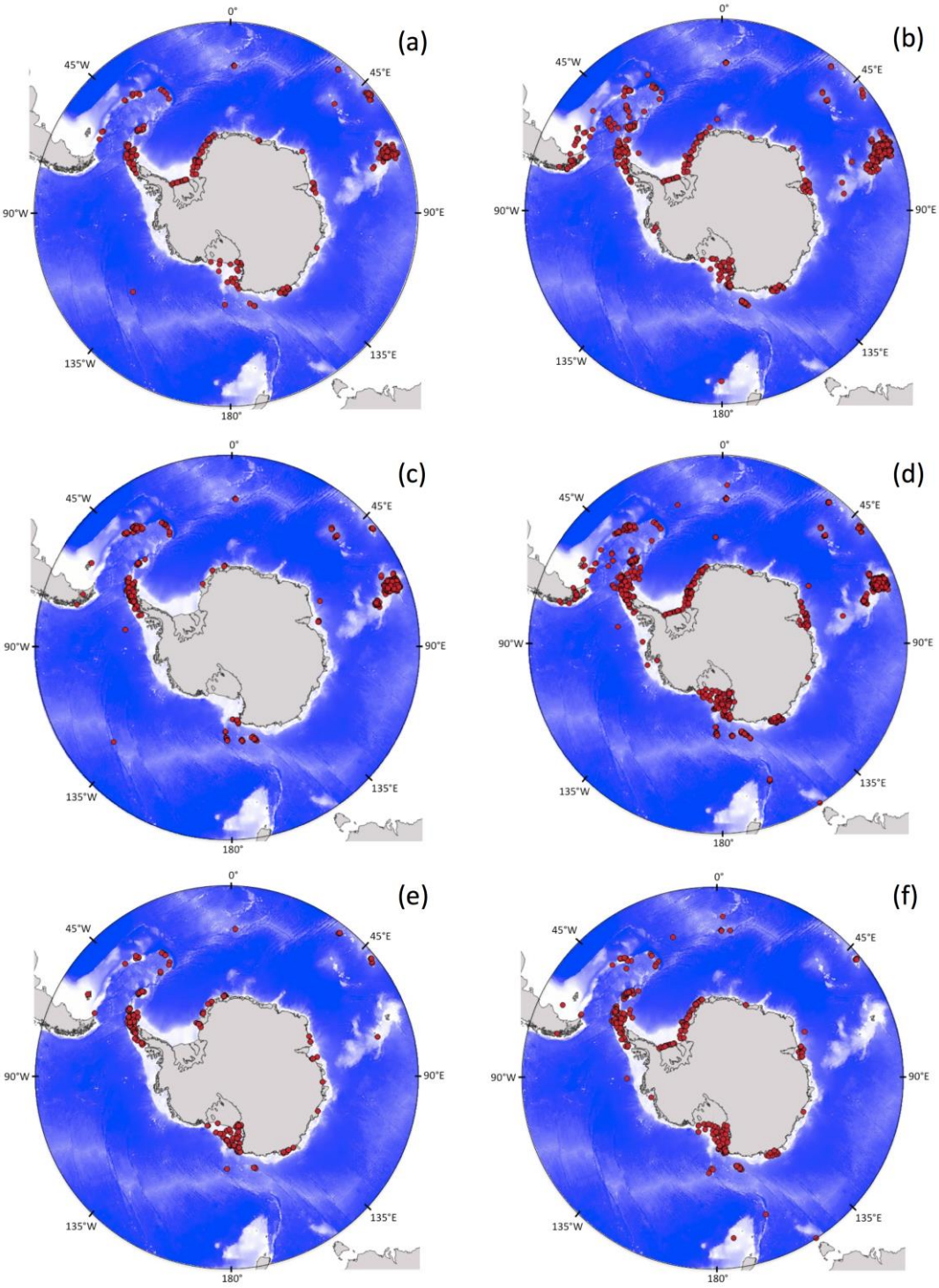


Figure S2.7. Presence-only records available for the six studied species (a) *Acodontaster hodgsoni* (n=297), (b) *Bathybiaster loripes* (n=585), (c) *Labidiaster annulatus* (n=373), (d) *Glabraster antarctica* (n=844), (e) *Odontaster validus* (n=309), (f) *Psilaster charcoti* (n=350). Bathymetry is represented by blue shaded background. The provided number of presence data available is given after removal of duplicate records on a same grid-cell pixel.

APPENDIX 2.8. Calibration of Boosted Regression Trees and parameter settings

BRT models are generated using the cross-validation procedure developed by Elith et al. (2008) that is, using the functions provided in their supplementary material and in the *gbm* R package (Ridgeway et al. 2006). We set the maximum number of trees to 10,000 and models are calibrated with the combination of parameters that minimizes the predictive deviance while producing the lowest number of trees (Fig. S2.8). Models are calibrated with all presence records available, using 1,000 background data randomly sampled in the area (restrained in depth for each species specifically) and according to the Kernel Density Estimate weighting scheme for the total Southern Ocean benthic samples (Guillaumot et al. 2019 - Chapter 2) and all the 58 environmental descriptors available. The following parameters are finally selected for each species: *Acodontaster hodgsoni* (tc=4, lr=0.007, bf=0.75), *Bathybiaster loripes* (tc=4, lr=0.012, bf=0.7), *Glabraster antarctica* (tc=4, lr=0.013, bf=0.75), *Labidiaster annulatus* (tc=4, lr=0.012, bf=0.75), *Odontaster validus* (tc=4, lr=0.007, bf=0.7), *Psilaster charcoti* (tc=4, lr=0.007, bf=0.7).

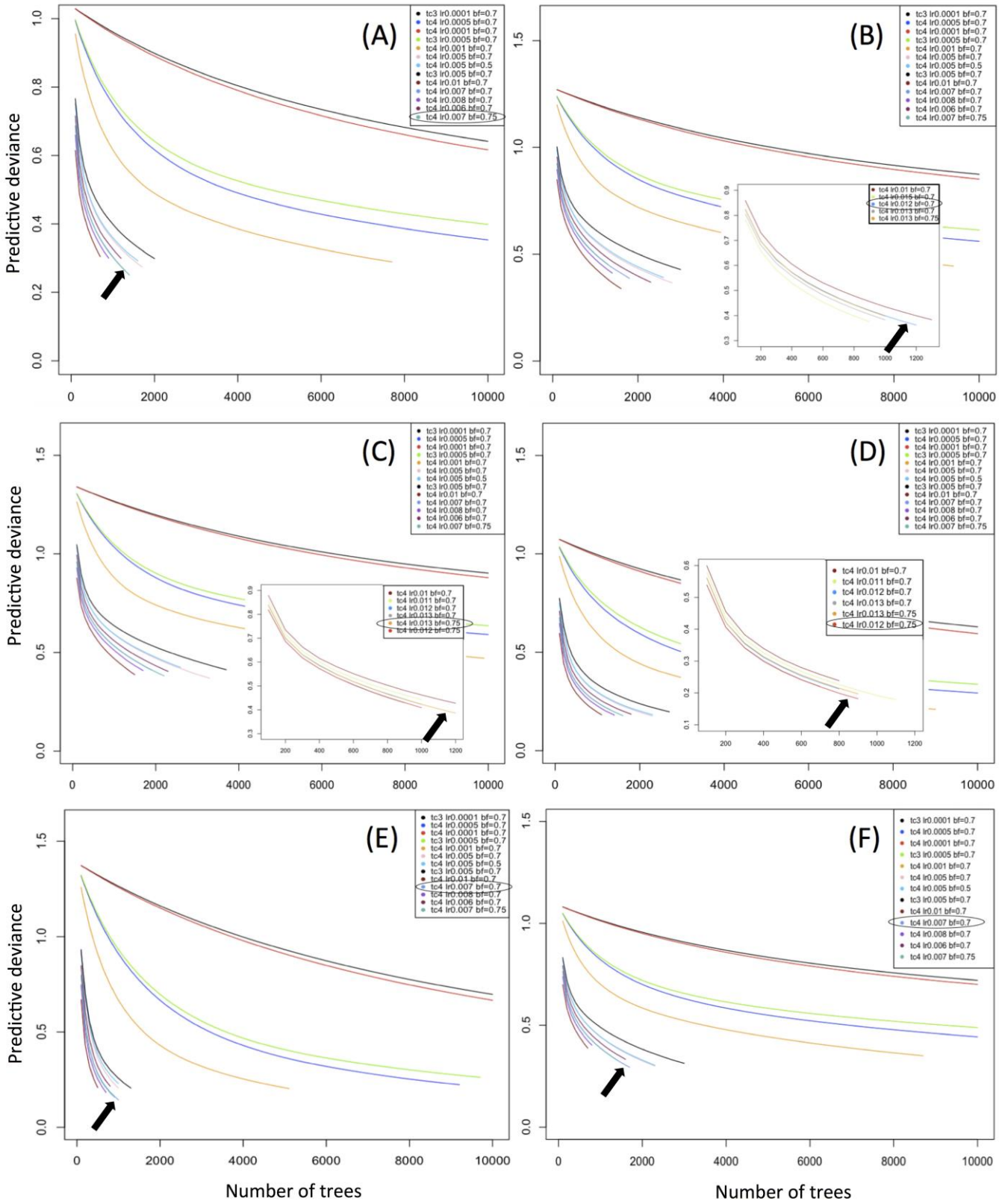


Figure S2.8. Comparison of model predictive deviance according to the number of trees used to build the models, for each species and for different parameter settings (tree complexity, tc; learning rate, lr; bag fraction, bf). Parameters for which the lowest predictive deviance is reached with the lowest number of trees are selected to generate the model (Elith et al. 2008). Species: (A) *Acodontaster hodgsoni*, (B) *Bathybiaster loripes*, (C) *Glabraster antarctica*, (D) *Labidiaster loripes*, (E) *Odontaster validus*, (F) *Psilaster charcoti*.

APPENDIX 2.9. Control of the number of environmental descriptors over modelling performances

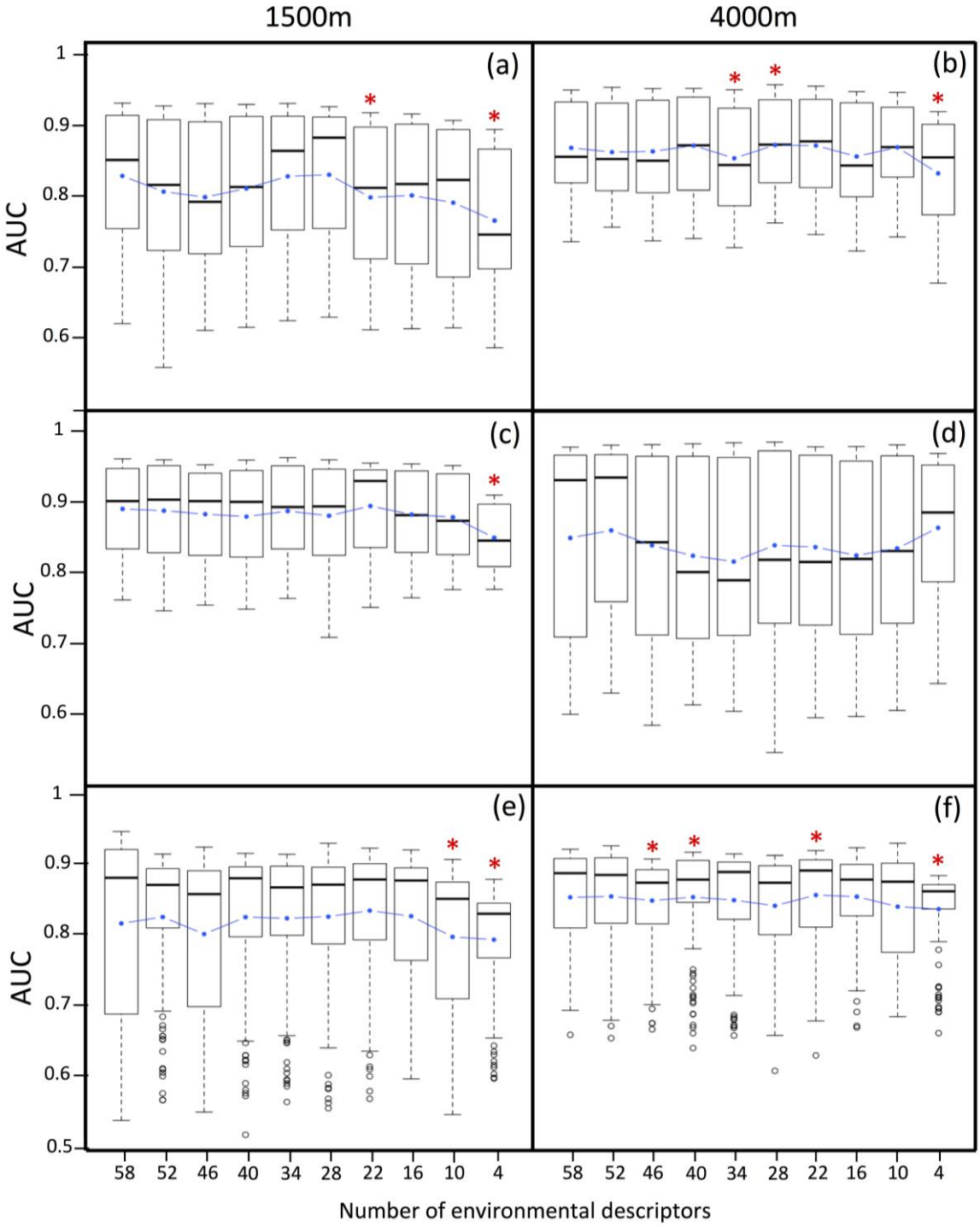


Figure S2.9A. Influence of the number of environmental predictors on SDM performance. Boxplot of 100 model replicates scores. Change in Area Under the Curve (AUC) values for (a) *Acodontaster hodgsoni*, (b) *Bathybiaster loripes*, (c) *Labidiaster annulatus*, (d) *Glabraster antarctica*, (e) *Odontaster validus*, (f) *Psilaster charcoti*. Average values are indicated in blue. Red stars indicate significant changes obtained in median values between two successive series (Wilcoxon Mann-Whitney rank paired test, $p < 0.05$). Left-side and right-side columns correspond to species for which models are respectively projected down to 1,500 m and 4,000 m depth.

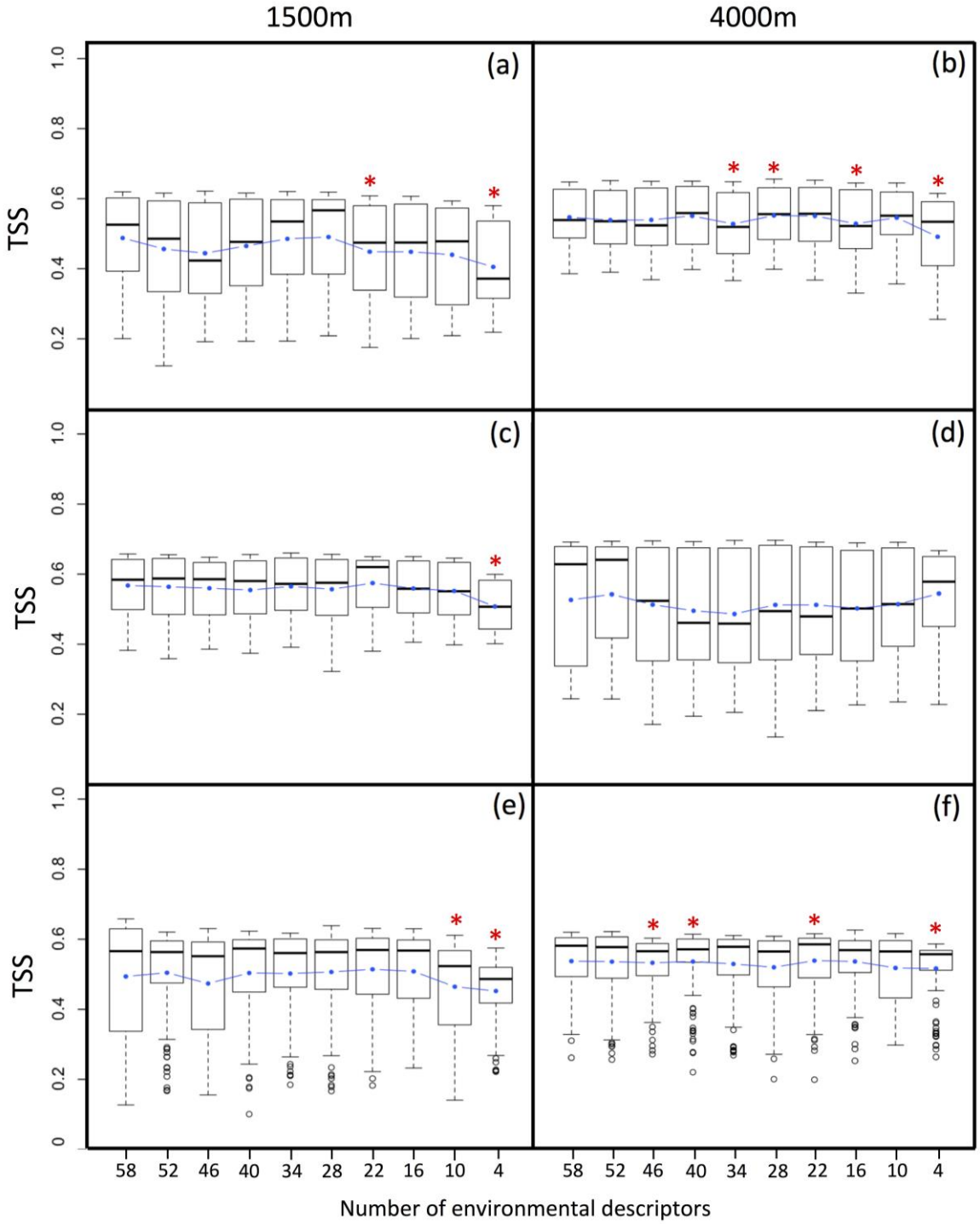


Figure S2.9.B Influence of the number of environmental predictors on SDM performance. Boxplot of 100 model replicates scores. Change in True Skill Statistics (TSS) values for (a) *Acodontaster hodgsoni*, (b) *Bathybiaster loripes*, (c) *Labidiaster annulatus*, (d) *Glabraster antarctica*, (e) *Odontaster validus*, (f) *Psilaster charcoti*. Average values are indicated in blue. Red stars indicate significant changes in median values between two successive series (Wilcoxon Mann-Whitney rank paired test, $p < 0.05$). Left-side and right-side columns correspond to species for which models are respectively projected down to 1,500 m and 4,000 m depth.

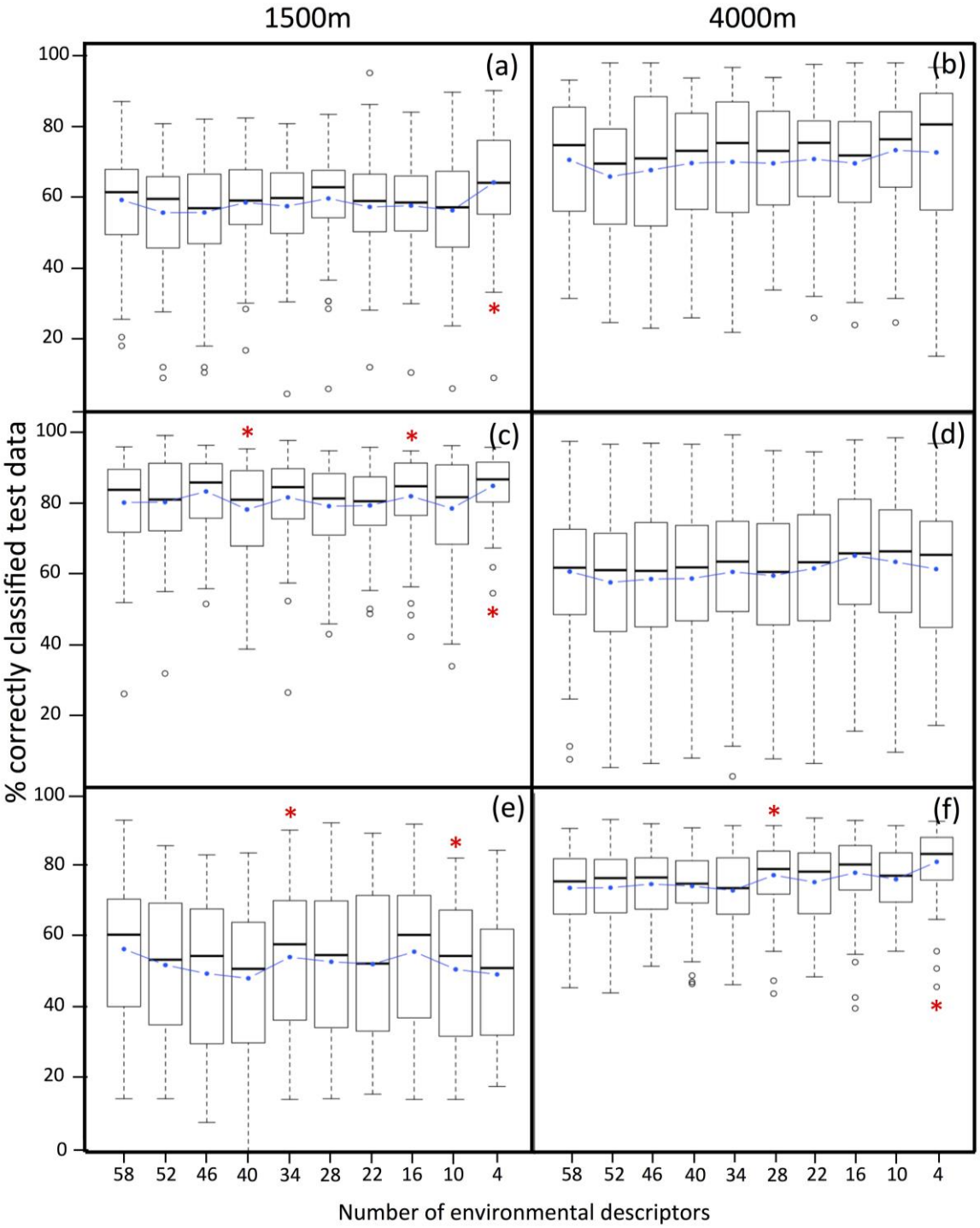


Figure S2.9.C. Influence of the number of environmental predictors on SDM performance. Boxplot of 100 model replicates scores. Change in the percentage of correctly classified test data (cross-validation procedure) for (a) *Acodontaster hodgsoni*, (b) *Bathybiaster loripes*, (c) *Labidiaster annulatus*, (d) *Glabraster antarctica*, (e) *Odontaster validus*, (f) *Psilaster charcoti*. Average values are indicated in blue. Red stars indicate significant changes in median values between two successive series (Wilcoxon Mann-Whitney rank paired test, $p < 0.05$). Left-side and right-side columns correspond to species for which models are respectively projected down to 1,500 m and 4,000 m depth.

APPENDIX 2.10. List of environmental descriptors selected to generate final models

Table S2.10. List of environmental descriptors selescted to generate final models, after removing distance descriptors, descriptors that always contribute less than 1% to species SDM (Table 2.5) and collinear descriptors (species-specific).

	Acodontaster hodgsoni	*Bathybiaster loripes*	*Glabraster antarctica*	*Labidiaster annulatus*	*Odontaster validus*	*Psilaster charcoti*
1	depth	depth	depth	depth	depth	depth
2	geomorphology	geomorphology	geomorphology	geomorphology	geomorphology	geomorphology
3	sediments	sediments	sediments	sediments	sediments	sediments
4	slope	slope	slope	slope	slope	slope
5	roughness	roughness	roughness	roughness	roughness	roughness
6	mixed layer depth	mixed layer depth	mixed layer depth	mixed layer depth	mixed layer depth	mixed layer depth
7	seasurface current speed	seasurface current speed	seasurface current speed	seasurface current speed	seasurface current speed	seasurface current speed
8	seafloor current speed	seafloor current speed	seafloor current speed	seafloor current speed	seafloor current speed	seafloor current speed
9	ice cover mean	ice cover mean	ice cover mean	ice cover mean	ice cover mean	ice cover mean
10	chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]
11	chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]
12	chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]
13	POC minimum [2005-2012]	POC minimum [2005-2012]	POC minimum [2005-2012]	POC minimum [2005-2012]	POC minimum [2005-2012]	POC minimum [2005-2012]
14	POC amplitude [2005-2012]	POC standard deviation [2005-2012]	POC standard deviation [2005-2012]	POC standard deviation [2005-2012]	POC standard deviation [2005-2012]	POC standard deviation [2005-2012]
15			Chlorophyll a minimum extreme events, minimun values	Chlorophyll a minimum extreme events, average values	Chlorophyll a minimum extreme events, minimun values	
16				Chlorophyll a minimum extreme events, minimun values		

APPENDIX 2.11. Modelling performance scores of final models

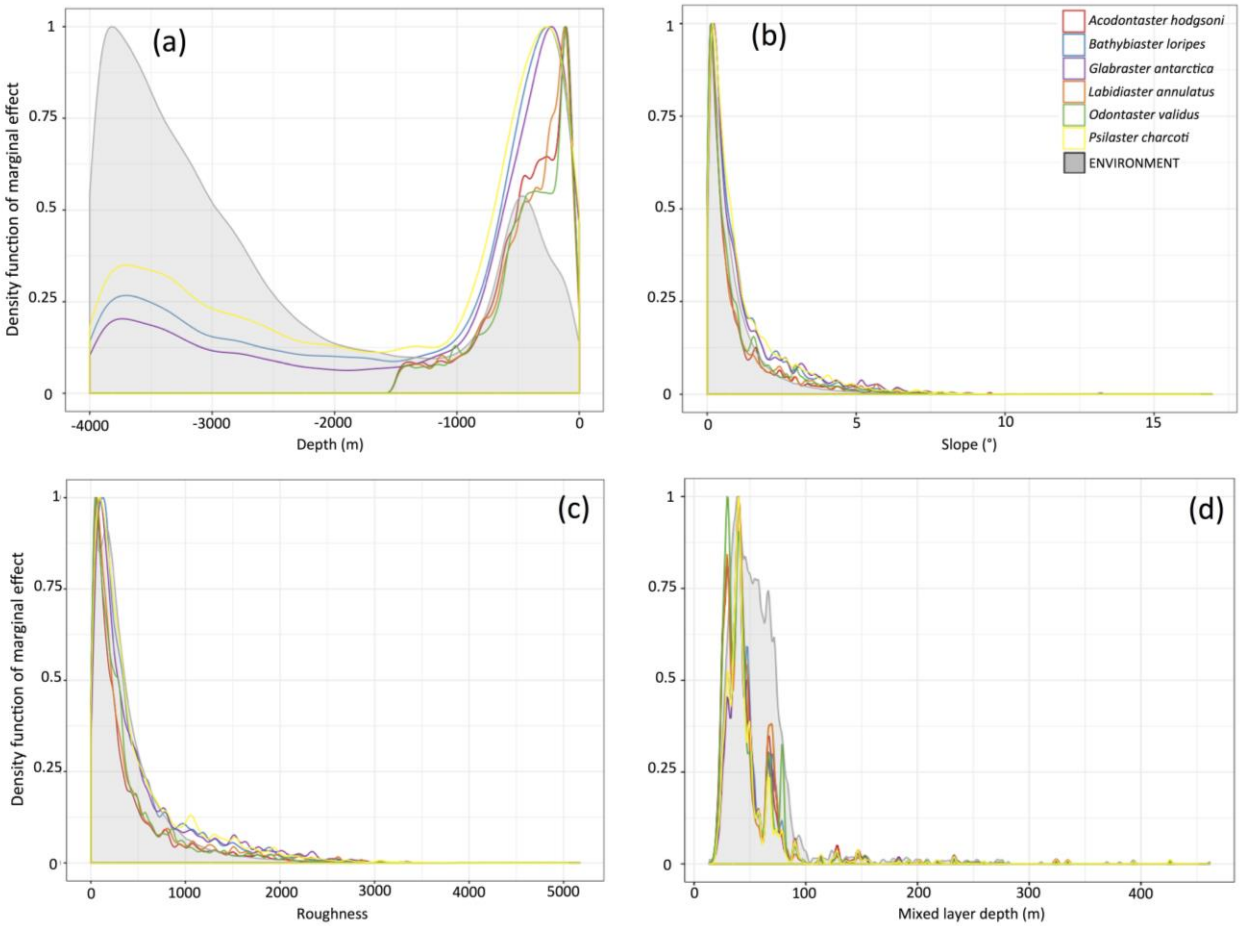
Table S2.11. Statistics (mean and standard deviation) measured for each species of models generated with the final set of environmental descriptors (Table S2.10). AUC: Area Under the Curve, COR: biserial Pearson correlation, TSS: True Skill Statistics. The percentage of correctly classified test data is defined by the proportion of test data that falls into pixels predicted as suitable (probability > maxSSS score).

	Acodontaster hodgsoni	*Bathybiaster loripes*	*Glabraster antarctica*	*Labidiaster annulatus*	*Odontaster validus*	*Psilaster charcoti*
AUC	0.810±0.09	0.871±0.07	0.872±0.07	0.837±0.117	0.830±0.09	0.868±0.05
TSS	0.461±0.121	0.546±0.08	0.545±0.09	0.492±0.146	0.489±0.120	0.543±0.06
COR	0.503±0.136	0.632±0.137	0.656±0.121	0.566±0.240	0.561±0.168	0.545±0.100
Correctly classified test data (%)	55.4±11.3	76.0±10.8	80.7±10.1	59.0±17.5	51.6±23.7	78.3±9.3

APPENDIX 2.12. Marginal effect of environmental descriptors

Partial dependence plots indicate the effect of an environmental descriptor on the model response after accounting for the average effects of all other descriptors in the model ("marginal effect", Elith et al. 2008). Results show weak contrasts between species and environmental descriptors.

Species preferences for slope, roughness, mixed layer depth, sea surface current speed are a consequence of the environmental preponderance of such conditions in the Southern Ocean environments. However, for other descriptors such as depth, seafloor current speed, average ice coverage, chlorophyll-a and POC concentrations species predicted preferences differ from what dominates in the environment (Fig. S2.12). This may be biased by sampling effort as it is exemplified by species distribution probabilities predicted in shallow areas (Fig. S2.12a) or areas with intermediate average ice cover values (Fig. S2.12g) or areas with intermediate. POC minimal concentrations preferences are contrasting between species. *A. hodgsoni*, *L. annulatus* and *O. validus* have preferences for high POC concentrations in comparison with *P. charcoti* and *G. antarctica* that prefer areas with low concentrations (Fig. S2.12k).



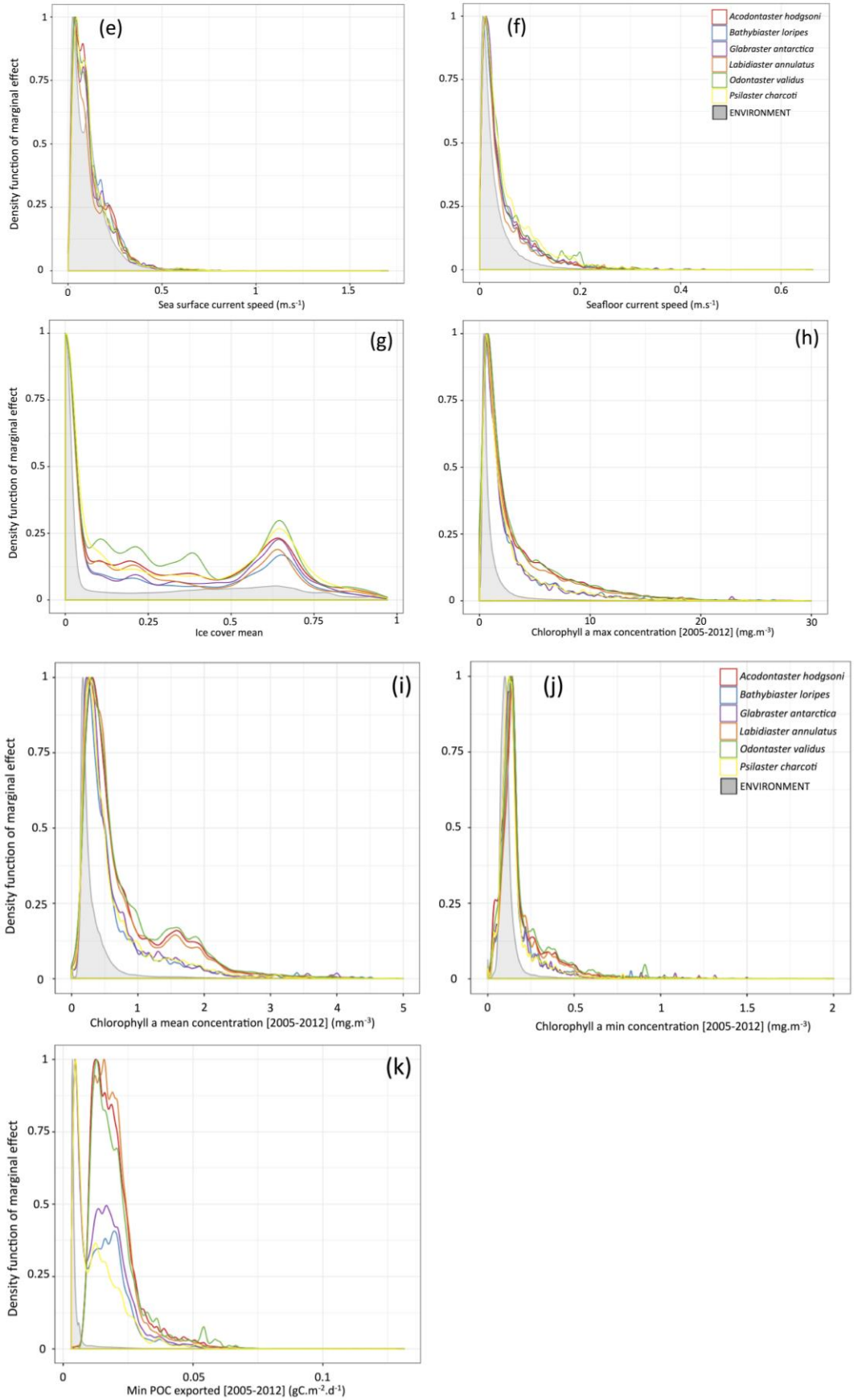


Figure S2.12. Partial dependence plots. Scaled density distributions of the marginal effect of environmental descriptors used to generate final models (Table S2.10) common to all species. Environmental values recorded in the entire Southern Ocean (<45°S, maximal m depth) are indicated in grey.

Extrapolation in species distribution modelling. Application to Southern Ocean marine species

Guillaumot Charlène[1,2], Moreau Camille[1,2], Danis Bruno[1], Saucède Thomas[2]

[1] Université Libre de Bruxelles, Marine Biology Lab. Avenue F.D. Roosevelt, 50. CP 160/15 1050 Bruxelles, Belgique
[3] UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, 6 bd Gabriel F-21000 Dijon, France

Progress in Oceanography, 188 (2020) 102438, Accepted September 9th, 2020

Abstract

Species distribution modelling (SDM) has been increasingly applied to Southern Ocean case studies over the past decades, to map the distribution of species and highlight environmental settings driving species distribution. Predictive models have been commonly used for conservation purposes and supporting the delineation of marine protected areas, but model predictions are rarely associated with extrapolation uncertainty maps.

In this study, we used the Multivariate Environmental Similarity Surface (MESS) index to quantify model uncertainty associated to extrapolation. Considering the reference dataset of environmental conditions for which species presence-only records are modelled, extrapolation corresponds to the part of the projection area for which one environmental value at least falls outside of the reference dataset.

Six abundant and common sea star species of marine benthic communities of the Southern Ocean were used as case studies. Results show that up to 78% of the projection area is extrapolation, i.e. beyond conditions used for model calibration. Restricting the projection space by the known species ecological requirements (e.g. maximal depth, upper temperature tolerance) and increasing the size of presence datasets were proved efficient to reduce the proportion of extrapolation areas. We estimate that multiplying sampling effort by 2 or 3-fold should help reduce the proportion of extrapolation areas down to 10% in the six studied species.

Considering the unexpectedly high levels of extrapolation uncertainty measured in SDM predictions, we strongly recommend that studies report information related to the level of extrapolation. Waiting for improved datasets, adapting modelling methods and providing such uncertainy information in distribution modelling studies are a necessity to accurately interpret model outputs and their reliability.

Key-words

Multivariate Environmental Similarity Surface (MESS), marine species, Antarctic, modelling relevance, conservation issues

ACKNOWLEDGEMENTS

This work was supported by a “Fonds pour la formation à la Recherche dans l’Industrie et l’Agriculture” (FRIA) and “Bourse Fondation de la Mer” grants to C. Guillaumot.

This is contribution no. 46 to the vERSO project (www.versoproject.be), funded by the Belgian Science Policy Office (BELSPO, contract n°BR/132/A1/vERSO). Research was also financed by the “Refugia and Ecosystem Tolerance in the Southern Ocean” project (RECTO; BR/154/A1/RECTO) funded by the Belgian Science Policy Office (BELSPO), this study being contribution number 23.

1. INTRODUCTION

Among the broad array of analytical tools developed for marine ecology studies over the last two decades, Species Distribution Modelling (SDM) has been increasingly used (Peterson 2001, Elith et al. 2006, Austin 2007, Gobeyn et al. 2019) and applied to Southern Ocean pelagic (Pinkerton et al. 2010, Freer et al. 2019), benthic organisms (Loots et al. 2007, Pierrat et al. 2012, Basher and Costello 2016, Xavier et al. 2015, Gallego et al. 2017, Guillaumot et al. 2018a - Appendix, 2018b, Fabri-Ruiz et al. 2019, Jerosch et al. 2019) and even marine mammals (Nachtsheim et al. 2017). SDM represents a complementary approach to individual-based modelling and eco-physiological experiments, quickly and synthetically identifying environmental correlates of species distribution (Brotons et al. 2012, Feng and Papes 2017, Feng et al. 2020). SDM is also used to define species distribution spatial range (Nori et al. 2011, Walsh and Hudiburg 2018) and can be used as decision criteria for conservation purposes (Guisan et al. 2013, Marshall et al. 2014). For instance, it is currently used in proposals developed by national committees of the CCAMLR (Commission for the Conservation of Antarctic Marine Living Resources) to support the definition and delineation of marine protected areas (Ballard et al. 2012, Arthur et al. 2018).

Applying SDM to Southern Ocean case studies is particularly challenging due to major constraints and biases that may reduce modelling performance. As for many oceanographic studies, access to environmental data with high temporal and spatial resolutions is difficult (Davies et al. 2008, Robinson et al. 2011). Antarctic coastal areas, in particular, are rarely accessed and documented due to logistical constraints, access being for example impossible during the austral winter due to sea ice cover (De Broyer et al. 2014). The availability of species absence records is also a limiting factor to modelling performances and model calibrations (Brotons et al. 2004, Wisz and Guisan 2009). Models are usually based on a limited number of presence-only records and limited number of sampling sites, which are both spatially aggregated in the vicinity of scientific stations, where access is frequent and datasets from different seasons, have been compiled over decades and even beyond (De Broyer et al. 2014, Guillaumot et al. 2018a - Appendix, Fabri-Ruiz et al. 2019, Guillaumot et al. 2019 - Chapter 2).

When generating a SDM, the model is fit to data with a given range of value for each environmental descriptor (i.e. the calibration range). When transferring model predictions, a portion of the environment may cover additionnal conditions that are outside this calibration range: these are non-analog conditions and the model extrapolates (Randin et al. 2006, Williams and Jackson 2007, Williams et al. 2007, Fitzpatrick and Hargrove 2009, Owens et al. 2013, Yates et al. 2018). Considering the limited number of species presence-only records occupied by each marine benthic species, and the poor quality and precision of environmental descriptors available for modelling Southern Ocean species distributions (Guillaumot et al. 2018a - Appendix, Fabri-Ruiz et al. 2019), a large proportion of cells might be expected to be extrapolations beyond the calibration range of the model.

The Multivariate Environmental Similarity Surface (MESS) approach analyses spatial extrapolation by extracting environmental values covered by presence-only records and estimates areas where environmental conditions are outside the range of conditions contained in the calibration area (Elith et al. 2010). The method considers that extrapolation occurs when at least one environmental descriptor value is outside the range of the environment envelop for model calibration (more details given in Appendix 2.16).
The MESS approach was initially used to determine the environmental barriers to the invasion of the cane toad in Australia, when facing new environments and under future conditions (Elith et al. 2010). Implemented in MaxEnt (Elith et al. 2011), MESS was subsequently used by several authors for defining the climatic limits to the colonisation of new environments by non-native species, such as the American bullfrog in Argentina (Nori et al. 2011), for studying contrasts between native and potential ecological niches like in the study of the spotted knapweed (*Centaurea stoebe*) (Broennimann et al. 2014), or for defining the limits to model transferability and predicting the distribution of trees under future environmental conditions (Walsh and Hudiburg 2018).

More recently, the MESS approach was used to define model uncertainties related to extrapolation (Escobar et al. 2015, Li et al. 2015, Cardador et al. 2016, Luizza et al. 2016, Iannella et al. 2017, Milanesi et al. 2017, Silva et al. 2019) and extrapolation areas where environmental conditions are non-analog to conditions of model calibration (Fitzpatrick and Hargrove 2009, Anderson 2013). Associating uncertainty information to model predictions has been acknowledged as a necessity for reliable interpretations of model predictions (Grimm and Berger 2016, Yates et al. 2018). It is also a requirement for specifying the level of risk associated with predictions and evaluating whether uncertainty can be mitigated to improve model outcomes (Guisan et al. 2013).

This study addresses the importance of extrapolation and associated uncertainties in SDMs generated at broad spatial scale for Southern Ocean species: an analysis that is seldom performed although important to characterise model reliability. Using the case study of six abundant and common sea star species in marine benthic communities, objectives of this work are to evaluate the importance of extrapolation proportions in wide projection areas, and to provide some methodological clues to mitigate the effects of extrapolation and improve model accuracy.

2. MATERIAL AND METHODS

2.1.Studied species and environmental descriptors

The distribution of six sea star species (Asteroidea : Echinodermata) was studied (Table 2.8). The six species, *Acodontaster hodgsoni* (Bell, 1908), *Bathybiaster loripes* (Sladen, 1889), *Glabraster antarctica* (Smith, 1876), *Labidiaster annulatus* Sladen, 1889, *Odontaster validus* Koehler, 1906 and *Psilaster charcoti* (Koehler, 1906) are abundant and common in benthic communities in the Southern Ocean. The biology, ecology and distribution of these species have been extensively studied and are relatively well documented (McClintock et al. 2008a, Mah and Blake 2012, Lawrence 2013). Presence-only records were compiled from a recently updated database, thoroughly scrutinised with the World Register of Marine Species (WoRMS Editorial Board 2016), to delete potential discrepancies, update taxonomy and correct for georeferencing errors (Moreau et al. 2018).
Models were generated for the different species using 298-851 presence-only records, and projected at different depth ranges (Table 2.8). The distributions of these presence-only records are contrasting between species (Appendix 2.13), with *A. hodgsoni, B. loripes* and *G. antarctica* having an Antarctic and sub-Antarctic distribution, with an important number of data available for *B. loripes* and *G. antarctica* but less data for *A. hodgsoni* (respectively 591, 851 and 298 presence-only records). *Labidiaster annulatus* has a distribution mainly gathered in the sub-Antarctic region with few data available (375 presence-only records). *Odontaster validus* and *P. charcoti* are mainly present on the coasts of the Antarctic shelf.

Table 2.8. Sea star species investigated in the present study. The number of presence-only records available was summed up after removal of duplicates from each grid cell pixel. Image sources: Brueggeman 1998, BIOMAR ULB database (P. Pernet), proteker.net, B121 expedition (Q. Jossart).

	Acodontaster hodgsoni (Bell, 1908)	*Bathybiaster loripes* (Sladen, 1889)	*Glabraster antarctica* (Smith, 1876)	*Labidiaster annulatus* Sladen, 1889	*Odontaster validus* Koehler, 1906	*Psilaster charcoti* (Koehler, 1906)
						
Presence-only records number	298	591	851	375	337	353
Model maximum depth	1500 m	4000 m	4000 m	1500 m	1500 m	4000 m

Environmental descriptors were selected from the dataset provided at https://data.aad.gov.au/metadata/records/environmental_layers. These are oceanography raster layers that mostly describe the physical and geochemical environment south of 45°S with a 0.1° grid-cell resolution (approximately 11km wide in latitude). Among the 58 environmental descriptors provided, only those that fulfilled the analysis performed by Guillaumot et al. (2020b - Chapter 2) were selected: 'distance' layers and 'extreme' layers were not selected because the interpretation of their respective contributions to niche models is complex or weak and collinear descriptors were also discarded for a Variance Inflation Factor (VIF) > 10 (Naimi et al. 2014). A set of 14-16 species-specific layers that characterise temperature, salinity, food availability and habitat characteristics were therefore used for model calibration (Table S2.14).

2.2. Model calibration

Species Distribution Models (SDMs) were generated using the Boosted Regression Trees (BRT), a machine-learning approach that was already calibrated for Southern Ocean case studies (Guillaumot et al. 2018b, Guillaumot et al. 2019 - Chapter 2) and was proved efficient to provide accurate models with good transferability performance, that is good ability to project model in space and time (Elith et al. 2008, Reiss et al. 2011, Heikkinen et al. 2012, Guillaumot et al. 2019 - Chapter 2). In order to minimalize the effect of presence-only records aggregation on model predictions, background data were randomly sampled in the environment following the probabilities defined by a Kernel Density Estimation (KDE) (see Phillips et al. 2009 for general principles, Guillaumot et al. 2018a, 2018b and Fabri-Ruiz et al. 2019 for applications). The number of background records was selected equal to the number of presence-only records (Barbet-Massin et al. 2012). The KDE was established based on the aggregation of benthos sampling effort provided in the Biogeographic Atlas of the Southern Ocean (De Broyer et al. 2014, map available in supplementary material of Guillaumot et al. 2019 - Chapter 2). One hundred SDMs were generated and averaged for each species, with background data randomly sampled following the KDE for each replicate.
SDMs were calibrated and reliability tested using a spatial cross-validation procedure. For each species, several procedures were compared following Guillaumot et al. (2019 - Chapter 2). The studied area was randomly subdivided into 2 to 6 areas of similar surfaces (longitude-split spatial folds), with presence and background data selected from one to three areas for model training and from the remaining areas for model testing. The "6-fold CLOCK" cross-validation approach was selected for *B. loripes*, *G. antarctica*, *L. annulatus* and *O. validus* and the "2-fold CLOCK" procedure was selected for *A. hodgsoni* and *P. charcoti*, according to the best percentage of test data correctly classified (Appendix 2.15).

The Maximum sensitivity plus specificity threshold (MaxSSS), considered the most appropriate threshold for presence-only SDM (Liu et al. 2013) was used to binarize models into suitable (>MaxSSS value) and unsuitable areas (<MaxSSS value). This threshold was used to measure the proportion of test data correctly classified. Modelling performances were also assessed using the three following metrics: Area Under the Receiver Operating Curve (AUC, Fielding and Bell 1997), the Point Biserial Correlation between predicted and observed values (COR, Elith et al. 2006) and the True Skill Statistics (TSS, Allouche et al. 2006).

Two analyses were performed: in Analysis #0 ('no-depth limited'), SDMs were projected on the entire Southern Ocean surface (south of 45°S) and in Analysis #1 ('depth limited'), SDM projections and background samplings were restricted to areas limited by a maximum depth threshold defined for each species based on the available species presence-only records (Table 2.8).

2.3. MESS calculation

The MESS was measured using the *dismo* R package (Hijmans et al. 2017) and following the guidelines provided in Elith et al. (2010). Pixels for which at least one environmental descriptor has a value that is outside the range of environmental values defined by presence-only records (calibration range) were considered to be extrapolation (i.e when MESS gets negative values, Appendix 2.16). The proportion of extrapolation areas (i.e. the proportion of cells defined as extrapolations over the total projection area) was calculated and compared between species. On SDM projection maps, extrapolated pixels were displayed in black.
Environmental parameters responsible for extrapolation were estimated by modifying the code provided in Elith et al. (2010). Detailed R scripts are available at https://github.com/charleneguillaumot/THESIS. Methodological details are provided in Appendix 2.16.

2.4. Influence of the number and distribution of presence-only records on extrapolation

The proportion of extrapolation areas may vary with presence-only sampling effort. In order to study the influence of the number and distribution of these presence-only records on the proportion of extrapolation areas, two analyses were performed. First, several SDMs were generated with different numbers of presence-only records, following the chronological addition of new presence-only records through time, from 1980 to 2016. Second, SDMs generated with 10-100% (10% increments, so 10 subsets) of the entire presence-only dataset were compared. In this analysis, in contrast to the previous one, presence-only records are randomly sampled among the datasets available.
In these two analyses, SDMs were projected on the environmental space limited by the maximum depth defined for each species (Table 2.8), 100 model replicates were generated and averaged in each case and spatial autocorrelation (SAC) was estimated to assess the influence of presence-only records aggregation on modelling performances. The significance of SAC was tested using the Moran's I index computed on model residuals (Luoto et al. 2005, Crase et al. 2012).

The relationship between the number of presence-only records used in SDM and the relative proportion of extrapolation areas was characterised using linear regressions. This allowed, for each model, estimation of the minimum number of presence-only records required to obtain a 'reasonable' proportion of extrapolation area arbitrarily set to a 10% threshold.

3. RESULTS

3.1. Extrapolation and the extent of projection areas

All generated SDMs are accurate and performant, with high AUC (AUC>0.91), TSS (TSS>0.559) and COR (COR>0.68) values, low standard deviations and good percentages of correctly classified presence-only test data (77-90 %) (Table 2.9). Descriptors that contribute the most to SDMs are depth (22-34%), minimum POC (6-21%), POC standard deviation (8-20%), mean ice cover depth (7-17%) and mixed layer depth (3-10%). Contrasts between species are in the

respective percentage of contribution of these descriptors. Descriptors that drive the most species distribution are similar between species (Appendix 2.17).

Models projected on the entire Southern Ocean (Analysis #0, 'no-depth limited') extrapolate on an area covering between 15-78% of the entire projection area, and 19-45% of the area initially predicted as suitable to the species distribution (Table 2.9, Fig. 2.16). Extrapolation areas cover more than 50% of the projection area for *A. hodgsoni* (78.6%), *P. charcoti* (67.8%), *L. annulatus* (64.8%) and *O. validus* (51.9%) and more than 30% of suitable areas (Table 2.9). For these four species, depth is responsible for 25-68% of extrapolation (Appendix 2.17). Geomorphology, mean ice cover and POC standard deviation are layers also contributing to 2-7% for extrapolation (Appendix 2.17). These descriptors that highly contribute to MESS also contribute to the model, and there are no descriptors for which the contribution to MESS is important whereas the contribution to the model is not substantial (Appendix 2.17).

In models projected on areas restrained in depth (Analysis #1, 'depth limited'), the percentage of extrapolation area sharply decreases from 59 to 18% according to the species (Table 2.9). However, model performances also decrease, with AUC values going down to 0.885, TSS values to 0.419 and COR values to 0.475. The percentage of correctly classified test data is much lower and more variable for the shallowest species *A. hogdsoni* (from 90 ± 6.26% to 45.5 ± 8.1%), *L. annulatus* (77.7 ± 15.2 % to 57.98 ± 20%) and *O. validus* (from 85.4 ± 9.6% to 57.68 ± 21%). For all species, predicted suitable areas increase two-fold.
Overall, descriptor contributions to the model remain unchanged between the two analyses, except for depth contribution that decreases to around 10% on average for all the species. In contrast, in Analysis #1, depth contribution to the MESS is very low (0.64-5.8%), except for *P. charcoti* (16.3%). Mean ice cover is the layer that contributes the most to extrapolation, extrapolation areas mainly corresponding to Weddell and Amundsen seas.

Table 2.9. Modelling performances for each species. Average and standard deviation values of the 100 model replicates. Pres. NB: number of presences-only records available for modelling (duplicates excluded); AUC: Area Under the Curve; TSS: True Skill Statistics; COR: Biserial Correlation.

Analysis #0, no-depth limited								
Species	Pres. NB	AUC	TSS	COR	Correctly classified test data (%)	Suitable area (% total area)	Extrapolation area (% total area)	Extrap. area (% suitable area)
Acodontaster hodgsoni	298	0.925 ± 0.02	0.579 ± 0.04	0.735 ± 0.06	90 ± 6.26	8.86	78.6	35.3 ± 4.1
Bathybiaster loripes	591	0.910 ± 0.02	0.559 ± 0.07	0.68 ± 0.09	80.6 ± 10.9	8.55	29.1	21.9 ± 4.4
Glabraster antarctica	851	0.929 ± 0.01	0.58 ± 0.05	0.719 ± 0.07	85.45 ± 6.34	7.95	15.73	19.9 ± 3.9
Labidiaster annulatus	375	0.95 ± 0.03	0.598 ± 0.07	0.730 ± 0.14	77.7 ± 15.2	3.33	64.83	42.1 ± 10.5
Odontaster validus	337	0.953 ± 0.01	0.605 ± 0.05	0.746 ± 0.09	85.4 ± 9.6	6.89	51.9	45.2 ± 5.65
Psilaster charcoti	353	0.911 ± 0.02	0.58 ± 0.03	0.723 ± 0.04	87.7 ± 4.8	8.90	67.9	32.5 ± 4.71

Analysis #1, depth limited								
Species	Pres. NB	AUC	TSS	COR	Correctly classified test data (%)	Suitable area (% total area)	Extrapolation area (% total area)	Extrap. area (% suitable area)
Acodontaster hodgsoni	298	0.823 ± 0.05	0.419 ± 0.1	0.475 ± 0.14	45.5 ± 18.1	17.49	40.6	27.5 ± 8.5
Bathybiaster loripes	591	0.887 ± 0.03	0.513 ± 0.08	0.607 ± 0.12	78.4 ± 11	15.75	18.2	20.8 ± 4.8
Glabraster antarctica	851	0.915 ± 0.01	0.537 ± 0.08	0.654 ± 0.1	81.8 ± 7.7	14.08	23.9	18.64 ± 3.5
Labidiaster annulatus	375	0.918 ± 0.03	0.482 ± 0.16	0.563 ± 0.25	57.98 ± 20	8.88	59.5	38.7 ± 14.6
Odontaster validus	337	0.908 ± 0.03	0.504 ± 0.13	0.586 ± 0.17	57.68 ± 21	11.64	51.5	38.3 ± 6.97
Psilaster charcoti	353	0.885 ± 0.02	0.546 ± 0.04	0.665 ± 0.06	83 ± 6.6	15.40	35.78	33.2 ± 5.1

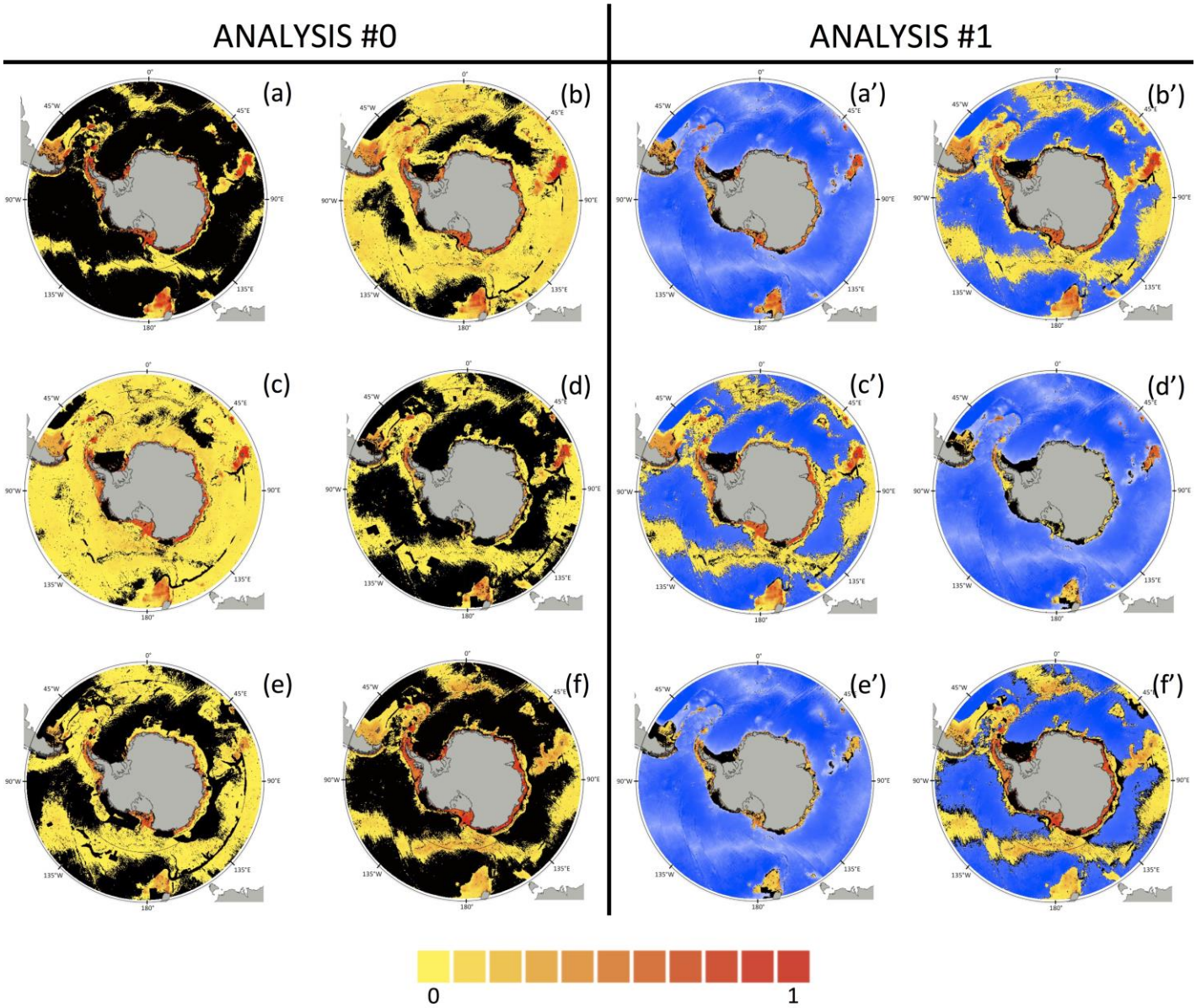


Figure 2.16. Maps of extrapolation areas covering SDM predictions, generated with all presence-only records available for the studied species. Left panel: projection area not limited in depth (Analysis #0), right panel: projection area limited to -1,500 m and -4,000 m depth (Analysis #1), according to the species (*A. hodgsoni*, *L. annulatus*, *O. validus* until 1,500 m; *B. loripes*, *G. antarctica*, *P. charcoti* until 4,000 m; Table 2.8). (a) *Acodontaster hodgsoni*, (b) *Bathybiaster loripes*, (c) *Glabraster antarctica*, (d) *Labidiaster annulatus*, (e) *Odontaster validus*, (f) *Psilaster charcoti*. Extrapolation areas displayed in black; pixels colored by the yellow-red color palette provide SDM distribution probabilities (comprised between 0 and 1); bathymetric chart in shades of blue.

3.2. Extrapolation and the number of presence-only records

Model performance and size of extrapolation area were compared between models run with different numbers of presence-only records, following the chronological addition of new samples (from 1980 to 2016). From 1980 to 2016, the number of presence-only records collected during oceanographic campaigns has increased from 1.9 to 3.3 times according to the species (1.9 times for *O. validus*, 3.3 times for *A. hodgsoni*) (Fig. 2.17A). Spatial autocorrelation between presence-only records varies between species, with the highest Moran's I scores obtained for *L. annulatus, O. validus* and *A. hodgsoni*. The highest Moran's I values were mainly calculated for the oldest presence-only subsets (1980), strenghtening the fact that the addition of new presence-only records with additional campaigns reduces spatial autocorrelation (Table S2.18).

Model performance increases (higher AUC scores) with the addition of new presence-only records, for all species except for models of *A. hodgsoni* and *B. loripes* for which AUC values are stable (Table S2.18). Similarly, the percentage of correctly classified test data presents important standard deviation values and improves with the addition of new presence-only records, except for *O. validus* (10% decrease) (Fig. 2.17B).

For all species, the addition of new data reduces the percentage of extrapolation over the total projection area (-30.7% for *A. hodgsoni*, -12.7% for *B. loripes*, -20.5% for *G. antarctica*, -17.6% for *L. annulatus*, -10.2% for *O. validus* and -11% for *P. charcoti,* i.e. differences between the two extrapolation % values) and over the species suitable area as well (Fig. 2.17C, Table S2.18).

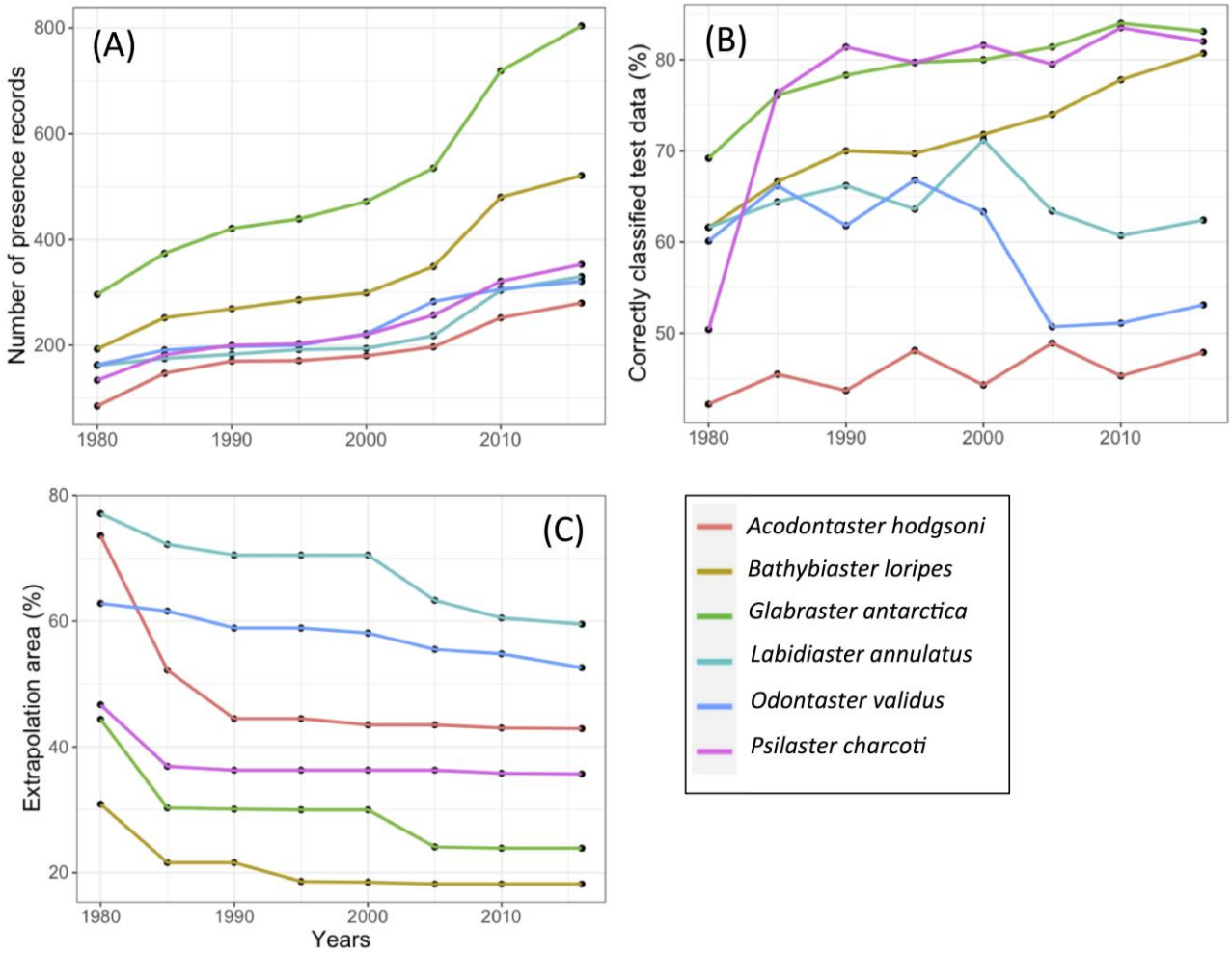


Figure 2.17. Evolution of model performances with the increase of data (chronological addition of presence-only records, by 5-year periods, from 1980 to 2016). (A) Number of presence-only records available to generate the model; (B) Mean correctly classified test data (%) (standard deviation values available in Table S2.18); (C) Proportion of grid-cell pixels of the projection area that are extrapolations (%). The maximal number of presence-only records present in Table 2.9 may not be reached here because some collection dates remain unknown.

The decrease of extrapolation with the addition of presence-only records was tested by running, for each species a series of models with different subsets of presence-only records randomly sampled from the total dataset. One hundred model replicates were progressively run with 10 to 100% of the total dataset and proportions of extrapolation areas were computed accordingly (Fig. 2.18, Table S2.19). Results confirm that the addition of presence-only records strongly reduces proportions of extrapolation areas. Proportions of extrapolation areas also vary between species models as a function of depth. Low proportions of extrapolation areas are obtained in models run for deep species and large datasets (e.g. 8.2% for 591 records in *B. loripes* and 23.9% for 851 records in *G. antarctica*). In contrast, models run for shallower species show higher proportions of

extrapolation areas (40.6% for 298 records in *A. hodgsoni,* 51.5% for 375 records in *L. annulatus* and 35.8% for 337 records in *O. validus*). For these last species, spatial autocorrelation values are also higher compared to other species (Table S2.19).

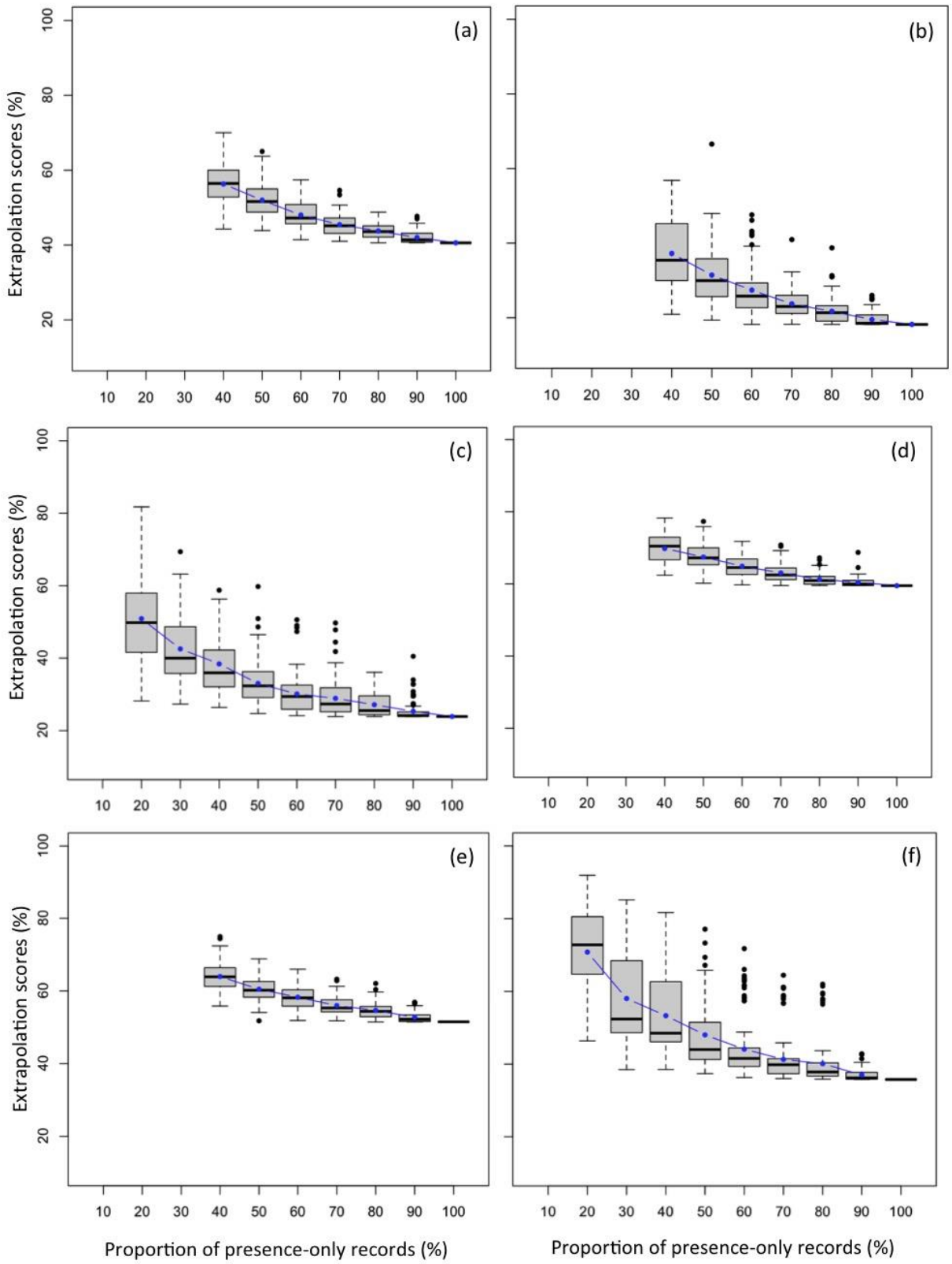


Figure 2.18. Boxplot diagrams representing the decrease of proportions of extrapolation areas (in % of the total projection area) with addition of presence-only records used to generate model replicates (in % of data available, see Table 2.8 and Table S2.19), for: (a) *Acodontaster hodgsoni*, (b) *Bathybiaster loripes*, (c) *Glabraster antarctica*, (d) *Labidiaster annulatus*, (e) *Odontaster validus*, (f) *Psilaster charcoti*. For each box, mean values (blue dots) and outliers (black dots) are shown for the 100 model replicates. Some boxes are

missing for low percentages of presence-only records (10-30%, corresponding to close or less than 100 presence-only records) that do not allow models to be generated.

A linear regression model was fit to the relationship between the number of presence-only records and proportions of extrapolation areas. For all species, regression coefficients are all negative and tested significant showing that proportions of extrapolation areas decrease with the addition of new records (Table 2.10). The intersection point between regression models and the (arbitrary) 10% extrapolation threshold was used to provide an estimate of the minimum number of records required for each species model to have an "adequate" proportion of extrapolation areas of 10%. This minimum number of presence-only records is reached for none of the studied species, and according to species, the number of presence-only records available should be increased at least by 1.6 to 3.3 times (Table 2.10).

Table 2.10. Equations of simple linear regressions between the number of presence-only records *X* and the average proportion of extrapolation areas *Y* (Table 2.9, significance levels: * p<0.1, ** p<0.05). The estimate of the number of presence-only records necessary to have a minimum "adequate" arbitrary proportion of extrapolation areas of 10% is given in the last column.

Species	Equation	R^2	Estimated Pres.NB. (with multiplier of actual Pres.NB. available)
Acodontaster hodgsoni	Y=-0.1358X + 73.616**	0.60	468 (x 1.6)
Bathybiaster loripes	Y=-0.0249X + 28.974*	0.42	762 (x 1.3)
Glabraster antarctica	Y=-0.0304X + 44.991**	0.61	1151 (x 1.4)
Labidiaster annulatus	Y=-0.0913X + 88.078**	0.85	855 (x 2.3)
Odontaster validus	Y=-0.0561X + 71.112**	0.93	1089 (x 3.2)
Psilaster charcoti	Y=-0.0301X + 44.613*	0.37	1150 (x 3.3)

4. DISCUSSION

4.1. Modelling performances and extrapolation

SDMs were generated for Southern Ocean sea star species, with contrasting distributions and different numbers of presence-only records available (Table 2.8, Appendix 2.13). Overall, species presence-only records are spatially concentrated in the most accessible and visited areas of the Southern Ocean. Most of the sea star samples were collected close to the coasts of the Western Antarctic Peninsula, the Ross Sea and sub-Antarctic Islands such as the Kerguelen Islands. Consequently, high spatial autocorrelation values were computed, for *L. annulatus* and *O. validus* in particular (Table S2.18).

Overall, models all show good performances (Table 2.9), the spatial cross-validation procedure ensuring a relevant evaluation of modelling performances when using spatially aggregated data (Muscarella et al. 2014, Dhingra et al. 2016, Guillaumot et al. 2019 - Chapter 2). However, models show high proportions of extrapolation areas, with extrapolation covering up to 78% of the projection area in *A. hodgsoni* model (Table 2.9). This means that even if models are evaluated as accurate, model extrapolation area can concern up to three quarters of the projection area! Assessing the proportion of the projection area for which models extrapolate is therefore necessary as a complementary statistic to adapt modelling methods and improve model predictions. Masking projections by extrapolation uncertainties is also important to perform accurate interpretations.

Extrapolation uncertainty maps have already been associated to SDM projections once in the context of the Southern Ocean, by Torres et al. (2015) in their study of the grey petrel *Procellaria cinerea*, performed at the scale of the Southern Ocean. More recently, the MESS approach has been introduced in the methodological paper of Guillaumot et al. (2019 - Chapter 2), showing an extrapolation area covering 64% of the projection area for the distribution model of the sea star *O. validus,* the most studied benthic invertebrate of the Southern Ocean. However, uncertainties associated to extrapolation were not provided in most model projections performed for Southern Ocean species studies. For instance, modelled distributions performed for the sea urchins *Sterechinus neumayeri* and *Sterechinus diadema* (Pierrat et al. 2012) were generated using a relative low number of presence-only records (241 and 332, respectively). Based on results of the present study, extrapolation could be expected to cover up to 60% of modelled distribution areas for these last two species. Further Southern Ocean species distribution models were generated with sometimes less than 100 presence-only records (see Guillaumot et al. 2018b and Fabri-Ruiz et al. 2019 for instance), suggesting that extrapolation could cover up to 70% of projection areas as visible in models of *A. hodgsoni* and *P. charcoti* performed in our study with few records (Fig. 2.17, Table S2.18-19).

In addition to model uncertainties associated to extrapolation, other biases can alter the performance of SDMs generated at broad spatial scales including the spatial and temporal aggregation of data (Hortal et al. 2008, Tessarolo et al. 2014, 2017), the selection and quality of environmental descriptors (Davies et al. 2008, Synes and Osborne 2011), the choice of modelling algorithms and the definition of model settings (Hartley et al. 2006, Marmion et al. 2009). Providing such uncertainty information, highlighted with some model statistics is very much encouraged here, as they are essential to model interpretation (Beale and Lennon 2012, Guisan et al. 2013, Yates et al. 2018).

4.2. How can we reduce model extrapolation? Enriching SDMs with knowledge of species ecology

One objective of this work was to provide some methods to mitigate the effect of extrapolation on model uncertainties. Our results show clear contrasts between models generated for “deep” and “shallow” species, with lower proportions of extrapolation areas computed for deep species models (29.1 and 15.73% respectively for *B. loripes* and *G. antarctica*). The model generated for *P. charcoti* departs from this general scheme, with extrapolation reaching 67.9% of the projection area. This is due to the strong spatial aggregation of records and the small presence-only record dataset available in deeper habitats. Depth is indeed responsible for 58.1% of the extrapolation for *P. charcoti* (Appendix 2.17). Indeed, the erroneous characterisation of species occupied space, due to an incomplete sampling, has been identified as a significant source of bias in SDM predictions (Hortal et al. 2007, 2008, Rocchini et al. 2011, Sánchez-Fernández et al. 2011, Titeux et al. 2017, El-Gabbas and Dormann 2018).

Limiting model projection areas to biogeographically, or ecologically “realistic” depth ranges can help reduce extrapolation as exemplified in the present study, for models of *A. hodgsoni* and *P. charcoti*, for which extrapolation was reduced from 78.6 to 40.6% and 67.9 to 35.8% respectively (Table 2.9). Restraining model projection areas based on species ecological or physiological tolerance thresholds is a common approach in ecological modelling using experimental data or field observations (Kearney and Porter 2009, Hare et al. 2012, De Villiers et al. 2013). Knowledge of species ecology and physiology can also be useful to delineate transferability areas (Feng and Papes 2017) and improve distribution models, as recently shown for Southern Ocean species (Guillaumot et al. 2018a, Guillaumot et al. 2019 - Chapter 2). Feng et al. (2020) developed a new modelling algorithm, called *Plateau*, which uses experimental data to define upper temperature conditions in distribution models. For temperature and salinity, physiological experiments and field observations can be used in models to determine species tolerance thresholds. This requires knowledge about the species ecology and physiology and the input from specialists, all conditions that remain difficult to meet, regarding deep-sea species of the Southern Ocean (Gage 2004, Gutt et al. 2010, De Broyer and Danis 2011). Moreover, several studies suggested that some Southern Ocean species might have found refuges in deep sea habitats in the past, during glacial maxima, which makes species depth range difficult to precise when deep and shallow populations have not

been differentiated into distinct taxonomic units yet (Rogers 2007, Arango et al. 2011, Havermans et al. 2011, Near et al. 2012).

4.3. How can we reduce model extrapolation? Improving sampling effort

Increased sampling effort over enlarged areas allows the production of larger datasets from which many records can be used to generate reliable models with reduced extrapolation areas. In this study, proportions of extrapolation areas proportionally decreased when increased numbers of presence-only records were used to generate models. The occurrence datasets were significantly augmented between 1980 and 2016, with a number of presence-only records multiplied by 1.9 -3.3 times according to the studied species, which allowed reduction of model extrapolation from 10.2 to 30.7% according to the species (Fig. 2.17, Table S2.18). However, results suggest that about twice the number of presence-only records actually available would be necessary to reduce extrapolation down to a "satisfactory" threshold of 10% of the projection area (Table 2.10).

Generating reliable and stable models using a sufficient number of presence-only records is essential. In this study, some models could not be run when the number of presence-only records was too low (approaching 150 presence-only records or less) compared to the broad extent of the projection area and the spatial aggregation of these data (Table S2.19). Considering that the spatial cross-validation procedure splits the initial dataset into training and test data, and that at each step, 75% of these training data are randomly sampled by BRT to iterately create a model tree (and generate stochasticity in the procedure), the final number of presence-only records available to describe the presence data - environment relationship becomes too low (around 37.5% of the initial number of presence-only records).
The lowest number of presence-only records required to build a reliable model is species-dependent as not all presence-only records are equally informative, due to species-specific relationships between records and the environment. When models are generated using BRT, records that bring no new environmental information to the model are dropped because they are not informative enough to improve the construction of BRT trees. Pruning non-informative data also reduces the total number of presence-only records available to generate a model (Elith et al. 2008). This is strongly related to prevalence that is, the ratio between the number of presence-only records and the size of the projection area (Jiménez-Valverde et al. 2009, Santika 2011, Barbet-Massin et al. 2012). In order to accurately describe a vast projection area and be able to create a model, it is necessary to gather a substantial amount of information about the geographic environmental conditions and about species known distribution. If a limited number of records is available and these data are aggregated in space (i.e. weakly informative), the first trees produced by BRT will contain most of the model deviance, but as no new information is provided, the model will quickly overfit because redundant information is provided by close presence-only records. Eventually, this will make the model collapse.
Increasing the number of presence-only records is proved an efficient alternative to generate more relevant models (Stockwell and Peterson 2002, Feeley and Silman 2011, van Proosdij et al. 2016), but the spatial distribution of these records is of importance as well (Yates et al. 2018). A uniform distribution of records over the entire projection area reduces spatial autocorrelation and optimizes the sampling and representativeness of environmental conditions under which species can thrive. In this study, the spatial aggregation of species records was particularly high for two species, *O. validus* and *L. annulatus*. It was estimated that the number of supplementary presence-only records necessary to reach a proportion of extrapolation areas of 10% should be twice as high as it is for other species (Table 2.10). Additional data are necessary to improve the establishment of the relationship between species distribution and the environment because species records are less informative when aggregated than when they are evenly distributed.

The Southern Ocean covers contrasting environmental conditions, biogeographic regions and ecoregions (Pierrat 2011, Fabri-Ruiz et al. 2020). Ideally, both species presence and absence should be recorded in each ecoregion for an accurate description of the occupied space (Torres et al. 2015). Because such a sampling effort is usually not achievable, nor realistic, alternatives would consist of (1) a relevant adjustment of projection areas, with for instance the combination of several SDM projections using different grid sizes according to what is available. Generating SDM

projections for large areas and combining results with projections zoomed in on areas where more environmental detail is available would provide more relevant and realistic modelled species distributions (Seo et al. 2009, Anderson and Raza 2010). (2) In order to compensate for the lack of presence-record availability, the 'ensembles of small models' approach is another alternative. This method fits a set of bivariate models (i.e. generated with two environmental descriptors only), within a hierarchic multi-scale framework (i.e. zooming in and out in space from local to regional predictions), and finally averages this ensemble of models with a weighted ensemble approach, which subsequently provides more accurate and robust model predictions (Lomba et al. 2010, Breiner et al. 2015, Habibzadeh and Ludwig 2019).

4.4. Some limitations to the MESS approach

The MESS approach can reveal parts of projection areas where models extrapolate. Extrapolation however can be over-estimated. Indeed, extrapolation is considered as soon as the value of a single environmental descriptor falls outside the range of the known species environmental requirements. But, some extreme values would not limit but can promote species presence: this is the case for descriptors relating to food resource availability (e.g. chlorophyll a, POC concentrations...), for which a high pixel value exceeding the range of values recorded based on species presences will be still considered as extrapolation, although more food usually means suitable conditions for species distribution.
Some fine-tuning of the MESS approach would imply to identify, for each pixel, which descriptor is responsible for extrapolation and filter the conditions for which the model should really extrapolate. Such an approach was developed by Owens et al. (2013), who used the MOP method (Mobility Oriented Parity). Based on multivariate analyses, they determined if pixels contain a combination of environmental conditions that should induce extrapolation. In contrast to the MESS approach, the MOP method can directly differentiate proportions of extrapolation areas according to the combination of descriptors responsible for extrapolation. Another complex alternative is the ExDet tool, developed by Mesgaran et al. (2014), which also accounts for multivariate extrapolation possibilities, i.e. extrapolation linked to novel combinations between covariates.
In this study, the MESS approach was favored as a more strict and conservative method to highlight the importance of extrapolation, the effect of data quantity and quality, and the relevance of the proposed corrections. The MESS is also simpler to apply and well suited to exploratory studies.

5. CONCLUSION

This study shows that when modelling species distribution on broad-scale areas, such as the Southern Ocean, important proportions of predicted distribution probabilities (suitable or not) are model extrapolations. This extrapolation uncertainty relies on the completeness of species sampling, and the definition of its occupied space to calibrate the model. Extrapolation occurs in areas where habitat suitability is unknown as no information on species presence or absence is provided.

Reducing extrapolation is possible by combining SDM with ecological and physiological knowledge of species requirements (e.g. depth range, temperature tolerance thresholds). Increased sampling effort over enlarged areas also allows the production of more reliable models with reduced extrapolation areas and our study shows that doubling the number of presence-only records available to generate the model would help reduce the extrapolation area down to 10% of the projected area.
While more data samples remain unavailable, some methods are increasingly developed to improve model performances, by adjusting the extent of the projection area or by generating and aggregating several ensembles of small models.

Finally, present results call for a widespread use of extrapolation maps and uncertainties associated to model predictions in model outputs, along with information about the quantity of presence-only records available, the quality and resolution of environmental descriptors and the state of our knowledge of species ecology. These are all essential information needed to support

model interpretations, as also stated in recent publications that review best practices in ecological modelling (Araújo et al. 2019, Zurell et al. 2020).

APPENDIX 2.13. Distribution of presence-only records

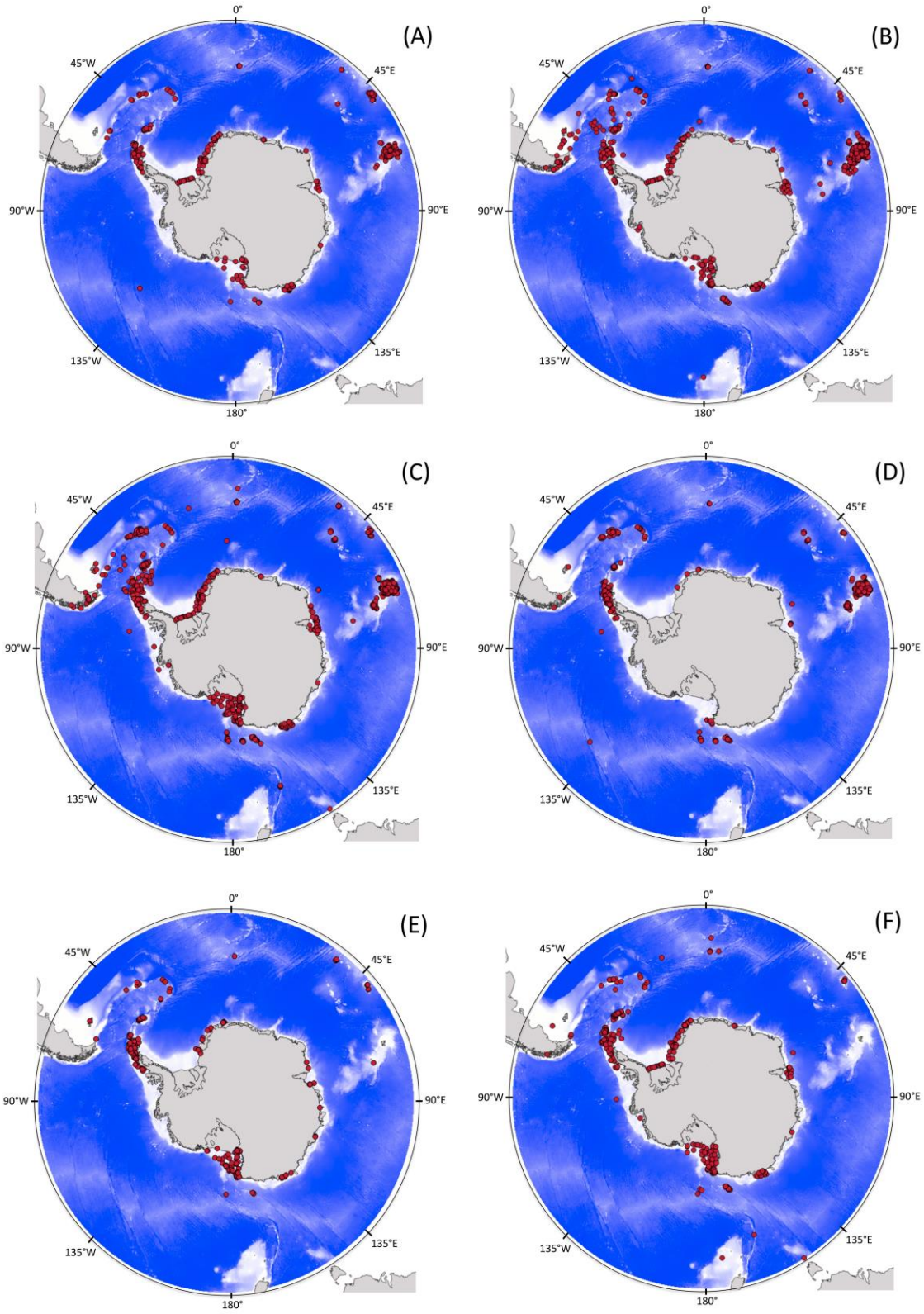


Figure S2.13. Distribution of presence-only records of the six sea star species studied in this work. (A) *Acodontaster hodgsoni*, (B) *Bathybiaster loripes*, (C) *Glabraster antarctica*, (D) *Labidiaster annulatus*, (E) *Odontaster validus*, (F) *Psilaster charcoti*. Presence-only record duplicates that fell on a same grid-cell pixel were removed from the analysis.

APPENDIX 2.14. List of environmental descriptors selected to generate the models

Table S2.14 List of species-specific environmental descriptors selected to generate final models after removal from the initial dataset of spatial distance descriptors, descriptors that always contribute less than 1% to SDMs (Guillaumot et al. 2020b - Chapter 2) and collinear descriptors. Extracted from the list of 58 layers available at https://data.aad.gov.au/metadata/records/environmental_layers (Guillaumot et al. 2020c).

Acodontaster hodgsoni	***Bathybiaster loripes***	***Glabraster antarctica***	***Labidiaster annulatus***	***Odontaster validus***	***Psilaster charcoti***
depth	depth	depth	depth	depth	depth
geomorphology	geomorphology	geomorphology	geomorphology	geomorphology	geomorphology
sediments	sediments	sediments	sediments	sediments	sediments
slope	slope	slope	slope	slope	slope
roughness	roughness	roughness	roughness	roughness	roughness
mixed layer depth	mixed layer depth	mixed layer depth	mixed layer depth	mixed layer depth	mixed layer depth
seasurface current speed	seasurface current speed	seasurface current speed	seasurface current speed	seasurface current speed	seasurface current speed
seafloor current speed	seafloor current speed	seafloor current speed	seafloor current speed	seafloor current speed	seafloor current speed
ice cover mean	ice cover mean	ice cover mean	ice cover mean	ice cover mean	ice cover mean
chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]	chlorophyll a max concentration [2005-2012]
chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]	chlorophyll a mean concentration [2005-2012]
chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]	chlorophyll a min concentration [2005-2012]
POC amplitude [2005-2012]	POC minimum [2005-2012]	POC minimum [2005-2012]	POC minimum [2005-2012]	POC minimum [2005-2012]	POC minimum [2005-2012]
POC minimum [2005-2012]	POC standard deviation [2005-2012]	POC standard deviation [2005-2012]	POC standard deviation [2005-2012]	POC standard deviation [2005-2012]	POC standard deviation [2005-2012]
		Chlorophyll a minimum extreme events, minimum values	Chlorophyll a minimum extreme events, average values	Chlorophyll a minimum extreme events, minimum values	
			Chlorophyll a minimum extreme events, minimum values		

APPENDIX 2.15. Spatial cross-validation procedure

The cross-validation procedure consists in using a subset of the total dataset to train the model, and the remaining part is used to test model predictions. In doing so, training and test data are independent whenever generating the model, which improves the reliability of model evaluation (Hijmans 2012).

The selection of training and test subsets is often done randomly, and most of the time, 70% of presence records are randomly chosen to train the model and 30% to test it (Fabri-Ruiz et al. 2019). However, when presence-only records are aggregated in space, splitting data at random would bias model evaluation and will inflate model performances. Splitting training and test data following a defined spatial pattern was shown to improve the relevance of model evaluation, in a context of aggregated data (Muscarella et al. 2014, Dhingra et al. 2016, Roberts et al. 2017, Guillaumot et al. 2019 - Chapter 2). Several methods were assessed and compared in Guillaumot et al. (2019 - Chapter 2). Here, we tested and selected the '2-fold CLOCK' method for models performed for *A. hodgsoni* and *P. charcoti* (Fig. S2.15a) and the '6-fold CLOCK' method (Fig. S2.15b) for other SDMs, based on the best AUC scores and percentage of correctly classified test data.

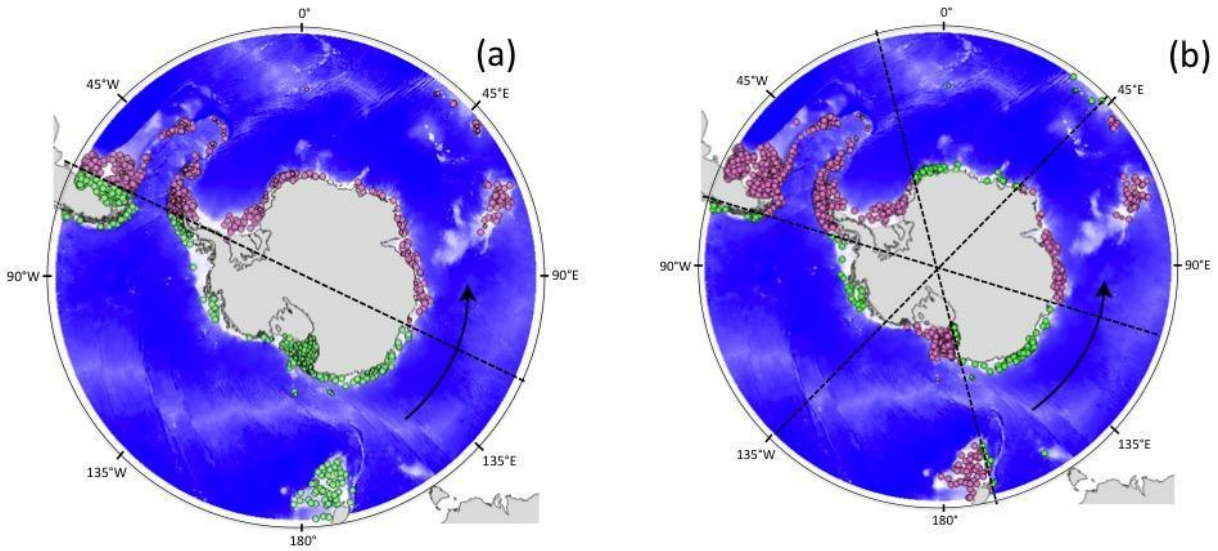


Figure S2.15. (a) '2-fold CLOCK' method and (b) '6-fold CLOCK' method. For each model replicate, the geographic space is split into 2 and 6 areas respectively, and test (green) and training (pink) presence and background data are selected in the defined areas. The model is built based on training data and model predictions are evaluated using presence test data. One hundred model replicates are generated and the average prediction calculated.

APPENDIX 2.16. Multivariate Environmental Similarity Surface principle

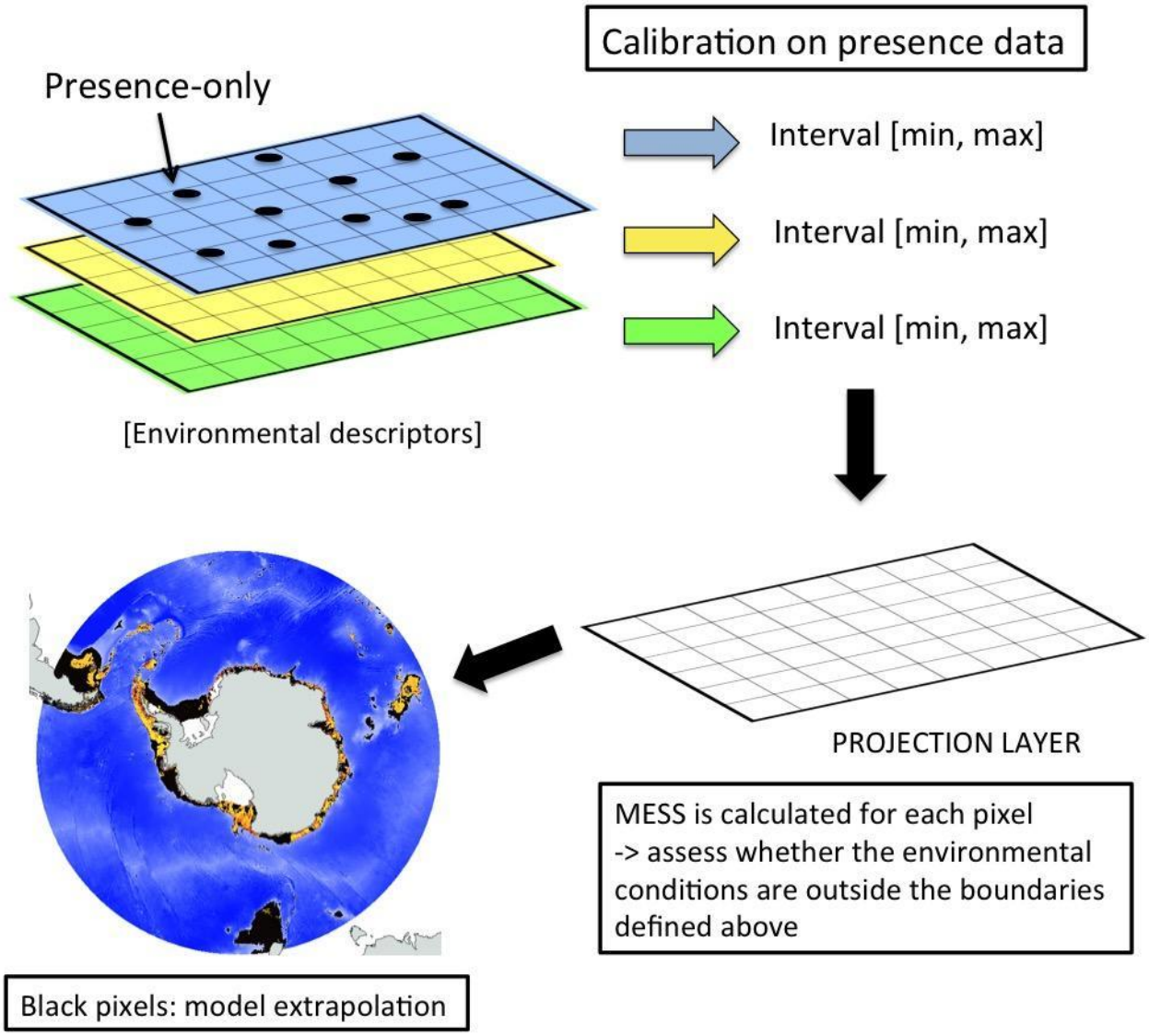


Figure S2.16. Illustrated principle of the Multivariate Environmental Similarity Surface approach.

In the 'dismo' R package (Hijmans et al. 2017), the *mess* function calculates the value of the MESS for each pixel of the rasterstack used for model projection (P). This rasterstack contains the environmental conditions into which the model is projected.
In the first step explained by Elith et al. (2010), the environmental conditions experienced by the presence data are extracted, and the minimal and maximal values define the boundaries of each descriptor (V_i).
1. Let min_i be the minimum value of descriptor V_i over the reference point set, and similarly for max_i.
Then, the environmental conditions of the projection layer P are extracted and compared to these minimal and maximal boundaries.
2. Let p_i be the value of descriptor V_i at pixel P_j.
3. Let f_i be the percent of reference points whose value of descriptor V_i is smaller than p_i.
4. Then the similarity of P_j with respect to descriptor V_i is:
$(p_i - min_i) / (max_i - min_i) * 100$ if $f_i = 0$
$2 * f_i$ if $0 < f_i \leq 50$
$2 * (100 - f_i)$ if $50 \leq f_i < 100$

$(max_i - p_i) / (max_i - min_i) * 100$ if $f_i = 100$
5. Finally, the multivariate similarity of P_j is the minimum of its similarity with respect to each descriptor.
This calculation is then applied to each pixel $P_{j...n}$

The final value of the MESS represents how similar pixel values of each descriptor (V1, V2, ...) are to the reference set of values defined by presence records. It allows negative values, and whenever the MESS is negative, it corresponds to the situation when at least one descriptor has a value that is outside the range of environments over the reference.
In this study, the MESS was estimated and all pixels for which the MESS value was negative were considered as extrapolation and colored in black.

The MESS calculation was also adapted for Elith et al. (2010) to be able to estimate for each pixel which descriptor is concerned with the extrapolation. The MESS was separately calculated for each layer of the rasterstack P. Whenever the MESS score calculated for pixel P_j was negative, it was considered that the model extrapolates at that specific pixel due to the specific layer studied. Results were compiled and the contribution of each descriptor to the extrapolation was assessed (Appendix 2.17).

APPENDIX 2.17. Descriptor contributions to the models and descriptors responsible for extrapolation

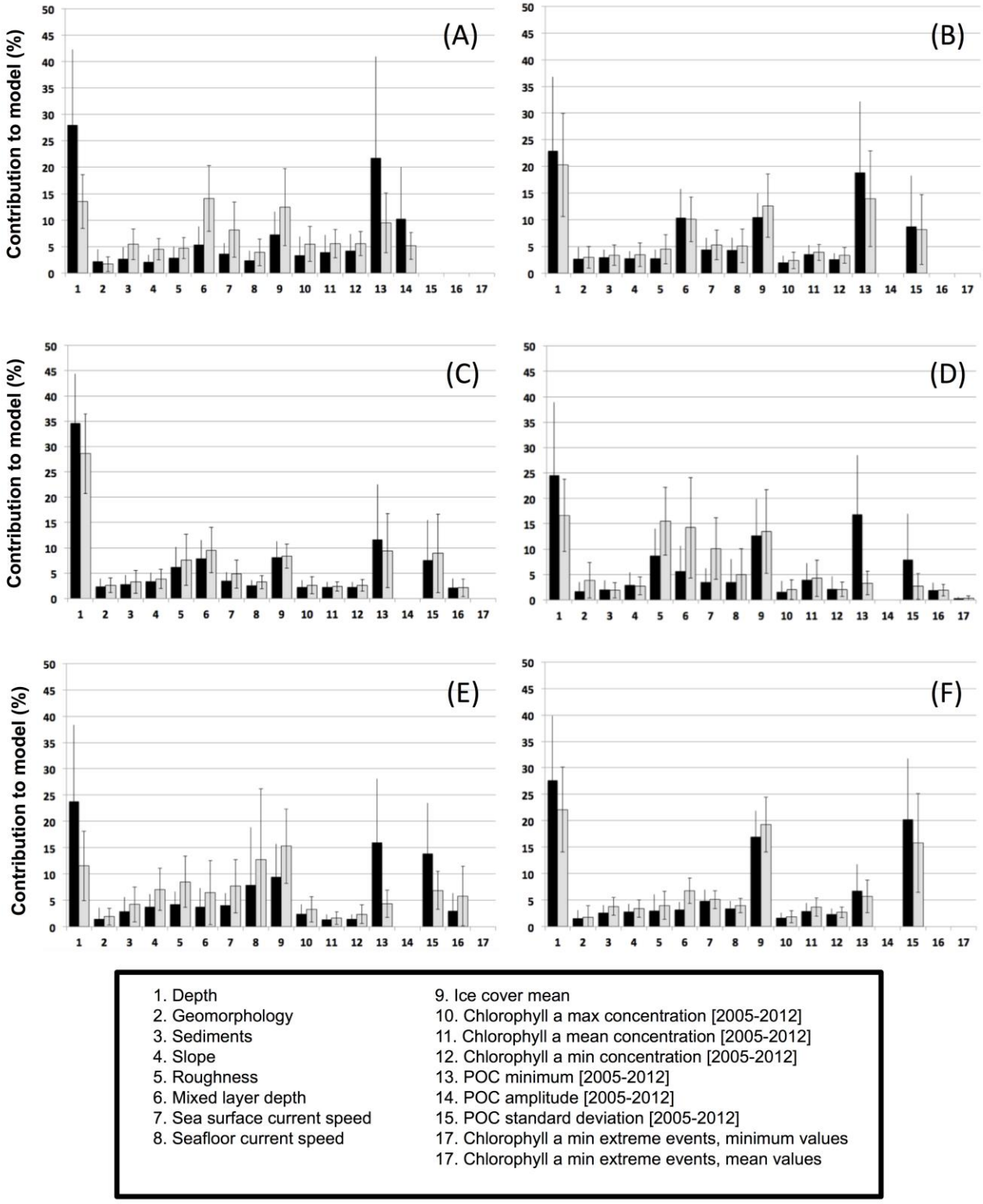


Figure S2.17.A. Influence of the different environmental descriptors on models (mean and standard deviation values calculated on the the 100 model replicates), for Analysis #0 (black bars) and Analysis #1 (grey bars). Analysis #0: models were projected on the entire Southern Ocean area. Analysis #1: the projection area was limited in depth according to each species distribution range (*A. hodgsoni*, *L. annulatus*, *O. validus* until 1,500 m; *B. loripes*, *G. antarctica*, *P. charcoti* until 4,000 m).

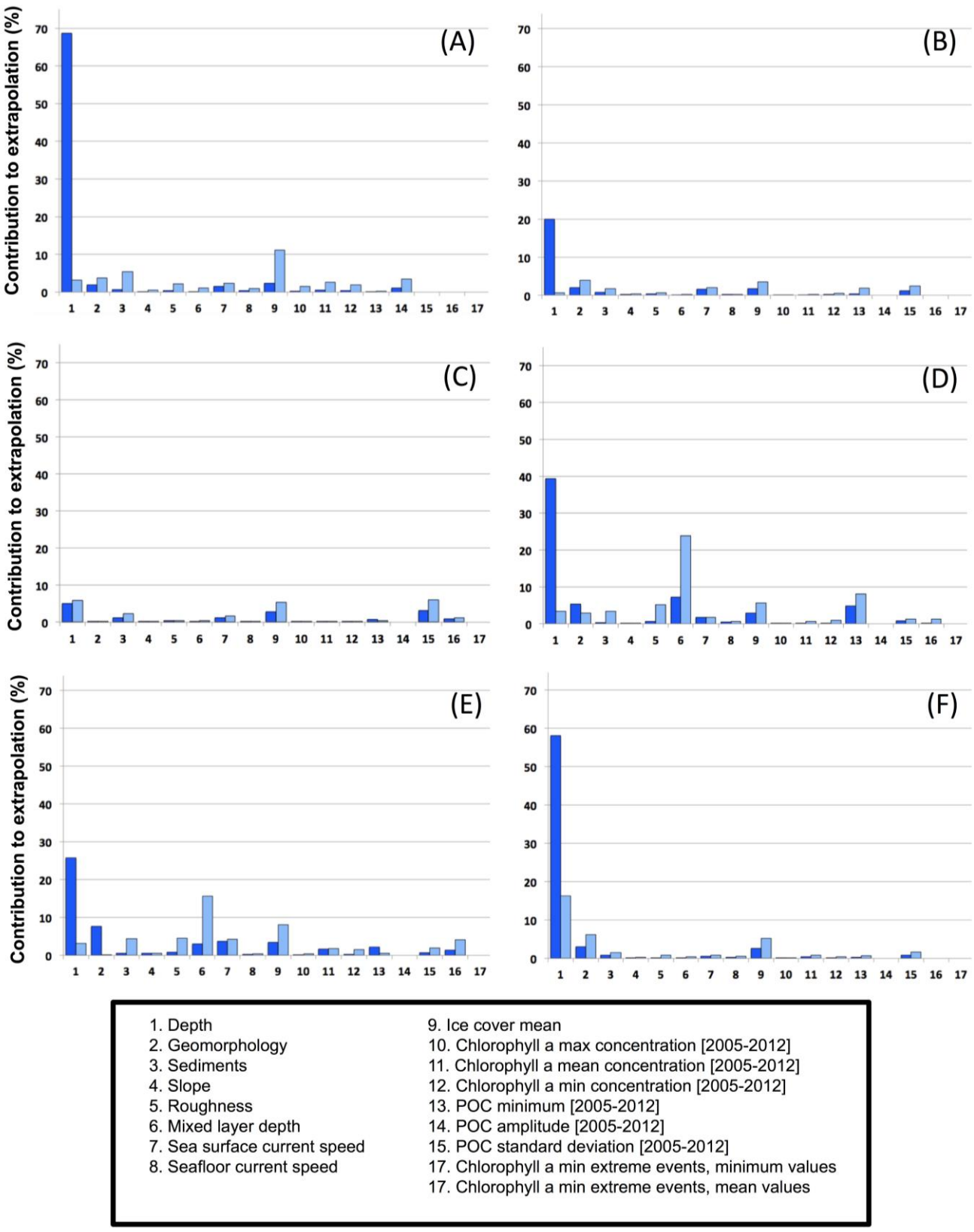


Figure S2.17.B. Influence of the different environmental descriptors on extrapolation (mean values calculated on the the 100 model replicates), for Analysis #0 (dark blue bars) and Analysis #1 (light blue bars). Analysis #0: models were projected on the entire Southern Ocean area. Analysis #1: the projection area was limited in depth according to each species distribution range (*A. hodgsoni*, *L. annulatus*, *O. validus* until 1,500 m; *B. loripes*, *G. antarctica*, *P. charcoti* until 4,000 m).

APPENDIX 2.18. Influence of the chronological addition of presence-only records on extrapolation area

Table S2.18. Evolution of model performances with the increase of data (chronological addition of presence-only records, by 5-year periods, from 1980 to 2016). Pres. NB: number of presence-only records used to generate the model; AUC: Area Under the Curve; Moran I index: spatial autocorrelation scores measured on model residuals (mean and standard deviation values are given). The maximal number of presence-only records present in Table 2.9 may not be reached here because some collection dates remain unknown.

Species	Year	Pres. NB	AUC	Correctly classified test data (%)	Moran I	Extrapolation (% total area)	Extrapolation (% suitable area)
Acodontaster hodgsoni	**1980**	85	0.843 ± 0.06	42.2 ± 21.6	0.24 ± 0.06	73.59	57.9 ± 9.6
	1985	147	0.836 ± 0.05	45.5 ± 17.5	0.14 ± 0.04	52.2	42.5 ± 7.5
	1990	170	0.822 ± 0.07	43.7 ± 22.9	0.13 ± 0.03	44.5	39.4 ± 9.6
	1995	171	0.835 ± 0.05	48.1 ± 18.7	0.13 ± 0.03	44.5	38.6 ± 8.7
	2000	180	0.827 ± 0.06	44.3 ± 19.9	0.12 ± 0.03	43.5	35 ± 8.1
	2005	197	0.836 ± 0.05	48.9 ± 20.7	0.11 ± 0.04	43.5	35.2 ± 8.6
	2010	252	0.829 ± 0.06	45.3 ± 16.9	0.10 ± 0.03	43	31.4 ± 8.1
	2016	280	0.821 ± 0.06	47.9 ± 15.3	0.10 ± 0.02	42.9	29.3 ± 7.3
	Year	**Pres. NB**	**AUC**	**Correctly classified test data (%)**	**Moran I**	**Extrapolation (% total area)**	**Extrapolation (% suitable area)**
Bathybiaster loripes	**1980**	193	0.860 ± 0.05	61.6 ± 16.2	0.13 ± 0.09	30.9	29.2 ± 11.1
	1985	252	0.855 ± 0.05	66.6 ± 14.5	0.12 ± 0.07	21.6	27.7 ± 6.6
	1990	269	0.849 ± 0.04	70 ± 13.1	0.10 ± 0.06	21.6	27.7 ± 6.7
	1995	286	0.854 ± 0.03	69.7 ± 13.6	0.10 ± 0.06	18.6	26.3 ± 6.4
	2000	299	0.850 ± 0.03	71.8 ± 13.3	0.10 ± 0.05	18.5	25.2 ± 5.8
	2005	349	0.869 ± 0.04	74 ± 12.6	0.10 ± 0.04	18.2	25.4 ± 5.2
	2010	480	0.878 ± 0.03	77.8 ± 11.3	0.09 ± 0.03	18.2	22 ± 4.3
	2016	521	0.879 ± 0.03	80.7 ± 9.1	0.10 ± 0.03	18.2	22.2 ± 4.3
	Year	**Pres. NB**	**AUC**	**Correctly classified test data (%)**	**Moran I**	**Extrapolation (% total area)**	**Extrapolation (% suitable area)**
Glabraster antarctica	**1980**	296	0.895 ± 0.03	69.2 ± 14.6	0.14 ± 0.06	44.4	30.4 ± 8.5
	1985	374	0.894 ± 0.04	76.1 ± 10.7	0.10 ± 0.04	30.3	24.8 ± 4.6
	1990	421	0.900 ± 0.03	78.3 ± 11.1	0.11 ± 0.04	30.1	25.8 ± 6.1
	1995	439	0.894 ± 0.03	79.7 ± 10.5	0.10 ± 0.03	30	23.9 ± 5.2
	2000	472	0.900 ± 0.02	80 ± 10.7	0.10 ± 0.03	30	23.9 ± 4.8
	2005	535	0.907 ± 0.02	81.4 ± 7.3	0.11 ± 0.03	24.1	22.7 ± 4.6
	2010	719	0.910 ± 0.02	84 ± 7.4	0.10 ± 0.03	23.9	18.9 ± 4
	2016	804	0.914 ± 0.02	83.1 ± 6.4	0.10 ± 0.03	23.9	19.2 ± 3.8

	Year	Pres. NB	AUC	Correctly classified test data (%)	Moran I	Extrapolation (% total area)	Extrapolation (% suitable area)
Labidiaster annulatus	**1980**	162	0.900 ± 0.04	61.6 ± 26.9	0.17 ± 0.08	77.1	56.2 ± 11.9
	1985	175	0.902 ± 0.04	64.4 ± 21.8	0.17 ± 0.07	72.2	52.4 ± 12.5
	1990	183	0.905 ± 0.03	66.2 ± 24.1	0.16 ± 0.07	70.5	47.7 ± 10.3
	1995	192	0.897 ± 0.03	63.6 ± 20.8	0.16 ± 0.07	70.5	48.6 ± 14
	2000	194	0.903 ± 0.03	71.2 ± 20	0.16 ± 0.07	70.5	45.4 ± 11.7
	2005	218	0.903 ± 0.04	63.4 ± 16.1	0.18 ± 0.09	63.3	47.5 ± 11.5
	2010	304	0.913 ± 0.05	60.7 ± 18	0.18 ± 0.09	60.5	44.8 ± 14
	2016	330	0.921 ± 0.03	62.4 ± 15.8	0.17 ± 0.08	59.5	41.5 ± 12.1
	Year	**Pres. NB**	**AUC**	**Correctly classified test data (%)**	**Moran I**	**Extrapolation (% total area)**	**Extrapolation (% suitable area)**
Odontaster validus	**1980**	163	0.860 ± 0.06	60.1 ± 16.6	0.17 ± 0.10	62.8	52.4 ± 7.5
	1985	191	0.883 ± 0.06	66.2 ± 16	0.15 ± 0.08	61.6	49.9 ± 6.9
	1990	198	0.875 ± 0.07	61.8 ± 17.7	0.16 ± 0.07	58.9	44.7 ± 8.2
	1995	200	0.873 ± 0.07	66.8 ± 16.1	0.16 ± 0.08	58.9	44.2 ± 7.3
	2000	222	0.856 ± 0.08	63.3 ± 15.4	0.13 ± 0.05	58.1	45.7 ± 9.4
	2005	283	0.922 ± 0.03	50.7 ± 23.5	0.13 ± 0.05	55.5	42.5 ± 6
	2010	306	0.920 ± 0.02	51.1 ± 24.9	0.12 ± 0.05	54.8	38.9 ± 6.9
	2016	321	0.914 ± 0.02	53.1 ± 24.1	0.13 ± 0.05	52.6	37.6 ± 7.5
	Year	**Pres. NB**	**AUC**	**Correctly classified test data (%)**	**Moran I**	**Extrapolation (% total area)**	**Extrapolation (% suitable area)**
Psilaster charcoti	**1980**	134	0.847 ± 0.05	50.4 ± 24.2	0.11 ± 0.07	46.7	39.4 ± 7.2
	1985	182	0.848 ± 0.05	76.4 ± 10.6	0.10 ± 0.06	36.9	37.8 ± 3.9
	1990	200	0.844 ± 0.05	81.4 ± 10.7	0.10 ± 0.06	36.3	39.5 ± 3.9
	1995	203	0.851 ± 0.04	79.7 ± 11.8	0.12 ± 0.07	36.3	38.1 ± 4.3
	2000	220	0.861 ± 0.04	81.6 ± 7.1	0.10 ± 0.05	36.3	37.4 ± 4.5
	2005	257	0.867 ± 0.03	79.5 ± 8.1	0.10 ± 0.05	36.3	36.4 ± 4.1
	2010	321	0.878 ± 0.03	83.5 ± 7.1	0.10 ± 0.04	35.8	33.8 ± 4.2
	2016	353	0.891 ± 0.02	82 ± 7.1	0.10 ± 0.04	35.7	32.9 ± 4.2

APPENDIX 2.19. Influence of the addition of presence-only records on extrapolation area

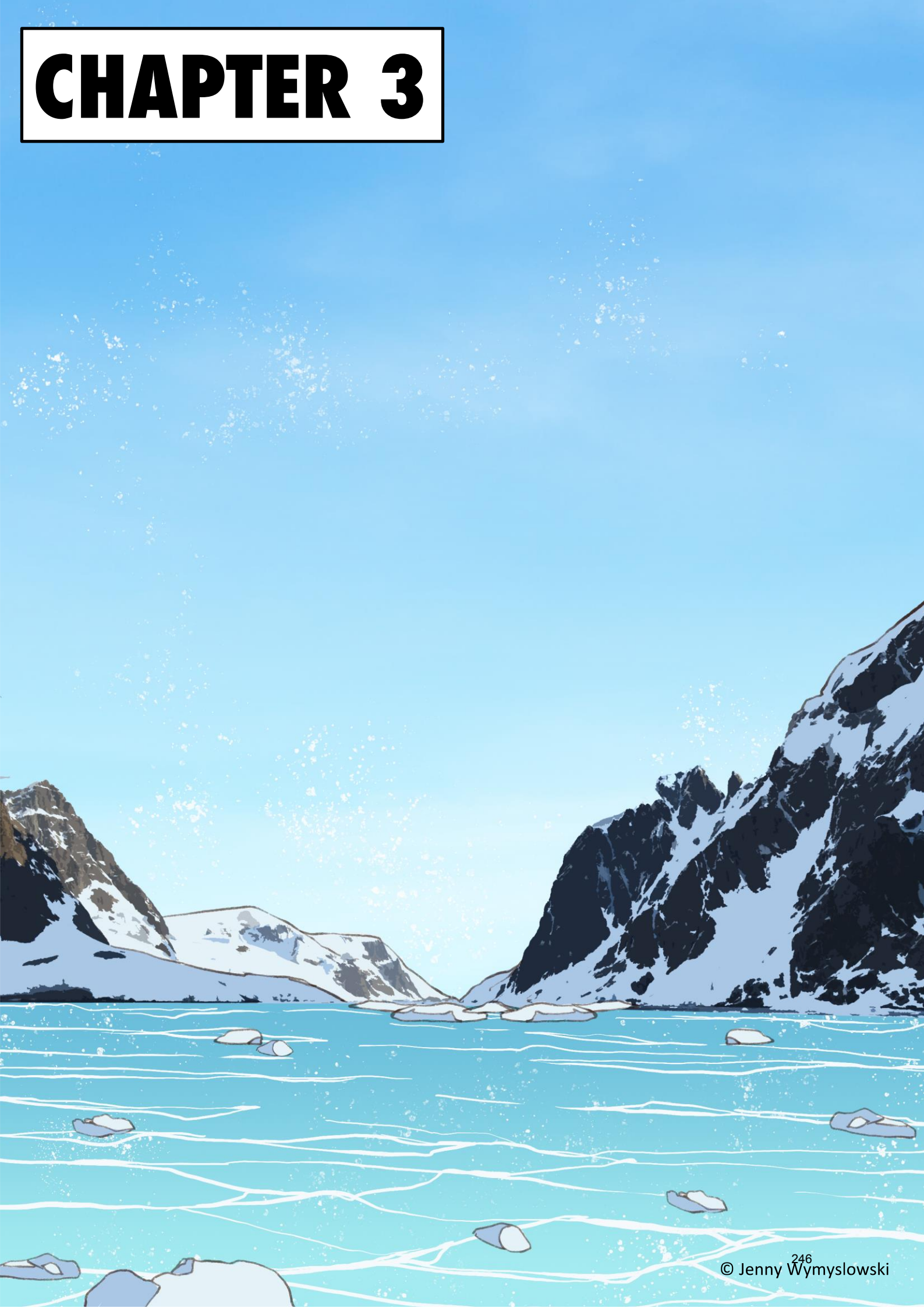
Table S2.19. Evolution of model performances with a random increase of data number (10 to 100% of the available presence datasets, randomly sampled). Measured average and standard deviation values. Pres. NB: corresponding number of presence-only records used to generate the model; AUC: Area Under the Curve; Moran I index: spatial autocorrelation scores measured on model residuals. The cells with no figure information correspond to models that could not be generated due to a too low number of presence records.

Species	%	Pres. NB	AUC	Correctly classified test data (%)	Moran I	Extrapolation (% total area)	Extrapolation (% suitable area)
Acodontaster hodgsoni	**10**	30	-	-	-	-	-
	20	60	-	-	-	-	-
	30	89	-	-	-	-	-
	40	119	0.808 ± 0.05	45.8 ± 19.4	0.10 ± 0.04	56.3 ± 5.1	43.1 ± 7.9
	50	149	0.821 ± 0.04	46.2 ± 16.2	0.10 ± 0.03	52 ± 4.4	40.4 ± 9.5
	60	179	0.812 ± 0.05	46.5 ± 16.9	0.10 ± 0.03	48 ± 3.4	36 ± 9
	70	209	0.818 ± 0.05	45.1 ± 16	0.10 ± 0.02	45.5 ± 2.9	32.3 ± 7.7
	80	238	0.821 ± 0.05	45.8 ± 18.1	0.09 ± 0.02	43.8 ± 1.9	30.7 ± 8.3
	90	268	0.832 ± 0.05	45 ± 17.8	0.10 ± 0.02	42 ± 1.5	28.1 ± 6.4
	100	298	0.823 ± 0.05	45.5 ± 18.1	0.09 ± 0.02	40.6	27.5 ± 8.5
	%	**Pres. NB**	**AUC**	**Correctly classified test data (%)**	**Moran I**	**Extrapolation (% total area)**	**Extrapolation (% suitable area)**
Bathybiaster loripes	**10**	59	-	-	-	-	-
	20	118	-	-	-	-	-
	30	177	-	-	-	-	-
	40	236	0.863 ± 0.06	73.8 ± 12.5	0.10 ± 0.04	37.2 ± 9	31.5 ± 7.3
	50	296	0.869 ± 0.04	75.5 ± 12	0.10 ± 0.04	31.4 ± 7.7	28.4 ± 7.2
	60	355	0.876 ± 0.03	75.6 ± 12.8	0.09 ± 0.04	27.4 ± 6.4	25.4 ± 5.7
	70	414	0.881 ± 0.03	75.7 ± 12.6	0.09 ± 0.03	23.7 ± 3.8	25 ± 5.6
	80	473	0.888 ± 0.02	76.8 ± 11.7	0.10 ± 0.04	21.7 ± 3.4	24.3 ± 5.6
	90	532	0.882 ± 0.03	76.5 ± 12.6	0.09 ± 0.02	19.5 ± 1.9	22.4 ± 5.5
	100	591	0.887 ± 0.03	78.4 ± 11	0.09 ± 0.02	18.2	20.8 ± 4.8
	%	**Pres. NB**	**AUC**	**Correctly classified test data (%)**	**Moran I**	**Extrapolation (% total area)**	**Extrapolation (% suitable area)**
Glabraster antarctica	**10**	85	-	-	-	-	-
	20	175	0.872 ± 0.04	82.7 ± 10.3	0.12 ± 0.05	50.9 ± 11.2	40.2 ± 7.7
	30	255	0.883 ± 0.03	82.8 ± 9.3	0.11 ± 0.04	42.6 ± 9.2	34.5 ± 5.8

	40	340	0.889 ± 0.03	84.5 ± 8.3	0.11 ± 0.04	38.4 ± 7.9	29.9 ± 6.4
	50	426	0.896 ± 0.03	83.7 ± 7.8	0.11 ± 0.04	33 ± 5.9	27.6 ± 5.7
	60	511	0.899 ± 0.02	81.2 ± 8.7	0.11 ± 0.04	30.1 ± 5.3	25 ± 5.5
	70	596	0.903 ± 0.02	81.2 ± 7.9	0.10 ± 0.03	28.9 ± 5.1	22.9 ± 5.1
	80	681	0.908 ± 0.02	80.7 ± 8.5	0.11 ± 0.03	27.1 ± 3.3	21.3 ± 4.5
	90	766	0.913 ± 0.02	80.2 ± 8.2	0.10 ± 0.02	25.4 ± 2.7	19.6 ± 4.3
	100	851	0.915 ± 0.01	81.8 ± 7.7	0.10 ± 0.03	23.9	18.64 ± 3.5
	%	**Pres. NB**	**AUC**	**Correctly classified test data (%)**	**Moran I**	**Extrapolation (% total area)**	**Extrapolation (% suitable area)**
Labidiaster annulatus	**10**	38	-	-	-	-	-
	20	75	-	-	-	-	-
	30	113	-	-	-	-	-
	40	150	0.850 ± 0.12	59.1 ± 23.2	0.17 ± 0.08	69.8 ± 3.8	46.7 ± 15.4
	50	188	0.897 ± 0.06	58.1 ± 20	0.17 ± 0.09	67.4 ± 3.5	48.1 ± 14.5
	60	225	0.898 ± 0.05	55.4 ± 19.4	0.15 ± 0.07	64.9 ± 3.1	45.5 ± 14.8
	70	263	0.903 ± 0.05	59.7 ± 18.7	0.18 ± 0.1	63 ± 2.5	44.2 ± 15.6
	80	300	0.918 ± 0.03	58.4 ± 20	0.16 ± 0.08	61.2 ± 1.6	39.7 ± 13.1
	90	338	0.923 ± 0.03	57.7 ± 18.7	0.15 ± 0.06	60.4 ± 1.3	38.9 ± 14.1
	100	375	0.918 ± 0.03	57.98 ± 20	0.15 ± 0.06	59.5	38.7 ± 14.6
	%	**Pres. NB**	**AUC**	**Correctly classified test data (%)**	**Moran I**	**Extrapolation (% total area)**	**Extrapolation (% suitable area)**
Odontaster validus	**10**	33	-	-	-	-	-
	20	67	-	-	-	-	-
	30	101	-	-	-	-	-
	40	135	0.873 ± 0.05	55.6 ± 25.4	0.13 ± 0.08	63.9 ± 4.1	54.3 ± 9.5
	50	169	0.878 ± 0.05	58.3 ± 21.1	0.13 ± 0.06	60.5 ± 3.5	49.9 ± 8.5
	60	202	0.896 ± 0.03	52 ± 23.3	0.14 ± 0.06	58.3 ± 3	45.3 ± 9.2
	70	236	0.899 ± 0.04	54.4 ± 23.2	0.13 ± 0.06	56 ± 2.5	44.9 ± 8.1
	80	270	0.900 ± 0.03	55.9 ± 22.8	0.13 ± 0.04	54.7 ± 2.2	38.9 ± 6.7
	90	303	0.911 ± 0.03	52.8 ± 23	0.12 ± 0.05	52.7 ± 1.4	38.1 ± 7.7
	100	337	0.908 ± 0.03	57.68 ± 21	0.12 ± 0.04	51.5	38.3 ± 6.97
	%	**Pres. NB**	**AUC**	**Correctly classified test data (%)**	**Moran I**	**Extrapolation (% total area)**	**Extrapolation (% suitable area)**
Psilaster charcoti	**10**	35	-	-	-	-	-
	20	71	0.837 ± 0.06	82.9 ± 14.7	0.09 ± 0.06	70.8 ± 12.1	55.7 ± 7.9

	30	106	0.844 ± 0.05	84.5 ± 9.7	0.10 ± 0.06	58 ± 11.4	49 ± 5.8
	40	141	0.861 ± 0.04	84.3 ± 9.5	0.11 ± 0.06	53.3 ± 10.3	45.9 ± 5.7
	50	176	0.863 ± 0.04	82.4 ± 9.8	0.11 ± 0.06	48 ± 9.7	41.9 ± 5.3
	60	212	0.870 ± 0.03	83.6 ± 8.8	0.11 ± 0.06	44.1 ± 8.1	39.7 ± 4.8
	70	247	0.875 ± 0.03	82 ± 8.6	0.10 ± 0.04	41.3 ± 6.3	38 ± 4.8
	80	282	0.876 ± 0.03	83.2 ± 7.6	0.10 ± 0.05	40.1 ± 6.3	36.7 ± 5.1
	90	316	0.885 ± 0.02	82.2 ± 8.1	0.10 ± 0.04	37 ± 1.65	34.4 ± 5.1
	100	353	0.885 ± 0.02	83 ± 6.6	0.09 ± 0.04	35.78	33.2 ± 5.1

CHAPTER 3



CHAPTER 3 INTEGRATED APPROACHES

Chapter 3 focusses on integrated approaches. Coupling SDM predictions with experimental results, *in situ* observations or results from other modelling approaches that detail species physiological tolerance, migratory potential or biotic interactions was shown to improve the relevance of species niche estimation. Such methods have however been rarely applied to Southern Ocean marine case studies. In this chapter, we studied the integration of SDM with physiological information.
DEB models characterise the species fundamental niche, by explicitly highlighting the influence of abiotic factors on species physiology. On the other hand, SDMs estimate the species realised niche. SDMs are indeed implemented using presence records, hence providing an implicit assessment of the influence of abiotic conditions but also dispersal barriers and biotic interactions on species distribution.

This chapter illustrates three case examples.

•The first study assessed the potential of the Patagonian crab *Halicarcinus planatus* (Fabricius, 1775) to survive in the Western Antarctic Peninsula using two approaches: experimental data that characterise the physiological boundaries of larvae and adult to temperature and salinities and SDMs that simulate species occupied space in present and future environmental conditions.

•In the second analysis, the case study of the sea urchin *Sterechinus neumayeri* (Meissner, 1900), distributed all around the Antarctic continent, was used to compare DEB model spatial projections and SDM predictions. Comparisons were performed for contrasting environmental conditions and future simulations.

•The third analysis used data from a long-term observing network located in the Kerguelen Islands, to implement for the first time in the Southern Ocean the integration of DEB and SDM models to predict the distribution of an endemic sub-Antarctic sea urchin, *Abatus cordatus* (Verrill, 1876) as a response to environmental drivers. We compared the performance of simple SDM and integrated approaches to predict *A. cordatus* distribution under seasonal variations. Two integrated approaches were studied and performed by either (1) including the spatial projection of the DEB model as an input layer inside the SDM or (2) using a Bayesian inference procedure to use DEB model outputs as priors of the Bayesian SDM.

•**Guillaumot C**/ López-Farrán Z (co-firstauthorship), Vargas-Chacoff L, Paschke K, Dulière V, Danis B, Poulin E, Saucède T and Gerard K (2021). Current and predicted invasive capacity of *Halicarcinus planatus* (Fabricius, 1775) in the Antarctic Peninsula. *Global Change Biology*. 00:1–18.

•Fabri-Ruiz S, **Guillaumot C,** Agüera A, Danis B and Saucède T (2021). Using correlative and mechanistic niche models to assess the sensitivity of the Antarctic echinoid *Sterechinus neumayeri* (Meissner, 1900) to climate change. *Polar Biology*.

•**Guillaumot C**, Buba Y, Belmaker J, Fourcy D, Danis B, Dubois P and Saucède T (submitted). Simple or hybrid ? Next generation ecological models to study the response of Southern Ocean marine species to changing environmental conditions. *Diversity and Distributions*.

Is the southern crab *Halicarcinus planatus* (Fabricius, 1775) the next invader of Antarctica?

[Zambra López-Farrán[1,2,3] / Charlène Guillaumot[4,5]], Luis Vargas-Chacoff[2,6], Kurt Paschke[2,7], Valerie Dulière[8], Bruno Danis[4], E. Poulin[1], Thomas Saucède[5], Jonathan Waters[9], Karin Gerard[3,10]

[1] LEM-Laboratorio de Ecología Molecular, Instituto de Ecología y Biodiversidad, Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile.
[2] Research Center Dynamics of High Latitude Marine, Ecosystem (Fondap-IDEAL), Universidad Austral de Chile, Valdivia, Chile
[3] LEMAS-Laboratorio de Ecología de Macroalgas Antárticas y Sub antárticas, Universidad de Magallanes, Punta Arenas, Chile
[4] Laboratoire de Biologie marine CP160/15, Université Libre de Bruxelles, Bruxelles, Belgium.
[5] Biogéosciences, UMR 6282 CNRS, Université Bourgogne Franche-Comté, Dijon, France.
[6] Instituto de Ciencias Marinas y Limnológicas, Laboratorio de Fisiología de Peces, Universidad Austral de Chile, Valdivia, Chile
[7] Instituto de Acuicultura, Universidad Austral de Chile, Puerto Montt, Chile
[8] Royal Belgian Institute Natural Sciences, Brussels, Belgium
[9] Otago Palaeogenetics Laboratory, Department of Zoology, University of Otago, Dunedin, New Zealand
[10] Centro de Investigación Gaia-Antártica, Universidad de Magallanes, Punta Arenas, Chile

Global Change Biology, accepted April 17th 2021.

Abstract

The potential for biological colonisation of Antarctic shores is an increasingly important topic in the context of anthropogenic warming. Successful Antarctic invasions to date have been recorded exclusively from terrestrial habitats. While non-native marine species such as crabs, mussels and tunicates have already been reported from Antarctic coasts, none have as yet established there. Among the potential marine invaders of Antarctic shallow waters is *Halicarcinus planatus* (Fabricius, 1775), a crab with a circum sub-Antarctic distribution and substantial larval dispersal capacity. An ovigerous female of this species was found in shallow waters of Deception Island, South Shetland Islands, in 2010. A combination of physiological experiments and ecological modelling was used to assess the potential niche of *H. planatus* and estimate its future southward boundaries under climate change scenarios. We show that *H. planatus* has a minimum thermal limit of 1°C, and that its current distribution (assessed by sampling and niche modelling) is physiologically restricted to the sub-Antarctic region. While this species is presently unable to survive in Antarctica, future warming under both 'strong mitigation' and 'no mitigation' greenhouse gas emission scenarios will favour its niche expansion to the Western Antarctic Peninsula (WAP) by 2100. Future human activity also has potential to increase the probability of anthropogenic translocation of this species into Antarctic ecosystems.

Key-words

Niche modelling, Southern Ocean, climate change, thermotolerance, survival, establishment, reptant crab, non-native species.

ACKNOWLEDGEMENTS

The research was supported by projects FONDECYT Regular 1161358 to KG and EP, INACh DG 14-17 and Chilean national doctoral scholarship CONICYT 21151192 to ZLF, Fondap-IDEAL 15150003 to ZLF and LV, PIA CONICYT ACT172065 to EP and KG, and FONDECYT 1160877 to LV. This work was supported by a "Fonds pour la formation à la Recherche dans l'Industrie et l'Agriculture" (FRIA) and "Bourse Fondation de la Mer" grants to C. Guillaumot. This work is contribution to the French Polar Institute and LTSER Zone Atelier Antarctique et sub-Antarctique (ZATA) program PROTEKER (n°1044).

This is contribution no. 48 to the vERSO project (www.versoproject.be), funded by the Belgian Science Policy Office (BELSPO, contract n°BR/132/A1/vERSO). Research was also financed by the "Refugia and Ecosystem Tolerance in the Southern Ocean" project (RECTO; BR/154/A1/RECTO) funded by the Belgian Science Policy Office (BELSPO), this study being contribution number 25.

We are thankful to Renato Borras, Daniel Ramirez, Yethro Henriquez and Eva Iglesias to help collecting individuals of Halicarcinus planatus for the physiological experiment (SCUBA diving work). Thanks to Alejandro Ortiz (IDEAL), Constanza Ceroni, Aurora Prado, Hermes Galo Andrade (IDEAL), Hans Bartsch (IDEAL), Camille Détrée (IDEAL) and Paola Muñoz for helping install and maintain experiments. Thanks to Hermes Galo Andrade, Jorge Navarro (IDEAL), research line Adaptation of the marine species (IDEAL), Angelica Saldivia (IDEAL) and Luis Villegas (IDEAL) for the technical support for this project. This article contributes to the SCAR Biology Programme AntEco (State of the Antarctic Ecosystem) for travel funding.

AUTHORS' CONTRIBUTIONS

-Zambra López-Farrán, first author of this paper, participated directly in experimental design, sampling, assembly of the experiments, maintenance of experiments, obtaining data, results analyses, writing of the manuscript.

-Charlène Guillaumot, co-first author of this manuscript participated in results analyses, writing, revision and correction of the manuscript.

-Karin Gerard participated directly in sampling, assembly of the experiments, writing, revision and correction of the manuscript.

-Luis Vargas-Chacoff participated directly in experimental design, assembly of the experiments, results analyses, revision and correction of manuscript.

-Kurt Paschke participated directly in experimental design, assembly of the experiments, results analyses, revision and correction of manuscript.

-Valerie Dulière participated in the analysis of the results, revision and correction of the manuscript.

Bruno Danis participated in the analysis of the results, revision and correction of the manuscript

-Thomas Saucède contributed to occurrence dataset, participated to the analysis of results, and to manuscript revision and correction.

-Elie Poulin participated directly in the project and experimental design, sampling, analysis of the results, revision and correction of the manuscript.

-Jonathan Waters participated to the analysis of results, and to manuscript revision and correction.

1. INTRODUCTION

Biological invasions are an important component of global change, and one of the most critical global threats to native biodiversity (Sax et al. 2005). According to (Richardson et al. 2000), a non-native species becomes an invasive species when a set of individuals is able to traverse natural barriers (whether geographical, environmental, or ecological) and subsequently establish in new habitats. While numerous anthropogenic activities can promote invasions, climate change may represent a particularly potent threat to natural ecosystems (Malcolm et al. 2006). Both the rate and dimension of biological invasions are likely to be influenced by global warming (Walther et al. 2009). Understanding the mechanisms and routes of such range shifts may help facilitate the design of strategies for controlling or preventing invasion (Estoup and Guillemaud 2010).

Notwithstanding the wide expanse of Southern Ocean waters isolating the southern tip of South America from other land masses, several non-native species have been reported in the Antarctic over recent decades (Smith and Richardson 2011). These examples include the invasive grass *Poa annua* (Molina-Montenegro et al. 2012, Chwedorzewska et al. 2015), seeds of the toad rush *Juncus bufonius* (Cuba-Díaz et al. 2013), the invasive mosquito *Trichocera maculipennis* (Potocka and Krzemińska 2018), and several South-American invertebrates (e.g. insects, worms, freshwater crustaceans; Hughes and Worland 2010, Hughes et al. 2015). Non-native species have also been reported in marine habitats and in the shallow subtidal zone, in particular in the south Shetland Islands (i.e. decapods and bivalves) and East Antarctica (i.e. bryozoans, hydrozoans, and tunicates) (McCarthy et al. 2019, Avila et al. 2020, Cárdenas et al. 2020). However, there is as yet no evidence for any non-native marine species having established in Antarctica.

Reaching Antarctic coasts requires dispersal across vast and deep biogeographical barriers that have isolated the continent for millions of years, including traversal of the westward flowing Antarctic Circumpolar Current (ACC) that apparently impedes latitudinal dispersal (Clarke et al. 2005, Rintoul 2009). The extreme cold temperatures of Antarctic waters (< +2°C) also imply a strong ecophysiological constraint to the survival and development of exotic marine species that have not adapted to near-zero and subzero temperatures (Marsh et al. 2001, Fraser et al. 2007, Peck 2016) that can reach down to -1.85°C in winter. Consequently, Antarctic marine communities have been considered among the most isolated and endemic on Earth and invasion by non-native species as unlikely (Clarke et al. 2005, Griffiths et al. 2009).

Human activities such as fisheries, tourism and scientific operations rely on direct maritime traffic between Antarctica and lower latitude coasts, including potential transport of alien organisms through ship hull fouling and larval propagules via ballast water (Lewis et al. 2003, 2005). With more than 50,000 tourists visiting the same west Antarctic spots each southern summer (McCarthy et al. 2019), and 4,000 scientists working in Antarctica during the summer and 1,000 in winter (Hughes and Convey 2014), tourism and science represent the main vectors of sub-Antarctic propagule pressure over Antarctic communities (Tavares and De Melo 2004, Meredith and King 2005, Lee and Chown 2007, Hellmann et al. 2008, Diez and Lovrich 2010, Galera et al. 2018, Avila et al. 2020). Consequently, the records of non-native species in Antarctica are increasing in number, with potential for establishment now primarily constrained by ecological and physiological limitations. As the climate continues to warm, the potential for successful marine invasions into Antarctica is projected to increase substantially (Richardson et al. 2000, Hellmann et al. 2008, Galera et al. 2018).

The Western Antarctic Peninsula (WAP) is the Antarctic region where the strongest climate warming has been recorded in the continent over the last 50 years (Convey et al. 2009, Turner et al. 2014, Gutt et al. 2015). Sea-water and air temperatures have increased by +1ºC and +7°C respectively in the past half-century (Meredith and King 2005, Schram et al. 2015), with particularly pronounced increases in winter air temperatures (King et al. 2003, Vaughan et al. 2003) and corresponding reductions in sea-ice cover (Stammerjohn et al. 2012, Ducklow et al. 2013, Turner et al. 2016, Schofield et al. 2017). Global climate change may cause typically sub-zero Antarctic waters to warm up to (and beyond) zero, potentially providing suitable conditions to the survival of non-native species along Antarctic coasts (Hellmann et al. 2008, Galera et al. 2018).

In February 2010, an ovigerous female of *Halicarcinus planatus* (Fabricius, 1775) (Brachyura, Hymenosomatidae) was found alive in shallow subtidal water of Deception Island (WAP; Aronson et al. 2014). Previous to this record, Stebbing (1914) reported this species in Macdougal Bay, South Orkney Islands; however, the reliability of this occurrence has been questioned, considering its circum-sub-Antarctic distribution (Thatje and Arntz 2004, Diez and Lovrich 2010, Aronson et al. 2015). *Halicarcinus planatus* is the only hymenosomatid crab that inhabits shallow waters (Garth 1958, Varisco et al. 2016) of southern South America and the sub-Antarctic Falkland/Malvinas, Marion, Crozet, Kerguelen and Macquarie Islands (Boschi et al. 1969, Melrose 1975, Richer De Forges 1977, Griffiths et al. 2013, Aronson et al. 2014). This small crab (carapace width up to 15 mm and 20 mm for female and male, respectively, in Punta Arenas; Fig. 3.1) is an opportunistic feeder (Boschi et al. 1969) commonly found sheltered under rocks in the intertidal and subtidal zones, in between holdfasts of the giant kelp *Macrocystis pyrifera* or sheltered in hydrozoans and mussel colonies (Richer De Forges 1977, Chuang and Ng 1994, Vinuesa and Ferrari 2008).

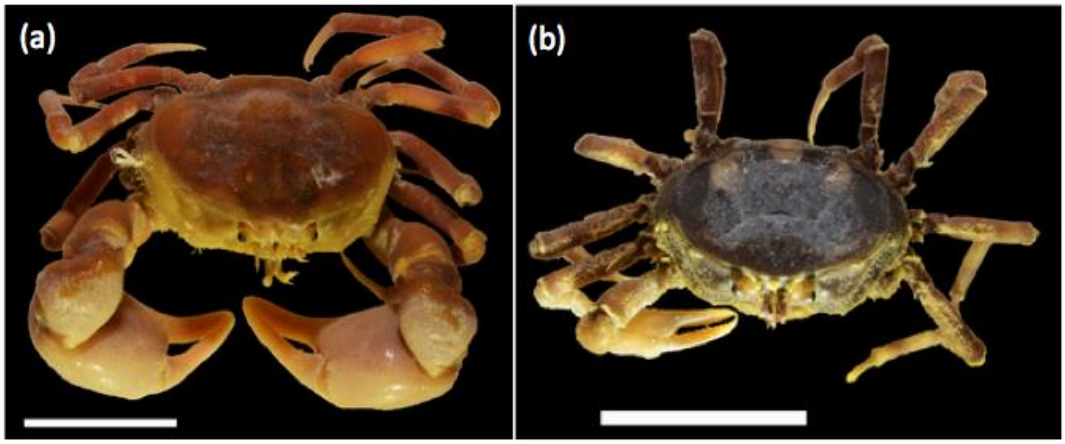


Figure 3.1. Male (a) and female (b) specimens of *Halicarcinus planatus* (Fabricius, 1775) collected in the Magellan Strait. Scale: 1 cm. Photograph credit to C. Ceroni and K. Gérard.

The potential of marine taxa to establish in Antarctic waters is likely heavily constrained by ecological and physiological adaptations. *H. planatus* has a strong dispersal potential mediated by an extended planktonic larval stage (Richer De Forges 1977, Diez and Lovrich 2010, Ferrari et al. 2011), lasting between 45 and 60 days (at temperatures of 11-13°C and 8°C respectively, in the laboratory) prior to benthic settlement (Boschi et al. 1969, Diez and Lovrich 2010). This species has the physiological capacity to withstand low temperatures. Most decapod taxa exposed to cold waters experience increased magnesium ion concentration in the hemolymph ($[Mg^{2+}]_{HL}$), reducing metabolic rates and aerobic activity, potentially leading to death (Frederich et al. 2001, Thatje et al. 2005a, Aronson et al. 2007, Diez and Lovrich 2010). However, *H. planatus* has the capacity to overcome these issues by reducing $[Mg^{2+}]_{HL}$ (Frederich et al. 2001), providing capacity for survival in cold waters like the Kerguelen Islands, where winter seawater temperatures range between +1.1 and +3.0°C (Féral et al. 2019). A broad analysis by Diez and Lovrich (2010) considering its broad sub-Antarctic distribution, high dispersal potential, and ability to live at low temperatures, concluded that *H. planatus* is the most likely future decapod invader of Antarctic shallow waters.

Following the recent discovery of a living specimen of *H. planatus* in Deception Island, we evaluate in this study the capacity of the species to settle and spread in the WAP and adjacent islands by combining experimental designs and a niche modelling approach. Correlative niche modelling approaches have long proved useful to project the distribution range of species for conservation purposes under stable environmental conditions (Richardson and Whittaker 2010). However, in the context of climate change, ecophysiological data are required to assess the capacity of organisms to survive under changing environmental conditions. In this study, we assessed experimentally the physiological capacity of *H. planatus* to tolerate extreme cold conditions in laboratory, and we evaluated the probability of the species to expand its distribution range southward using a Species Distribution Model (SDM). The modelled distribution of *H. planatus* was first projected under

current climatic conditions in order to evaluate its distribution range in sub-Antarctic and Antarctic regions. Then the species distribution was modelled under the 'strong mitigation' and 'no mitigation' scenarios (RCP 2.6 and RCP 8.5 respectively) for 2050 and 2100 to determine the probability that *H. planatus* will colonize Antarctic shallow-water habitats in the future. RCP scenarios assess the evolution of the atmospheric radiative forcing towards 2300, and correspond to the level of the projected radiative forcing in 2100, expressed in W/m2 (RCP 2.6 corresponds to 2.6 W/m^2 in 2100; https://sos.noaa.gov/datas ets/catal og/datasets/air?order ing=name).

2. MATERIAL AND METHODS

2.1. Experimental design

Ethical Protocol.
All experiments were performed in compliance with bioethics guidelines established by the *Comisión Nacional de Ciencia y Tecnología de Chile* (CONICYT) and the CICUA from Universidad de Chile (*Comité Institucional de Cuidado y Uso de Animales*).

Thermotolerance experiments.
One hundred and twenty adult specimens of *H. planatus* were collected alive in the subtidal zone by SCUBA diving at *Rinconada Bulnes* (RB) (53°35'49.91"S, 70°56'5.19"W, south to Punta Arenas, Chile) on April 9, 2018. Individuals were transported to the IDEAL-CENTER laboratory (Punta Arenas) and distributed in six containers for the experiment. In each container (Appendix 3.1), 15 females and five males were isolated individually in a 1-dm^3 glass jar of seawater containing a 2-cm-long PVC tube (2.5 cm diameter). This unequal sample size between gender reflected to the disproportional sex ratio in nature (Vinuesa and Ferrari 2008, Diez et al. 2011), at the time of collection, 30% of the crabs were males and 70% were females. A plastic container of seawater was used for water replacement. Each jar and container were aerated and temperature was controlled by a cooler exchanger (Alpha RA12 and RA8, Lauda- Koenigshofen®, Germany). Individuals were acclimated for 15 days with temperature, salinity and photoperiod adjusted to the sampling location (9°C, 30 PSU, 11hrs light/13hrs dark, on April 9, 2018). Individuals were fed every four days with thawed and chopped mussels and polychaetes. The next day, 30% of seawater was removed from each jar, sucking the bottom to eliminate faeces and food debris. Recipients were then refilled with clean seawater at the exact same temperature and salinity from the plastic seawater container. The latter was then refilled with new seawater, which had the time to reach the specific temperature before the next refill. After acclimation, temperature was reduced by 0.5°C every day, until it reached a threshold value set at 5°C (control; minimal seawater temperature in Punta Arenas), 2°C, 1°C, 0°C, -1°C or -1.8°C, depending on the experiment, which was conducted for 90 days following (Vargas-Chacoff et al. 2009). The different temperature threshold values used in the experiment correspond to subtidal temperatures recorded in Fildes Bay (62°12'11.95"S 58°56'37.00"W; King George Island, South Shetland Islands, WAP), which ranged between -1.9ºC and 2.1ºC; summer average 1.2ºC (-0.2ºC to 2.1ºC) and winter average -1.6ºC (-1.9ºC to -1.1ºC) in 2017 (data from IDEAL-CENTER, published by Cárdenas et al. 2020). The 90 days simulate the duration of winter. Survival was checked each morning, and dead specimens were removed and preserved in 96% ethanol.

Salinity and larval experiments.
To assess survival at different salinities, adult individuals of *H. planatus* were collected at the same location (RB) on July 5, 2018, transported to the laboratory and separated in containers. Eighteen females and four males were isolated in a recipient of 10-dm^3 filled in with seawater. After a 15-day acclimation period at the same temperature, salinity and photoperiod as the sampling location (5°C, 30 PSU, 8:16 L:D), individuals were submitted to different salinities of 30 PSU (control 1), 23 PSU, 18 PSU, 11 PSU and 5 PSU for 39 days at 5ºC. In parallel, some individuals submitted to natural 18-PSU seawater collected in Skyring Sound (52°33'48.07"S, 71°34'15.54"W) were used as a second control. The previously detailed protocol for feeding and cleaning was followed. Survival was checked every morning and dead specimens were removed and preserved in 96% ethanol.

During the salinity experiment, at 30 PSU, some individuals released larvae which were subsequently collected and placed in a 1-dm^3 glass jar (200 larvae in each) filled with seawater at 5ºC, 2ºC and 1ºC for 12 days. Crab larvae were fed daily with newly hatched nauplii. Their survival was checked on days 1, 3, 6, 8, 10 and 12 and on the cleaning day which consists in the complete seawater replacement. Dead individuals were removed and preserved in 96% ethanol.

2.2. Species Distribution Modelling

Species Distribution Models (SDM) are used to project the distribution of organisms based on the statistical analysis of spatial relationships between environmental conditions and species records (Elith et al. 2006, Peterson 2003, Peterson et al. 2011). SDMs have been widely used in the past decades for various applications among including assessing species potential distribution (Reiss et al. 2011, Nachtsheim et al. 2017, Guillaumot et al. 2018b) and evaluating potential changes in predicted suitable areas under environmental shifts (Berry et al. 2002, Pearson and Dawson 2003, Thomas et al. 2004, Engler et al. 2009, Meier et al. 2011).

2.3. Occurrence dataset

The study had a limited geographical extent where occurrence records have been reported (Longitude: 70.5°E to 75.5°W, Latitude: 36°S to 70.5°S). Presence and absence data were collected during different sampling expeditions carried out between 2015 and 2019 (PROTEKER 1, 4, 5 and 6, INACH ECA 53, 54 and 55), obtained from collaborators, and retrieved from IOBIS and GBIF databases, and from the scientific literature (Appendix 3.2). The georeferencing of each occurrence was verified and for this study repeated geographical points were removed; the identification of collected specimens was checked following current taxonomy (Boschi 1964). Occurrences located north of 34°S in Chile were not considered, since these points were outside the distribution range of the species and could not be corroborated.

A DarwinCore-compliant dataset was built using presence and absence data of *H. planatus* occurring on sub-Antarctic islands and South America between 1948 and 2019. Four types of records were included: individualized by specimen, by groups, records obtained from bibliographic reviews and absence records. The dataset was published in GBIF (López-Farrán et al. 2020).

Distribution models were built using 314 presence records of both adults and larvae, and 57 absence records (Fig. 3.2, Appendix 3.2).

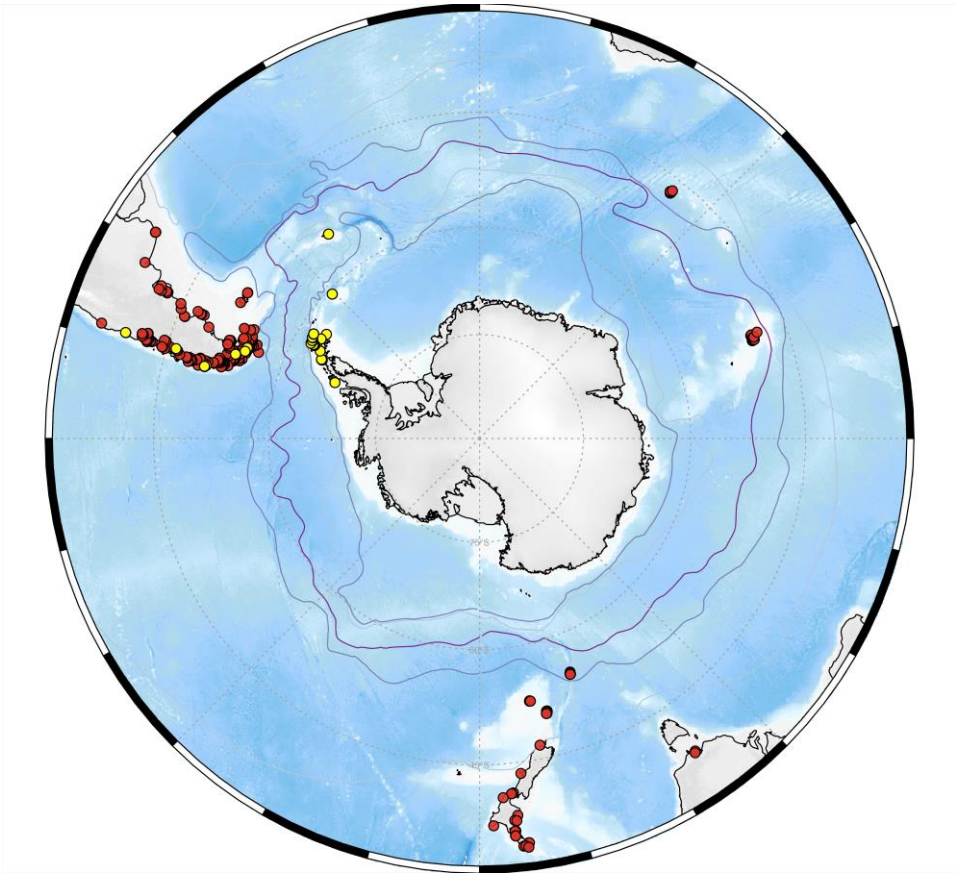


Figure 3.2. Presence (red dots) and absence (yellow dots) records of *Halicarcinus planatus* in the Southern Ocean used in the present study.

2.4. Environmental datasets

The distribution of *H. planatus* was modelled using 16 environmental parameters as descriptors of the crab habitat (Table 3.1). Depth and its derivatives (slope and roughness) were taken from GEBCO (Table 3.1). Other descriptors were compiled from the Bio-ORACLE marine layers dataset and obtained from pre-processed global ocean re-analyses, combining satellite and *in situ* observations at regular two- and three-dimensional spatial grids (Assis et al. 2018). Minimal, maximal and mean values were used as descriptors and combined as suggested in the literature (Franklin 2010a, Bucklin et al. 2015). Environmental layers provide average monthly values for the present decade [2000-2014] at a spatial resolution of 5 arc-minutes (about 8 x 8 km) and describe monthly averages for the period 2000-2014.

Species distribution was also modelled according to two greenhouse gas emission scenarios, RCP 2.6 and RCP 8.5 scenarios for future decades 2040-2050 and 2090-2100 (IPCC 2014). Maps of projected changes in ocean conditions were downloaded from Bio-ORACLE (https://www.bio-oracle.org/index.php; Table 1). The RCP 2.6 scenario (Appendix 3.3) predicts an increase of mean seafloor temperatures of up to +0.7°C along the Argentinian coasts by 2100, +1.3°C in the Weddell Sea region, and +1.3°C on the northern Kerguelen Plateau. The RCP 8.5 scenario (Appendix 3.3) for decade 2040-2050 predicts that seafloor waters will warm up by +1 °C along the southern South American coasts and in the Weddell Sea, and for decade 2090-2100 predicts an increase of seafloor mean temperatures of up to +4 °C along the Argentinian coasts, +0.5 to +1 °C in the WAP, up to +3 °C on the northern Kerguelen Plateau and a predicted decrease of -0.5 to -1 °C in insular regions such as South Georgia and the South Orkney Islands. Salinity is predicted to decrease in the sub-Antarctic and Antarctic regions from -0.1 to -0.2 PSU unit for 2050 and 2100 scenarios respectively, with close tendencies between RCP 2.6 and RCP 8.5. Sea-ice thickness is predicted to reduce in some areas from a few centimetres to 0.6 m in RCP 2.6 scenario and up to

1.2 m for RCP 8.5 scenario, resulting in an expansion of ice-free areas in the Weddell Sea region (Fig. S3.3.A).

Primary production and oceanographic current speed for decades 2040-2050 and 2090-2100 were considered unchanged and similar to present-day conditions as there were no predictions available for these parameters.

Table 3.1. Environmental descriptors used for modelling and sources. Spatial resolution set at 5 arc minutes (around 8 km).

Descriptors	Present	Future	Source
Depth	-	-	GEBCO[1]
Roughness	-	-	Modified from Depth layer, 'raster' R package function *terrain*
Slope	-	-	Modified from Depth layer, 'raster' R package function *terrain*
Seafloor mean temperature	2000-2014	RCP 2.6 and 8.5 for 2050 and 2100	BioOracle[2]
Seafloor min temperature	2000-2014	RCP 2.6 and 8.5 for 2050 and 2100	BioOracle[2]
Seafloor max temperature	2000-2014	RCP 2.6 and 8.5 for 2050 and 2100	BioOracle[2]
Seafloor mean salinity	2000-2014	RCP 2.6 and 8.5 for 2050 and 2100	BioOracle[2]
Seafloor min salinity	2000-2014	RCP 2.6 and 8.5 for 2050 and 2100	BioOracle[2]
Seafloor max salinity	2000-2014	RCP 2.6 and 8.5 for 2050 and 2100	BioOracle[2]
Seafloor mean primary productivity	2000-2014	Same as present conditions	BioOracle[2]
Seafloor min primary productivity	2000-2014	Same as present conditions	BioOracle[2]
Seafloor max primary productivity	2000-2014	Same as present conditions	BioOracle[2]
Ice mean thickness	2000-2014	RCP 2.6 and 8.5 for 2050 and 2100	BioOracle[2]
Ice min thickness	2000-2014	RCP 2.6 and 8.5 for 2050 and 2100	BioOracle[2]
Ice max thickness	2000-2014	RCP 2.6 and 8.5 for 2050 and 2100	BioOracle[2]
Seafloor mean current	2000-2014	Same as present conditions	BioOracle[2]

1. https://download.gebco.net/, accessed February 2020
2. https://www.bio-oracle.org/index.php, accessed February 2020

In order to spot and remove extrapolation errors, the Multivariate Similarity Environmental Estimate (MESS, Elith et al. 2010) was computed based on the presence records (Guillaumot et al. 2019, 2020c - Chapter 2). The MESS provides an estimate of the range of environmental conditions under which species occurrences were found and used to calibrate the model. It is then used to select areas where model projections will be calculated, dismissing areas where environmental conditions are not met, and where the model extrapolates. This was helpful to prevent from projecting the model far from the conditions in which the species can be found (noteworthy for depth).

2.5. Model calibration

Species distribution models were generated using the Boosted Regression Trees (BRT) algorithm with the following settings, learning rate 0.005, bag fraction 0.9 and tree complexity 4. These settings minimize the model predictive deviance according to the tests generated following Elith et al. (2008) (Appendix 3.4). The R package 'gbm' was used to run the model (Ridgeway 2006, Elith et al. 2008). Models were calibrated using presence and absence data. Modelling performance was assessed using a spatial random cross-validation procedure adapted from Guillaumot et al. (2019 - Chapter 2) for model calibration using absence records (instead of background records). Also considering the limited number of occurrence records available and their patchy distribution at broad spatial scale, the occurrence dataset was randomly split into five spatial parts, with 80% (four parts) of the dataset used as a training subset and, 20% (one part) used as a test subset (Elith et al. 2008). The procedure was repeated 20 times to generate a set of 100 model replicates. The proportions of presence and absence data falling into areas predicted as suitable and unsuitable for the species distribution was evaluated to assess modelling performance. Modelling performance was also assessed using the Area Under the Curve (AUC, Fielding and Bell 1997), the True Skill Statistics (TSS, Allouche et al. 2006) and the Biserial Correlation metrics (COR, Elith et al. 2006).

2.6. Model outputs

Model predictions were projected on the entire study area (Longitude: -76°E to 178°W, Latitude: -35°S to -68°S) with a focus on areas where the species is mainly reported presently and where it may be expected in the future, in southern South America, the Scotia Arc and the WAP, the WAP alone, and the Kerguelen Plateau.

3. RESULTS

3.1. Survival rate in the temperature experiment

One individual died the next day after reaching the target temperature in the -1.8°C temperature experiment. Survival rate at -1.8°C reached 0% on day 11. Survival reached 0% on day 15 at -1.0 ºC. Survival rate at 0 ºC was 52% on day 27 and 0% on day 59. Survival rates were 60% at 1°C, 75% at 2°C and 95% at 5°C on day 90 (Fig. 3.3).

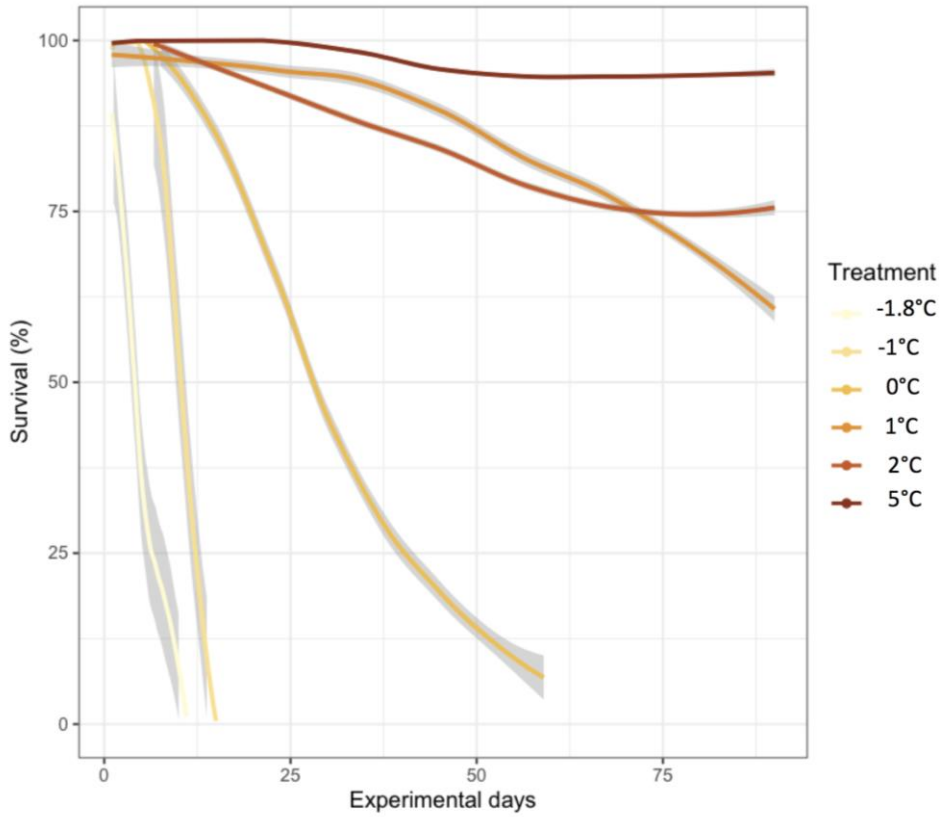


Figure 3.3. Survival rates of adults of *Halicarcinus planatus* at different temperatures over 90 days.

3.2. Survival rate of adults and larvae in the salinity experiment

Survival rate in the salinity experiment at 5 PSU was 0% on day 2. Survival rate at 11 PSU was 0% on day 14. Survival rate was 50% on day 36 and 36% on day 39 at 18 PSU. Interestingly, survival rate was over 50% (67%) for the experiment at 18 PSU performed with seawater from Skyring Sound. Survival rates were 95% on day 39 at 25 PSU and 30 PSU (Fig. 3.4).

Females collected on July 5 were ovigerous and released larvae at the end of August at 5 ºC, 30 PSU and 25 PSU. The survival rate of larvae at 1 ºC was 62.5% on day 12. Survival rates at 2 ºC and 5 ºC were 85% and 92.5%, respectively, on day 12 (Fig. 3.5).

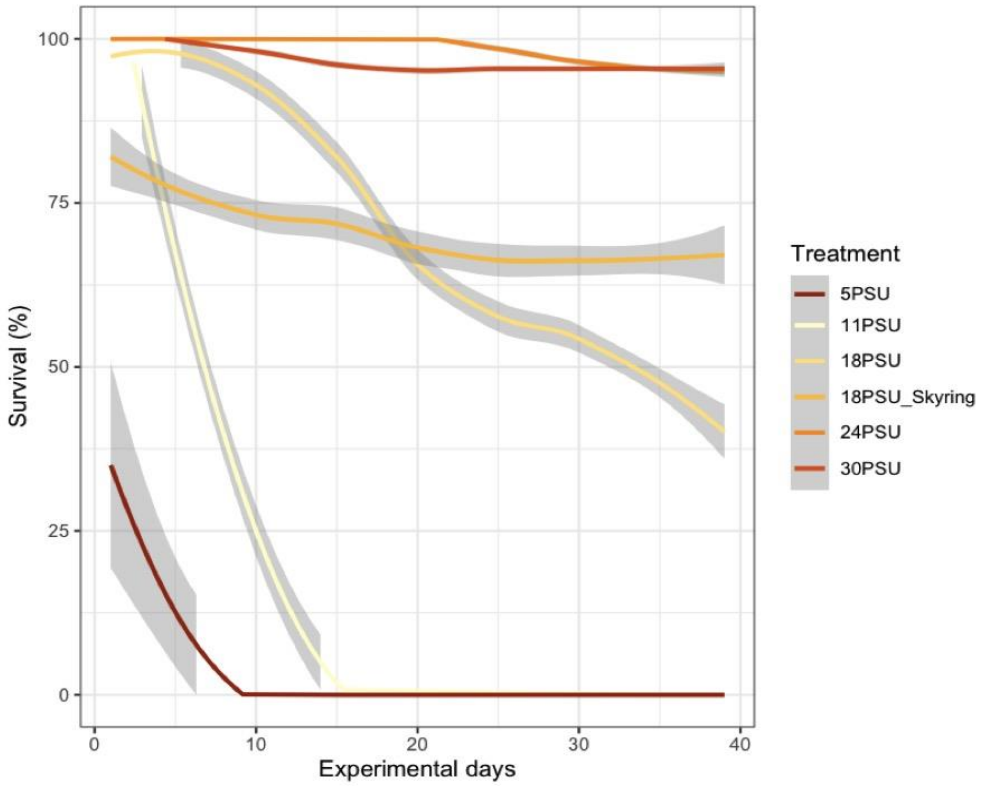


Figure 3.4. Survival rates of adults of *Halicarcinus planatus* at different salinities over 39 days.

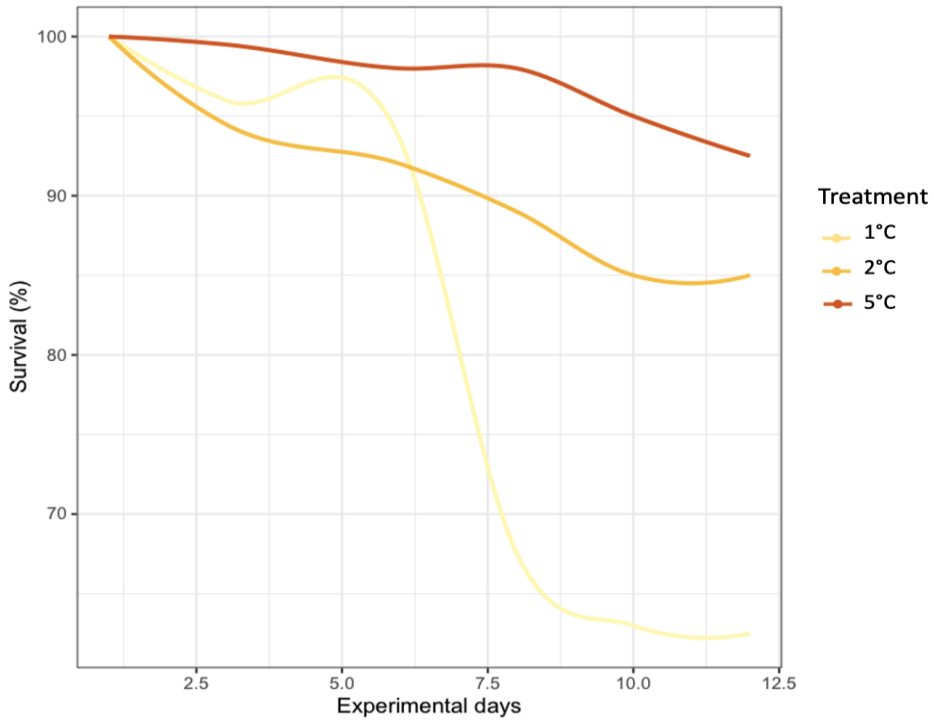


Figure 3.5. Survival rates of larvae of *Halicarcinus planatus* for 12 days at different temperatures.

3.3. SDM predictions under current environmental conditions [2000-2014]

SDMs showed high AUC scores of 0.947 ± 0.059, TSS of 0.795 ± 0.123 and COR of 0.873 ± 0.070. Correctly classified test data also reach high scores (89.9 ± 0.3 % for presence test records and 92.9 ± 2.2 % for absence test records). The proportion of areas where the model extrapolates

is very high (86.3%, Fig. 3.7) highlighting again the relevance of using the MESS method as recommended by Guillaumot et al. (2019, 2020c - Chapter 2).
Mean sea-ice thickness (40.1 ± 3.2%) and seafloor temperatures are the two main drivers of the species distribution (mean, maximal, and minimal seafloor temperatures with 37.8 ± 3.7, 7.6 ± 1.9 and 6.9 ± 2.4% contribution to the model, respectively; Table 3.2), with suitable areas corresponding to low sea-ice coverage (<0.1%) and minimum temperatures over +2°C (Fig. 3.6). These environmental values match perfectly with the latitudinal partition in the distribution of *H. planatus*, with warmer temperatures (> +2°C) and lower ice coverage (< 0.1%) at the lower latitudes associated with most presence records and few absences, and in contrast, colder temperatures (< +2°C) and thicker sea ice coverage (> 0.1%) associated with absence records. Interestingly, primary production is not a good predictor of the species distribution (<1%).
As occurrence records are mainly distributed in coastal shallow-water areas, depth does not contribute much to the model as no contrast in bathymetry values are present in the dataset. Slope and roughness have probably more contrasting values in deep-sea habitats and consequently do not significantly contribute to the model (< 0.2%).

Table 3.2. Average contribution values and standard deviation (SD) of the 16 environmental descriptors to model predictions.

Descriptor	**Mean ± SD (%)**	**Descriptor**	**Mean ± SD (%)**
Mean Ice thickness	40.1 ± 3.2	Mean seafloor primary production	0.8 ± 0.1
Mean Seafloor temperature	37.8 ± 3.7	Max seafloor primary production	0.5 ± 0.02
Max Seafloor temperature	7.6 ± 1.9	Depth	0.5 ± 0.05
Min Seafloor temperature	6.9 ± 2.4	Slope	0.2 ± 0.06
Min Seafloor salinity	1.4 ± 0.2	Roughness	0.1 ± 0.03
Mean Seafloor salinity	1.4 ± 0.1	Max Seafloor salinity	0.1 ± 0.03
Mean Seafloor current speed	1.3 ± 0.2	Max seafloor primary production	0.001 ± 0.001
Max Ice thickness	1.1 ± 0.1	Min Ice thickness	0

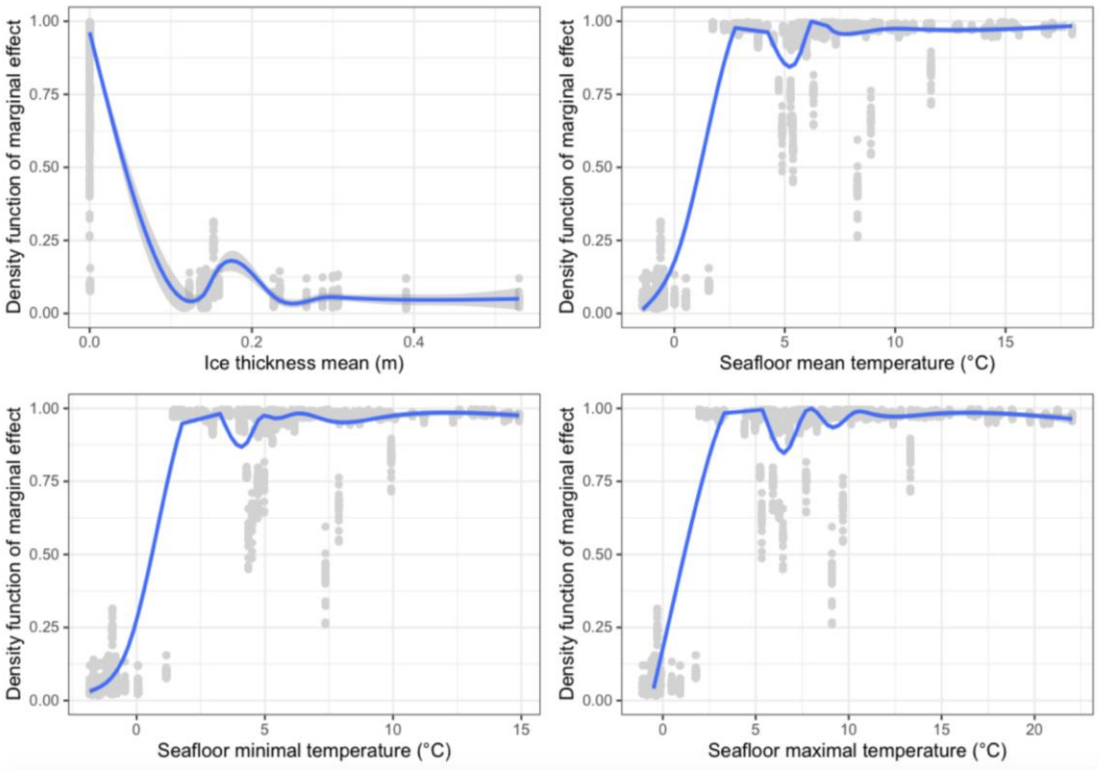


Figure 3.6. Partial dependence plots for the four environmental descriptors that contribute the most to the model. Scaled density distribution of the marginal effect of the descriptors to the model, data points (grey) fitted with a generalized additive model (GAM, blue line).

The extrapolation mask importantly reduces the projected area to shallow habitats (Fig. 3.7). Distribution probabilities predicted by the model were the highest in southern South America, New Zealand and Australia and most sub-Antarctic Islands (Kerguelen, Heard, Marion, Bouvet and South Sandwich Islands; Fig. 3.7A). Interestingly, the model predicts an intermediate probability of distribution in South Georgia, for which a single absence was reported (Fig. 3.2), and a high probability on Heard Island, where no occurrence data have been reported yet. The WAP is predicted as unsuitable to the survival of *H. planatus*, as in the case of Deception Island (Fig. 3.7C).

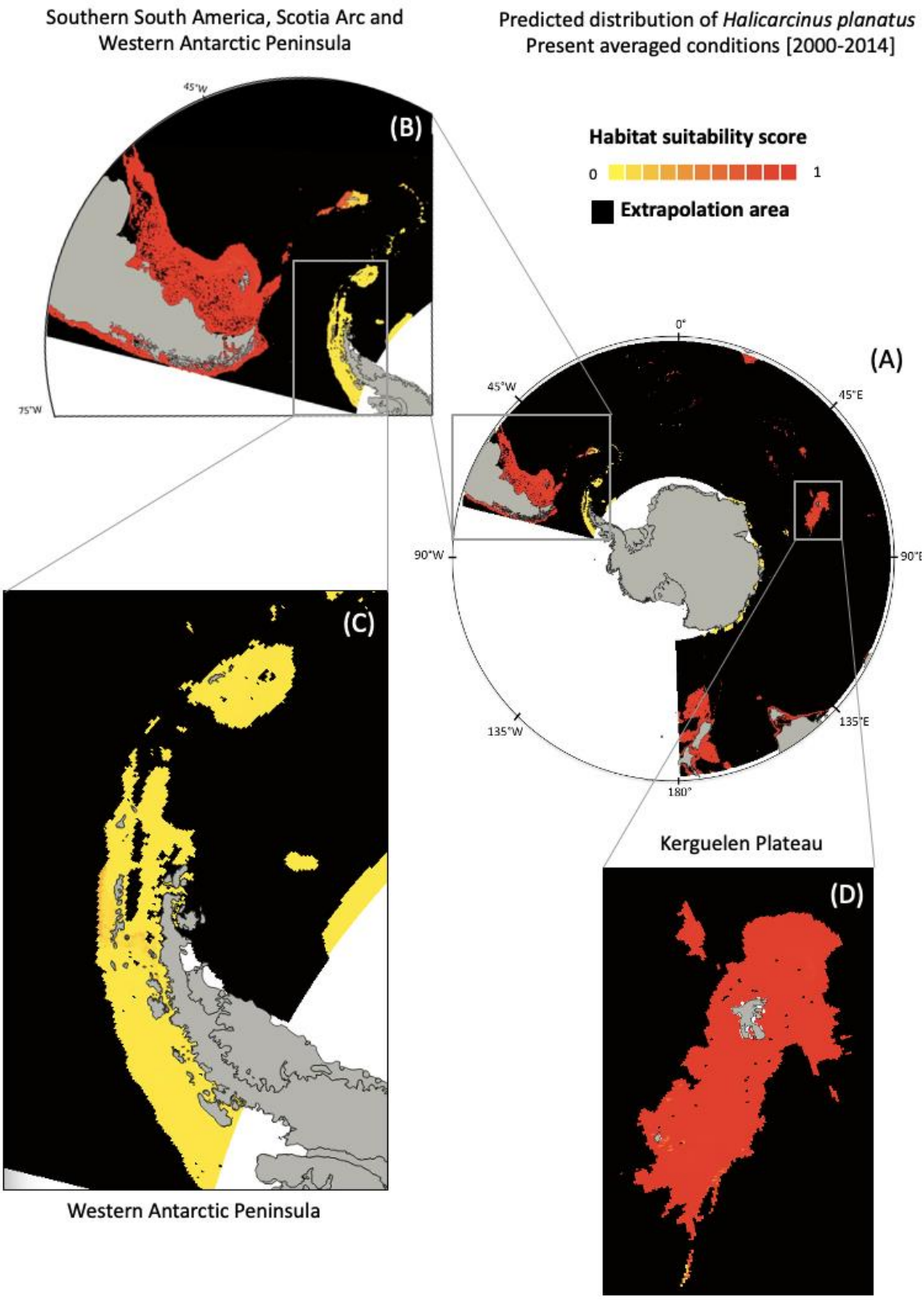


Figure 3.7. SDM predictions of presence probability (contained between 0 and 1) for *Halicarcinus planatus,* projected under current environmental conditions [2000-2014] for the entire Southern Ocean (A), and with a focus on southern South America, the Scotia Arc and the Western Antarctic Peninsula (WAP) (B), the WAP alone (C), and the Kerguelen Plateau (D). Black pixels correspond to extrapolation areas for which predictions are not reliable and were removed from projection (according to the Multivariate Environmental Similarity Surface index, MESS).

3.4. SDM predictions under future environmental conditions

SDM future predictions under RCP 2.6 in decades 2040-2050 and 2090-2100 predict low and intermediate probability of *H. planatus* to settle in South Georgia, Elephant Island and the WAP respectively (Fig. 3.8, 3.9). The RCP 8.5 scenario shows an increase in probability for *H. planatus* to survive in the WAP (Fig. 3.10, 3.11). Models predict higher presence probabilities compared to present-day predictions in South Georgia and the South Shetland Islands for both decades 2040-2050 and 2090-2100, with the highest values predicted in the northern tip of the South Shetland Islands. The South Orkneys are not predicted as suitable by 2040-2050, but some patches of suitable areas appear by 2090-2100.

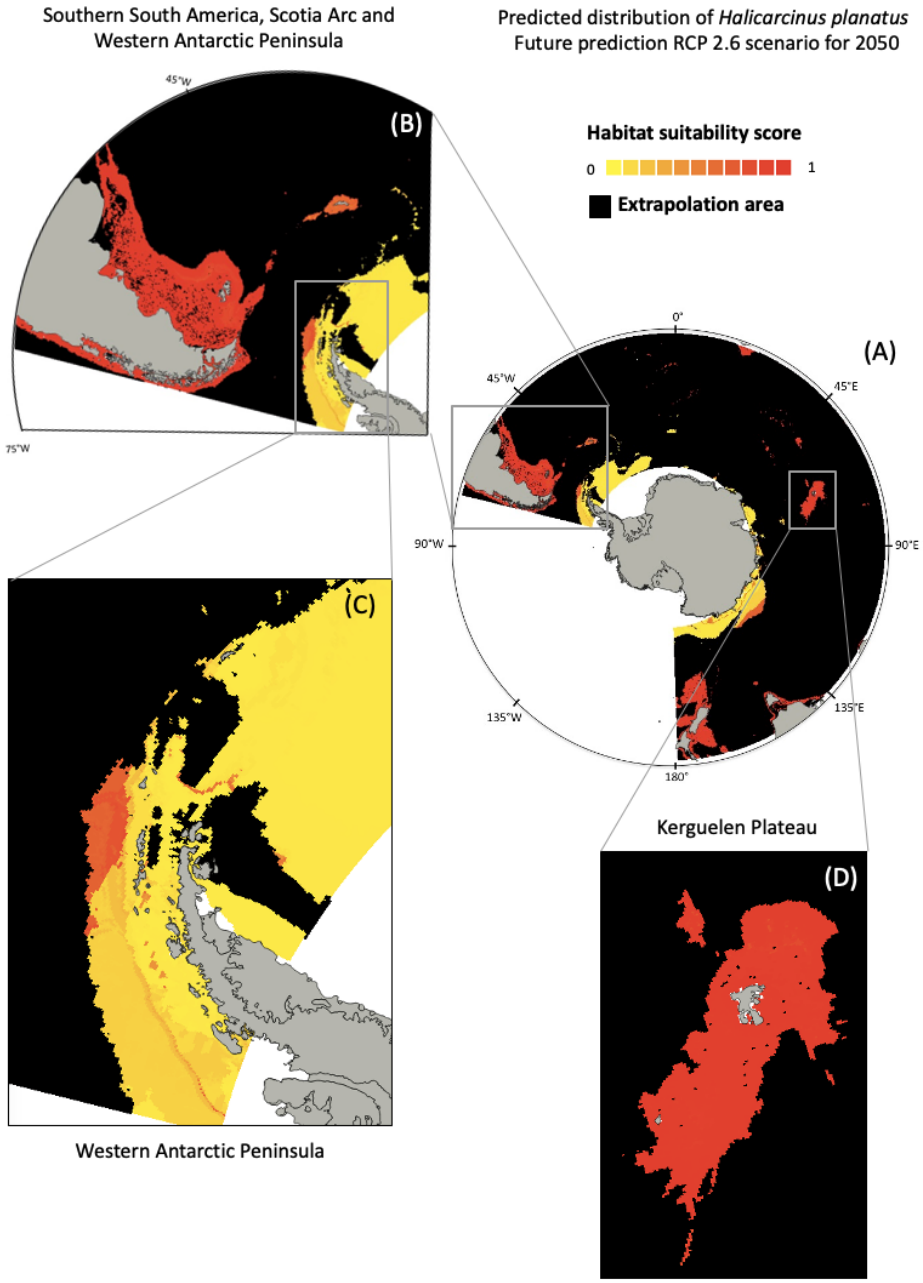


Figure 3.8. SDM predictions of presence probability (between 0 and 1) for *Halicarcinus planatus,* projected under environmental conditions IPCC RCP 2.6 climate scenario for 2050 for the entire Southern Ocean (A), with focus on southern South America, the Scotia Arc and the Western Antarctic Peninsula (WAP) (B), the WAP alone (C), and the Kerguelen Plateau (D). Black pixels correspond to extrapolation areas for which predictions are not reliable and were removed from projection (according to the Multivariate Environmental Similarity Surface index, MESS).

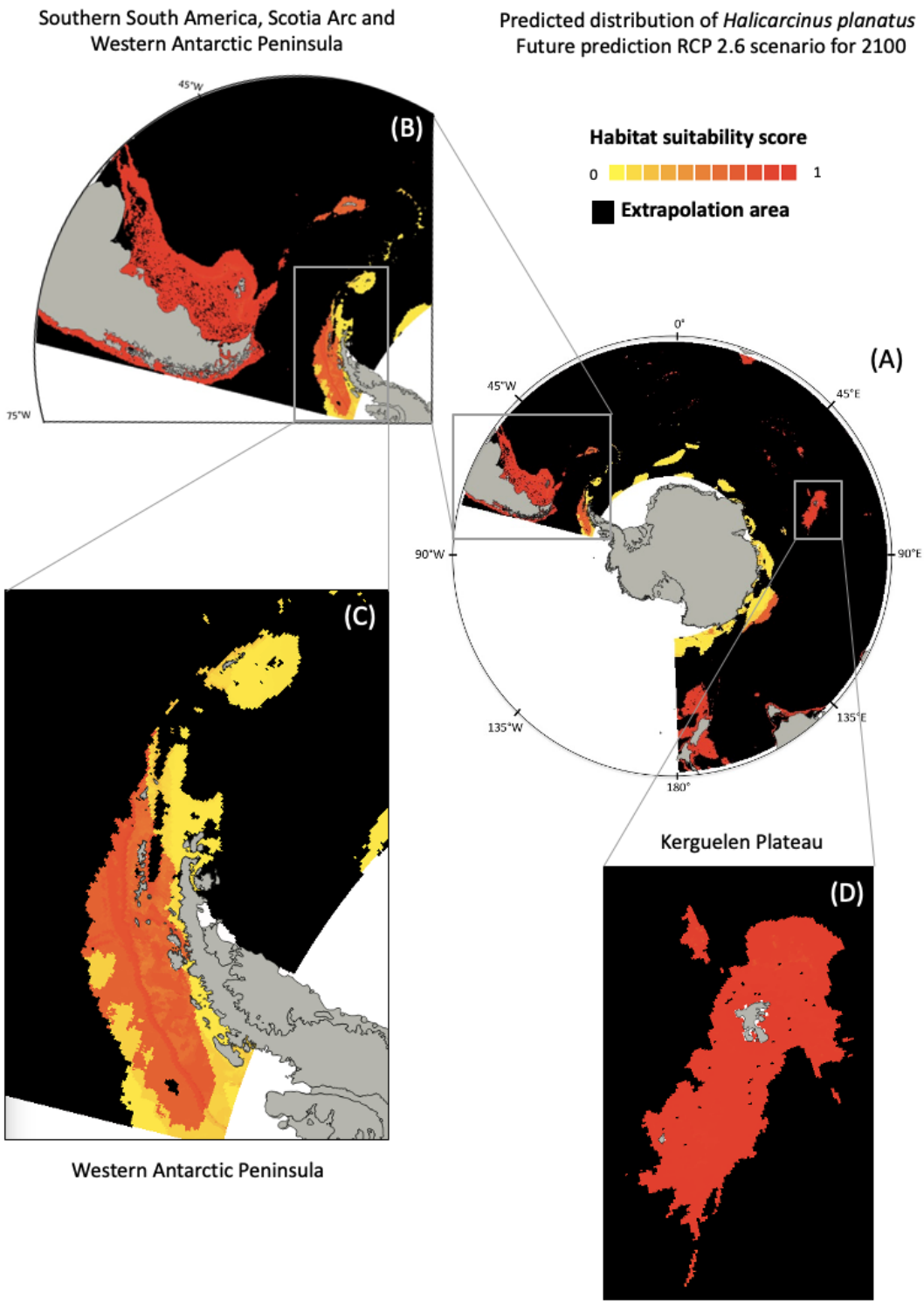


Figure 3.9. SDM predictions of presence probability (between 0 and 1) for *Halicarcinus planatus,* projected under environmental conditions IPCC RCP 2.6 climate scenario for 2100 for the entire Southern Ocean (A), with focus on southern South America, the Scotia Arc and the Western Antarctic Peninsula (WAP) (B), the WAP alone (C), and the Kerguelen Plateau (D). Black pixels correspond to extrapolation areas for which predictions are not reliable and were removed from projection (according to the Multivariate Environmental Similarity Surface index, MESS).

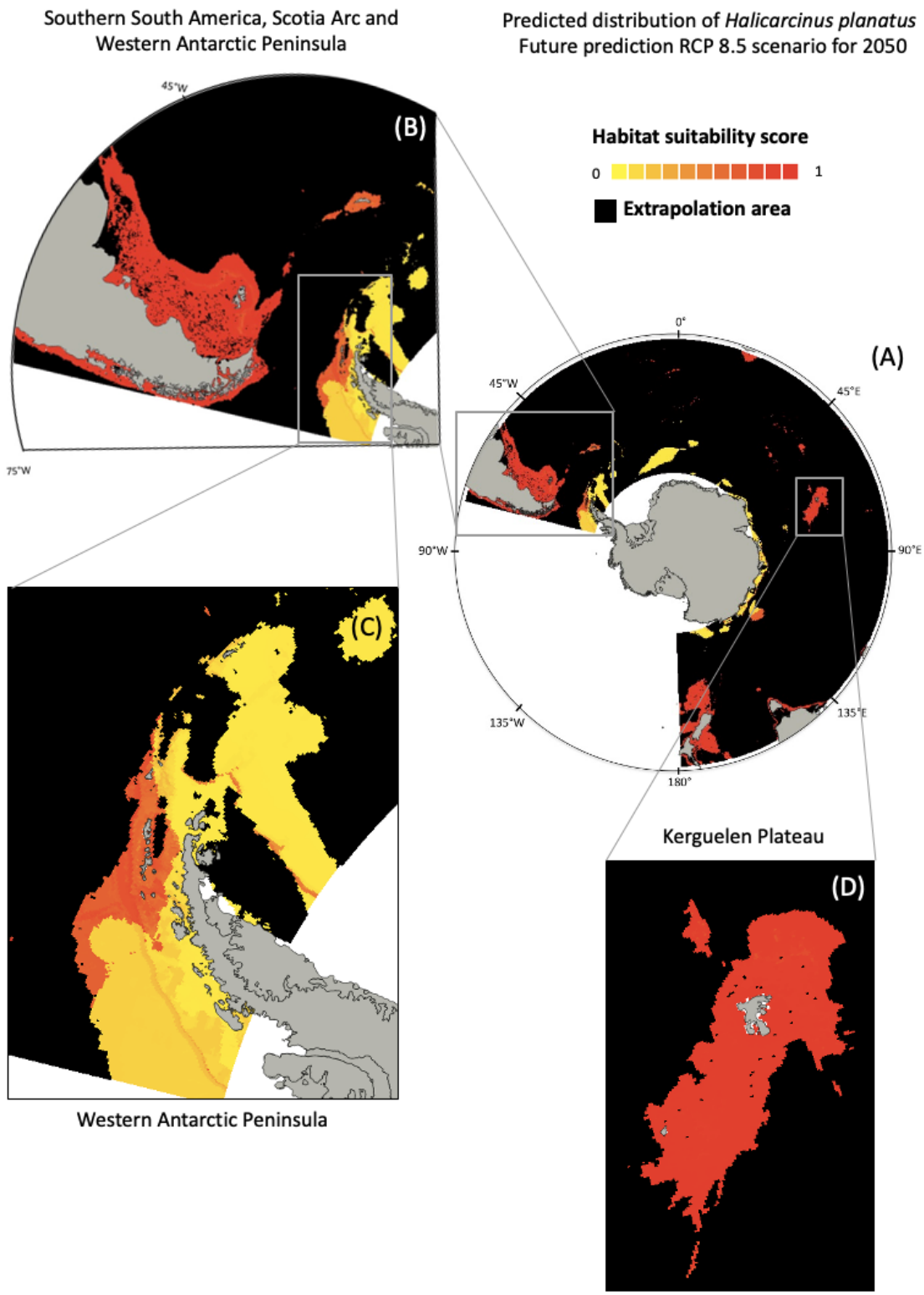


Figure 3.10. SDM predictions of presence probability (between 0 and 1) for *Halicarcinus planatus,* projected under environmental conditions IPCC RCP 8.5 climate scenario for 2050 for the entire Southern Ocean (A), with focus on southern South America, the Scotia Arc and the Western Antarctic Peninsula (WAP) (B), the WAP alone (C), and the Kerguelen Plateau (D). Black pixels correspond to extrapolation areas for which predictions are not reliable and were removed from projection (according to the Multivariate Environmental Similarity Surface index, MESS).

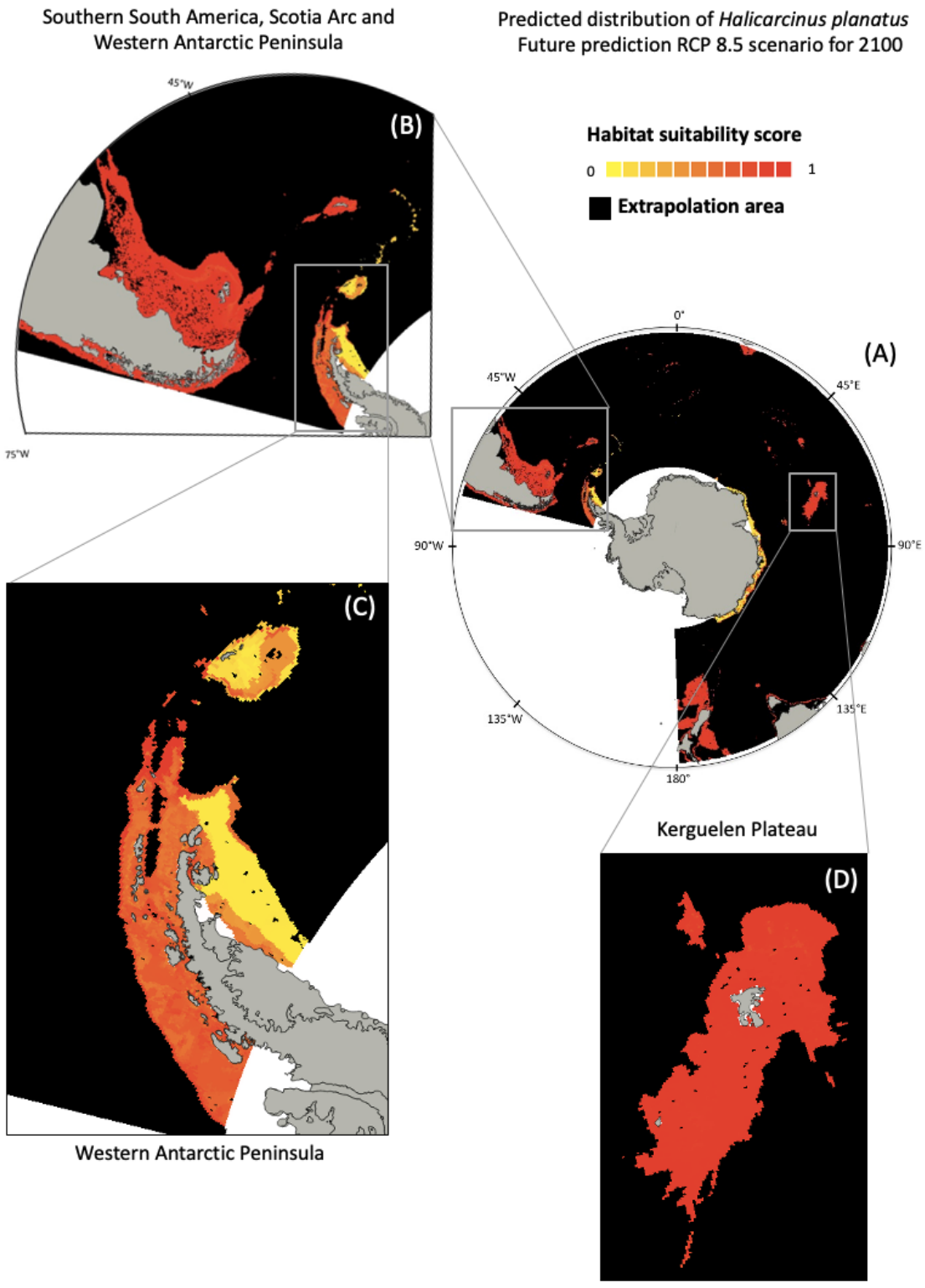


Figure 3.11. SDM predictions of presence probability (between 0 and 1) for *Halicarcinus planatus,* projected under environmental conditions IPCC RCP 8.5 climate scenario for 2100 for the entire Southern Ocean (A), with focus on southern South America, the Scotia Arc and the Western Antarctic Peninsula (WAP) (B), the WAP alone (C), and the Kerguelen Plateau (D). Black pixels correspond to extrapolation areas for which predictions are not reliable and were removed from projection (according to the Multivariate Environmental Similarity Surface index, MESS).

4. DISCUSSION

This study combines physiological and ecological modelling approaches to highlight the increased risk of marine incursions into Antarctic coastal ecosystem over the coming century. Specifically, we reveal that the widespread sub-Antarctic decapod *H. planatus* has significant potential to establish in Antarctic waters under realistic climate change scenarios in the coming decades. More broadly, this prospect of future marine introductions into Antarctic ecosystems potentially has crucial implications for the conservation of endemic Antarctic coastal assemblages. Indeed, over recent decades an increasing number of non-native marine taxa has been reported from Antarctic ecosystems, including: *Rochinia gracilipes* in the South Shetland Islands (Griffiths et al. 2013); *Bugula neritina* (Bryozoa) off Dronning Maud Land (East Antarctica) in the 1960s (McCarthy et al. 2019); *Hyas araneus* (Decapoda) from Elephant Island in the 1980s (McCarthy et al. 2019); *Ectopleura crocea* (Hydrozoa) off Dronning Maud Land and off Queen Mary Land (East Antarctica); and *Ciona intestinalis* (Ascidiacea) off Dronning Maud Land (East Antarctica) in the 1990s (McCarthy et al. 2019). Newer records since 2000 include *Emerita* sp. and *Pinnotheres* sp. (larval stage) in King George Islands in the 2000s (Thatje and Fuentes 2003); *H. planatus* from Deception Island (Aronson et al. 2014); *Membranipora membranacea* (Bryozoa) and *Macrocystis pyrifera* (Brown algae) from Deception Island (Avila et al. 2020); *Durvillaea antarctica* (Brown Algae) from King George Island (Fraser et al. 2018) and Livingston Island (Avila et al. 2020); and *Mytilus* cf. *platensis* (Bivalvia) in King George Island (Cárdenas et al. 2020) in the 2010s.

There are potentially several different modes of dispersal for species to reach Antarctica. Fraser et al. (2018) and Avila et al. (2020) identified dispersal by rafting on buoyant kelps as a possible mechanism for the arrival of non-native species to Antarctica. The former study also included a Lagrangian analysis to show that particles released from South Georgia and the Kerguelen Islands were able to drift across the Polar Front and reach Antarctic coasts following strong storm events. According to this model, storm conditions may enable buoyant kelps to reach the WAP. Such conditions may not be rare, as remains of the kelp *D. antarctica* were observed onshore in the WAP in 2019 and 2020 (López-Farrán personal observation). Direct observations (from southern New Zealand) of *Halicarcinus* adult individuals associated with *D. antarctica* holdfasts, and also in detached, drifting *D. antarctica* at sea (Waters unpublished data) imply rafting as a direct mechanism for adults of this decapod taxon into Antarctic waters. Anthropogenic activities may also be potential dispersal vectors for this decapod (Avila et al. 2020, Cárdenas et al. 2020 ; e.g. via ship hulls, ballast waters, outdoor and personal equipment of tourists or oceanographic equipment of scientists).

No established non-native marine species have as yet been observed in Antarctica, suggesting that physiological barriers may be key in preventing such invasions (Richardson et al. 2000). In this study, we combined two independent approaches to define the environmental and geographical boundaries of *H. planatus* distribution under present and future environmental conditions. SDM provides an estimate of a species' 'realised niche' (Hutchinson 1957, Soberón and Peterson 2005, Soberón 2010). The thermal limit of *H. planatus* established in this study, corresponds to the coldest conditions of its sub-Antarctic distribution, located in the Kerguelen Islands, where subtidal temperature ranges between +1.1 and +3.0 °C during the Austral winter (Richer De Forges 1977, Lucas 1980, Féral et al. 2019). This species can therefore potentially endure summer conditions in WAP (1ºC and above) in a wide range of salinity (between 18 PSU and 30-33 PSU), but would not survive during the cold winter months. Our experimental results may indicate that Antarctic seawater temperatures may impede larval development even during the summer, suggesting that this species is not able to complete its development in Antarctica under present conditions. In parallel, the survival rates among larvae were 65%, 85% and 92% at 1, 2 and 5°C, respectively, thus coinciding with adult rates, and confirming the sensibility to low temperature mentioned by Pörtner and Farrel (2008), confirming that at that stage larvae and adults can survive during Antarctic summer only.
Halicarcinus planatus has previously been highlighted as a potential invader of Antarctica (Diez and Lovrich 2010), because of its potential to live in cold waters, through regulation of $[Mg^{2+}]_{HL}$. However, the present results demonstrated that this physiological characteristic is not sufficient to survive the sub-zero temperatures that typify current Antarctic winters (Fig. 3.3). The finding that

brachyuran crabs cannot currently establish in Antarctica may also help to explain their extinction from shallow Antarctic habitats from the mid-Miocene, ~14 million years ago, when ACC intensification led to cooling and the establishment of a perennial sea-ice cover in the region (Thatje et al. 2005b, Zachos et al. 2008, Hansen et al. 2013, Crampton et al. 2016). Numerous marine lineages including brachyurans, lobsters and sharks disappeared from Antarctic waters, along with most teleosteans except for cold-adapted nototheniids and liparids (Aronson and Blake 2001, Clarke et al. 2004, Aronson et al. 2007). The simultaneous extinction of these diverse taxa was presumably driven by their lack of physiological tolerance to cold conditions (Frederich et al. 2001, Clarke et al. 2004, Aronson et al. 2007). Together, these data may highlight the crucial role of thermal barriers in preserving the integrity of Antarctic coastal ecosystem.

Under future warming scenarios with increased seawater temperatures and shortened sea-ice seasons, physiological barriers to Antarctic incursions are projected to weaken. For example, near Palmer Station the ice season decreased by 92 days from 1979/80 to 2012/13 (Meredith and King 2005, Ducklow et al. 2013). According to IPCC RCP scenarios the WAP will continue to warm (Appendix 3.3), facilitating the establishment of alien species already arriving. *Halicarcinus planatus* is not able to establish in the WAP under present conditions because it is not a suitable environment (Table 3.2, Fig. 3.6, 3.7), however this may change in the future. In the South Shetland Islands, the worst scenario RCP 8.5 predicts a decrease in ice thickness, the expansion of ice-free areas (Appendix 3.3) and a 1 to 2°C increase of seafloor temperature in 2100, leading to suitable conditions for *H. planatus* establishment. SDM predictions indicate the highest suitability for *H. planatus* presence in South Georgia and some places of the WAP (Fig. 3.11B,C). The most optimistic climate change scenario RCP 2.6 predicts in 2100 a rise of seafloor water temperature of 0.4 ºC in the South Shetland Islands, resulting in intermediate SDM predictions in the WAP and South Georgia (Fig. 3.9B,C). Thus, according to these future scenarios, it is just a matter of time before the WAP would reach suitable environmental conditions for *H. planatus.*

Survival is not the only requirement for the establishment of a species in a new area. A successful invasion also implies developing, reproducing and then dispersing to new places (Richardson 2000), and active behaviour to escape, feed and mate (Frederich et al. 2001). According to SDM predictions and the thermotolerance experiment, a successful invasion would be possible in an environment at +2ºC. Deception Island is the most active volcanic island of the South Shetland Islands, where many subtidal hydrothermal points and geothermal activity offer various temperatures that could favour the establishment of non-native species (Agusto et al. 2004), converting Port Foster into a key location for alien species colonization (Aronson et al. 2014, Avila et al. 2020). During three scuba diving campaigns between 2017 and 2019, we searched for *H. planatus* in several places in the WAP, including where it was collected in 2010 - shallow waters off Baily Head outside the caldera of Deception Island (Aronson et al. 2014) - and other active sites (within the caldera of Deception Island, in Penguin Island (South Shetland Islands), and Paulet Island in the Weddell Sea), or inactive sites like King George Island (South Shetland Islands), Doumer Island, Roberts Island, Coppermine Peninsula, Chile Bay in Greenwich Island, among other places, and none were found. This absence agrees with our results, but contradicts the presence of the ovigerous female in Deception Island (Aronson et al. 2014), which would need at least two years to reach pubertal molt, the time required in the Kerguelen Islands (Richer De Forges 1977). This female certainly would not have grown up *in situ*; this place being on the outer coast under full Antarctic conditions (without geothermal activity or hydrothermal influence typical of the interior of Deception Island). Our results suggest its arrival at the mature stage or maybe the ovigerous stage, implying that its development was completed elsewhere. An arrival through rafting is also unlikely. Early stages of *H. planatus* have been observed in floating kelps (*Macrocystis pyrifera*) in the Internal Sea of Chiloé (Hinojosa et al. 2010), and kelps have been reported in Deception Island (Avila et al. 2020). However, the journey from the sub-Antarctic area to the WAP implies two years across the SO riding kelps, which is highly improbable. This female was more probably brought through the Drake by ship during the southern 2009-2010 summer; the extraction of an adult crab together with kelps frond and holdfast wrapped around an anchor is quite likely (K. Gérard pers. observation).

The establishment of non-native marine species in Antarctica is an issue that is becoming more pressing. The composition of the community may change dramatically according to which species establishes. Antarctica is characterised by the absence of durophagous predators (bony and cartilaginous fishes and brachyurans) and short food webs. Therefore, according to Aronson et al. (2007), the arrival of a reptant crab may affect Antarctic ecology and the biodiversity of the shallow Antarctic. With the arrival of the invasive red king crab *Paralithodes camtschaticus* in the Barents Sea, reductions of diversity and benthic biomass were observed as a result of the predation pressure (Falk-Petersen et al. 2011), as well as shifts in interspecific competition (David et al. 2017, Britton et al. 2018) and infection of native species by parasites associated with invaders (Bevins 2019). Although the effects of invasive species are impossible to predict, the return of durophagous predators such as decapods, chondrichthyans and teleosteans in Antarctic shallow waters has been feared, because they may cause shifts in benthic communities, modifying trophic relationships and homogenizing the Antarctic ecosystem (Aronson et al. 2007, 2015). However, *H. planatus,* with its small size, opportunistic feeding behaviour and soft exoskeleton, is definitely not a top predator (Boschi et al. 1969). It feeds on phytoplankton remains accumulated at the bottom, such as carrion, detritus, mucopolysaccharides from algae and small soft individuals, even of its own species (López-Farrán, personal lab. observation). They are prey for fishes (as *Harpagifer bispinis*, *Patagonotothen tessellata* and *Austrolycus depressiceps*; Diez et al. 2011), birds, crabs and sea stars, among others, and look for refuge among rocks and kelp holdfasts to survive (Richer De Forges 1977, Chuang and Ng 1994, Vinuesa and Ferrari 2008). *Halicarcinus planatus* is part of the sub-Antarctic ecosystem, playing a fairly important role in food webs (Richer De Forges 1977, Diez et al. 2011). However, as it is not considered as a keystone or a bioengineer species, its establishement would not affect the Antarctic ecological community significantly. Although the effects of introduced non-native species are impossible to project, *H. planatus* may just incorporate into the already well-represented detritivorous guild of the WAP shallow benthic ecosystems. Under warmer conditions (2°C), the increase of seawater temperature would affect the WAP ecosystem more intensively than the arrival of a small soft-shelled detritivorous brachyuran such as *H. planatus* (Turner et al. 2014, Ashton et al. 2017, Clark et al. 2019). An example of a bioengineer species that would change the intertidal and shallow subtidal in the WAP is *Mytilus* cf. *platensis*, a non-native species recorded in 2019 (Cárdenas et al. 2020). Mussels have the capacity to provide dense three-dimensional matrices (Alvarado and Castilla 1996) that persist for long periods, constituting a micro-habitat which reduces desiccation during low tides, offering a stress-free space for small fish, invertebrate and alga species (Prado and Castilla 2006).

Antarctic water temperature continues to rise and stirs up the debate on the potential establishment of incoming species through transport on ship hulls, in ballast waters or on floating kelps (Aronson et al. 2014, Avila et al. 2020). Maritime traffic and tourism have increased the footprint and intensity of human activity within Antarctica (Kruczek et al. 2018, Hughes et al. 2019), raising the pressure of propagules in marine Antarctica, and probably this will continue to increase in next years (Kruczek et al. 2018). However, the involuntary introduction of non-native species to the Antarctic region and the movement of species and/or individuals within Antarctica from one zone to any other are among the highest priority issues considered for the Committee for Environmental Protection (CEP) and the Scientific Committee for Antarctic Research (SCAR). Therefore, a strong effort has been invested to improve the ballast water management of ships in Antarctica and to develop a strategy for biofouling (MEPC 2011).

Regardless of whether *H. planatus* individuals are able to reach the WAP by themselves or not, the SDM projected under conditions of IPCC RCP 2.6 or 8.5 climate scenarios indicate that individuals could survive and settle, either sooner (Fig. 3.8, 3.10) or later (Fig. 3.9, 3.11) in the future depending on the warming rapidity. *H. planatus* is highly abundant around Punta Arenas and Ushuaia, two frequently used harbours for the ships with WAP destination (Cárdenas et al. 2020). Therefore, if the vectors of *H. planatus*, ship or rafting, persist (Hinojosa et al. 2010, Aronson et al. 2014, Avila et al. 2020), some stages (larval, juvenile or adult) may reach the WAP, survive and settle.

SDMs are tuned to generate a simple spatial representation of the occurrence of a species based on environmental variables (Guisan and Zimmermann 2000 Mateo et al. 2011). Our results rely on

models that simplify of complex facts (Mateo et al. 2011), and make assumptions on future conditions. Beside temperture, ice thickness and salinity, there are many other variables that may vary over time and influence species distribution, such as primary production and ocean currents. Although niche models do not include eco-evolutionary parameters such as adaptation, gene flow or dispersal capacity, they are widely used to provide an insight into present and future species distribution (Thuiller et al. 2004, Titeux et al. 2017). Combining such results with information on biological interactions, physiology, anthropic influence on individual introductions or a complete evaluation of the dispersal capacities of *H. planatus* using a spatial and dynamic approach would fill knowledge gaps about their real invasive capacities in future environmental conditions.

In conclusion, our results suggest that *H. planatus* cannot presently establish in WAP waters, but this situation has a very strong probability to change under projected climate change in the 21st century. While the full consequences of Antarctic warming are yet to be realised, some changes in the distribution and composition of communities have already been observed (Turner et al. 2014, Ashton et al. 2017, Clark et al. 2019). The key for future studies will be to track species distribution and demographic shifts directly as warming continues, to help understand and mitigate marine biological impacts on Antarctic coastal ecosystem.

APPENDIX 3.1. Survival experiments

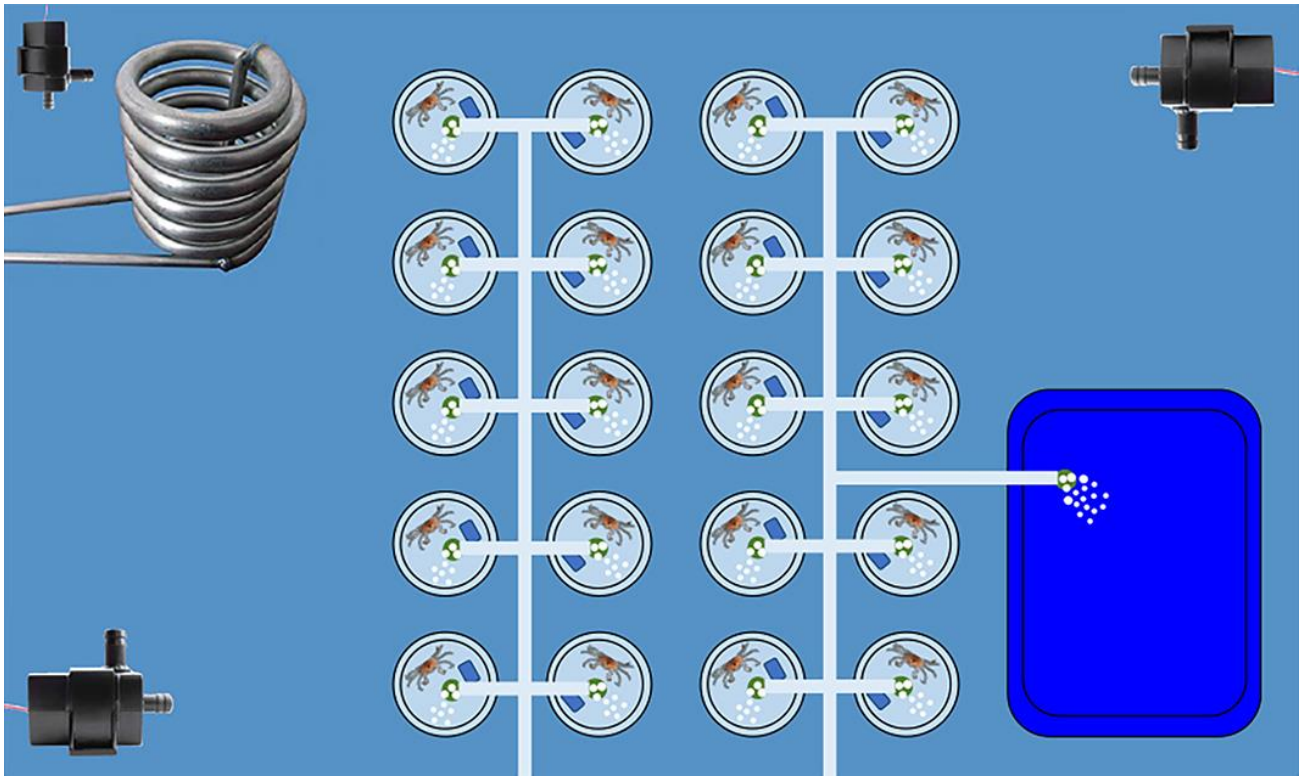


Figure S3.1. Schematic representation of the 6 containers (and their content) for the thermo-tolerance experiment on adult specimens. Big containers (light blue) were filled with seawater until 10 cm depth. A cold exchanger (top-left metallic curled object) sets the temperature of the whole container. Three water pumps (top-left, top-right and bottom-left black symbols) spread the cold water across the whole container. Each of the 20 glass jars contained one *H. planatus* specimen, a refuge (2 cm-long PVC tube) and an aerator. A smaller container (dark blue) of 10L with an aerator was used as clean water supply at the corresponding temperature.

APPENDIX 3.2. Occurrence distribution and sampling effort

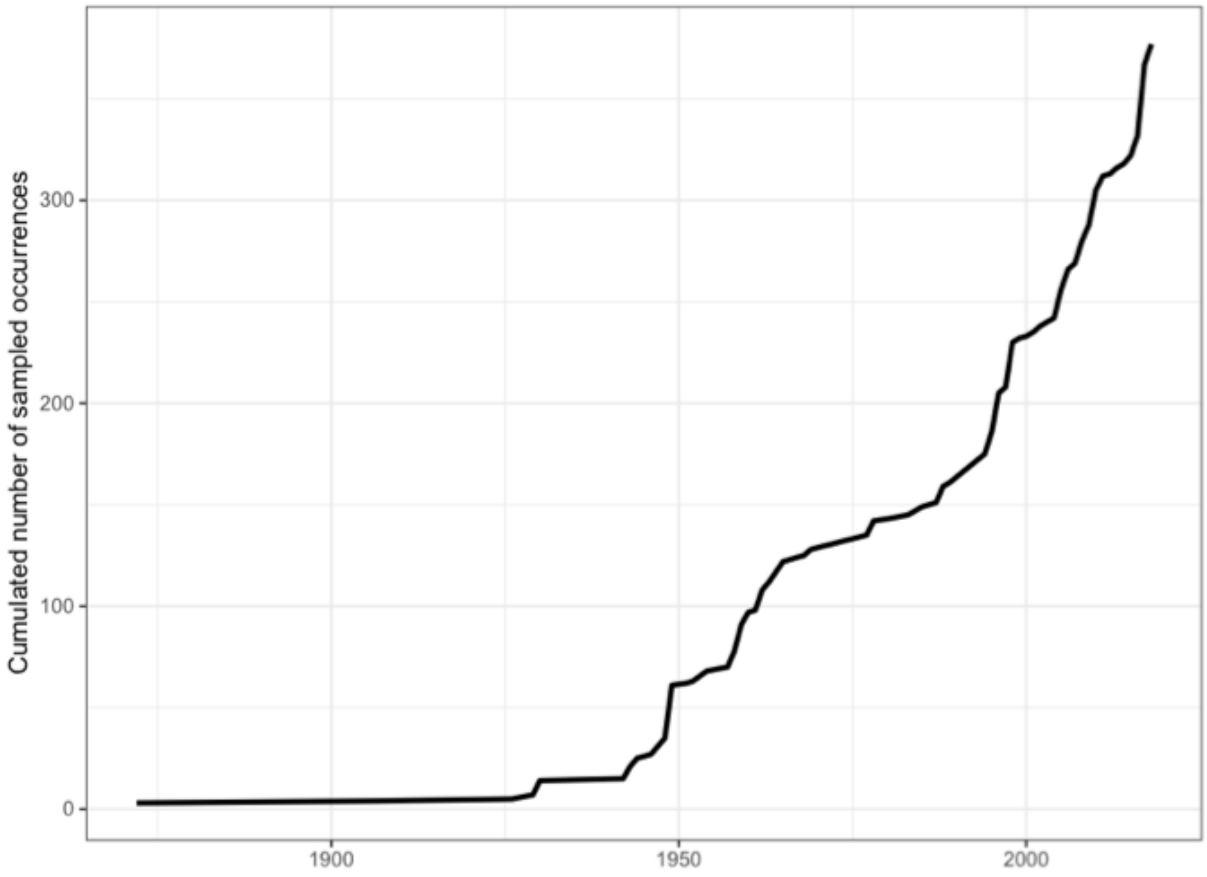


Figure S3.2. Increase of occurrence records through sampling and human observations through time (years).

APPENDIX 3.3. IPCC climate scenarios

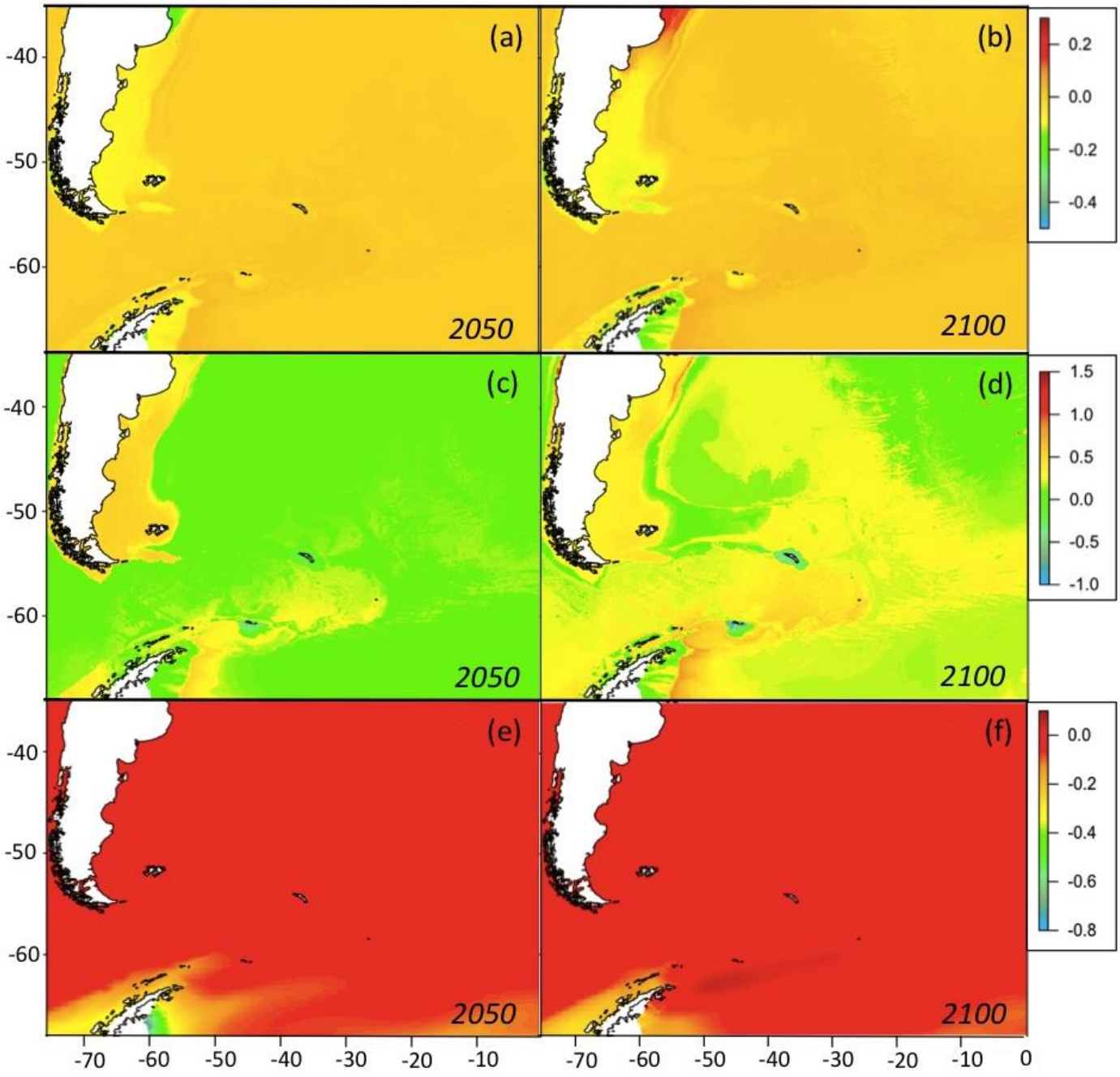


Figure S3.3.A. Focus on the Western Antarctic Peninsula and southern South America. Differences in seafloor salinities in PSU (a-b), seafloor temperatures in °C (c-d) and ice thickness in metre (e-f); between predicted future scenarios RCP 2.6 (mean values) for 2050 and 2100 and present environmental conditions (mean maximal values recorded between 2000 and 2014).

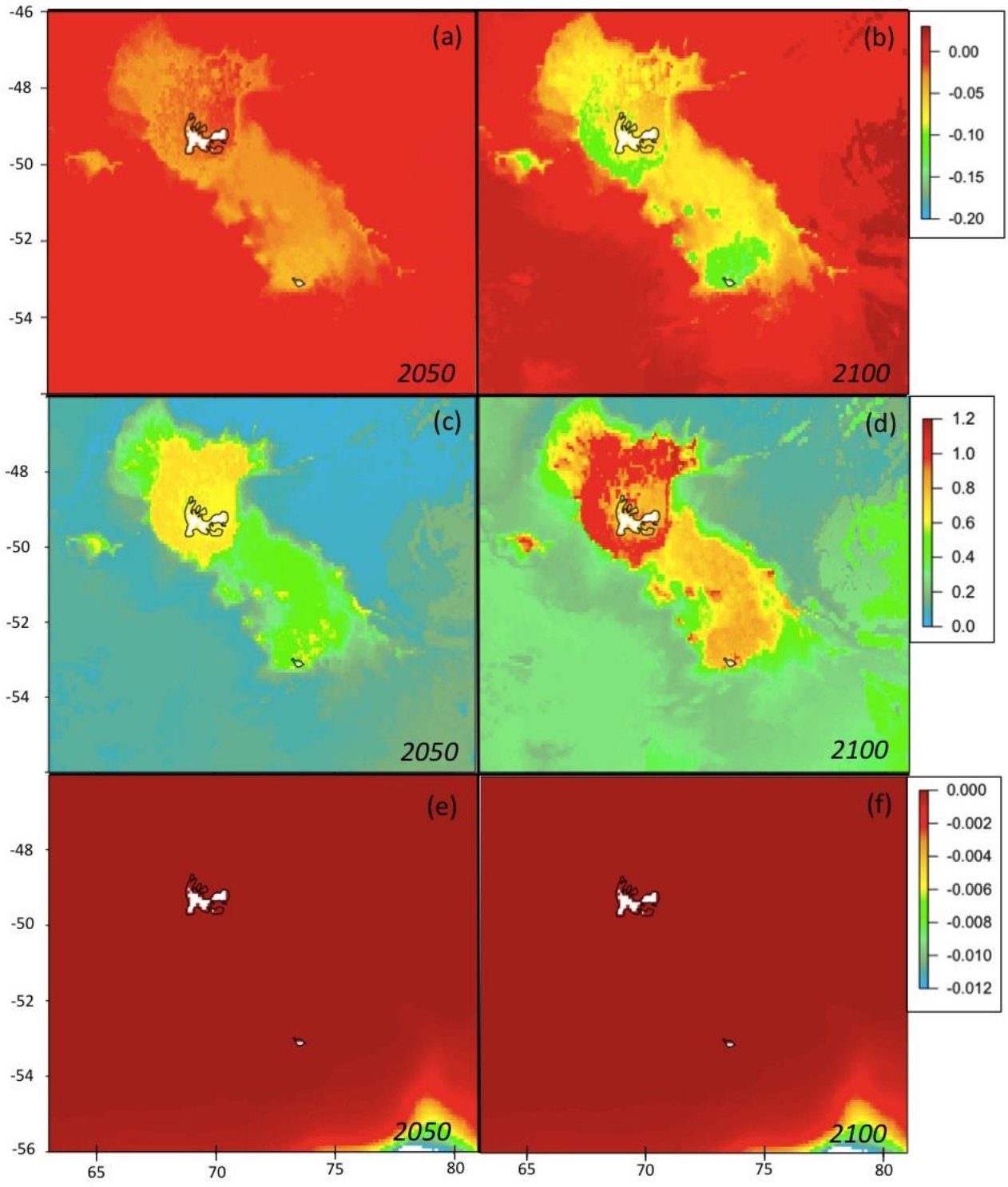


Figure S3.3.B. Focus on the Kerguelen Plateau (Kerguelen and Heard islands). Differences in seafloor salinities in PSU (a-b), seafloor temperatures in °C (c-d) and ice thickness in metre (e-f); between predicted future scenarios RCP 2.6 (mean values) for 2050 and 2100 and actual environmental conditions (mean maximal values recorded between 2000 and 2014).

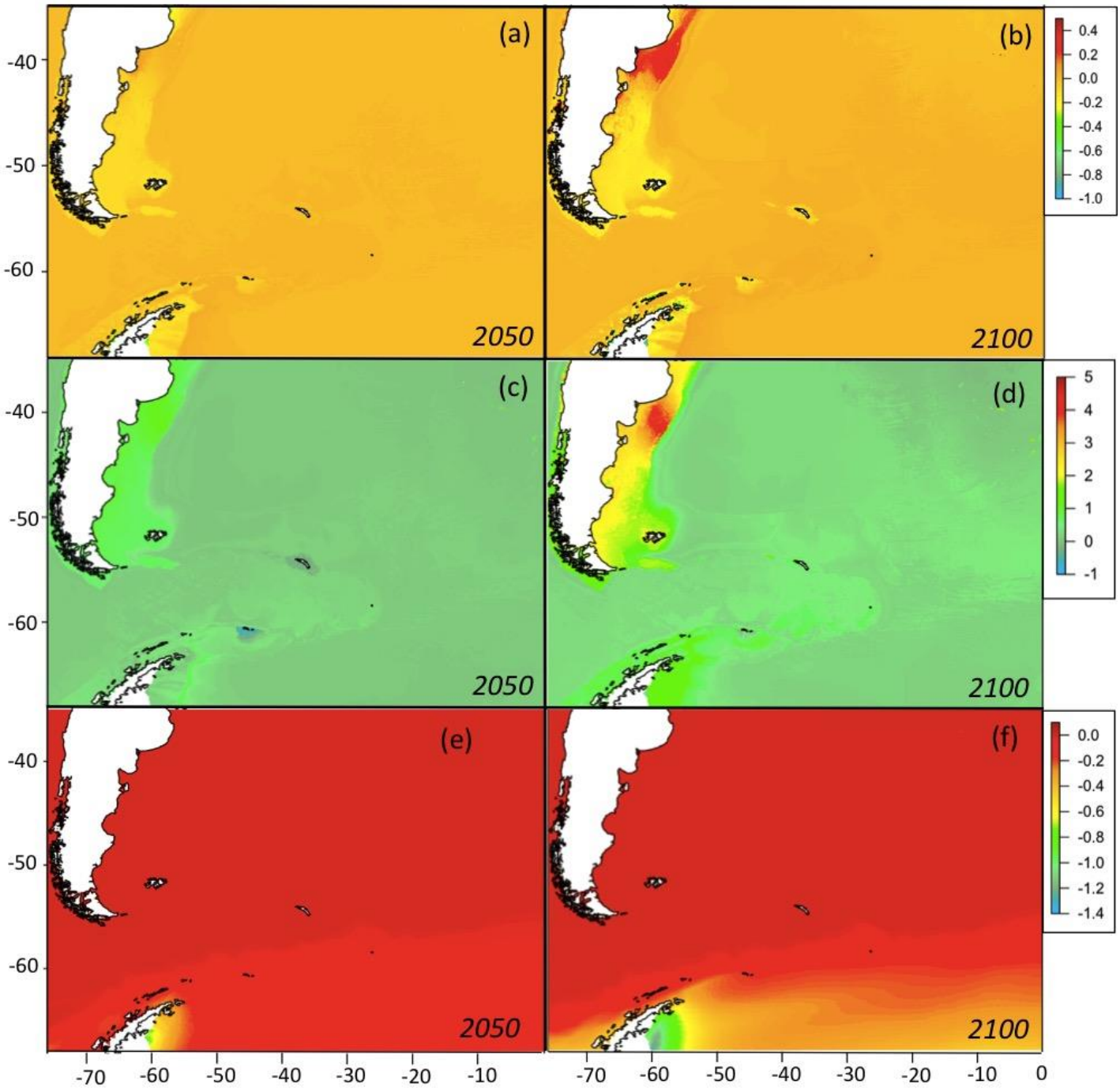


Figure S3.3.C. Focus on the Western Antarctic Peninsula and Southern America. Differences in seafloor salinities in PSU (a-b), seafloor temperatures in °C (c-d) and ice thickness in metre (e-f); between predicted future scenarios RCP 8.5 (mean values) for 2050 and 2100 and actual environmental conditions (mean maximal values recorded between 2000 and 2014).

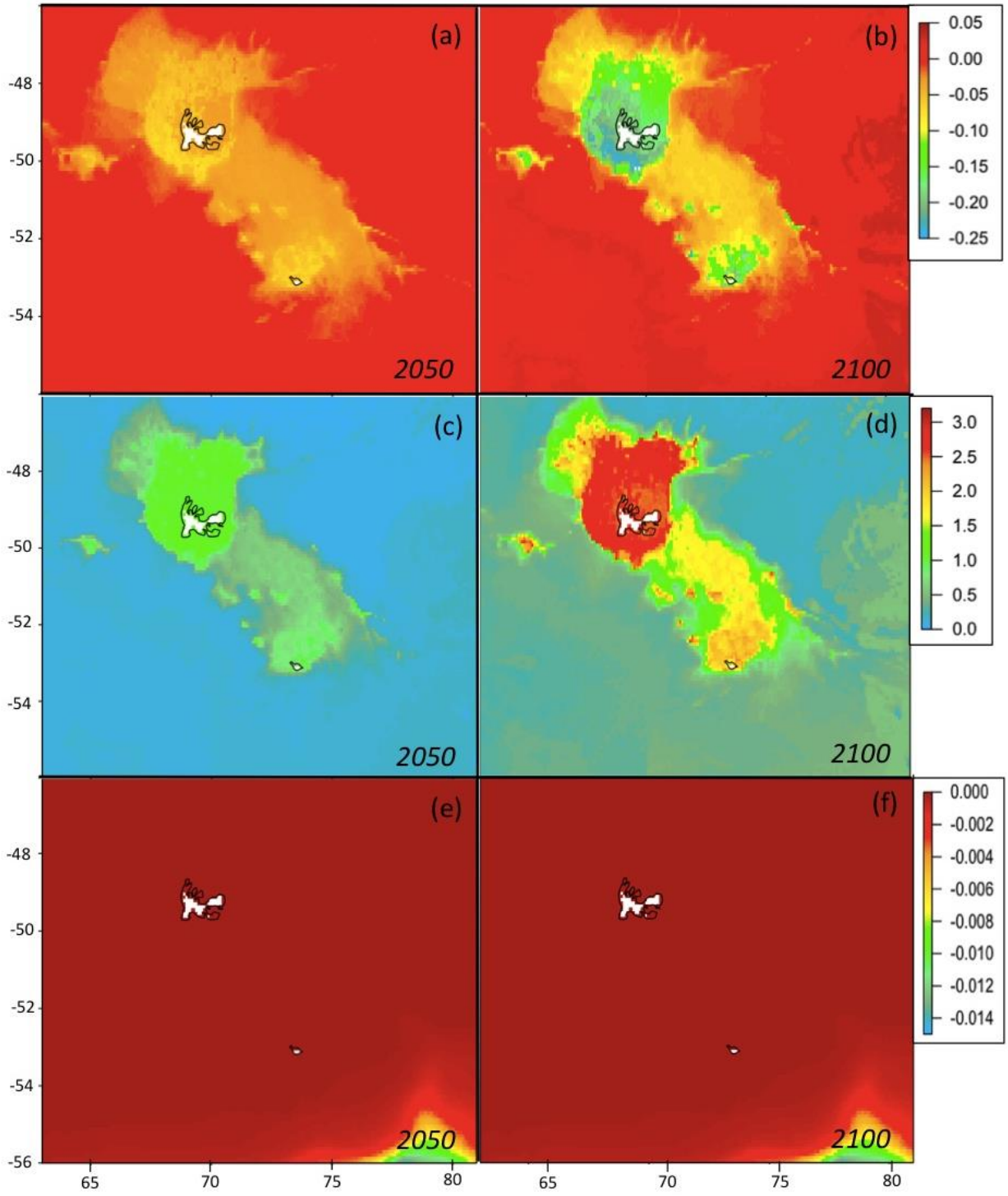


Figure S3.4.D. Focus on the Kerguelen Plateau (Kerguelen and Heard islands). Differences in seafloor salinities in PSU (a-b), seafloor temperatures in °C (c-d) and ice thickness in metre (e-f); between predicted future scenarios RCP 8.5 (mean values) for 2050 and 2100 and actual environmental conditions (mean maximal values recorded between 2000 and 2014).

APPENDIX 3.4. BRT calibration

BRTs were generated using the cross-validation procedure of Elith et al. (2008) and the *gbm* R package (Ridgeway et al. 2006) with codes provided in their supplementary material. A maximum number of 10,000 trees was set and models were calibrated with the combination of parameters that minimises the predictive deviance to the test data while producing the lowest number of trees (Fig. S3.4). Parameters finally selected to generate the models are: tree complexity tc = 4, learning rate lr = 0.005, and bag fraction bf = 0.9.

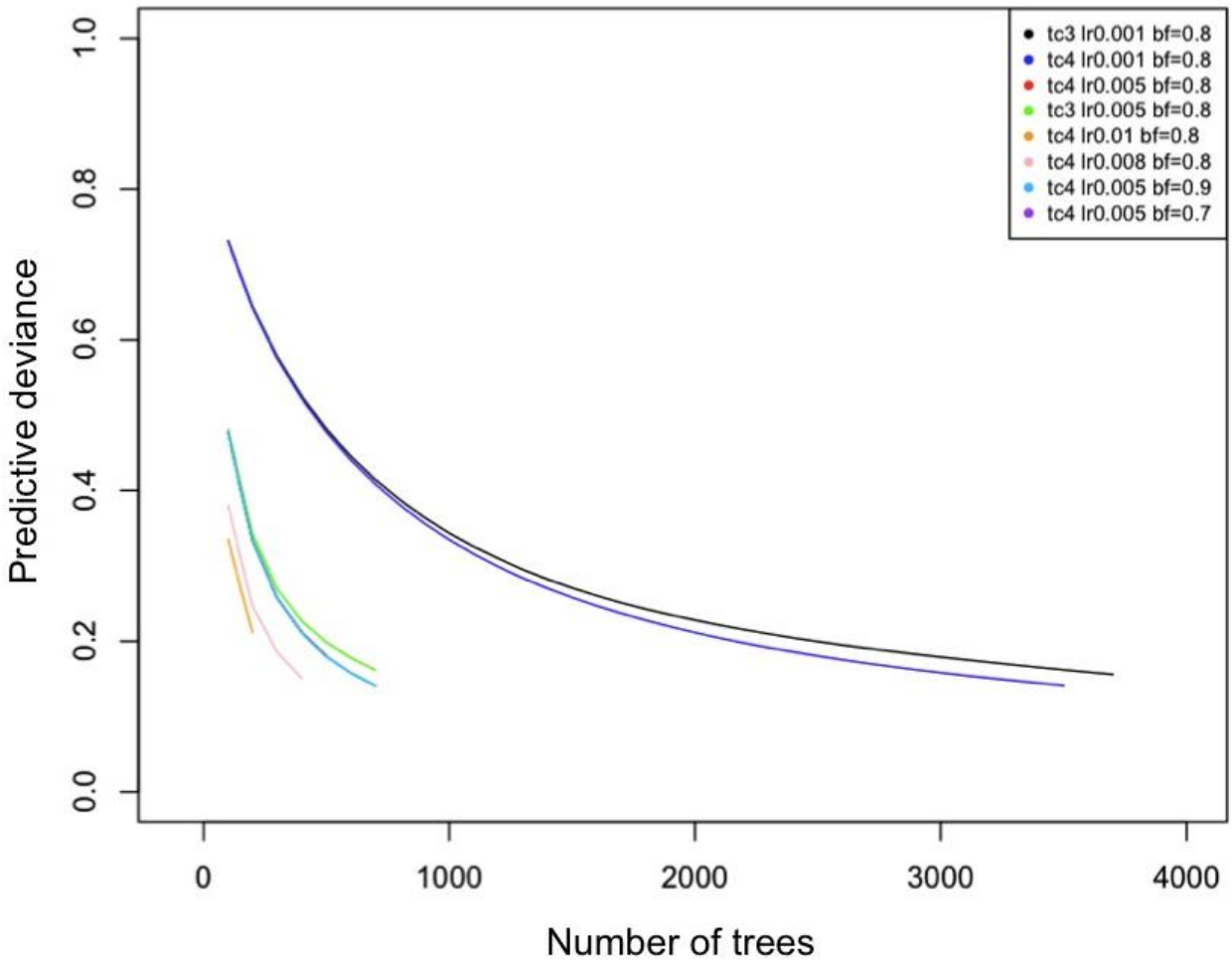


Figure S3.4. Comparison between model predictive deviance using different combinations of parameters. The one that reaches the minimal predictive deviance while requiring the lower number of trees to build the model is favoured (light blue curve). Tc: tree complexity, lr: learning rate; bf: bag fraction (see Elith et al. 2008 for details).

Using correlative and mechanistic niche models to assess the sensitivity of the Antarctic echinoid *Sterechinus neumayeri* (Meissner, 1900) to climate change

Fabri-Ruiz Salomé [1,2], Guillaumot Charlène[2,1], Agüera Antonio[3], Danis Bruno[2], Saucède Thomas[1]

1 UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, EPHE, 6 bd Gabriel F-21000 Dijon, France
2 Laboratoire de Biologie Marine, Université Libre de Bruxelles, Avenue F.D.Roosevelt, 50. CP 160/15. 1050 Bruxelles, Belgium
3 Institute of Marine Research, Austevoll Station, Sauganeset 16, 5392 Austevoll, Norway

Polar Biology, accepted May 21st 2021.

Abstract

The Southern Ocean is undergoing rapid environmental changes that are likely to have a profound impact on marine life, as organisms are adapted to sub-zero temperatures and display specific adaptations to polar conditions. However, species ecological and physiological responses to environmental changes remain poorly understood at large spatial scale owing to sparse observation data. In this context, correlative ecological niche modelling (ENMc) can prove useful. This approach is based on the correlation between species occurrences and environmental parameters to predict the potential species occupied space. However, this approach suffers from a series of limitations amongst which extrapolation and poor transferability performances in space and time. Mechanistic ecological niche modelling (ENMm) is a process-based approach that describes species functional traits in a dynamic environmental context and can therefore represent a complementary tool to understand processes that shape species distribution in a changing environment. In this study, we used both ENMc and ENMm projections to model the distribution of the Antarctic echinoid *Sterechinus neumayeri*. Both models were projected according to present [2005-2012] and future IPCC scenarios RCP 4.5 and 8.5 for [2050-2099].

ENMc and ENMm projections are congruent and predict suitable current conditions for the species on the Antarctic shelf, in the Ross Sea and Prydz Bay areas. Unsuitable conditions are predicted in the northern Kerguelen Plateau and South Campbell Plateau due to observed lower food availability and higher seawater temperatures compared to other areas. In contrast, the two models diverge under future RCP 4.5 and 8.5 scenarios. According to ENMm projections, the species would not be able to grow nor reach sexual maturity over the entire ocean, whereas the Antarctic shelf is still projected as suitable by the ENMc. This study highlights the complementarity and relevance of ENM approaches to model large-scale distribution patterns and assess species sensitivity and potential response to future environmental conditions.

Keywords

Ecological Niche model, Dynamic Energy budget, Species Distribution Modelling

ACKNOWLEDGEMENTS
This research is respectively contribution no. 44 and no. 19 to the vERSO and RECTO projects (http://www.rectoversoprojects.be), funded by the Belgian Science Policy Office (BELSPO). We thank Dina Lika for her help for the DEB model validation. We are grateful to Rose Stainthorp and Simon Morley for providing physiological data on *Sterechinus neumayeri*. We thank the three anonymous reviewers for their thoughtful comments and efforts towards improving our manuscript.

AUTHORS' CONTRIBUTIONS
This S.F.R, C.G, T.S and B.D conceived the idea and designed the manuscript. S.F.R, C.G and A.A provided data and validated the methodology. S.F.R and C.G. compiled and analysed the data. All authors equally contributed to the interpretation of analyses. S.F.R. C.G. and T.S. wrote the manuscript with contributions and inputs from all authors.

1. INTRODUCTION

Polar regions -and the Southern Ocean in particular- are increasingly affected by climate changes (Stammerjohn et al. 2008, 2012, Schofield et al. 2010, Turner et al. 2014). Temperature records over the previous decades unambiguously show an overall warming of water masses within the Antarctic Circumpolar Current area, from the surface down to 2,000 m depth, at a more rapid pace than average shifts measured in the global ocean (Gille 2002, Böning et al. 2008, Giglio and Johnson 2017). Contrasts however exist between regions of the Southern Ocean. For instance, a 1°C rise in sea water temperature has been recorded down to 25 meters in the water column at Potter Cove (King George Island, Antarctic Peninsula) over 19 years, with a decrease in sea ice extent (Meredith and King 2005). At the same time, sea ice has significantly been increasing in the Ross Sea both in concentration, extent (Comiso and Nishio 2008) and duration (Stammerjohn et al. 2012).

In the last report (IPCC 2015) of the Intergovernmental Panel on Climate Change (IPCC), CMIP5 (Coupled Model Intercomparison Project) climate models predict a global warming of the entire water column south of the Polar Front by the end of the century under either moderate (RCP 4.5) or business-as-usual Representative Concentration Pathway scenarios (RCP 8.5) (Turner et al. 2009, 2014, Liu and Curry 2010). Associated to this overall warming, changes in the extent and duration of the Antarctic seasonal sea ice and water freshening close to glacier melting sources are also expected (Meredith and King 2005, Bracegirdle et al. 2008, Stammerjohn et al. 2012). The Antarctic sea ice plays a crucial role in ecosystem functioning and regulates the timing of primary production (Petrou et al. 2016). Changes in sea ice regimes will impact the dynamics of phytoplankton blooms. Primary production constitutes an essential food intake for the benthos (Smith et al. 2006, Lohrer et al. 2013, Petrou et al. 2016, Schofield et al. 2017). Therefore, changes in phytoplankton dynamics could have a profound effect on the structure and functioning of benthic ecosystems.
The tectonic, climate and glacial history of the Southern Ocean (waters below 60°S in latitude) have conditioned the evolution of the Antarctic marine biota through various adaptive radiations, speciation, dispersal and extinction events. Associated to the isolation of the Antarctic continent, this led to the evolution of an original benthic fauna unparalleled in other parts of the world's ocean (Arntz et al. 1997, Clarke et al. 2005, Linse et al. 2006, Barnes and Griffiths 2007, Griffiths et al. 2009, Pearse et al. 2009, Rogers et al. 2012, David and Saucède 2015). High Antarctic marine benthic invertebrates have adapted to sub-zero temperatures and their feeding strategies have been conditioned by the seasonality in food availability due to the variation of sea ice dynamics (Knox 2006). Antarctic species commonly exhibit low metabolic and growth rates associated with a high longevity compared to temperate and tropical species (Pearse and Giese 1966, Brey 1991, Nolan and Clarke 1993, Peck and Bullough 1993, Brey et al. 1995, Peck et al. 2016). Most of the marine species present on the Antarctic shelf are consequently stenothermic (Peck 2002, 2006) and very sensitive to seawater warming and temperature variations (Peck et al. 2009b). Temperature changes can affect their physiological performance, phenology and distribution (Peck et al. 2009b, Morley et al. 2009a, 2010, 2011).
Along Antarctic coasts, marine benthic communities are at the southernmost boundary of the temperature latitudinal gradient of the marine biome (Peck et al. 2006). Consequently, in a context of warming temperatures, species are spatially limited and cannot easily migrate or find refuges to survive (Peck and Conway 2000).

Monitoring and predicting the response of Antarctic species to environmental change is challenging as gaps still persist in our knowledge of Antarctic marine species distribution (Kaiser et al. 2013, Kennicutt et al. 2014, 2019, Gutt et al. 2018), despite the significant efforts led during the International Polar Year and the Census of Antarctic Marine Life (Schiaparelli et al. 2013, Fabri-Ruiz et al. 2019). Data collection and experimental setups are strongly conditioned by financial and technological limitations in such a remote and hard-working region (extreme climate conditions,

difficult to access) (Gutt et al. 2012). Ecological Niche Modelling (ENM) can represent an alternative to overcome this issue.

Correlative Ecological Niche Models (ENMc) can be used to predict species distribution based on the statistical relationship between species occurrence records and abiotic conditions (Guisan and Thuiller 2005, Pearson 2007, Elith and Leathwick 2009). ENMc provide a spatial representation of the species realised niche under the assumption of equilibrium between species distribution and the abiotic environment (Guisan and Zimmermann 2000, Pearson and Dawson 2003). In contrast, mechanistic Ecological Niche Models (ENMm) use eco-physiological data and life history traits to describe organisms' physiology. They can predict species capabilities to survive, grow and reproduce under changing environmental conditions and describe a part of the species fundamental niche (Brown et al. 2004, Kearney et al. 2008, 2009, Sousa et al. 2008, Cabral and Kreft 2012).

ENMc have been widely developed for the study of Antarctic marine organisms such as pelagic plankton and fish (Pinkerton et al. 2010, Duhamel et al. 2014), deep-water shrimps (Basher and Costello 2016), cirripeds (Gallego et al. 2017), molluscs (Xavier et al. 2015), echinoids (Pierrat et al. 2012, Fabri-Ruiz et al. 2019, 2020), or sea stars (Guillaumot et al. 2019b). In contrast, ENMm (such as the projection of Dynamic Energy Budget models, DEB, Kooijman 2010) have never been developed for Antarctic species case studies so far, due to the more important amount of data required to implement the DEB model (eco-physiological data on the different species life stages; van der Meer 2006, Kearney and Porter 2009), and the novelty of the DEB projection method (Thomas and Bacher 2018).
Once created, DEB models are published in the Add-my-Pet collection (https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/about.html), that already provides a list of 37 Antarctic marine and terrestrial species. Among them, the most commonly found in communities and well-studied Southern Ocean benthic invertebrates are the sea star *Odontaster validus* (Agüera et al. 2015), the bivalve *Laternula elliptica* (Agüera et al. 2017 - Appendix), the echinoid *Abatus cordatus* (Guillaumot 2019c, Arnould-Pétré et al. 2020 - Chapter 1), the gastropod *Nacella concinna* (Guillaumot 2020b, Guillaumot et al. 2020a - Chapter 1) and the bivalve *Adamussium colbecki* (Guillaumot 2019a). DEB models have also been developed for pelagic species such as the Antarctic krill *Euphausia superba*, the salp *Salpa thompsoni* (Jager and Ravagnan 2015, Henschke et al. 2018) and also for marine mammals such as the elephant seal *Mirounga leonina* (Goedegebuure et al. 2018).

Providing relevant projections of the impact of climate change on biodiversity is crucial to conservation biology (McMahon et al. 2004, Gotelli et al. 2009, Gutt et al. 2012, Evans et al. 2015, Pertierra et al. 2019). Usually, ENMc and ENMm are independently used to study the relationship of a species with its environment (Dormann et al. 2012a). Combining both approaches has only recently emerged in link with computing advances (Kearney et al. 2010, Buckley et al. 2011, Dormann et al. 2012a, Meineri et al. 2015, Briscoe et al. 2016, Enriquez-Urzelai et al. 2019, Pertierra et al. 2019). This combination was proved efficient to improve predictions compared to simple models, as ENMm can address the deficits of ENMc by explicitly including processes, offering the opportunity to describe, within and without the predicted suitable boundaries of the ENMc predictions, the process-based causes of the species distribution (Kearney and Porter 2009, Dormann et al. 2012a). It can also provide more insight into drivers that shape species current distribution and potential distribution shifts under changing environmental conditions (Kearney and Porter 2009, Buckley et al. 2011, Ceia-Hasse et al. 2014, Meineri et al. 2015).

The echinoid *Sterechinus neumayeri* (Meissner, 1900) is abundant, common and endemic to the Antarctic continental shelf. It has widely been studied in various fields such as reproductive biology, embryology, toxicology, ecology and physiology (Bosch et al. 1987 - McMurdo; Stanwell-Smith and Peck 1998 - Signy Island; Marsh et al. 1999, 2001 - McMurdo; Tyler et al. 2000 - Rothera; Brockington and Peck 2001 - Rothera; Pace and Manahan 2007 - McMurdo; Moya et al. 2012 - Bellingshausen Sea; Yu et al. 2013 - McMurdo; Lister et al. 2015 - McMurdo; Alexander et

al. 2017 - Peterson Channel). Widely distributed all around Antarctica (Fig. 3.12), its distribution ranges from the subtidal zone to 800-m depth with most records found in shallow waters of the continental shelf above 400-m depth (David et al. 2005). Recent molecular studies showed that the species combines a unique genetic entity all around the Antarctic continent (Díaz et al. 2011, 2018). It plays an important ecological role in structuring benthic communities. The "grazing" pressure exerted by *S. neumayeri* is believed to control the local distribution of bryozoans and spirorbid annelids and could therefore have a negative feedback on the recruitment of some sessile species (McClintock 1994, Bowden 2005, Figuerola et al. 2013). Adult specimens are omnivorous and mainly feed on bryozoans, foraminifera, polychaetes, diatoms and macro-algae (McClintock 1994, Amsler et al. 1999, Jacob et al. 2003, Michel et al. 2016). As in many other Antarctic species, the development rate of *S. neumayeri* is low (Bosch et al. 1987), longevity can exceed 40 years (Brey 1991, Brey et al. 1995) and the feeding period is seasonal (Brockington and Peck 2001). *S.neumayeri* is a broadcast spawner, planktonic larvae can drift in the water column for more than 8 months before metamorphosis takes place on the seabed (Pearse and Giese 1966) (see details in Appendix 3.5 and 3.6). The test of adult specimens can reach a final size of seven centimeters in diameter (Brey et al. 1995).

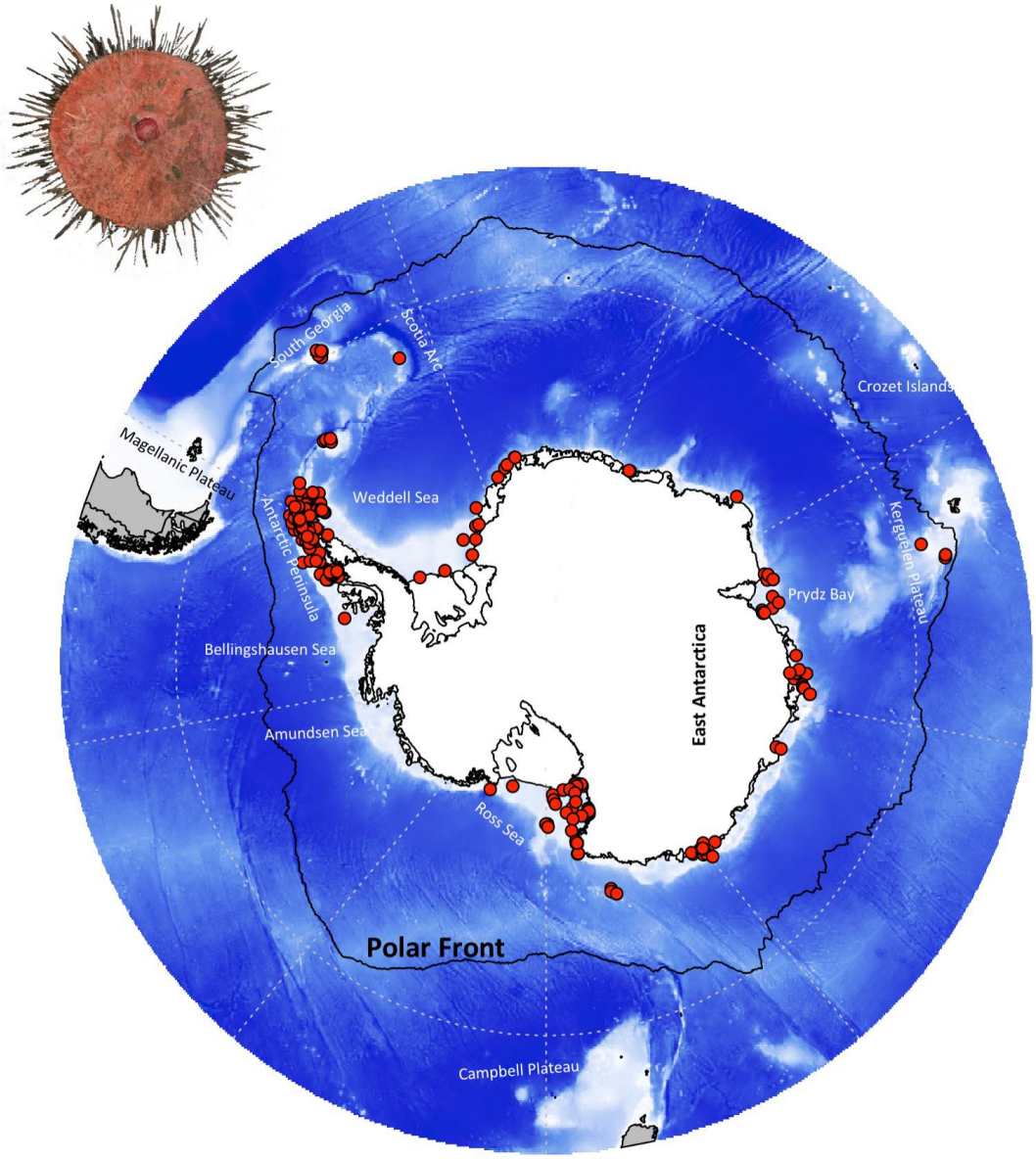


Figure 3.12. *Sterechinus neumayeri* occurrence data extracted from Fabri-Ruiz et al. (2017a). Illustration of *Sterechinus neumayeri* © J-G. Fabri.

In the present work, we used both ENMc and ENMm approaches to project the distribution response of *S. neumayeri* to present-day conditions and to future IPCC scenarios of climate change RCP 4.5 and RCP 8.5. ENMc were generated to predict species distribution in these environmental conditions using the Random Forest algorithm (Breiman 2001). The DEB model created for *S. neumayeri* was spatially projected (i.e. ENMm model) in these three environmental scenarios. The results of both ENMc and ENMm models were compared to get more insight into the physiological processes and mechanisms that constrain the species distribution, and assess model performances and ecological significance under present-day conditions and future scenarios of climate change.

2. MATERIAL AND METHODS

2.1. Correlative ecological niche models (ENMc)

Occurrence data and environmental predictors.
An ENMc was generated using georeferenced presence-only data of *S. neumayeri* extracted from an extensive Southern Ocean echinoid distribution database (Fabri-Ruiz et al. 2017a) that includes field samples collected between 1901 and 2015 (Fig. 3.12). Considering the broad spatial scale of the analysis and the congruence between historical and present-day presence records (David et al. 2005, Fabri-Ruiz et al. 2019), it is here assumed that the species distribution did not significantly change over the last century at the scale of the entire Southern Ocean.
Environmental predictors used in the study were extracted from Fabri-Ruiz et al. (2017b) (Appendix 3.7). Predictors were selected based on their ecological relevance for explaining the distribution of *S. neumayeri* (Pierrat et al. 2012, Saucède et al. 2014, Fabri-Ruiz et al. 2019). Collinearity between descriptors was tested to limit possible biases in predictor contributions and model predictive performances and the presence of spatial autocorrelation (Dormann et al. 2012b). For this purpose, we performed a Spearman pairwise correlation test between descriptors that were iteratively removed for correlation values of $r_S > 0.8$ (Dormann et al. 2012b). Over 26 possible descriptors, 13 were used to run the models. The physical habitat was described using the following descriptors: depth, geomorphology, slope, sea surface temperature range, seafloor temperature range, mean seafloor temperature and sea ice cover. Summer chlorophyll-a concentration was used as a proxy of food resources and habitat chemistry was described based on seafloor salinity, seafloor salinity range, sea surface salinity range, sea surface salinity and seafloor oxygen (Appendix 3.7). Predictor 'range' is here defined as the difference between winter and summer mean values.

Future projections were based on IPCC scenarios RCP 4.5 and RCP 8.5 (IPCC 2015, Appendix 3.8) extracted from the NOAA database (https://www.esrl.noaa.gov/psd/ipcc/ocn/ [accessed on 2019-12-19]). Future projections were not available for seafloor oxygen conditions under IPCC scenarios. The descriptor was therefore considered unchanged (present conditions) in future models.

ENMc calibration.
The distribution of *S. neumayeri* was modelled using Random Forests algorithm (RF) (Breiman 2001) computed with the *biomod2* R package (Thuiller et al. 2009). In a former study, RF was proved relevant to model the distribution of *S. neumayeri*, models showing high and stable predictive performances, and appropriately captured the species environmental envelope (Fabri-Ruiz et al. 2019). Here, the ENMc was parameterized with 500 classification trees, a tree number that minimizes the difference in predictive performance between models. This number was selected by testing different values of tree number (50, 100, 500 and 1000). Five node size (minimum size of the final node of any tree) and mtry =13 (the number of candidate variables to include at each split) was tuned using the 'tuneRF' function from the *caret* package (Kuhn 2012).

The occurrence dataset was randomly split into a 70% subset used to train the model and a 30% subset to test model predictions. As only presence data were available, pseudo-absences were randomly generated following Barbet-Massin et al. (2012) with a number of pseudo-absences equal to the number of presences. Fifty pseudo-absence replicates were generated and for each, ten evaluation runs were computed.

Spatial sampling bias is generally pervasive in species occurrence data, which were typically not evenly sampled across the ocean (De Broyer and Koubbi 2014). This may generate strong spatial autocorrelation in model residuals, that is, the fact that close observations in geography will be more similar than random (Legendre 1993). The presence of spatial autocorrelation breaks the assumption of «independent errors» when significant (Dormann et al. 2007) and leads to unreliable model evaluation (Phillips et al. 2009, Kramer-Schadt et al. 2013, Warren et al. 2014).
To limit this bias, pseudo-absence data were sampled following the same sampling pattern as all Antarctic echinoid records available in the Southern Ocean. A Kernel Density Estimation map established from all Antarctic echinoid records using Spatial Analyst in ArcGIS v10.2 was used to target the pseudo-absence sampling accordingly (Phillips et al. 2009, Guillaumot et al. 2018a - Appendix). In total, 50 pseudo-absence replicates were generated and spatial autocorrelation was quantified for each pseudo-absence replicate using the Moran I index computed with the *ape* R package (Paradis et al. 2008). Moran I measures the average correlation value of a variable between values taken at close localities. It is an easy correlation index to interpret, that varies between -1 (negative spatial autocorrelation: values at close localities are opposite compared to the mean value) and +1 (positive spatial autocorrelation: values at close localities are similar), with 0 for an absence of spatial autocorrelation. The significant values of spatial autocorrelation statistic are indicated by a p-value. Over the 50 replicates of pseudo-absences, we selected thirty replicates showing $p > 0.5$ (with p, the p-value of the significance of Moran's I), other pseudo-absences replicates have depicted a p-value less than 0.5.

The wide extent of the study area implies that a wide range of environmental conditions may be used to fit the models and leads to overestimate and extrapolate the species modelled niche (Giovanelli et al. 2010, Barve et al. 2011, Anderson 2013, Guillaumot et al. 2020c - Chapter 2). To limit extrapolation, the modelling area was limited to the maximum species registered depth (800 m, David et al. 2005) for model calibration and projection.

Model predictive performances were assessed with the TSS metric (True Skill Statistics) (Allouche et al. 2006) that is the sum of the sensitivity (proportion of correctly predicted presences) and the specificity (proportion of correctly predicted absences) minus one (*sensitivity* + *specificity* -1). The contribution of environmental predictors to the models was provided as “contribution permutation” available under the *biomod2* R package (Thuiller et al. 2009). For each predictor, contribution permutation was calculated as the Pearson correlation coefficient between model predictions by randomly permuting the predictors. For this purpose, we performed ten permutation runs. The higher the value, the more the predictor contributes to the model. Response plots were provided to show the relationship between habitat suitability for *S. neumayeri* and environmental predictors.

ENMc projections.
ENMc projections were generated using three sets of environmental predictors: for the present time [2005-2012], for scenario RCP 4.5 [2050-2099] and scenario RCP 8.5 [2050-2099]. Presence probability maps of *S. neumayeri* were produced with values close to zero indicating low presence probabilities, and values close to one indicating high presence probabilities.

2.2. Mechanistic ecological niche model ENMm (spatial projection of Dynamic Energy Budget models)

Model description.
DEB models provide a mechanistic and quantitative description of the energy fluxes in an organism that assimilates and uses energy for its maintenance, growth and reproduction throughout its entire life cycle (Kooijman 2010). DEB theory aims at describing how species energy fluxes change according to environmental conditions (i.e food and temperature) and can help estimate the species fundamental niche (Kearney and Porter 2004). DEB models rely on physiological and experimental data/traits (Kearney and Porter 2004, van der Meer 2006). This approach models a part of the species fundamental niche.

In DEB models, energy flows between four state variables: reserve (E), structure (V), maturation (E_H) and reproductive buffer (E_R) (Fig. 3.13).

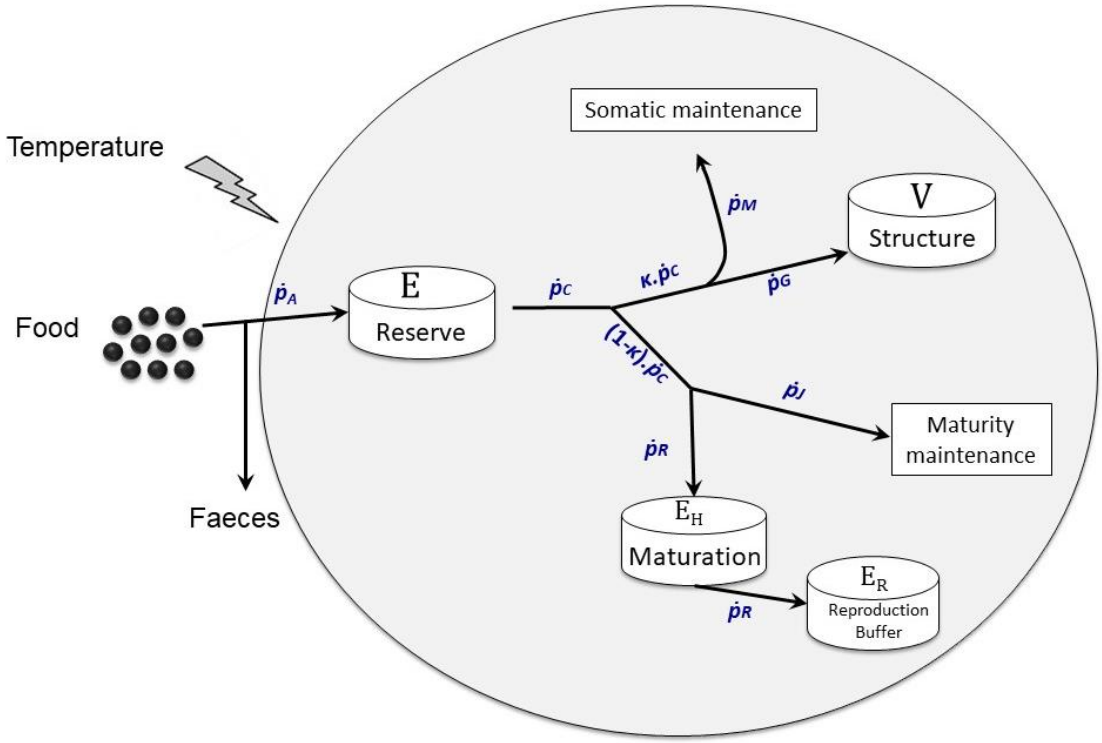


Figure 3.13. Conceptual representation of the standard Dynamic Energy Budget model. Arrows show energy flows (J.d^{-1}) involved in the dynamics of the four state variables (represented by boxes: reserve (E), structure (V), maturation (E_H) and reproductive buffer (E_r). $\dot{p}_A$ is the assimilation rate into the reserve, $\dot{p}_C$ is the energy rate leaving the reserve which is divided in two branches: $\kappa.\dot{p}_C$ allocated to the somatic maintenance ($\dot{p}_M$) and growth ($\dot{p}_G$) and the fraction $(1-\kappa).\dot{p}_C$ allocated to maturity maintenance ($\dot{p}_J$), maturation and reproduction ($\dot{p}_R$).

Energy enters into the body by food (X) ingestion at a rate $\dot{p}_X$.

$\dot{p}_X = \{\dot{p}_{Xm}\}.f.L^2$ [Eq. 1], with:

• $f = \frac{X}{X+X_K}$ corresponding to the food functional response [Eq.2].
X is the amount of available resources (mg.m^{-3}) and X_k the half-saturation parameter (mg.m^{-3})

- $\{\dot{p}_{Xm}\}$ = max. surface area-specific ingestion rate (J.cm^{-2}.d^{-1})
- L = individual's length (in cm)

DEB models use a version of a Hollings' type II functional response. The functional response *f* changes when the resource (X) is different. The *f* value varies between 0 and 1 (van der Meer 2006).

Chlorophyll-a concentration was considered as a proxy of food resources for *S. neumayeri* (McClintock 1994, Jacob et al. 2003, Michel et al. 2016). In Cape Evans (McMurdo), Pearse and Gierse (1966), based on gut content, has emphasized that food of *S. neumayeri* could be mainly constituted of diatoms which is also highlighted by Brockington et al. (2001).

Sea surface chlorophyll-a concentration data (X in Eq.2) and gut content (*f* in Eq. 2) were obtained from a long-term experiment conducted at Rothera Station (Western Antarctic Peninsula) in 1997-1998 (Brockington and Peck 2001). A non-linear least squares regression was performed to adjust the functional response (Eq. 2) using chlorophyll-a concentration and gut content (Appendix 3.9 for more details). The estimation gives a value of 2.95 mg.m^{-3} for the half-saturation coefficient (X_k).

After food ingestion, the energy is assimilated and stored into the reserve compartment at a given rate expressed in Joules per time ($\dot{p}_A$). The energy leaving the reserve ($\dot{p}_C$) is subdivided according to the "kappa-rule" (κ-rule) in between somatic maintenance ($\dot{p}_M$), growth ($\dot{p}_G$), maturity maintenance ($\dot{p}_J$), maturation, and reproduction ($\dot{p}_R$, proportion 1-κ) (van der Meer 2006, Kooijman 2010). Maturity does not contribute to body mass. The amount of energy contained in the maturity compartment thresholds the different life stages of the species during its life cycle (birth: ability to feed, puberty: ability to reproduce) (Jusup et al. 2017, Appendix 3.6). Once puberty is reached, the species is considered to be a fully developed adult, and the energy initially allocated to maturation begins to be used for reproduction.

There is no competition between the two branches of the κ-rule, which means that an organism can continue to grow and reproduce at the same time. However, energy is still primarily allocated to maintenance to prioritize body functions that are essential to the organism survival (i.e. maintenance of cell concentration gradients, protein turnover, enzyme functioning, mucus production, osmoregulation) and the maintenance of maturity (maintenance of the structure complexity).
Reserve compounds do not need maintenance as energy is continuously used. Growth corresponds to the increase of the body structure and maturation is the energy dissipated or expended by the body in the increase of maturity.

Estimation of DEB model parameters.
The DEB model was parameterized using literature data from field and experimental works mainly led at McMurdo and Rothera stations, Antarctica (Table 3.3, Appendix 3.10, 3.11).
Zero-variate data correspond to single measurements at a given time (characterised by specific food and temperature conditions) and uni-variate data are relationships between two variables (e.g. mass, oxygen consumption etc. against duration, temperature, *etc.*). From these data, DEB parameters were estimated using the covariation method (Lika et al. 2011a, 2011b, Marques et al. 2018) that aims at looking for the combination of parameters (Table 3.3) that minimizes the difference between observations and predictions (i.e. minimizing the loss function). The evaluation of the parameter estimation is assessed by calculating the Mean Relative Error (MRE) which can vary between 0 and ∞, with MRE=0 meaning a perfect match between observations and predictions. For each univariate and zero-variate data the relative error was computed as the ratio of the absolute error value to the variate value.

Table 3.3. DEB parameter values estimated by the covariation method (Lika et al. 2011a, 2011b, Marques et al. 2018)

DEB parameter	Unit	Value
z, zoom factor	–	1.364
$\delta_{M.emb}$, shape coefficient embryos	–	0.487
$\delta_{M.lrv}$, shape coefficient larvae	–	0.505
δ_M, shape coefficient	–	0.612
$\{\dot{F_m}\}$, max. specific searching rate	$L.d^{-1}.cm^{-2}$	6.5
κ_X, digestion efficiency of food to reserve	–	0.83
$\dot{v}$, energy conductance	$cm.d^{-1}$	0.033
κ, allocation fraction to soma	–	0.722
κ_R , reproduction efficiency	–	0.95
$[\dot{p_M}]$, vol-specific somatic maintenance	$J.cm^{-3}.d^{-1}$	24.42
$\dot{k_J}$, maturity maintenance rate coefficient	d^{-1}	$2.5.\ 10^{-3}$
$[E_G]$, specific cost for structure	$J.cm^{-3}$	2350
E_H^b, energy maturity at birth	J	$4.5.\ 10^{-3}$
E_H^j, energy maturity at metamorphosis	J	0.3
E_H^p, energy maturity at puberty	J	2266
$\ddot{h_a}$, Weibull aging acceleration	d^{-2}	$2.\ 10^{-8}$
S_G, Gompertz stress coefficient	–	$1.\ 10^{-4}$

The description of methods on temperature sensitivity using Arrhenius temperature and changes in body shape using post-metamorphic shape coefficient is provided in Appendix 3.12. All analyses were conducted under Matlab 2016 using the *DEBtools* repository (https://github.com/add-my-pet/DEBtool_M/).

Rothera data were used to perform sensitivity analysis of DEB model estimation (Appendix 3.13). For this purpose, marginal confidence intervals of the estimated parameters were computed to provide the uncertainty related to the parameter estimations using the covariation method (Stavrakidis-Zachou et al. 2019). The profile method (Marques et al. 2019) was used to build the profile of the loss function of each parameter and estimate the level of the loss function that corresponds to the uncertainty. A total of 1,000 Monte-Carlo datasets was generated by adding a constant centered log-normal scatter to the predictions of each zero and uni-variate data. The threshold value of the loss function Fc that is used to assess the uncertainty level was obtained from P (X < Fc) = 0.9, with 0.9 being the confidence level initially chosen in the procedure. The marginal confidence interval of each parameter is the interval of values for which the loss function is below the threshold value Fc.

Spatial projection of the DEB model.
For each pixel of the study area, food (i.e. summer chlorophyll-a concentration converted into f [0-1] according to the procedure explained above) and temperature were both used as input into the DEB model that consequently calculated how energy is used and allocated to the different metabolic processes, given these environmental conditions. Projections of the DEB model were performed according to present-day conditions [2005-2012] and future RCP 4.5 and RCP 8.5 scenarios [2050-2099] (environmental layers are displayed on Appendix 3.14). Different simulations were carried out for temperature or food changes only.

A first projection provides the maximum size reached by individuals, which gives some information about the species ability to survive and to invest energy into growth. It also provides a quantitative estimate of the stress experienced by *S. neumayeri* at large spatial scale, the smaller individuals, the less suitable the environment. According to DEB theory, the somatic maintenance has priority over reproduction and growth to ensure survival. In order to identify regions where individuals are able to survive from an energetic point of view, the somatic maintenance flow $\dot{p}_M$ was calculated according to the given food and temperature conditions and compared to the values of the total energy available from the reserve $\dot{p}_C$. When somatic maintenance values are higher than the energy available in the reserve compartment ($\dot{p}_M > \dot{p}_C$), it suggests that individuals do not have enough energy to maintain their soma and should die (Fig. 3.13). $\dot{p}_M$ values were also compared to the flow $\kappa.\dot{p}_C$, that corresponds to the proportion of the mobilized energy from the reserve that is invested into growth and the somatic maintenance. The organism survives if $\dot{p}_M < \kappa.\dot{p}_C$. On the other hand, if $\dot{p}_M < \dot{p}_C$ but $\dot{p}_M > \kappa.\dot{p}_C$, the organism will have difficulties to maintain its soma and a part of the energy allocated to maturation, reproduction and growth will be redirected to somatic maintenance.

A second projection provides suitable areas for reproduction that is, areas in which environmental conditions allow the species to invest energy into growth and reproduction. In DEB theory, the organism can reproduce when enough energy has been invested into maturity ($E_H > E_H^p$), passing from the juvenile to the adult life stage ('puberty' threshold). To assess whether individuals can invest energy into reproduction, we first calculated the size (L_p) at which individuals reach puberty (Eq. 3) for each pixel of the projection map. The DEB parameter shape coefficient δ_M estimated by the model is used to translate physical measurements taken from experimental data to the structural length used by the model (Appendix 3.12).

$$L_p = \frac{L_m.l_p}{\delta_M} \quad \text{[Eq. 3]}$$

- L_m: Maximum structural size (cm)
- l_p: Standardized size at sexual maturity (= puberty) (unitless)
- δ_M: Shape coefficient of post-metamorphic individuals (unitless)

Considering the body length at puberty (L_p), we then identified if somatic maintenance could be ensured at puberty ($\dot{p}_c > \dot{p}_M$ and $\kappa.\dot{p}_C > \dot{p}_M$). The total cost of maintenance ($\dot{p}_M + \dot{p}_J$) was also compared to the outflow from the reserve $\dot{p}_C$, with $\dot{p}_C > \dot{p}_M + \dot{p}_J$ meaning that individuals can invest energy into reproduction. All DEB models were computed from R functions available at https://github.com/Echinophoria/DEB/.

3. RESULTS

3.1. Species distribution models under present-day conditions

Correlative ecological niche model (ENMc).
For the ENMc generated under present-day environmental conditions, the average predictive accuracy of model replicates is good (TSS = 0.64 ± 0.078), which indicates a relatively good match between presences and predictions. High species presence probabilities (p > 0.8) are predicted south of the Polar Front: over the Antarctic shelf, along the Western Antarctic Peninsula and the Scotia Arc region (Fig. 3.14a). The highest values are in the northern tip of the Western Antarctic Peninsula, in East Antarctica and in the Ross Sea. Medium values (p ~ 0.5) are mainly located in the Amundsen and Bellingshausen seas, the Weddell Sea and in South Georgia. Regions located north of 55°S latitude such as the Kerguelen, Magellanic, and Campbell plateaus are mostly predicted as unsuitable areas (p < 0.2). Environmental predictors that most contribute to the model are seafloor temperature, geomorphology, slope, sea ice cover, and depth, in decreasing order of importance (Fig. 3b). Chlorophyll-a concentration was used as an indirect proxy of food supply but it does not contribute much to the model (ranked seventh most contributing predictor). Parameters such as seafloor oxygen concentration, seafloor temperature range, seafloor salinity, seafloor salinity range and sea surface salinity do not contribute much to the model.

Curves of the species response to main environmental predictors allow visualizing conditions that are the most suitable for species distribution (Fig. 3.14c). These are shallow areas (< 400 m depth) represented in geomorphology as banks, coastal terranes, seamounts and volcanoes (Appendix 3.16) with positive slope values (> 0.05°), cold water sea floor temperatures (< 1°C), and weak sea ice coverage (< 60%) (Fig. 3.14c, Appendix 3.15). The response curve to chlorophyll-a concentration values shows little variation, the highest probability values corresponding to low chlorophyll-a concentrations (< 2mg/m^3, Fig. 3.14c).

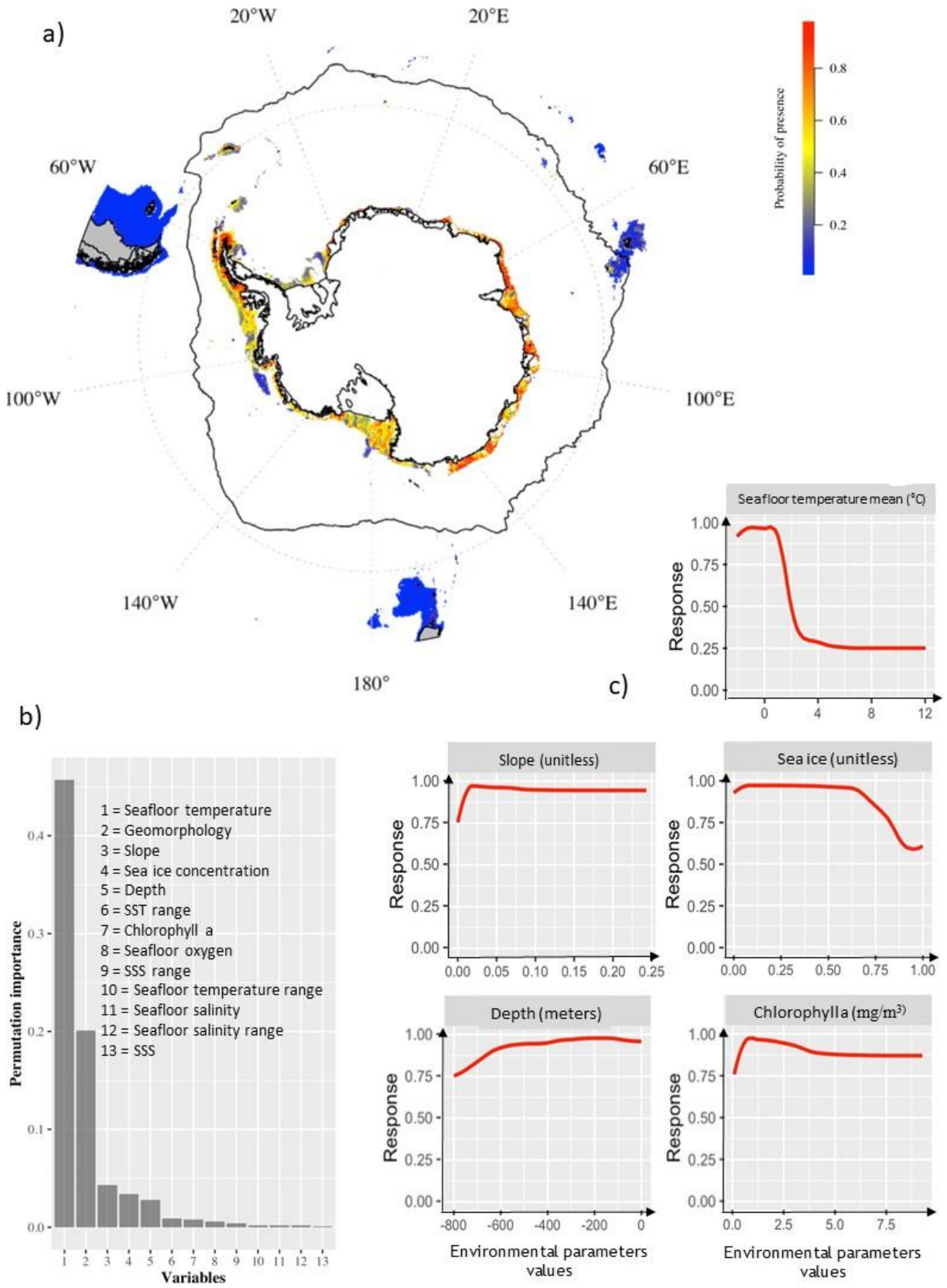


Figure 3.14 (a) Spatial projection of the ENMc under present-day conditions in the Southern Ocean with (b) the respective contributions of environmental descriptors to the model and (c) the species response (distribution probability) to the main contributing predictors (mean seafloor temperature, slope, sea ice coverage and depth) and for chlorophyll-a concentration (as a proxy of food supply). No response curve can be displayed for geomorphology, which is a categorical variable (see Appendix 3.16).

Projection of the Dynamic Energy Budget model (ENMm).

Experimental data available for the different life stages of *S. neumayeri* allow a robust prediction of DEB parameters (Appendix 3.10, 3.11) with a total goodness of fit resulting in relative low error values (MRE = 0.095). For comparison, the values fall within the range of median values usually obtained for DEB models (median MRE < 0.1; Marques et al. 2018).

Most zero-variate and uni-variate data are accurately described by the estimated model parameters with low error values. For uni-variate data, the highest relative error values are obtained for the C:N mass of fertilized egg (RE = 0.29) and the uni-variate data Ash Free Dry Mass (AFDM, g) vs. O_2 consumption in µmol/h in summer (RE= 0.27) (Appendix 3.10, 3.11). The pre-metamorphic larval size is slightly underestimated in the model but the error is low (RE= 0.093) (Appendix 3.10). The prediction of the adult size-age relationship also shows a low error value (RE= 0.13) (Appendix 3.10) as for the weight-size data (RE= 0.05) (Appendix 3.10). Models of winter and summer oxygen consumption ~ weight data have similar patterns (Appendix 3.10) with a shift in oxygen consumption values for individuals of 0.2 g (AFDM), which corresponds to a transition stage between the embryo and the pre-metamorphic larvae.

Model validation gives low marginal confidence intervals for each parameter (Appendix 3.13), which means that the DEB model is stable.

The predicted suitable areas were projected for the different size classes (Fig. 3.15a). Overall, the Antarctic shelf is suitable to the largest individuals (> 5 cm), while the Magellanic Plateau is predicted as suitable for individuals < 4 cm. Suitable areas for individuals of the maximum size class are restricted to regions of East Antarctica (Prydz Bay, the Amundsen-Bellingshausen and the Ross seas) and in the Western Antarctic Peninsula. Areas predicted as unsuitable to the species survival are the South Campbell and northern Kerguelen plateaus. Small individuals (< 2 cm) are predicted to survive at all latitudes south of 45° south, from the Magellanic Plateau to the Antarctic shoreline but individuals of 1 to 2 cm are restricted to the Kerguelen Plateau, the Western Antarctic Peninsula and some regions in East Antarctica.

Reproduction is possible when individuals grow over 3 cm in diameter, that are individuals able to invest energy into reproduction (Fig. 3.15b). Suitable areas for the species to reproduce are mainly located on the Magellanic Plateau and East Antarctica, in Prydz Bay and the Amundsen-Bellingshausen and the Ross seas. The Kerguelen and Campbell plateaus are predicted as unsuitable to the species reproduction as hypothetical individuals present in these areas would never reach sexual maturity.

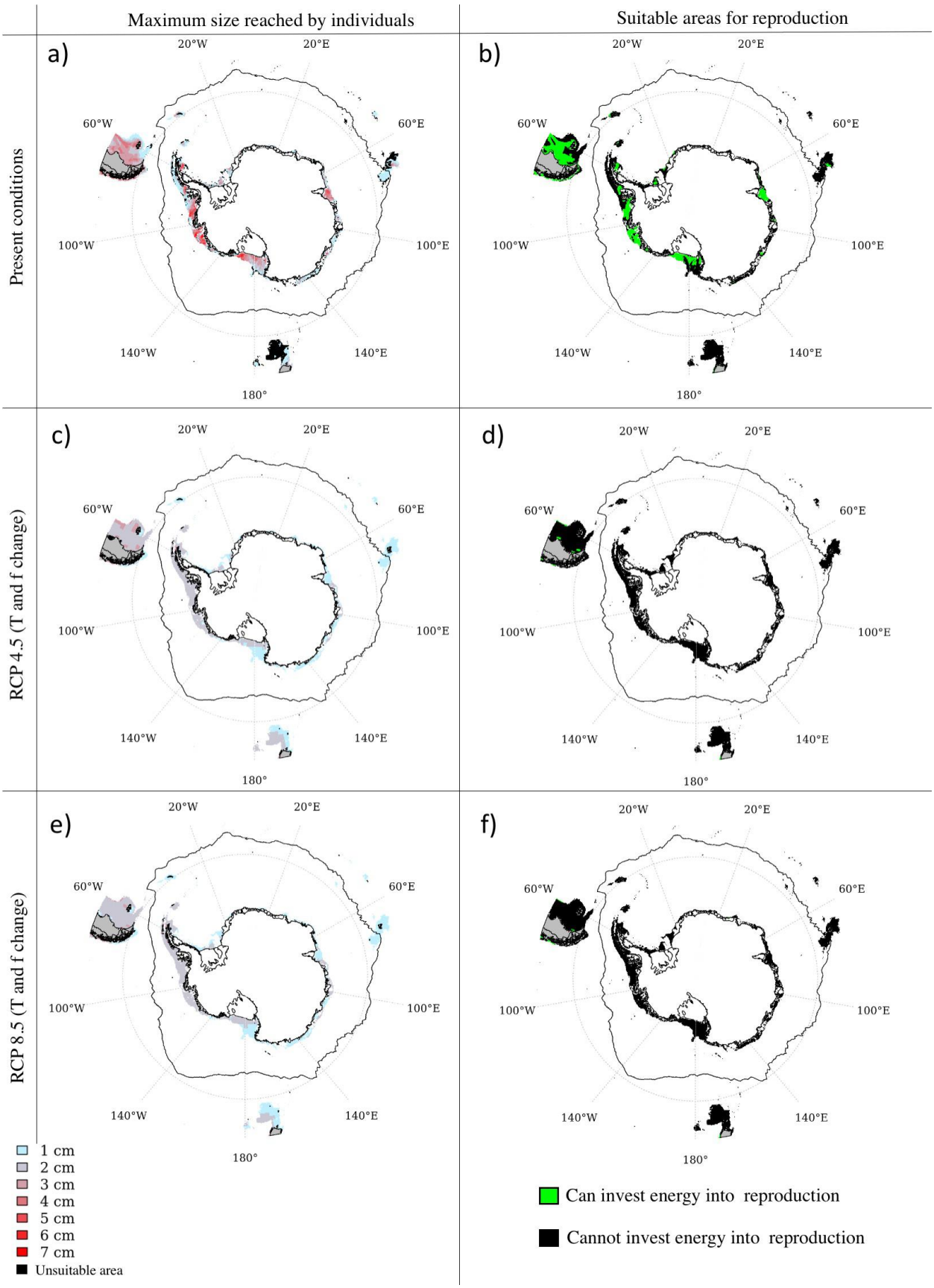


Figure 3.15. Projections of the mechanistic ecological niche model (ENMm, DEB). (a,c,e) classes of maximum size reached by individuals and (b,d,f) suitable areas for reproduction under present-day (a,b), RCP 4.5 (c,d) and RCP 8.5 (e,f) scenarios. Future projections were modelled for both food and temperature changes.

3.2. Projections under IPCC scenarios of climate change

Correlative ecological niche model (ENMc).
Projections of ENMc of *S. neumayeri* according to IPCC scenarios RCP 4.5 and RCP 8.5 (Fig. 3.16) display few changes compared to present-day maps (Fig. 3.14a), and both scenarios give very similar results. Areas predicted as suitable under future conditions are mainly predicted in the Ross Sea and in East Antarctica. In contrast, the species presence probabilities are low in the Bellingshausen and Amundsen seas compared to present-day projections. All areas located north of the Polar Front are predicted as unsuitable with very low presence probabilities ($p < 0.2$).

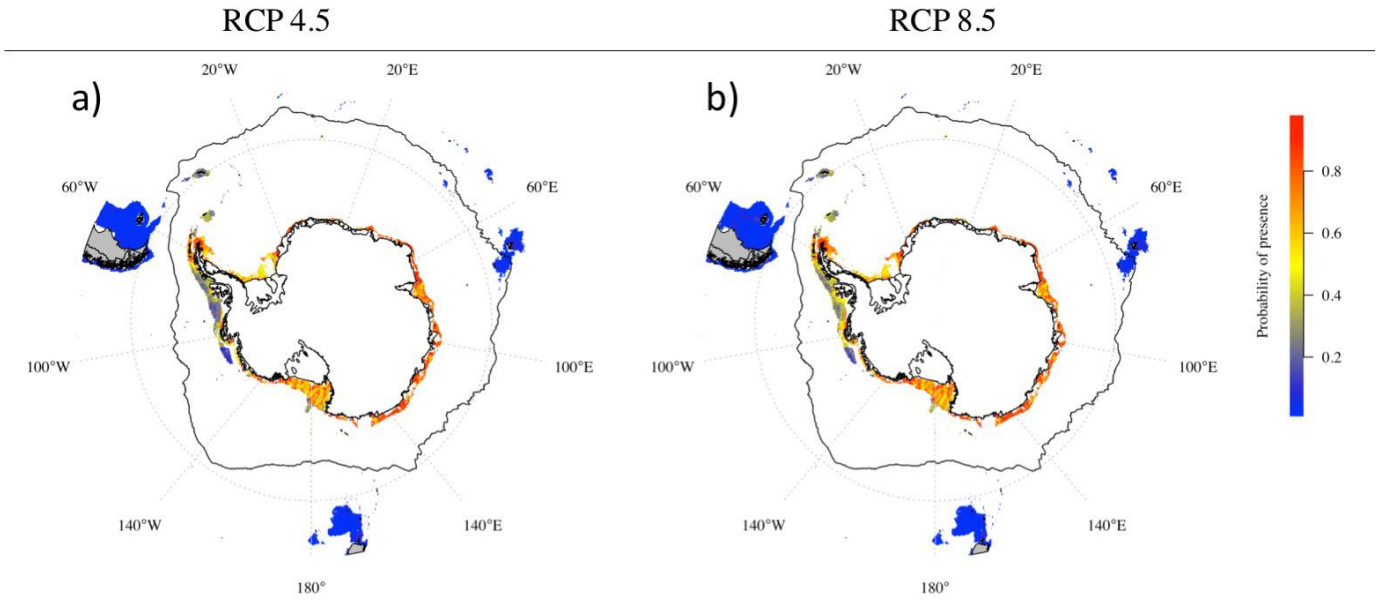


Figure 3.16. Projections of the correlative model under (a) RCP 4.5 (left panel) and (b) RCP 8.5 (right panel) scenarios [2050-2099].

Projection of the mechanistic Ecological Niche Model (ENMm).
Three projections were performed for each IPCC scenario according to (1) both food availability and temperature changes (Fig. 3.15c,d,e,f), (2) temperature only (Fig. 3.17a,b, Fig. 3.18a,b) and (3) food availability only (Fig. 3.17c,d, Fig. 3.18c,d). "Food and temperature" and “food only” projections give similar model outputs under both IPCC scenarios for maximum size and reproduction areas (Fig. 3.15c,d,e,f, Fig. 3.17c,d, Fig. 3.18c,d). The main differences with present-day models are located on the Antarctic shelf and Magellanic Plateau, which are mostly predicted as unsuitable to the species. In contrast “temperature only” projections (Fig. 3.17a,b, Fig. 3.18a,b) show no noticeable change with present-day models, and model outputs are identical under both IPCC scenarios of climate change.

Projections of "food and temperature" (Fig. 3.15c,e) and “food only” (Fig. 3.18c,d) models predict that individuals may reach very small sizes over the entire species distribution range, with a maximum size predicted to reach 1 cm only in the Weddell and Ross seas, in East Antarctica and on the Kerguelen and Campbell plateaus. Size is also predicted to be small (< 2 cm) along the Antarctic Peninsula and on the Magellanic Plateau. As a consequence, reproduction is predicted as impossible over the entire species distribution range under future IPCC scenarios, the model predicting that no energy would be available for maturity, maintenance and reproduction (Fig. 3.15d,f, Fig. 3.17c,d).

The "temperature only" model (Fig. 3.18a,b) predicts unsuitable areas for growth over the Kerguelen Plateau and some areas in East Antarctica (Prydz Bay excepted). In contrast, large individuals (> 4 cm) are predicted in the Bellingshausen-Amundsen seas, the Ross Sea and on the Magellanic Plateau. Suitable areas for the species reproduction match with areas where

individuals can reach up to 2 cm in size that is, in the Bellingshausen-Amundsen seas, the Ross Sea and Prydz Bay areas (Fig. 3.17a,b).

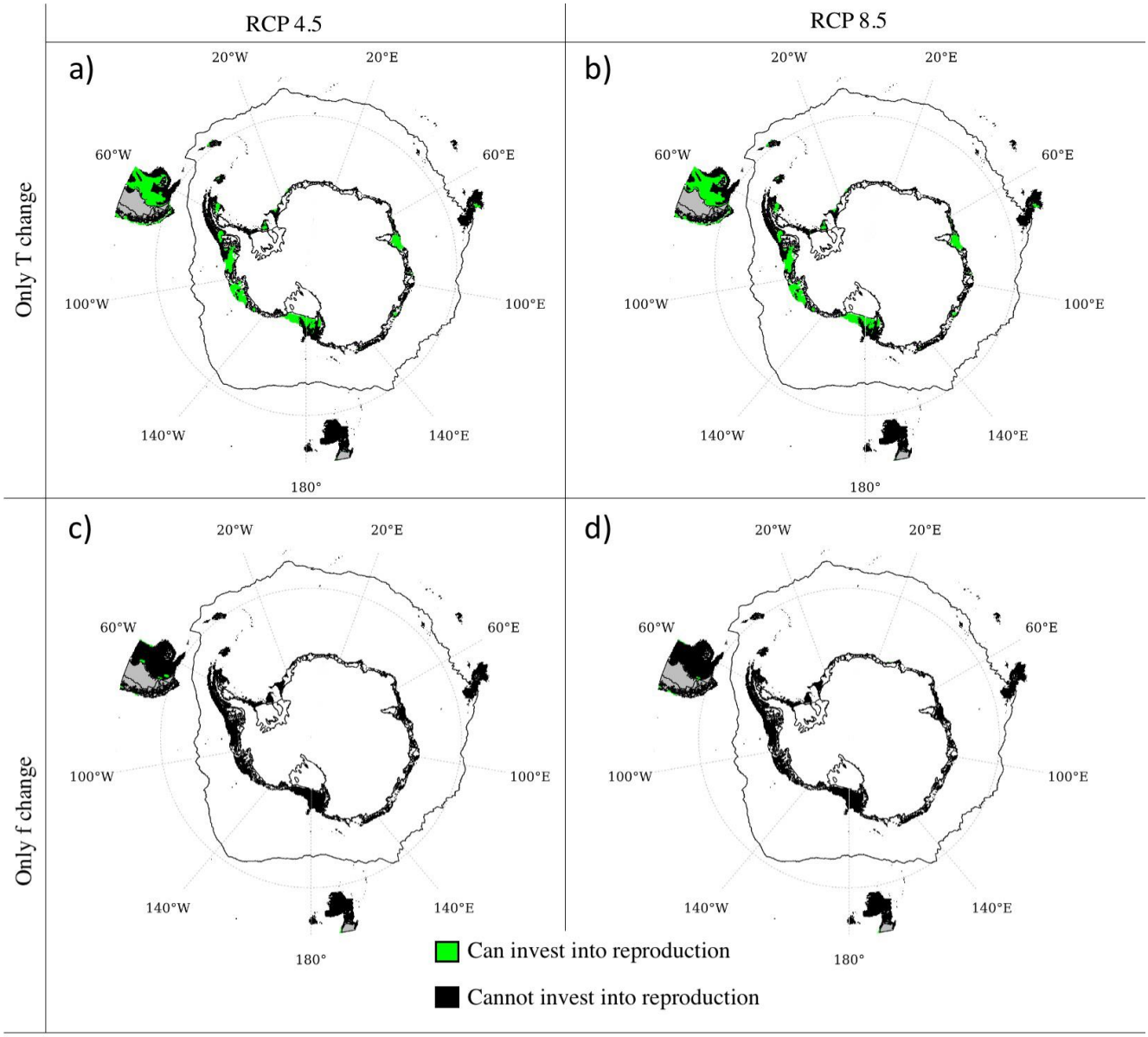


Figure 3.17. Projections of the DEB ENMm under future conditions: predicted suitable areas to the species reproduction under IPCC scenarios RCP 4.5 (left panel) and RCP 8.5 (right panel). Predictions were modelled for temperature change only (top panels) and food availability change only (bottom panels), respectively.

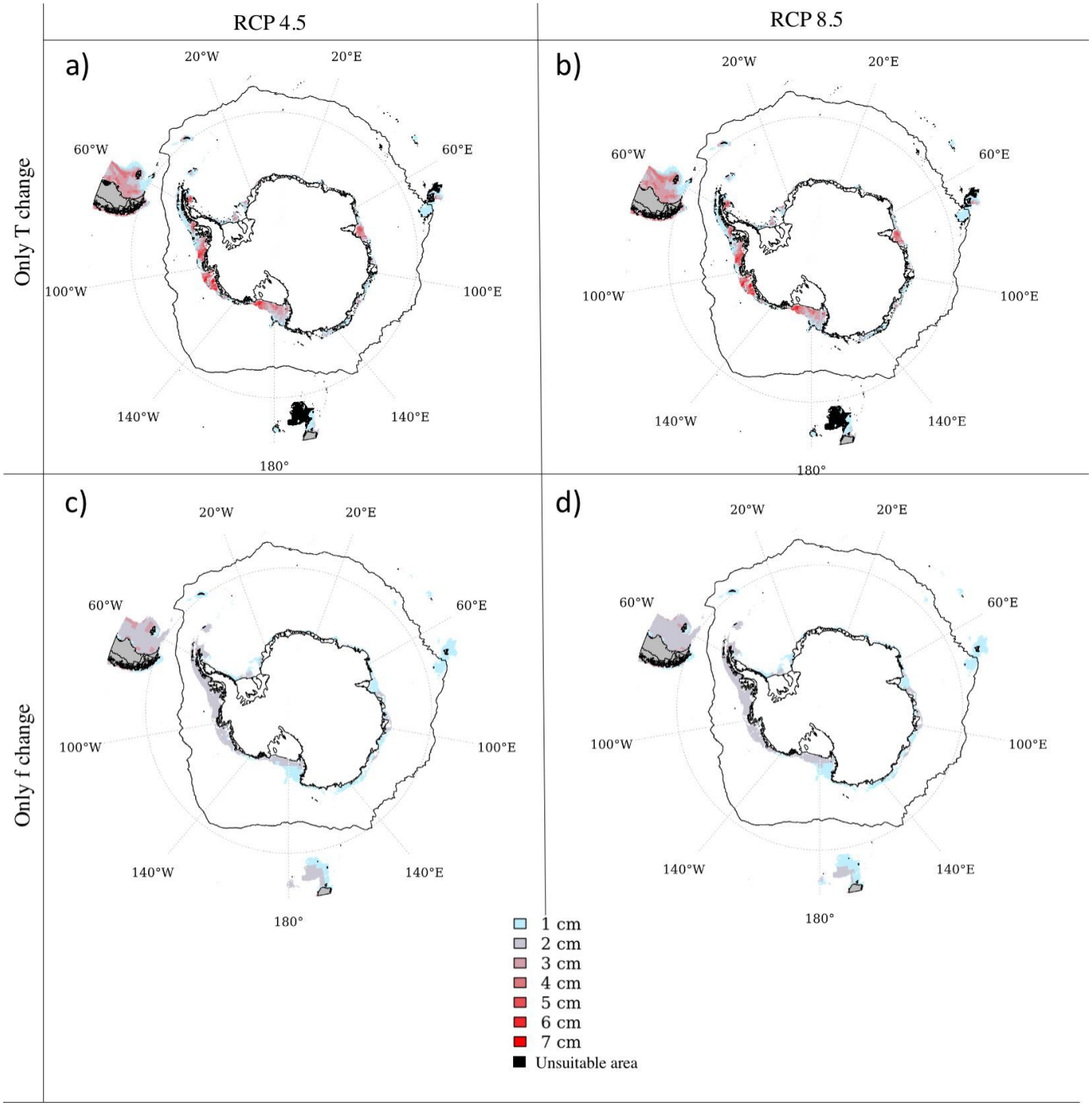


Figure 3.18. Projections of the DEB ENMm under future conditions: maximum size reached by individuals under IPCC scenarios RCP 4.5 (left panel) and RCP 8.5 (right panel). Predictions were modelled temperature change only (top panels) and food availability change only (bottom panels), respectively.

4. DISCUSSION

4.1. Model projections and their ecological significance

Present-day projections.

The ENMc predicts suitable conditions to *S. neumayeri* in Antarctic cold waters south of the Polar Front for the present time period (temperature < +2°C, Fig. 3.13, Appendix 3.15). This is in line with our knowledge of the species biogeography, which is endemic to the Antarctic continental shelf (Pierrat et al. 2012, Fabri-Ruiz et al. 2019, 2020). Temperature is usually a major driver of species distribution as already shown in former studies on Antarctic echinoid species (Saucède et

al. 2014, 2017, Guillaumot et al. 2018b, Fabri-Ruiz et al. 2019). Along with geomorphology, slope and depth, these variables are related to main habitat characteristics (Appendix 3.16) and are considered to have a dominant role in the structure and composition of benthic communities (O'Brien et al. 2009, Kaiser et al. 2013, Post et al. 2014).
In addition to the importance of the environment, the endemicity of Antarctic benthic fauna is also believed to be favored by the presence of the Antarctic Circumpolar Current acting as a biogeographic barrier to dispersal towards the north (Arntz et al. 1997, Linse et al. 2006, Barnes and Griffiths 2007, Griffiths et al. 2009). For instance, 68% of Antarctic echinoids species (Saucède et al. 2014), 74% of gastropods (Schiaparelli and Linse 2014) and 57% of bivalves (Linse 2014) were reported to be endemic to the Antarctic continental shelf.

Stable DEB models were produced (Appendix 3.13) and projections also show that under present conditions, Antarctic regions such as the western part of the Ross Sea, Prydz Bay area, the East Antarctic Peninsula, and the Bellingshausen-Amundsen seas are predicted to be suitable for the species growth and reproduction (Fig. 3.15). This is line with observed data in these regions where *S. neumayeri* is adapted to low temperatures with display of low aerobic scopes (Peck and Conway 2000, Peck 2002, Pörtner and Knust 2007). Previous works focused on the development rate of embryos and data were provided on the range of suitable temperatures for planktonic larvae to grow. Stanwell-Smith and Peck (1998) showed an increase in development rates between -2°C and +2°C, with low and stable rates between +0.2°C and +1.7°C. Development rates do not increase for temperatures above +2°C. Bosch et al. (1987) and Pauline et al. (2013) reported the onset of larval development between -0.8°C and +0.5°C, and between -1.8°C and -0.9°C respectively. Kapsenberg and Hofmann (2014) reported a larval upset at -0.7°C. Finally, food supply is also reported as sufficient for individuals to survive and allocate energy to reproduction (Appendix 3.14).
In contrast, the Kerguelen Plateau, the Western Antarctic Peninsula, East Antarctica (except Prydz Bay) and eastern part of the Ross Sea were modelled as suitable areas but for small individuals only (< 2cm). In these regions, the energy available and stored into the reserve compartment ($\dot{p}_c$) is only sufficient to ensure somatic maintenance ($\dot{p}_M$) but cannot cover energy costs related to growth and/or reproduction ($\dot{p}_M > \kappa.\dot{p}_c$) as the somatic maintenance has priority over processes in the model. In these regions, the maintenance of species populations would exclusively depend on larval supply from other areas. This could be possible *via* the Antarctic Circumpolar Current that is a major vector of larval dispersal in the Southern Ocean (Pearse et al. 2009, Moon et al. 2017, González-Wevar et al. 2018a) but this hypothesis remains to be tested and supported by field data.

Projections under future scenarios of climate change.
Future projections of ENMc showed few changes in the species potential distribution over the Antarctic shelf. This can be explained by the important contribution of physical descriptors, geomorphology, slope and depth to the model, three variables that were considered unchanged in a near future in the model, being here considered that predictions of sea level rise should have little effect on model outputs at large, ocean-wide scale (De Conto and Pollard 2016). Local shifts in the species distribution probabilities are however predicted, compared to the present-day model. They are mainly localized in the Bellingshausen-Amundsen seas and are triggered by future predictions of temperature rise and reduction in sea ice coverage (Appendix 3.14). A reduction in sea ice coverage will have serious impacts on the seasonal production of food supply and will also result in a reduction of the protection of shallow benthic organisms from UV-B induced damages (Lister et al. 2010). Changes in ice regime are also expected to have multiple impacts in the region due to ice shelf melting and collapses. This will result in the freshening of Antarctic waters and associated changes in water biogeochemistry, and to an increase in the intensity of iceberg scouring on seabeds in shallow water, coastal areas (Meredith and King 2005, Bracegirdle et al. 2008, Stammerjohn et al. 2012). This phenomenon was shown to have serious effects on the structure of benthic communities, (Gutt 2001, Gutt and Starmans 2001, Gutt and Piepenburg 2003), resulting in a decrease in habitat heterogeneity and local (alpha) diversity (Brown et al. 2004, Barnes and Souster 2011).
In projections of the ENMm performed for future conditions, the combined effect of "temperature and food change" on individual physiology is predicted to induce important shifts in energy

availability (Fig. 3.15c,d,e,f). The allocation of energy into reproduction is predicted to become impossible anymore and growth rates are predicted to strongly decrease in the entire Southern Ocean. These results suggest a high sensitivity of *S. neumayer*i to environmental changes under RCP 4.5 and 8.5 scenarios. Overall, this also stresses the important impact of food availability for benthic species. The seasonal phytoplankton bloom is known to constitute an important source of food for many species (Brockington and Peck 2001, Ahn et al. 2003, Jacob et al. 2003, Michel et al. 2016, Agüera et al. 2017), and predicted shifts and decrease in this resource might have important consequences for marine communities.
In the ENMm, the future "only temperature change" projection (Fig. 3.17a,b, 3.18a,b) is identical to the present-day projection. Medium size (~ 4cm) to large (> 5 cm) individuals as well as suitable areas for reproduction are predicted north of the Polar Front for both periods. We could expect a synergetic and cumulative effect on growth and reproduction under "temperature and food change" (Fig. 3.15) than under "only food change" (Fig. 3.17c,d, Fig. 3.18c,d) or "only temperature change" (Fig. 3.17a,b, Fig. 3.18a,b). On the contrary, our results suggest similarities between "only food change" and "temperature and food change" projections. Metabolic rates of Antarctic species increase with temperature, as does the oxygen consumption. If temperature rises and oxygen supply are insufficient to meet the organism metabolic needs, the organism switches to an anaerobic metabolism (Peck and Conway 2000, Peck 2002, Pörtner and Knust 2007). The ability of individuals to survive depends on their ability to maintain an anaerobic metabolism over time. As a result, rising temperatures should lead to changes in the survival and resilience of Antarctic marine invertebrates.

S. neumayeri occurs in shallow waters compared to other *Sterechinus* species (David et al. 2005, Díaz et al. 2011). The hypothesis of a possible in-depth migration to colder water areas may be considered. In the future, warmer temperatures could occur in deeper areas corresponding to optimal temperature window of the species and decrease in sea-ice cover could also lead to higher exposure to UV-B in shallow waters. However, studies suggest that it may compete in these environments with *Sterechinus diadema*, its sister species living in deeper habitats (Jacob et al. 2003, Díaz et al. 2011). Moreover, pressure increase with depth reduces the thermal optimal window for the development of eggs and embryos, generating a new physiological stress and reducing the species fitness and survival (Tyler et al. 2000). It can therefore be assumed that current environmental changes are expected to lead to a potential reduction in the distribution of *S. neumayeri*.

4.2. Model comparison and complementarity

Model comparison.
Overall, ENMc and ENMm run for present-day conditions provide congruent projections (Fig. 3.14, 3.15a,b). For the Antarctic shelf, in regions such as the Ross Sea and the Prydz Bay area in particular, the ENMm predicts the prevalence of large (> 4cm) and sexually mature individuals and the ENMc shows high presence probabilities. These regions are characterised by cold temperatures and high food availability (f > 0.5), which are favorable conditions for the species development and survival. In contrast, in the northern Kerguelen Plateau and the Campbell Plateau, low presence probabilities are modelled by the ENMc due to warm water temperatures (> 4°C) (Fig. 3.14, Appendix 3.14), and small (< 1 cm) and sexually immature individuals are predicted by the ENMm due to low food availability limiting growth and reproduction (Appendix 3.14). Model projections however do not match for certain areas. For instance, small and sexually immature individuals are predicted along the Antarctic Peninsula in the ENMm, whereas the ENMc predicts high presence probabilities. In the sub-Antarctic, the ENMm predicts suitable conditions for the species growth (> 3cm) and reproduction on the Magellanic Plateau, whereas this area is predicted unsuitable in the ENMc. *S. neumayeri* is known to be endemic to the Antarctic Peninsula and East Antarctic shelf (David et al. 2005, Saucède et al. 2014), which suggests that the ENMm projection may not predict the species current distribution properly. This can be explained by the lack of eco-physiological data documenting the species response to variations in food resources and temperature (Bosch et al. 1987, Stanwell-Smith and Peck 1998, Marsh et al. 1999, 2001, Tyler et al. 2000, Brockington and Peck 2001, Alexander et al. 2017). On the other hand, temporal

scales of physiological experiments are over a limited time frame and different from the temporal scale of the used environmental layers, which characterise overall climate conditions.

In ENMm, the Arrhenius temperature is the parameter that determines the metabolic rate as a function of temperature variation (Appendix 3.12). In the present model, the Arrhenius temperature was estimated based on three measurement points only (Bosch et al. 1987), which may induce a lack of precision in the simulation of the species metabolic rate. In addition, lower and upper lethal temperatures could not be entered in the model due to the absence of relevant physiological data (Appendix 3.10) and the species optimal temperature range could not be determined precisely. As a consequence, the modelled physiological performance of the species tends to increase constantly with temperature and partly outreaches the biological optimum.

Only data on chlorophyll-a concentration and on the gut content were available to model the functional response of *S. neumayeri* to food resources (McClintock 1994, Jacob et al. 2003, Michel et al. 2016). Therefore, in the model, sea surface chlorophyll-a concentration in summer was used as a proxy of food resources for *S. neumayeri* (Appendix 3.9), which is an opportunistic, omnivorous feeder. The species does not feed directly on chlorophyll-a but is indirectly dependent on this food supply as it feeds on various sources of particulate organic matter deposited on the sea floor as well as some suspension feeders (Smith et al. 2006, Lohrer et al. 2013, Petrou et al. 2016, Schofield et al. 2017). In addition, winter conditions are known as periods of low chlorophyll-a concentrations in Antarctic surface waters (Thomalla et al. 2011, Deppeler and Davidson 2017), which could not be used as input in the model projection due to the lack of satellite data for this season. In a DEB model developed for the Antarctic bivalve *Laternula elliptica* (King and Broderip 1832), Agüera et al. (2017) - Appendix showed that reserve is seasonal and that low food availability generated a 25% loss in the species body mass, also delaying gonadal development. In *S. neumayeri*, post-metamorphic individuals do not feed in winter (Brockington and Peck 2001) but no quantitative data on energy allocation are available for this season. Additional works would be useful to refine the present DEB model. Complementary data based on new eco-physiological experiments describing the effect of different levels of food supplies, abundant, limited, or starvation, on the metabolic rate should contribute to improving model accuracy (Sarà et al. 2013, Augustine et al. 2014, Hamda et al. 2019).

Complementarity between modelling approaches.

The two modelling approaches mainly differ in their scientific objectives. To run the ENMc, 13 abiotic parameters were used to describe part of the species realised niche, the effect of biotic interactions and biogeographic constraints also indirectly acting on model outputs through the position of observed occurrences and the spatial correlation between abiotic descriptors, biotic factors and biogeographic barriers. Projections therefore partly fit to the species realised distribution because they partly take into account the multi-dimensions of the species realised niche. Parameters of the physical habitat such as geomorphology were shown to have an important role in the structuring and composition of Antarctic benthic communities (O'Brien et al. 2009, Kaiser et al. 2013, Post et al. 2014); such parameters were not considered to run the ENMm. In contrast, the ENMm integrates the effect of temperature and food resources on the species physiology, focusing on two dimensions of the species fundamental niche, whatever its distribution and realised niche. The ENMm provides biological insights for understanding the physiological processes that underpin the observed species distribution.

Major differences between models show up when it comes to run future projections under IPCC RCP scenarios. ENMm models predict unsuitable conditions for the species growth and reproduction over the entire ocean. In contrast, ENMc models predict the species persistence on the Antarctic shelf, the Bellingshausen and Amundsen seas excepted. The ENMc uncertainties increase when species' responses to environmental conditions are extrapolated out of the range of values for which the model was trained (Guillaumot et al. 2020c - Chapter 2). This holds particularly true for future conditions that do not prevail in present-day environments yet (Fitzpatrick and Hargrove 2009, Elith et al. 2010, Jiménez-Valverde et al. 2011, Dormann et al. 2012a) so that the ENMc may fail to predict as unsuitable environmental conditions that would exceed the species physiological tolerance (Anderson 2013). Moreover, without presence-absence

or abundance data, habitat suitability is partly biased because all presences are treated equally. With presence-absence and if possible, abundance data, more discrimination of suitable habitat is gained, which is beneficial when ENMc are used to project species distribution across space and time. Adding absence data is known to provide greater ability to delineate species range boundaries and produce more accurate models (Howard et al. 2014, Yates et al. 2018).

Such discrepancies between the two modelling approaches in a context of climate change were already highlighted in previous studies. For instance, Buckley et al. (2011) showed that ENMm predicted much greater migrations with climate change than ENMc in a study on Lepidotera. Further, Kearney et al. (2008) predicted that toad species survival in southern Australia would no longer be possible due to global warming according to ENMm, while the ENMc still predicted the region as suitable.
In the present study, while suggesting unrealistic projections on the Magellanic Plateau under present-day conditions, future projections of the ENMm are more in line with a majority of works suggesting that climate change would induce unsuitable conditions to the survival of Antarctic benthic marine ectotherms (Peck et al. 2014, Hawkins et al. 2018). All these results highlight the necessary complementarity of ENMc and ENMm approaches for providing independent and relevant projections, relying either on biogeographic (ENMc) or physiological (ENMm) data (Morin and Thuiller 2009, Kearney et al. 2009). Comparing and combining projections from different modelling approaches provide more insight on both species present-day distributions and sensitivity to future projections (Guisan and Zimmermann 2000, Elith and Graham 2009, Elith et al. 2010).

4.3. Future prospects

The present work underlines ENMc as a useful and powerful approach to predict current species distribution. ENMc are relatively simple to implement and do not require a deep knowledge of population dynamics nor of ecological processes linking organisms to their abiotic environment. They can be applied to a large number of taxa (Guisan and Zimmermann 2000, Elith and Graham 2009, Elith et al. 2010) and are often used upstream to address conservation issues (Evans et al. 2015). However, ENMc do not imply any inference on causal relationships between species distribution and environmental descriptors, and such relationships may also imply indirect responses to collinear variables that are not entered in the model (Guisan and Thuiller 2005).

In a context of environmental changes, extrapolation represents a serious limitation to ENMc that have limited capacities to transfer model outputs both in space and time (Yates et al. 2018, Guillaumot et al. 2020c - Chapter 2). In the present study, such a limitation is highlighted by the mismatch between ENMm and ENMc future projections of *S. neumayeri*. ENMm appear to be more informative than the ENMc when it comes to describe species distribution under changing environmental conditions. However, few Antarctic species have been the focus of detailed eco-physiological studies and few mechanistic models were developed, considering the important amount of physiological data required. Mechanistic models are therefore generally used when species physiology has been relatively well studied (Kearney et al. 2008, Buckley et al. 2011, Evans et al. 2015, Thomas and Bacher 2018) and our knowledge of marine species physiology is usually biased toward 'model' species that most interest the public and researchers (Clark and May 2002, Sousa-Silva et al. 2014, Feng and Papeş 2017). Many authors have stressed the importance and benefits of considering mechanistic approaches for conservation purposes and the implementation of management plans (Cooke and O'Connor 2010, Cooke et al. 2012, 2014, Evans et al. 2015). If the integration of biological data into open-access databases has significantly increased with multiple initiatives such as TRY, Globtherm, FSRD, Anage, GenBank, add-my-pet (De Magalhaes and Costa 2009, Kattge et al. 2011, Karányi et al. 2013, Bennett et al. 2018, Marques et al. 2018), there is still no data portal devoted to describing species physiological traits. Mining such data through experiments and the literature to perform mechanistic models remains a complex, time-consuming task, limiting the integration of ENMm into conservation strategies (Evans et al. 2015). In contrast, ENMc are mainly based on occurrence or abundance data that are made available through international databases allowing open-access data sharing (Pearse et al. 2018, Wüest et al. 2020). Common databases would be particularly valuable to address ecological

issues linking patterns to processes across spatial and temporal scales, and improving our knowledge of ecosystem functioning in a context of climate change (Sutter et al. 2015).

4.4. Conclusion

The present study highlights the complementarity of correlative and mechanistic ENM to predict species present distributions and sensitivity to changing environmental conditions. Overall, congruent projections were obtained with the two modelling approaches for present-day conditions. In contrast, different models were generated under future scenarios. Both models agree on the fact that *S. neumayeri* is circum-polar in distribution with suitable areas restricted to the Antarctic continental shelf area (< 400m), with low temperatures (< 2°C), limited sea ice concentrations (< 50%) and high food availability (f > 0.7). The ENMm approach provided an additional understanding of physiological processes determining the species distribution with regards to growth and sexual maturity as a function of temperature and food availability.

The combination of ecological modelling, ENMm and ENMc, with satellite remote sensing and climate models provides a valuable approach to study large-scale responses of marine species to climate change (Guisan and Thuiller 2005, Pearson 2007, Kearney and Porter 2009, Elith and Leathwick 2009, Buckley et al. 2011, Thomas and Bacher 2018, Rodríguez et al. 2019). Multiple challenges however remain to be overcome. Eco-physiological data are still needed to produce reliable mechanistic DEB models, including data on Arrhenius temperatures. In addition, ENMm do not take into account extrinsic factors that shape species distribution such as biogeographic barriers, physical habitats and biotic interactions (predation/competition/facilitation). Combining correlative and mechanistic models in an integrative approach therefore constitutes a promising perspective, which has already been developed for certain terrestrial and marine organisms (Elith et al. 2010, Dormann et al. 2012a, Roos et al. 2015, Mathewson et al. 2017, Rodríguez et al. 2019), and could prove particularly relevant to predict the sensitivity of Antarctic organisms to a fast changing environment.

APPENDIX 3.5. Life cycle of *Sterechinus neumayeri*

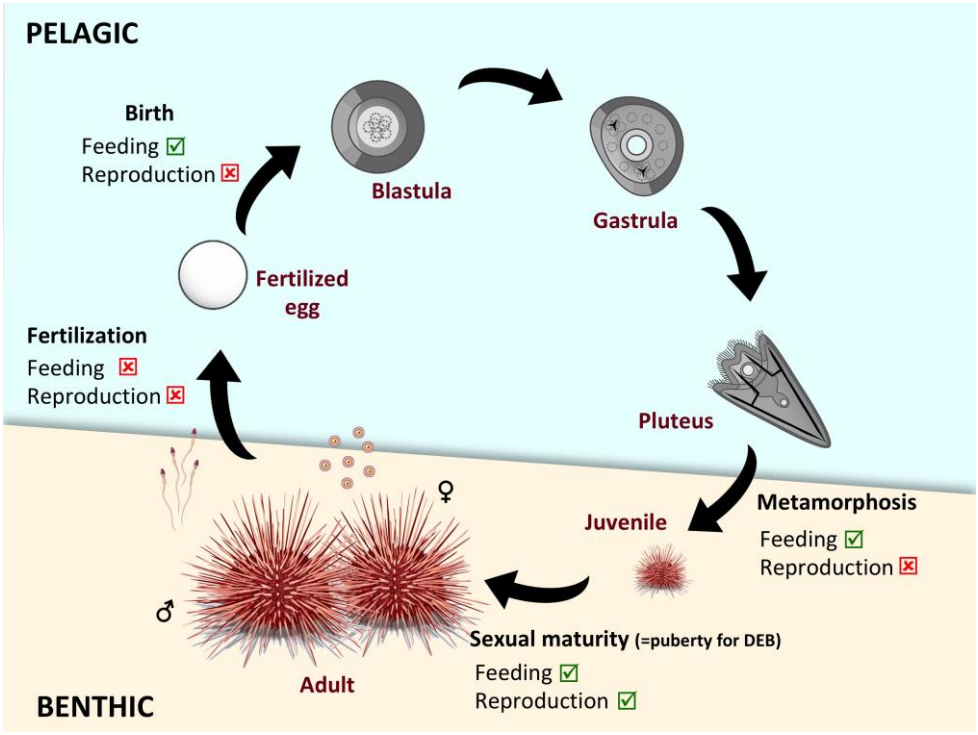


Figure S3.5. Life cycle of *Sterechinus neumayeri*. Red crossed boxes: inactive functions at different life stages; green ticked boxes: active functions at different life stages. From Fabri-Ruiz (2018).

Life cycle of *Sterechinus neumayeri* starts with sexual mature individuals. The vitellogenic cycle lasts from 18 to 24 months, with oocytes starting to develop during the first winter and achieving development during the second winter (Brockington et al. 2007). The gametes are then expelled in the water column, where fertilization occurs, and the onset of the embryonic stage takes place between mid-November and December (Pearse and Giese 1966, Bosch et al. 1987). Pre-metamorphic larvae appear between late December and early March when they are able to feed (i.e. DEB 'birth' stage), taking advantage of the summer phytoplankton bloom (Chiantore et al. 2002). The larval recruitment on the sea bed corresponds to the metamorphosis stage and mainly occurs between the end of February and March of the following year (Bosch et al. 1987).

APPENDIX 3.6.

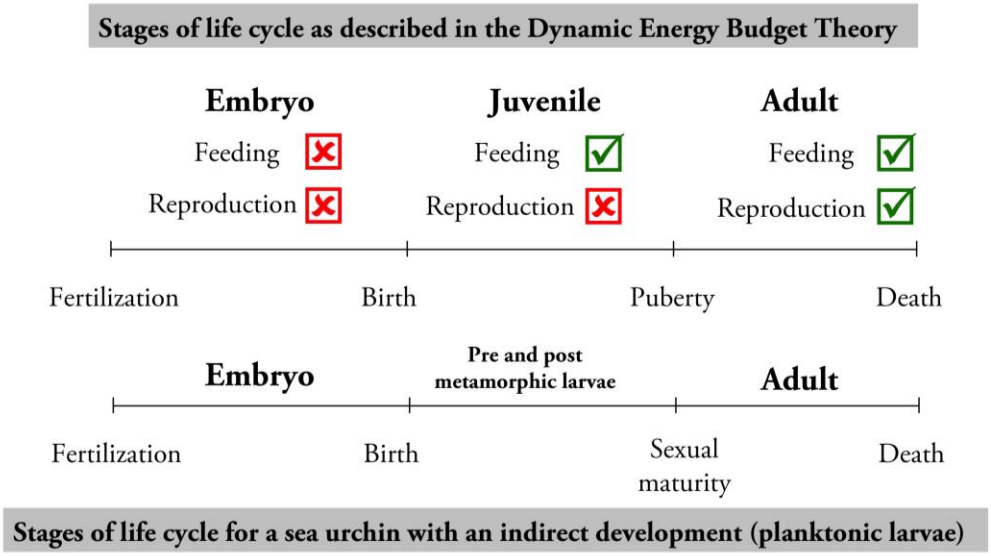


Figure S3.6. Reproduction and feeding functions represented over a theoritical life cycle according to DEB theory and correspondence with the life cycle of *Sterechinus neumayeri*. Red crossed boxes: inactive functions at different life stages ; green ticked boxes: active functions at different life stages.

In a standard DEB model (Kooijman 2010), the organism is isomorphic, i.e. it maintains the same shape throughout its entire life cycle. This life cycle is characterised by three stages that are distinguished by their energy flow: the embryo, juvenile and adult (Fig. S3.6). The embryo does not assimilate food and relies on reserves. The juvenile stage happens after birth (i.e. according to DEB theory, it corresponds to the moment when the organism is able to feed). The transition between the embryo and juvenile stages occurs when the individual has reached a particular threshold of energy invested into its development.
At this point, the individual is complex enough to start feeding and uses the energy gained from food to continue its development, growth and maintenance but it does not provide energy into reproduction. The other stage of the organism development is the transition from the juvenile to the adult stage called puberty. After this stage, when the organism becomes an adult, it stops allocating energy into its development and redirects the energy towards reproduction and the production of gametes.

APPENDIX 3.7.

Table S3.7. Environmental descriptors used to build ENMc models for the current period. Predictor 'range' is here defined as the difference between winter and summer mean values.

Environmental data	Years	Units	Sources
Depth		meter	http://topex.ucsd.edu/WWW_html/mar_topo.html
Geomorphologic features		categorial	ATLAS ETOPO2 2014 (Douglass et al. 2014)
Sea surface salinity	2005-2012	PSS	https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
Sea ice concentration		unitless	(Mormède et al. 2014c)
Sea surface salinity	2005-2012	PSS	https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
Sea surface temperature range	2005-2012	°Celsius degrees	https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
Seafloor oxygen concentration	1955-2012	mL/L	https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
Seafloor salinity	2005-2012	PSS	https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
Seafloor salinity range	2005-2012	PSS	https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
Seafloor temperature	2005-2012	°Celsius degrees	https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
Seafloor temperature range	2005-2012	°Celsius degrees	https://www.nodc.noaa.gov/OC5/woa13/woa13data.html
Slope		unitless	(Mormède et al. 2014c)
Summer chlorophyll concentration	2002-2009	mg/m^3	(Mormède et al. 2014c)

APPENDIX 3.8.

Table S3.8. Environmental descriptors used to build the ENMc models for future IPCC scenarios (RCP 4.5 and RCP 8.5). Predictor 'range' is here defined as the difference between winter and summer mean values.

Environmental data	Years	RCP	Units	Sources
Sea ice concentration	2050-2099	RCP 4.5 / RCP 8.5		https://www.esrl.noaa.gov/psd/ipcc/ocn
Sea surface salinity	2050-2099	RCP 4.5 / RCP 8.5	PSS	https://www.esrl.noaa.gov/psd/ipcc/ocn
Sea surface salinity range	2050-2099	RCP 4.5 / RCP 8.5	PSS	https://www.esrl.noaa.gov/psd/ipcc/ocn
Sea surface temperature range	2050-2099	RCP 4.5 / RCP 8.5	Celsius degrees	https://www.esrl.noaa.gov/psd/ipcc/ocn
Seafloor salinity	2050-2099	RCP 4.5 / RCP 8.5	PSS	https://www.esrl.noaa.gov/psd/ipcc/ocn
Seafloor salinity range	2050-2099	RCP 4.5 / RCP 8.5	PSS	https://www.esrl.noaa.gov/psd/ipcc/ocn
Seafloor temperature range	2050-2099	RCP 4.5 / RCP 8.5	Celsius degrees	https://www.esrl.noaa.gov/psd/ipcc/ocn
Summer chlorophyll concentration	2050-2099	RCP 4.5 / RCP 8.5	mg/m^3	https://www.esrl.noaa.gov/psd/ipcc/ocn

APPENDIX 3.9.

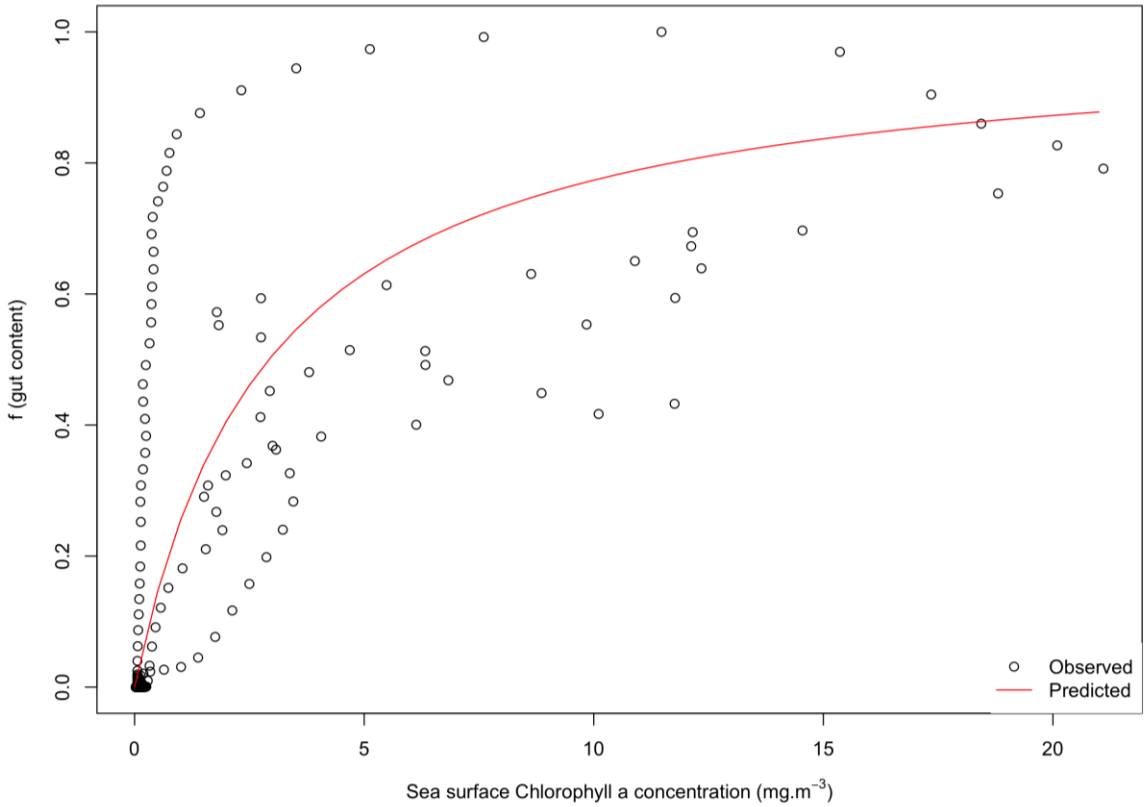


Figure S3.9. Observed values (circles) and projection (line), based on a type II feeding functional response $f = \frac{X}{X+X_K}$. The estimated value for the half-saturation parameter X_k is the food density at which feeding rate is *half* of its maximum value, here X_k 2.95 mg.m^{-3}.

Data for sea surface chlorophyll-a concentration and gut content were extracted from Brockington and Peck (2001). Chlorophyll-a concentration and gut content data do not have the same time interval. These data were not calibrated to the same time interval. Data were splined according to time to get regular time intervals. Then, we used a moving average, which estimates the trend-cycle at time *t* by averaging values within *k* periods of *t*. This method removes transient fluctuations and keeps an overall trend. The same computation has been done for temperature data according to time. These analyses were performed using the *castr* package (https://github.com/jiho/castr). Data for the gut content were corrected by temperature and scaled to values comprised between 0 and 1. A non-linear least squares regression model was then applied to adjust the functional response.

APPENDIX 3.10.

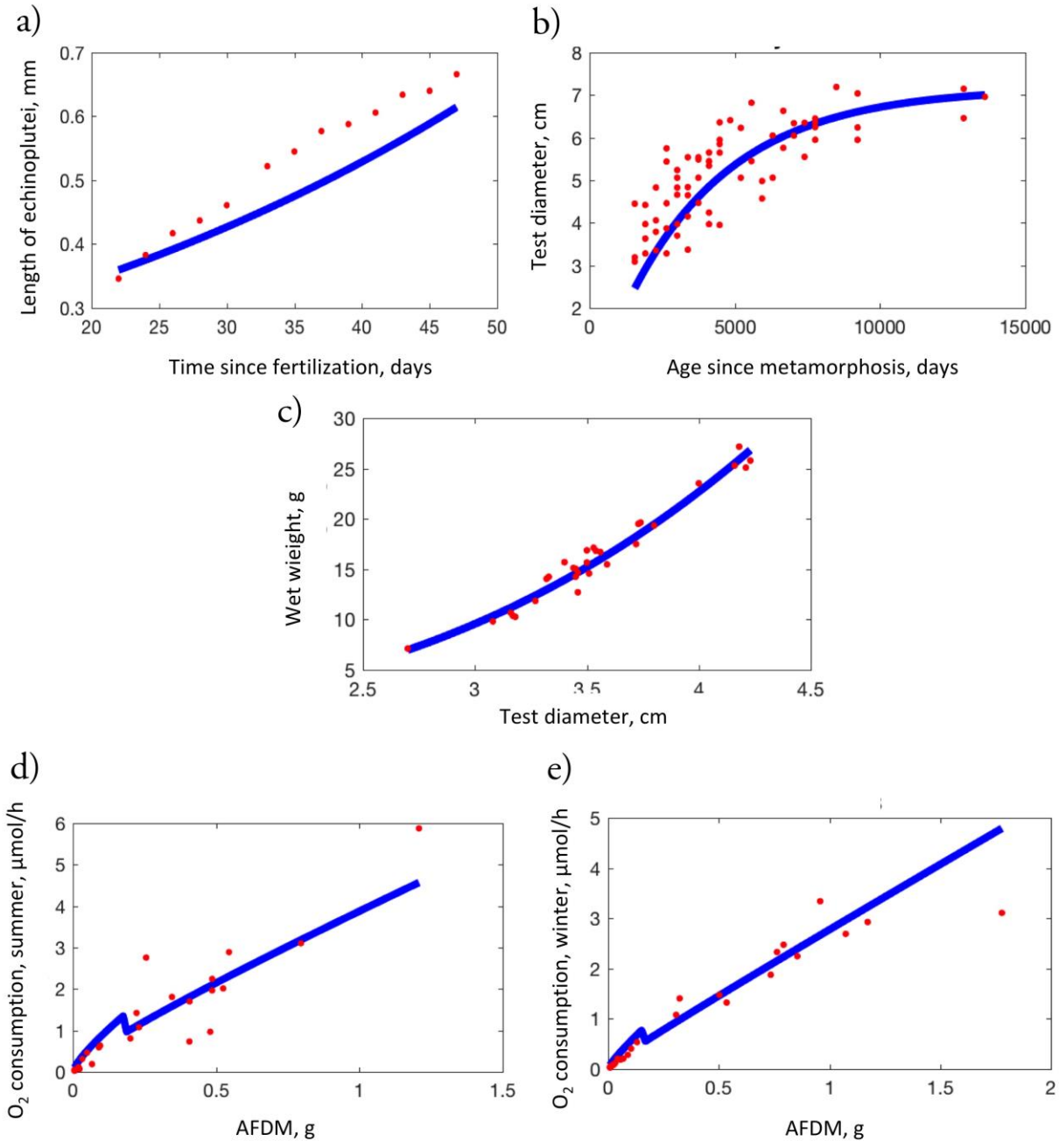


Figure S3.10. DEB model fit (blue curves) and experimental values (red dots) for univariate data: (a) size of larvae as a function of time since fertilization (Marsh et al. 1999), (b) test diameter according to age since metamorphosis (Brey et al. 1995), c) adult wet weight according to test diameter (S. Morley, com. pers), (d) summer oxygen consumption according to weight (AFDM: Ash Free Dry Mass) for post-metamorphic individuals (Souster et al. 2018), (e) winter oxygen consumption according to weight (AFDM) for post-metamorphic individuals (Souster et al. 2018).

APPENDIX 3.11.

Table S3.11. Experimental and predicted DEB modeled values for zero- and univariate data.

Zero-Variate data	Observation	Prediction	Relative Error	Source
Age at birth at -1.35°C (day)	21	20.53	0.023	Bosch et al. 1987
Age at birth at at 0°C (day)	17	17.66	0.039	Bosch et al. 1987
Time since birth at metamorphosis (day)	103	92.64	0.10	Bosch et al. 1987
Life span (day)	$1.46.10^{-4}$	$1.46.10^{-4}$	$8.64.10^{-9}$	Bosch et al. 1987
Length from aboral apex to tips of postoral arms of echinoplutei (cm)	0.035	0.035	$7.59.10^{-9}$	Bosch et al. 1987
Length of the pluteus before metamorphosis (cm)	0.12	0.13	0.11	Bosch et al. 1987
Diameter of test at puberty (cm)	2.04	2.06	$9.\ 10^{-3}$	Guessed *
Ultimate diameter of test (cm)	7.02	7.17	0.02	Brey et al. 1995
C :N mass of fertilised egg (g)	$4.49.10^{-7}$	$5.78.10^{-7}$	0.29	Marsh et al. 1999
C :N mass of pluteus larva (g)	$5.21.10^{-7}$	$4.05.10^{-7}$	0.22	Marsh et al. 1999
Wet weight adult including gonads (g)	129	130.6	0.012	Pearse and Giese 1966
Gonadal somatic index (g)	0.1	0.096	0.042	Pearse and Giese 1966
Univariate data				
Time since fertilization (day) vs. length of echinoplutei (mm)	See Figure S3.10		0.093	Marsh et al. 1999
Age since metamorphosis day) vs. test diameter (mm)			0.13	Pearse and Giese 1966
Test diameter (mm) vs. wet weight (g)			0.053	British Antarctic Survey unpublished data
Ash Free Dry Mass (g) vs. O_2 consumption in µmol/h in winter			0.21	Souster et al. 2018
Ash Free Dry Mass (g) vs. O_2 consumption in µmol/h in summer			0.27	Souster et al. 2018

* based on same relative length compared to *Echinus affinis* (https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/entries_web/Gracilechinus_affinis/Gracilechinus_affinis_res.html)

APPENDIX 3.12. Temperature sensitivity and post-metamorphic shape coefficient

1) Temperature sensitivity

a. The Arrhenius temperature

The Arrhenius temperature (T_A) provides information on metabolic rate variations as a function of temperature and can be calculated from observed, experimental values at different temperatures. In this study, T_A was estimated using embryonic development time data at different temperatures (Bosch et al. 1987). Values were standardized for a value of 1 at a reference temperature of 273 K (0°C). The Arrhenius temperature was then obtained by fitting the Arrhenius function (Eq. A) to the scaled values using a linear least squares regression (Agüera et al. 2015, 2017) with the R package *minpack.lm* (Elzhov et al. 2013).

$$\dot{k}(T) = \dot{k}(T_1).e^{\left\{\frac{T_A}{T_1}-\frac{T_A}{T}\right\}} \quad \text{(A)}$$

$\dot{k}(T)$: Rate at temperature T(K)
$\dot{k}(T_1)$: Rate at reference temperature (T_1)
T_A : Arrhenius temperature (K)
T_1: Reference temperature (K)

b. Thermo-tolerance window

When data are sufficiently detailed, the DEB model can also include the description of the influence of body temperature on physiological rates over the temperature range in which enzymes are assumed to be active and delimited by the parameters T_L and T_H (Eq. B). Above and below the thermo-tolerance window enzymes become inactive, leading to a decline in physiological rates, which can be traced by the parameters T_{AL} and T_{AH}, respectively. These five parameters fully define an organism thermal performance curve, in accordance to the formula:

$$\dot{k}(T) = \dot{k}_1 \cdot \exp\left\{\frac{T_A}{T_1} - \frac{T_A}{T}\right\} \cdot \left(1 + \exp\left\{\frac{T_{AL}}{T} - \frac{T_{AL}}{T_L}\right\} + \exp\left\{\frac{T_{AH}}{T_H} - \frac{T_{AH}}{T}\right\}\right)^{-1} \quad \text{(B)}$$

where $\dot{k}(T)$ is the value of the physiological rate at a given body temperature T and $\dot{k}_1$ is the known value of the metabolic rate at a reference temperature T_1.

2) Post-metamorphic shape coefficient

From birth to metamorphosis, *S. neumayeri* larvae grow exponentially, conducting to an increase of the assimilation flux $\dot{p}_A$ (energy flow from assimilated food into the reserve) and mobilization $\dot{p}_C$ (outflow from the reserve) (Kooijman 2010). In addition, during recruitment (planktonic larval stage to benthic juvenile) (Fig. S3.5, Fig. S3.6), individuals undergo metamorphosis that is, a change in body shape. These changes during the life cycle are included in the DEB model, using the shape parameter δ_M.

The relation between size (L) and the volume of the structure (V) is provided by the following relation:

$$V = \delta_M.L^3 \text{ (C)}$$

The shape coefficient δ_M is used to convert size into the structural volume (V, *i.e* the cube of volumetric length). DEB theory partitions the body mass into the abstract quantities of structural volume *V* and the reserve E. The structure is the 'permanent' biomass such as proteins and membranes proportional to structural volume. The structural volume is the key feature that allows body size to be included in the complete budget of the organism. Several shape factors were used, $\delta_{M.larv}$ for the pre-metamorphic larva and δ_M for the post-metamorphic larva and the adult. An acceleration factor s_M allows taking into account changes in the parameters related to the exponential growth period between birth and metamorphosis.

APPENDIX 3.13.

Table S3.13. Parameter estimate of the DEB model and marginal confidence intervals obtained with the profile method. Threshold value Fc: 0.088.

Parameter	Estimate	Marginal confidence interval
$\dot{v}$	0.03301	[0.0183 ; 0.0454]
κ	0.7221	[0.596; 0.805]
$[\dot{p}_M]$	24.42	[18.69; 42.09]
E_H^b	0.004515	[0.0023; 0.0078]
E_H^j	0.3	[0.086 ; 0.665]
E_H^p	2266	[526.55; 8386.15]
z	1.364	[1.11; 1.76]
δ_M	0.612	[0.486; 0.687]

APPENDIX 3.14.

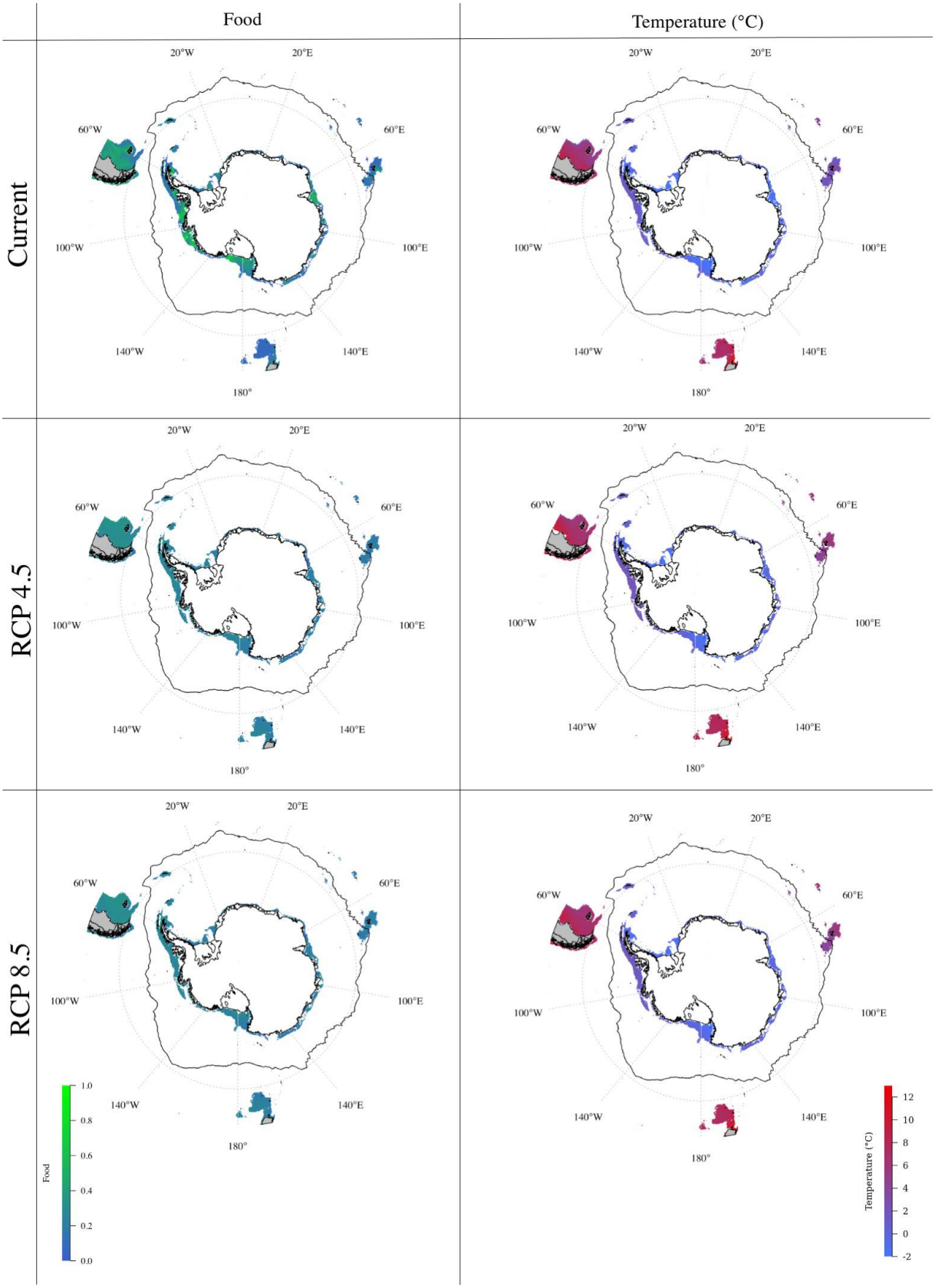


Figure S3.14. Current and future environmental layers (food and temperature) used to project DEB model outputs.

APPENDIX 3.15.

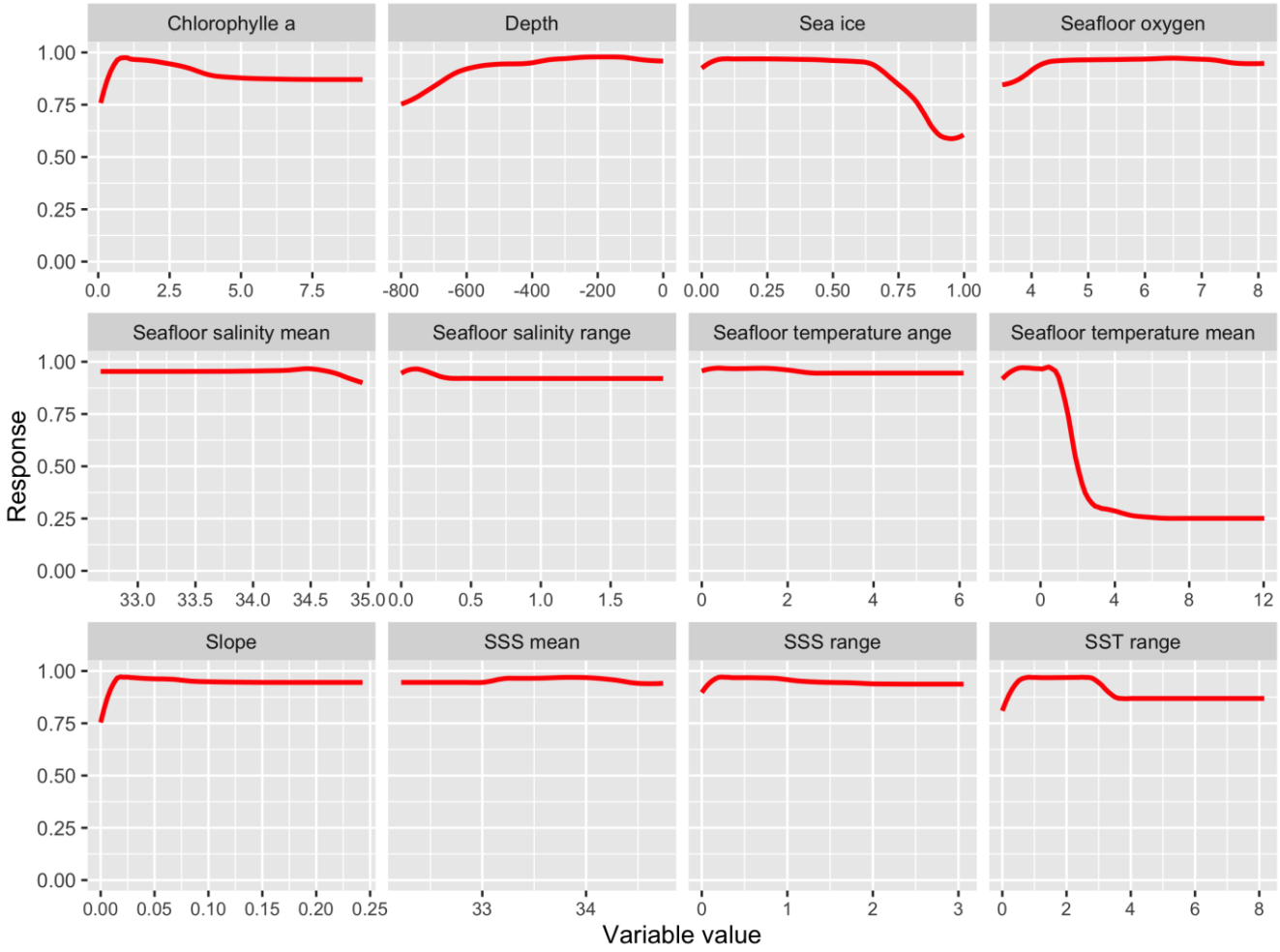
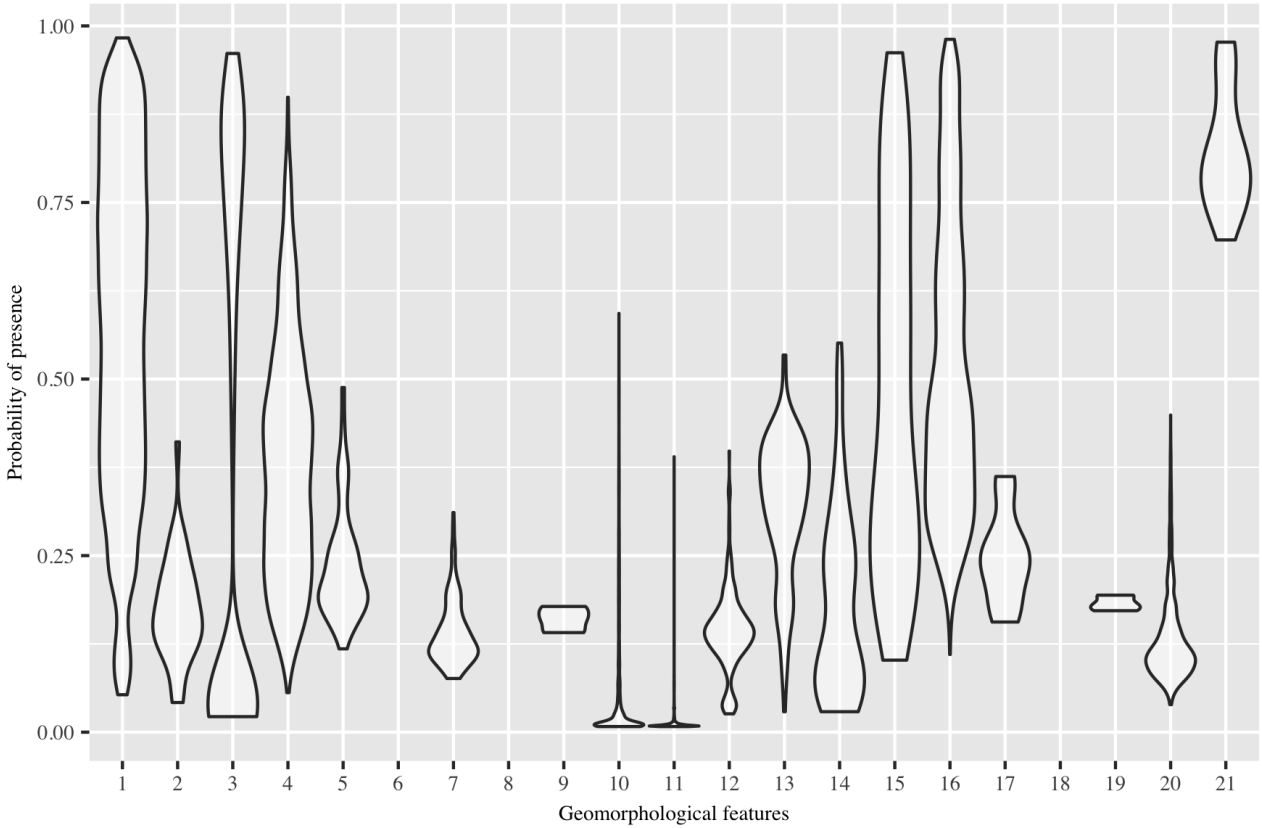


Figure S3.15. Response curve of all predictors used in the correlative niche model approach. The response curves show the relationship between the distribution probability for a species and each environmental variable.

APPENDIX 3.16.



Geomorphological features	Corresponding numeros code
Bank	1
Canyon	2
Coastal (Rugged) Terrane	3
Cross Shelf Valley	4
Island Arc	5
Lower Slope	6
Margin Ridges	7
Marginal Plateau	8
Ocean Trough	9
Plateau	10
Plateau Slope	11
Ridge	12
Rugose Ocean Floor	13
Seamount Ridges	14
Seamount	15
Shelf Deeps-Depressions	16
Structural Slope Region	17
Trench	18
Trough Mouth Fans	19
Upper Slope	20
Volcano	21

Figure S3.16. Presence probabilities for each geomorphological category in the ENMc.

Simple or hybrid? The performance of next generation ecological models to study the response of Southern Ocean species to changing environmental conditions

Guillaumot Charlène[1,2], Belmaker Jonathan[3], Buba Yehezkel[3], Fourcy Damien[4], Dubois Philippe[1], Danis Bruno[1], Saucède Thomas[2]

1 Laboratoire de Biologie Marine, Université Libre de Bruxelles, Avenue F.D.Roosevelt, 50. CP 160/15. 1050 Bruxelles, Belgium
2 UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, EPHE, 6 bd Gabriel F-21000 Dijon, France
3 School of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel.
4 ESE, Ecology and Ecosystem Health, INRAE, 35042, Rennes, France

Diversity and Distributions, submitted.

Abstract

Ecological modelling is widely used in the various fields of ecology but models usually require large datasets, a serious limitation to the approach for application to organisms of remote and little studied regions such as polar seas. Correlative and mechanistic modelling approaches are usually used independently in distinct studies. Using both approaches in integrative, hybrid models however can help better estimate the species realised niche, as mechanistic and correlative models complement each other very well, giving more insights into species potential response to fast changing environmental conditions. In this study, we implemented for the first time an hybrid, correlative and mechanistic model to predict the response of a marine invertebrate endemic to the Southern Ocean, the sea urchin *Abatus cordatus* (Verrill, 1876). We compared the respective performance of simple and hybrid models by analyzing the effect of seasonality on species distribution, a key feature of ecosystem functioning at high latitudes. Higher performances were obtained for the 'integrated Bayesian' approach compared to simple mechanistic and correlative models. The hybrid model more precisely predicts the effect of seasonality on habitat suitability. Such results are promising and show that hybrid approaches can be applied to case studies for which limited datasets are available.

Keywords

Kerguelen Islands, sea urchin species, Species Distribution Modelling, integrated approaches, Bayesian inference

ACKNOWLEDGEMENTS

We acknowledge the use of imagery from the NASA Worldview application (https://worldview.earthdata.nasa.gov), part of the NASA Earth Observing System Data and Information System (EOSDIS). We are thankful to Bas Kooijman for his helpful comments to adapt the DEB model. This work was supported by a “Fonds pour la formation à la Recherche dans l’Industrie et l’Agriculture” (FRIA) and “Bourse Fondation de la Mer” grants to C. Guillaumot. Ph. Dubois is a Research Director of the National Fund for Scientific Research (Belgium). This is contribution no. 50 to the vERSO project and no. 27 to the RECTO project (www.rectoversoprojects.be), funded by the Belgian Science Policy Office (BELSPO, contracts no. BR/132/A1/vERSO and no. BR/154/A1/RECTO). This study is a contribution to program PROTEKER (No. 1044) of the French Polar Institute and LTSER Zone ATelier Antarctique (ZATA, France). Work in the field also benefited from the support of the National Nature Reserve of the French Southern Territories. We are particularly indebted to Sebastien Motreuil, Gilles Marty, and to the crew of the boat Le Commerson for their invaluable help and support in the field.

1. INTRODUCTION

For the last two decades, an ever-growing number of ecological studies have used modelling approaches to highlight the main ecological drivers of species distribution and evaluate the response of species to changing environmental conditions and anthropogenic stressors (Elith et al. 2006, Elith and Leathwick 2009, Franklin 2010b). The overall tendency is to use these models across groups of organisms and regions (Gutt et al. 2012) to inform stakeholders and conservation policies (Thuiller et al. 2013, Mouquet et al. 2015, Singer et al. 2016).
Current developments are focused on the integration of distinct modelling methods (i.e. hybrid modelling) that has long been considered as a way to improve the understanding of ecosystem functioning (Gutt et al. 2012, Dormann et al. 2018, Guillaumot et al. 2018b, Benito Garzón et al. 2019). For instance, combining correlative methods, which rely on spatial relationship between species occurrence records and the environment (e.g. Species Distribution Models, SDMs), with ecophysiological approaches (e.g. mechanistic models) was shown to improve the modelling performance compared to single correlative methods (Elith et al. 2010, Singer et al. 2016, Pertierra et al. 2019, Schouten et al. 2020). Correlative models statistically assess the main drivers of species distribution (Elith et al. 2006, Peterson 2011), and are used to estimate the realised ecological niche (Elith and Leathwick 2009, Soberón 2010). As a consequence, SDMs perform well when species distribution and the environment are in equilibrium, in static systems, a prerequisite that is not verified in highly dynamic ecosystems subject to environmental changes or in studies addressing environmental rapid changes (Loehle and Leblanc 1996, Schouten et al. 2020, Fabri-Ruiz et al. in press - Chapter 3).
Mechanistic models can evaluate the effect of environmental conditions on the physiological performance of individuals or populations (Kearney and Porter 2009). Such models typically require a greater level of biological knowledge, but, in contrast to static, correlative approaches, they explicitly include dynamic processes, offering the opportunity to describe process-based causes of species distribution change (Kearney and Porter 2009, Dormann et al. 2012a), even in non-equilibrium systems (Kearney et al. 2008, Keith et al. 2008). They include a set of mathematical functions relating to species' functional traits (morphology, behaviour, physiology) or associated life history (development, growth, reproduction) and then evaluate the effect of environmental drivers on species physiological traits (Dormann et al. 2012a, Kearney and Porter 2009), which leads to estimating the species' fundamental niche (Kearney and Porter 2009).

Several methods have been developed to integrate correlative and mechanistic models. For instance, mechanistic models can be spatially-projected and used as a input predictor in SDMs (Elith et al. 2010, Buckley et al. 2011, Mathewson et al. 2017, Rodríguez et al. 2019). Other close approaches consist in defining absence records from the mechanistic model and use the set of presence-absence records to implement SDMs (Elith et al. 2010, Feng and Papes 2017) or to fine-tune thresholds for lethal conditions from the mechanistic approach and associate uncertainty estimates to SDM predictions accordingly (Woodin et al. 2013). Bayesian inference methods have also been widely used (Ellison 2004, Brewer et al. 2016, Talluto et al. 2016, Feng et al. 2020, Gamliel et al. 2020), following the development and better accessibility of high-performance computers and programs (Van Dongen 2006). They were proved interesting to optimize the estimation of species habitat suitability (Zurell et al. 2016), to better assess the effect of seasonality in predictions and highlight critical tipping points in changing ecosystems (Oberle et al. 2013, Zhao et al. 2019) providing accurate uncertainty estimates (Zhao et al. 2019). Bayesian methods combine the information of a prior belief (i.e. the prior distribution, for instance our knowledge of species physiology) with new information (i.e. the distribution probabilities) to produce a posterior estimation (Van Dongen 2006). These two steps therefore update the probability of the hypothetical distribution as more evidence or information on species physiology is available (Van Dongen 2006).

Many regions of the Southern Ocean, either in Antarctic or sub-Antarctic zones (Convey et al. 2009, Féral et al. 2019), are currently exposed to fast environmental changes (Cook et al. 2016, Turner et al. 2016, Convey and Peck 2019), including increasing seawater temperatures and shifting seasonality (Bers et al. 2013, Schofield et al. 2017, Henley et al. 2019), glacier melting, changing wind speed (Meredith and King 2005, Cook et al. 2016), which in turn have an impact on

food chains, organic matter production and processes of the bentho-pelagic coupling (see Convey and Peck 2019, Henley et al. 2019 as reviews). Climate changes together with the ever-increasing maritime traffic (i.e. fisheries, tourism and science) boost the introduction of non-native species in Southern Ocean coastal areas, a major threat to polar ecosystems usually characterised by high levels of endemic species (McCarthy et al. 2019, Hughes et al. 2020). These combined issues strongly urge the need to fill the gaps in our knowledge of ecological processes and ecosystem dynamics (Kennicutt et al. 2015).
Due to remoteness and harsh weather conditions, above all in winter, access to the field and data collection in the Southern Ocean are strongly limited (De Broyer et al. 2014), resulting in missing data, spatial and temporal aggregations of observations and difficulties to conduct biological experiments (see Guillaumot et al. in press - Chapter 2 as a review). However, research on marine life of the Southern Ocean has recently benefited from a significant coordinated and international effort with the emergence of oceanographic campaigns and international scientific programs such as the International Polar Year (IPY 2007-2008), the Census of Antarctic Marine Life (CAML 2005-2010) or the Scientific Committee on Antarctic Research, Evolution and Biodiversity in Antarctica (SCAR-EBA 2006-2013) (non-exhaustive list) (Schiaparelli et al. 2013, De Broyer et al. 2014). Several studies have used correlative approaches to characterise the relationship between environmental conditions and the distribution of Southern Ocean species (Pinkerton et al. 2010, Bombosch et al. 2014, Freer 2018, Fabri-Ruiz et al. 2019) or used physiological models to evaluate the influence of environmental conditions on organisms' physiological performances (Agüera et al. 2015, Jager and Ravagnan 2015, Agüera et al. 2017) and population dynamics (Groeneveld et al. 2015, Goedegebuure et al. 2018, Arnould-Pétré et al. 2020 - Chapter 1). However, surprisingly, no study has used integrated modelling approaches despite their considerable potential for analyzing dynamic, complex and ill-known systems.

In this study, we used data from on-going research on a sea urchin species, *Abatus cordatus* (Verrill 1876), in the *Golfe du Morbihan*, the most visited area of the otherwise highly remote archipelago of the Kerguelen Islands (French sub-Antarctic islands). We tested the performance of integrated modelling approaches to deal with limited datasets for a study on a Southern Ocean marine species and compare model outputs with other 'simple' correlative (SDM) and mechanistic (Dynamic Energy Budgets) approaches. In addition, we integrate the effect of seasonality, a fundamental feature of ecosystem functioning in high latitudes and a key to understand the functioning of marine life in the Southern Ocean. Dealing with seasonality was here chosen to test the performance of different modelling procedures in a dynamic context.

2. MATERIAL AND METHODS

2.1. Study species

The heart urchin *Abatus cordatus* (Verrill, 1876) is a shallow deposit-feeder and sediment swallower restrained to soft sediment habitats (De Ridder and Lawrence 1982, Poulin 1996) (Fig. 3.19A,B). Endemic to the Kerguelen Plateau, the species is distributed from shallow subtidal (< 2 m depth) to deep shelf areas exceeding 500 m depth (Poulin 1996). In coastal zones, populations of *A. cordatus* can locally reach densities of up to 280 individuals per square meter (Magniez 1980, Poulin 1996). High population densities along with the species endemicity were interpreted as a consequence of the species reproduction strategy and direct development that includes no dispersal larval stage (Mespoulhé 1992, Poulin and Féral 1995). Females brood their young on the aboral side of the test, inside four brood chambers formed by the sunken paired ambulacra, until juveniles exit the pouch and reach the sea bottom at proximity of their mothers (Fig. 3.19B, Magniez 1983). Depth, temperature and primary production were identified as major environmental drivers of the distribution of *A. cordatus* (Poulin 1996). In shallow-water areas, the species was shown to be tolerant to environmental stressors induced by high variations in salinity, as a result of fresh-water runoffs (Guille and Lasserre 1979), and sudden temperature shifts including heat waves in the austral summer (Motreuil et al. 2018).

2.2. Study area, environmental predictors and seasonality

The study area focusses on the *Golfe du Morbihan*, a 700 km^2 silled basin 50 m deep on average, located in the east of the Kerguelen Islands (sub-Antarctic) (Fig. 3.19C, Fig.3.20A). Since the 1960s, the area has been recurrently studied by marine ecologists who conducted research programs in biological oceanography including studies of micro- and macrobenthic communities (Delille et al. 1996, Poulin 1996).
In addition to depth, sea surface temperature and primary production were used as environmental predictors of the distribution of *A. cordatus*. Seasonality was assessed by focusing on environmental contrasts between the austral summer and the austral winter, here assessed by differences in monthly values between February (summer) and August (winter), the warmest and coldest months (http://www.proteker.net/) associated to the highest and the lowest values of primary production, respectively (Delille et al. 1996).

Bathymetry.
The bathymetric chart was obtained from Beaman and O'Brien (2011), available at http://www.deepreef.org/publications/reports/99-kergdem.html (Fig. 3.20A), with a resolution of 0.001*0.001 arc-degree grid-cell pixels (equivalent to about 100 m). It was updated from Sexton (2005) using new single beam echosounder data from commercial fishing and research voyages, and some new multibeam swath bathymetry data. Satellite-derived datasets were used to provide island topography and to fill in no data areas (see Beaman and O'Brien 2011).

Chlorophyll-a concentration.
As a deposit-feeder, *A. cordatus* feeds upon organic grain coatings and particles present in sediments (Pascal et al. in press). Sea water chlorophyll-a concentration was used as a proxy of food availability, because data on the exact organic content of sediments is not available at the scale of the entire bay (Arnould-Pétré et al. 2020 - Chapter 1). Values were retrieved using imagery from Operational Land Imager (OLI) and Thermal InfraRed Sensor (TIRS) of Landsat 8 obtained from USGS (United States Geological Survey, 2019, https://earthexplorer.usgs.gov/, accessed on May 2020). Chlorophyll-a concentration was derived from OLI data using the Case-2 Regional Coast Colour processor (C2RCC) (Brockmann et al. 2016) for the SentiNel Application Platform (SNAP 2020). Main processing steps are described in Appendix 3.17. Due to the near-permanent cloud cover, only images taken on 2017/02/09 and 2017/08/20 could be retained to depict the contrasting conditions prevailing in the austral summer and winter, respectively; assuming that these two days are each representative of overall seasonal conditions.

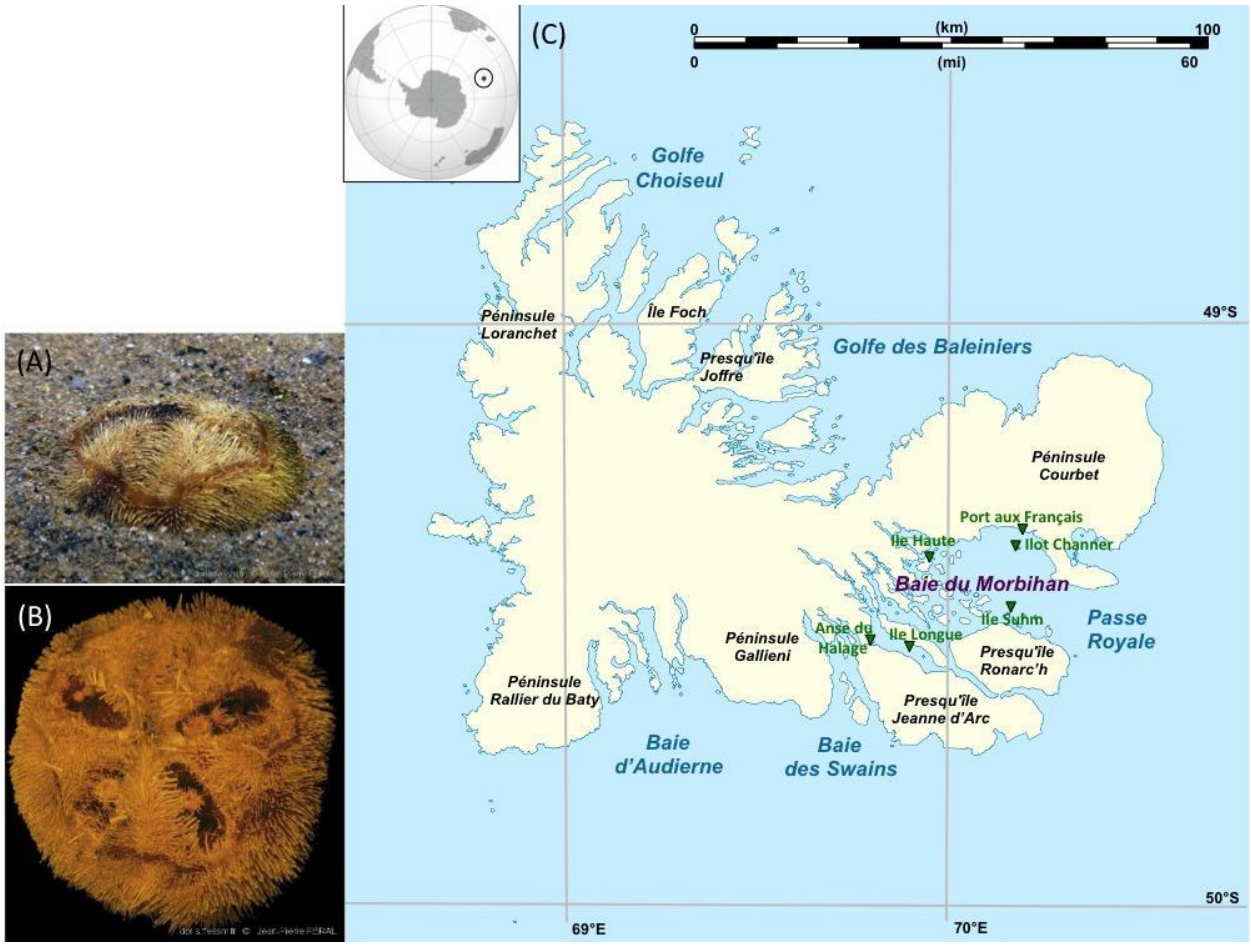


Figure 3.19. (A) Specimen of *Abatus cordatus* half buried into the sand with aboral side emerging from the sediment surface, (B) Aboral view of a female brooding its young in the incubating pouches. Adults can reach a maximum diameter of 4.9 cm (Mespoulhé 1992) © Féral J-P. (C) Location of the *Golfe du Morbihan* in the east of the Kerguelen Islands. Monitoring sites of program PROTEKER (http://www.proteker.net/?lang=en) are indicated in green. Mean daily temperature records measured at these sites were used to cross-validate sea surface temperature values derived from satellite data.

Sea water temperature.

We used satellite-derived sea surface temperature (SST) data from the level 4 Multi-Scale Ultra-High-Resolution Global Foundation Sea Surface Temperature Analysis (MUR, 2015). The MUR SST v4.1 data are based upon night time skin and subskin SST observations from several instruments and are interpolated on a global 0.01 degree grid. Data are produced by the Group for High-Resolution Sea Surface Temperature (GHRSST) and were downloaded from The Physical Oceanography Distributed Active Archive Center (PO.DAAC, https://worldview.earthdata.nasa.gov, accessed May 2020).

SST data were downloaded for 2017/02/09 and 2017/08/20, the two dates retained for chlorophyll-a concentration data. The accuracy of satellite-derived SST data was verified by the close similarity obtained with local *in situ* measurements performed at five distant stations of the bay (program PROTEKER, Appendix 3.17). The spatial resolution of satellite-derived chlorophyll-a and SST data was resampled at 0.001° by a neighbor joining approach to fit with the resolution of the bathymetric chart.

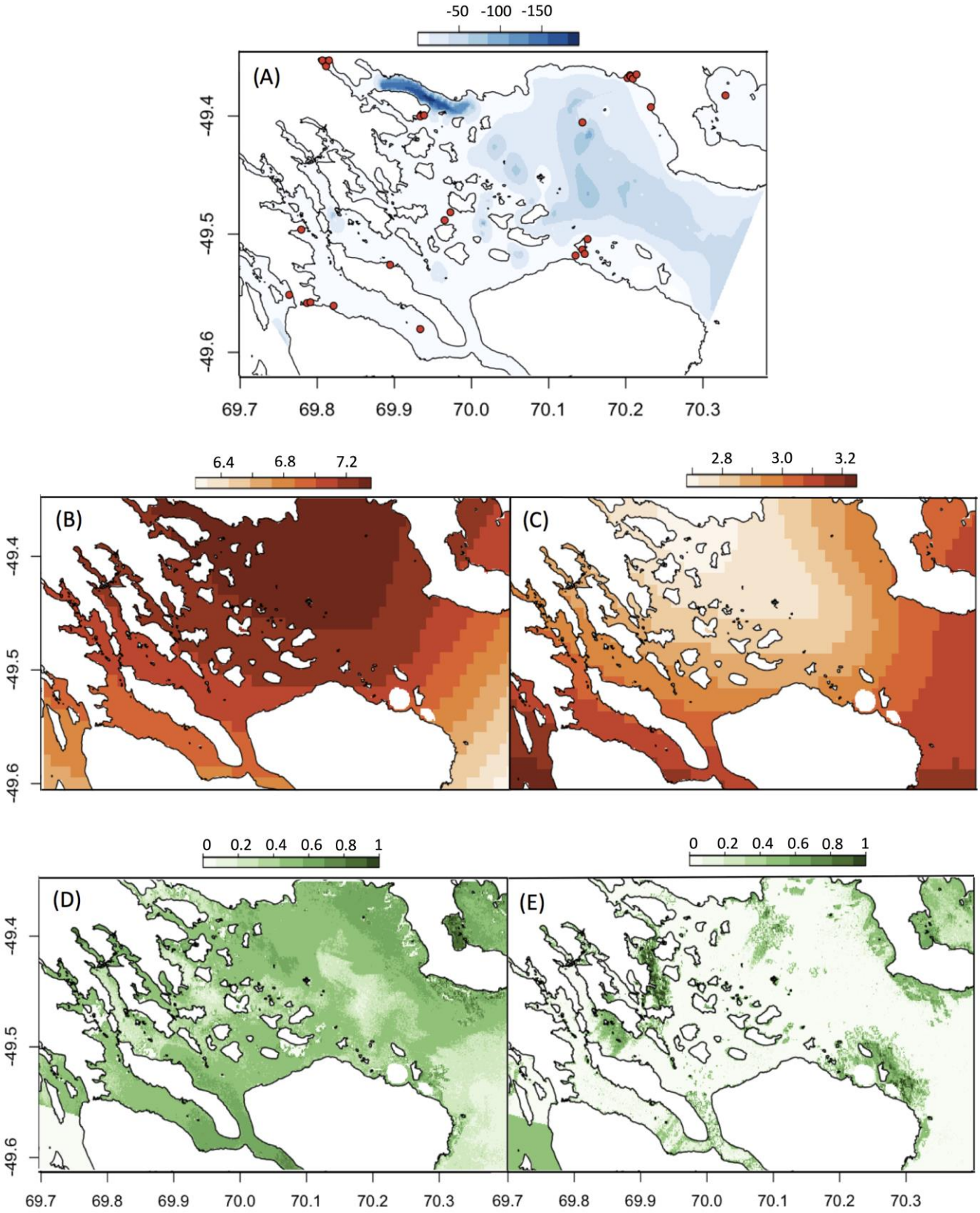


Figure 3.20. (A) Bathymetry (in meters, red dots show presence records of *A. cordatus*). (B) Sea surface temperature in February, on 2017/02/09 and (C) in August, on 2017/08/20. (D) Food availability (scaled between 0 and 1, see section 3) in February, on 2017/02/09 and (E) in August, on 2017/08/20, in the *Golfe du Morbihan*. Water is colder in August (temperatures range between 2.7 and 3.3°C) and food availability much lower than in February, with the richest environments located nearshore.

2.3. Dynamic Energy Budget (DEB) model

The DEB theory defines individuals as dynamic systems and provides a mathematical framework for the life cycle of an organism, from the start of the embryo development to the death. It describes the physiological processes with four primary state variables: reserve, structure, maturity and reproduction buffer (the latter for adults only), directly linked to mass and energy flows and influenced by two forcing environmental variables: temperature and food resources availability (Fig. 3.21, Kooijman 2010). DEB theory relies on key concepts such as first laws of thermodynamics for conservation of mass, energy and time (Jusup et al. 2017) and assumes that the various energetic processes, such as assimilation and maintenance rates are dependent either on surface area or on body volume (van der Meer 2006).

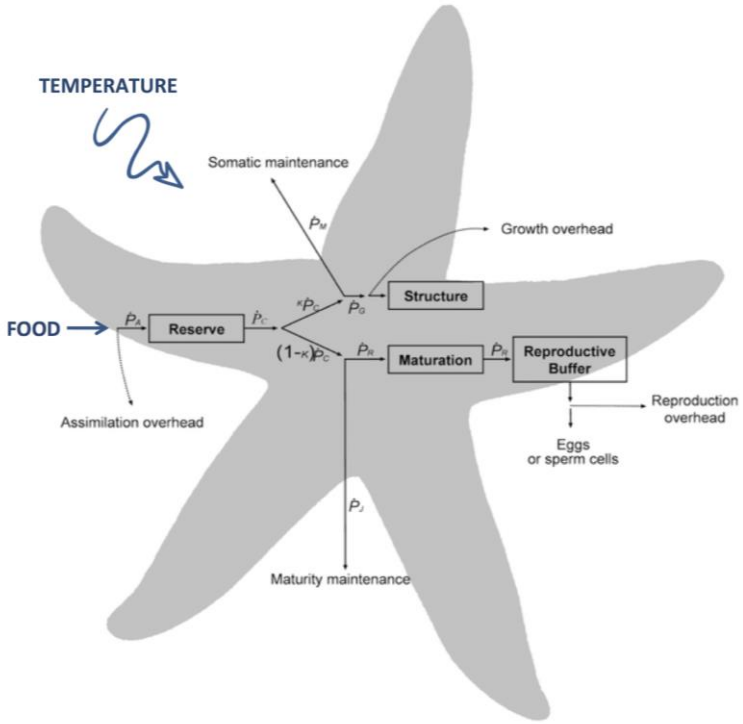


Figure 3.21. Conceptual scheme of the basic parameters and theoretical compartments of the DEB theory. Figure modified from Monaco et al. (2013). Each DEB parameter has a defined link with physiological processes (van der Meer 2006) and the combination of these parameters covers the different energetic processes of the organism (feeding, digestion, storage, maintenance, growth, development, reproduction, ageing) (Marques et al. 2018).

The model was built using zero-variate (single data) and uni-variate (x~y relationship data) datasets extracted from the literature or obtained from experiments set-up purposely for the DEB model. These data were recorded at different life stages of the individual, with food and temperature conditions recorded and informed in the model (list of data available in Guillaumot 2019c, Arnould-Pétré et al. 2020 - Chapter 1).
The model was validated by estimating the goodness of fit using the mean relative error (MRE) which quantifies the overall model performance. MRE values can have values from 0 to infinity, with 0 value meaning that predictions match observation data exactly (Marques et al. 2018). The MRE of *A. cordatus* DEB model is 0.121 (Arnould-Pétré et al. 2020 - Chapter 1).

DEB model forcing by food.
Food resource is included in the DEB model by the scaled functional response *f* parameter, with values between 0 and 1. As previously stated, chlorophyll-a concentration was used to characterise food availability in the area. However, the linkage between chlorophyll-a concentration and food availability for *A. cordatus* is indirect. Chlorophyll-a concentration constitutes a proxy of

primary production and food resources for the zooplankton. Then, the organic matter produced in the upper water layers will sink in the water column and reaches the seafloor to be consumed by the benthos following a process that can take several days (Turner 2002). Sediments also contain organic matter remineralized by bacteria (Nixon 1981, Becquevort et al. 2001, Jacquet et al. 2008), which constitutes a background food storage for the sea urchin (Pascal et al. in press).
In order to estimate *f*, the minimal food availability necessary in sediments for the survival of *A. cordatus* individuals, growth rates were simulated for different *f* values (Appendix 3.18). Results showed that below *f* = 0.45, the sea urchin will never reach the reproduction size (1.9 cm) within its expected average life span (8-9 year-old) (Guillaumot 2019c). We therefore hypothesized that *f* = 0.45 is a minimum threshold value for populations of *A. cordatus* to occur and reproduce (Appendix 3.18). According to Delille and Bouvy (1989), Port-aux-Français is the station in the bay area where a minimum level of food is available in August (austral winter) that is, 1.35 $mg.m^{-3}$ according to satellite-derived data. It was therefore assumed that *f* = 0.45 is a minimum in areas where chlorophyll-a concentrations reach at least 1.35 $mg.m^{-3}$. It was also assumed that *f* < 0.45 in areas where chlorophyll-a concentrations are lower than 1.35 $mg.m^{-3}$. Highest population densities of *A. cordatus* and food quantity were recorded at Anse du Halage station in the summer (Delille and Bouvy 1989, Poulin 1996), which corresponded to chlorophyll-a concentrations of 10.9 $mg.m^{-3}$ on 2017/02/09. This value was used to define the maximum value of *f* = 1 and to scale *f* values comprised between *f* = 0.45 (1.35 $mg.m^{-3}$) and *f* = 1 (10.9 $mg.m^{-3}$). Final *f* maps showed important spatial and seasonal contrasts in food availability in the bay area (Fig.3.20D,E).

DEB model forcing by temperature.
The DEB model generated by Guillaumot (2019c) for *A. cordatus* was complemented with experimental data obtained in the Kerguelen Islands (Motreuil et al. 2018), which provided individual respiration rates of sea urchins at 5, 6, 7 and 9°C. Along with data from the literature, these results were used to define the maximum temperature range for survival *A. cordatus*, comprised between -1°C and + 12°C, with an optimal metabolic performance observed at +6°C and a performance decrease above +8°C. These results were used to determine five Arrhenius parameters in the calculation of a temperature correction factor, integrated into the model to take into consideration the influence of temperature on metabolic rates.
Considering *Tau*, a given metabolic rate, the following equation was applied:
$Tau(T=T_i) = Tau(T) * T_C$, where T_i is the environmental temperature and T_C the temperature correction factor, with:

$$Tc = \exp\left\{\frac{T_A}{T_1} - \frac{T_A}{T}\right\} . \left(1 + \exp\left\{\frac{T_{AL}}{T_1} - \frac{T_{AL}}{T_L}\right\} + \exp\left\{\frac{T_{AH}}{T_H} - \frac{T_{AH}}{T_1}\right\}\right) . \left(1 + \exp\left\{\frac{T_{AL}}{T} - \frac{T_{AL}}{T_L}\right\} + \exp\left\{\frac{T_{AH}}{H} - \frac{T_{AH}}{T}\right\}\right)^{-1}$$

with T_{ref} the reference temperature (4°C), T_L and T_H the lower and higher boundaries of the optimal tolerance range and T_{AL} and T_{AH} the Arrhenius temperatures beyond the lower and higher temperatures, respectively (Thomas and Bacher 2018).

Spatial projection of the DEB model.
Outputs of the DEB model were projected over the entire bay area by estimating the species physiological performance for each pixel of the map, using pixel-specific values of food availability and temperature (Thomas and Bacher 2018, Fabri-Ruiz et al. in press - Chapter 3). Reproduction and survival capacities were estimated by comparing somatic maintenance $\dot{pM}$ and maturation maintenance $\dot{pJ}$ costs over the total energy available from the reserve compartment $\dot{pC}$ (Fig. 3.21). According to DEB theory, the somatic maintenance $\dot{pM}$ has priority over growth and reproduction to ensure survival. Maturity maintenance $\dot{pJ}$ has priority over reproduction (Kooijman 2010). These conditions imply that if the energy available in the reserve compartment $\dot{pC}$ is not sufficient to pay for the required maintenance costs ($\dot{pC} < \dot{pM} + \dot{pJ}$), the organism cannot reproduce, and will progressively starve and die.

2.4. 'Simple' Species Distribution Modelling (SDM)

A set of 26 presence-only records of *A. cordatus* sampled from 1898 to 2015 in the *Golfe du Morbihan* was compiled from Guillaumot et al. (2016), checked for georeferencing errors and complemented with data from Poulin and Féral (1995) (Fig. 3.20A). Data are homogeneously distributed in the area with a Moran's I score of -0.01 (p-value= 0.15). Consequently, background records were randomly sampled in the area without any targeted sampling approach as the effect of spatial autocorrelation was not significant (Phillips et al. 2009, Guillaumot et al. 2018a - Appendix). In order to sample environmental conditions prevailing in the study area as precisely as possible, while being close to the number of presence-only records available, 200 background records were sampled across the entire projection area (Barbet-Massin et al. 2012).
Generalized linear models (GLMs) (here referred as 'simple SDM') were used to relate species occurrences with the three environmental predictors previously described (depth, food availability, sea surface temperature and their square forms, Fig. 3.20). In this approach, presence and background data are treated as Bernoulli trials, where p is the probability of finding *A. cordatus*. A non-informative normal distribution (μ=0, sigma= 10,000) was used as a prior for the regression coefficients. The model was run using a burn-in period of 4,000 samples, followed by 4,000 additional MCMC samples to estimate the posterior distribution of regression coefficients. The procedure was replicated for 50 background records sampling, and average species distribution probabilities were predicted on a map. Posterior parameters were saved and used afterwards to initiate the 'integrated Bayesian' approach (section 2.6).

Model extrapolation areas were defined using the Multivariate Environmental Similarity Surface index (MESS, Elith et al. 2010). Extrapolation areas correspond to all grid-cell pixels where descriptor values are not contained within the range of environmental conditions for which presence-only data are recorded. Extrapolation is defined for negative values of MESS, and the environmental predictor responsible for extrapolation was evaluated (for further details see Elith et al. 2010, and Guillaumot et al. 2020c - Chapter 2).

2.5. Integrated 'SDM-DEB' model

Integrating correlative and mechanistic models was first tested by using the spatial projection of the DEB model (section 2.3) as an environmental predictor in the SDM (Elith et al. 2010, Buckley et al. 2011, Mathewson et al. 2017, Rodríguez et al. 2019). The procedure is similar to the 'simple' SDM model approach (section 2.4), except that the DEB layer (i.e. '$\dot{pC}$ > ($\dot{pM}$+$\dot{pJ}$)?', Fig. 3.22) was added to the initial set of environmental predictors (depth, temperature, food availability). Similarly, the procedure was replicated for 50 background records sampling, and average distribution probabilities were predicted on a map.

2.6. 'Integrated Bayesian' model

The method developed by Talluto et al. (2016), and applied by Gamliel et al. (2020) was used to develop an 'integrated Bayesian model' (physiology-SDM model). For this purpose, the 'simple SDM' (section 2.4) was combined with some physiological information obtained by a physiological submodel (detailed below). This combination was performed with a Bayesian approach by using the posterior distributions of the physiological submodel as priors for the SDM to create 'integrated Bayesian model' coefficients (see also the detailed method in Talluto et al. (2016) supplementary material).

Using DEB equations and parameters (Eq. 1), average growth rates were calculated for individuals measuring from 2.5 to 4.5 cm, according to food availability (for all values available in the projection area, Fig. 3.20) and a random selection of temperatures within the range of values of the considered season. This constitutes the 'physiological submodel' that therefore takes into account both food availability and temperature. Twenty-five replicates of individual sizes and temperature selection were performed. The growth rate was calculated with the following DEB equation (Kooijman 2010): $\dot{pG}$= (*kap* * $\dot{pC}$ - $\dot{pM}$) / $\dot{kM}$ / T_C (Eq.1)

with *kap* being the fraction of energy directed towards complexity (-), $\dot{pC}$ the mobilisation flux (energy.time^{-1}), $\dot{pM}$ the somatic maintenance rate (energy.time^{-1}), $\dot{kM}$ the somatic maintenance rate coefficient (time^{-1}) and T_C the temperature correction factor (-).

A Bayesian *beta* regression, with food availability as a predictor and growth performance probability as a response, was applied to estimate how physiology changes with food conditions. A total of 4,000 MCMC samples were used for burn-in and the posterior distribution was estimated using 4,000 additional samples. The physiological submodel coefficients were initiated with Gaussian priors, with the mean taken from the maximum likelihood estimation to improve convergence and a vague prior set on the variance (set at 1,000).

Posterior priors of this physiological submodel were then used as priors to represent food availability *f* and its square form f^2 in the 'integrated Bayesian model'. As for the other priors (intercept, depth, temperature and temperature^2), they were attributed the posterior priors of the 'simple SDM', with their variance arbitrarily fixed at 100, as we considered them as vague priors (Gamliel et al. 2020). The detail of prior values is given in the result section (Table 3.4).
Similarly, 50 model replicates (i.e. background samplings) were generated, averaged and plotted for comparisons.

2.7. Model performance

Model predictions for all approaches were evaluated by measuring the Area Under the Curve (AUC) (Fielding and Bell 1997, Allouche et al. 2006, Elith et al. 2006) using the R package ROCR (Sing et al. 2005). In complement, the percentage of correctly classified presence data was measured by extracting prediction values over the position of each presence data and compared to the MaxSSS threshold (Maximum Sensitivity plus Specificity threshold), highlighted to be the best threshold to characterise predicted suitable (>MaxSSS value) and unsuitable areas (<MaxSSS value) for presence-only models (Liu et al. 2013). Standard deviations of model replicates were used as uncertainty maps (Buisson et al. 2010, Swanson et al. 2013).

Partial dependence plots were used to represent the relationship between model predictions and environmental values and compared between models. They are built by plotting model prediction values of each grid-cell pixel (y axis) against the value of the environment at the same pixel (x axis; each partial dependence plot is specific of a single environmental layer).

R codes developed for this study are available at https://github.com/charleneguillaumot/THESIS.

3. RESULTS

3.1. Spatial projection of the DEB model

Spatial projections of DEB model outputs show important contrasts between the two seasons (Fig. 3.22). In February, when temperatures are higher than 6°C and food availability homogeneously higher than 0.5 over the entire bay area (Fig. 3.20), high species survival and reproduction are predicted almost everywhere (Fig. 3.22A), except in some areas where food availability is very low (Fig. 3.20). Nearly four times more energy is predicted to be contained in the reserve compartment of *A. cordatus* in February compared to August (Appendix 3.19), an energy available for individuals' maintenance and development.
In contrast, in August, the DEB model predicts maintenance costs of up to three times higher than in February while the energetic load available is lower (Appendix 3.19), leading to reduced reproduction and survival abilities in the majority of the study area. Individual survival is modelled to be higher closer to the shoreline due to higher food availability (Fig. 3.22B).

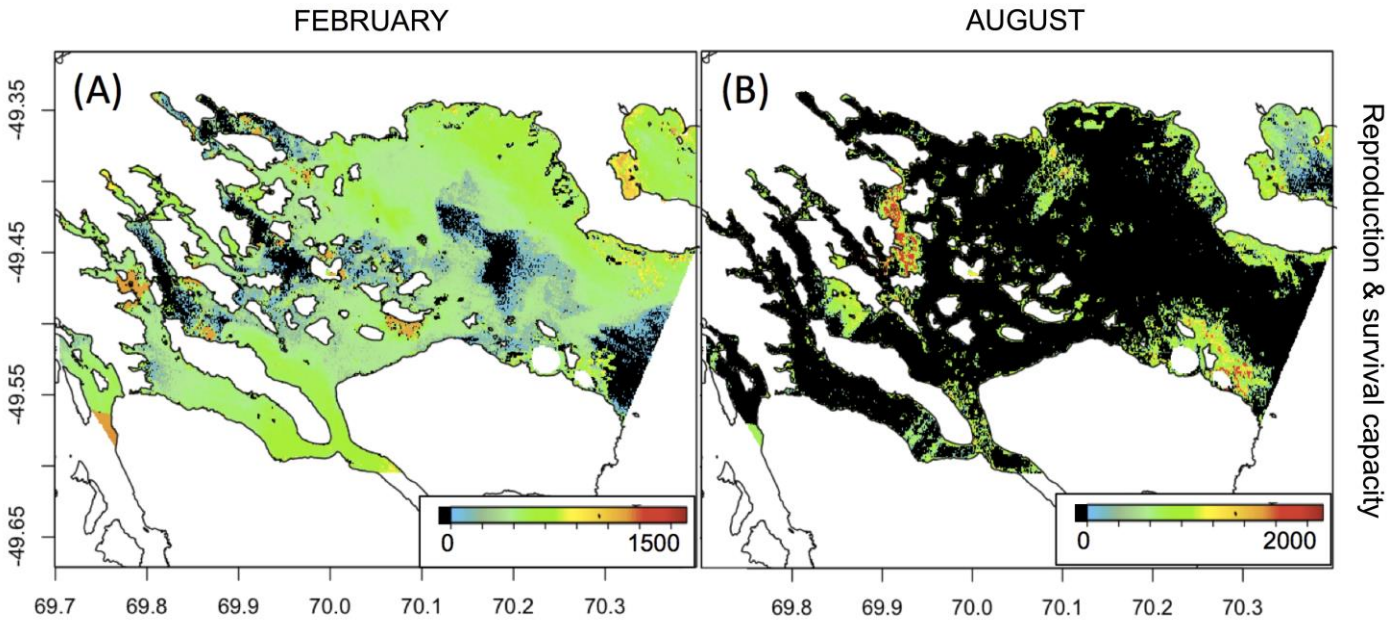


Figure 3.22. Spatial projections of the DEB model in February (A) and August (B). Reproduction and survival capacity is given by $\dot{pC}$ - ($\dot{pM}$+$\dot{pJ}$) (energy.time^{-1}), with 'Reproduction and survival possible' for values >0 (color bar) ; 'Not possible' for values <0 (black).

3.2. Simple Species Distribution Models (SDMs)

Overall distribution probabilities predicted by 'simple SDMs' are low (<0.5, Fig. 3.23A,B) for the entire area and both seasons, and standard deviations are comparatively high (homogeneously close to 0.45 for February and more contrasted in space but coastal areas reaching 0.45 too for August, Fig. 3.23C,D), stressing an important variability between model replicates.
Average predictions are more contrasting in August than in February (Fig. 3.23A,B). For August, the model predicts the highest distribution probabilities (around 0.5) near the shoreline, in shallow-water areas, and the lowest probabilities (around 0.2) in the center of the bay and in a northwestern fjord characterised by deep waters (Fig. 3.23B). In February, distribution probabilities are homogeneous in all the area and close to 0.4 (Fig. 3.23A).
Areas where model extrapolation occurs correspond from 36 (in February) to 37.8% (in August) of the total surface of the projection area and is mainly to be related to depth and to temperature in large patches for February (black and grey patches, Fig. 3.23E,F).

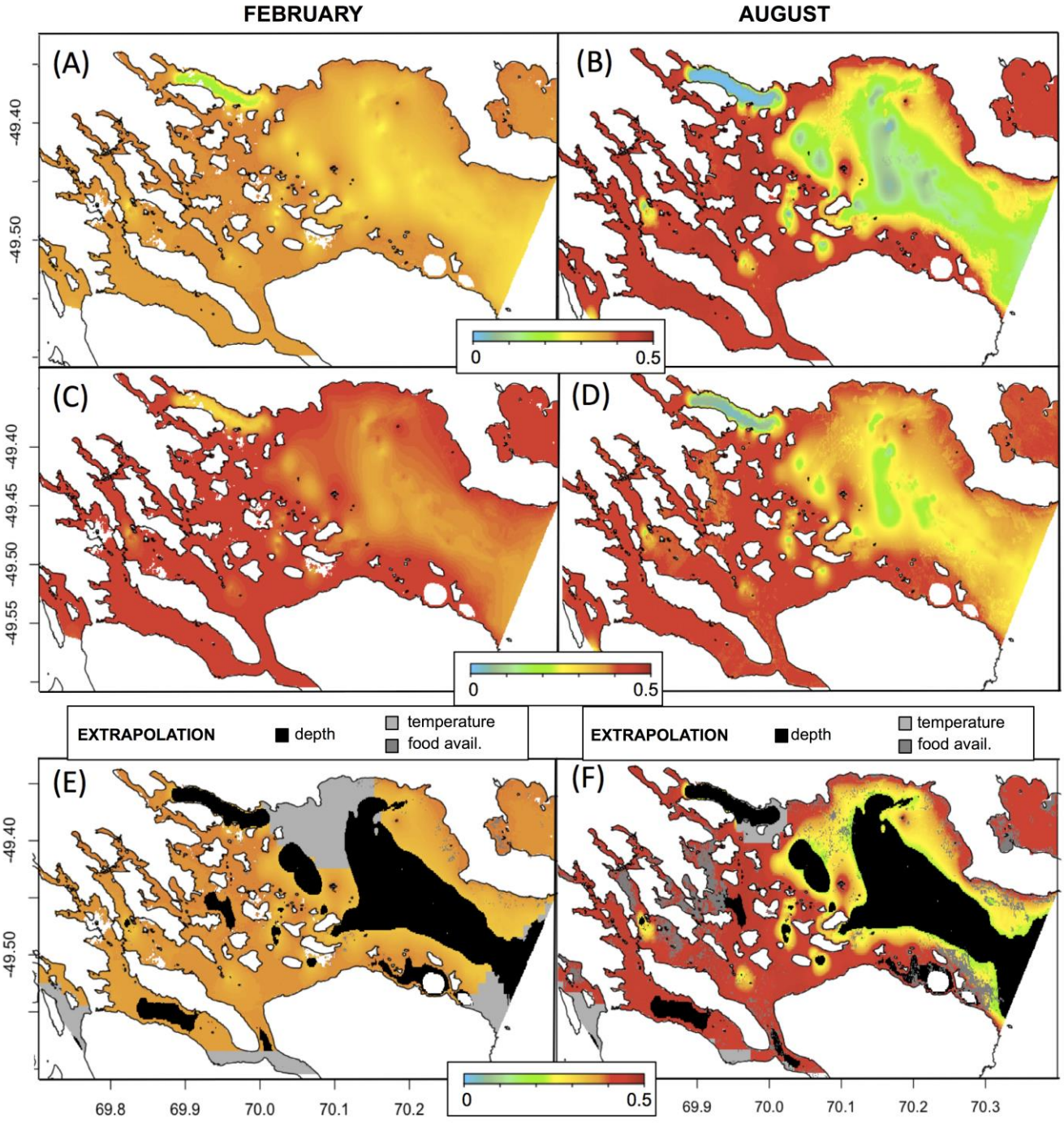


Figure 3.23. Spatial projections of the 'simple SDM' for February (A,C,E) and August (B,D,F), average of 50 model replicates. Average distribution probabilities (A,B), standard deviations (C,D) and average distribution probabilities with extrapolation areas associated to for each environmental descriptor (depth, temperature or food availability) (E,F).

3.3. 'Integrated SDM-DEB' model

Model predictions are highly contrasting between February and August according to the 'integrated SDM-DEB' model (Fig. 3.24A,B). In February, distribution probabilities are close to 0.55 over the entire area, except for some patches located in the center of the bay and in coastal zones with predictions of up to 0.85. In contrast, low prediction scores are evenly predicted over the entire area for August (0.33 maximum, Fig. 3.24B). Standard deviations are higher in August than in February in coastal areas (0.4 vs. 0.3 for August and February, respectively) and reach the same range of values (around 0.3) in the deep central area of the *Golfe du Morbihan* (Fig. 3.24C,D).

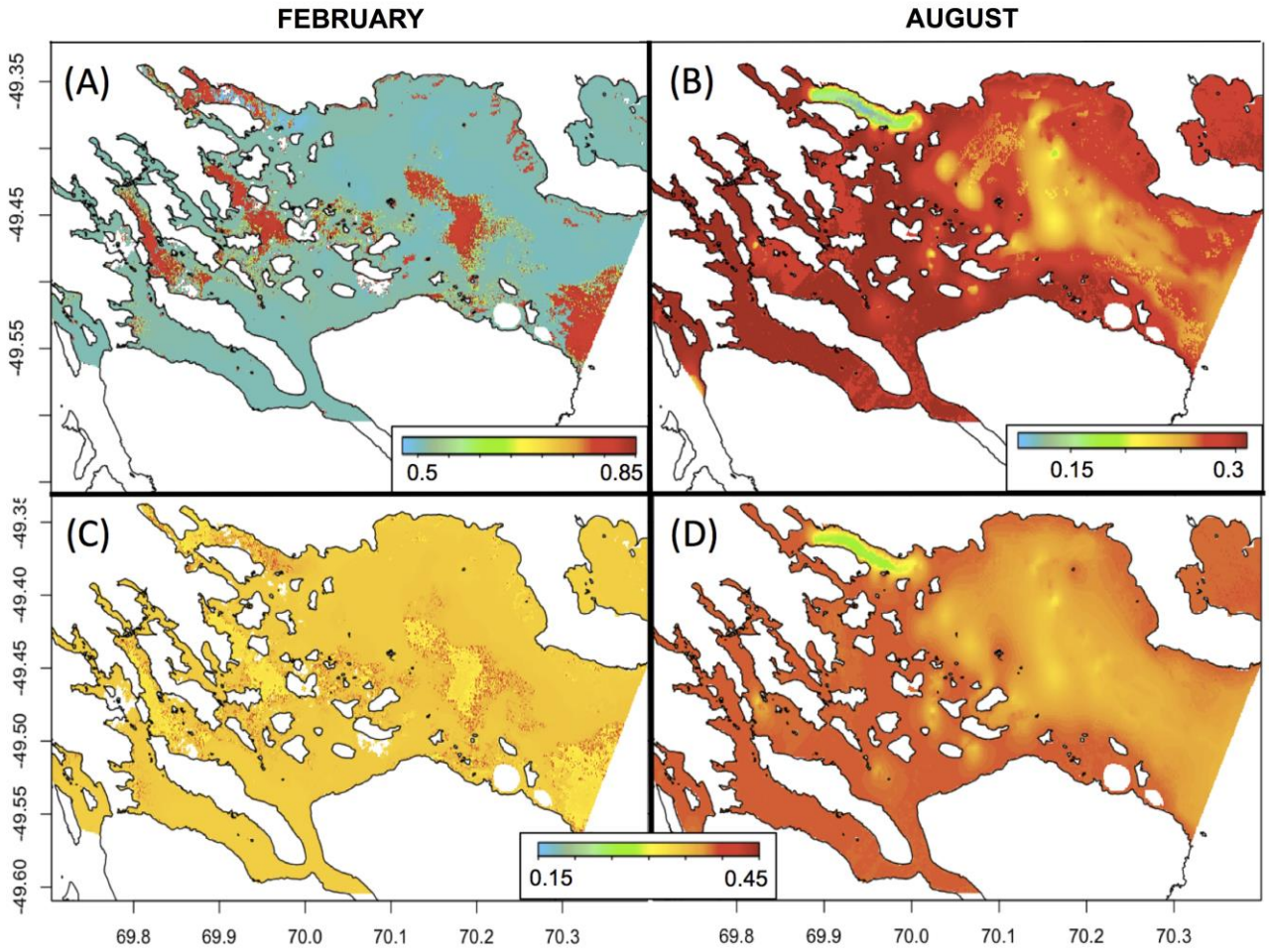


Figure 3.24. Spatial projections of the 'integrated SDM-DEB' models for February (A,C) and August (B,D), averaged of 50 model replicates. Average distribution (A,B) and associated standard deviations (C,D). The available energy after paying off the somatic and maturity maintenances is integrated in the model as a predictor that assesses for each pixel the value of $\dot{pC}$ - ($\dot{pM}$ - $\dot{pJ}$), with $\dot{pC}$ the amount of energy contained in the reserve compartment, $\dot{pM}$ the amount of energy required for somatic maintenance and $\dot{pJ}$ the amount of energy required for maturity maintenance.

3.4. 'Integrated Bayesian' model

'Integrated Bayesian' models were implemented using the following set of parameters as priors (Table 3.4). The coefficient values of *f* and f^2 are high compared to the other parameters (average and tau scores), increasing the influence of food availability in final model outputs (Table 3.4). In August, the coefficient value of the *f* parameter is eight times higher than in February (8.43 compared to -0.89) but f^2 is twice lower (11.38 compared to 27.78) (Table 3.4).

Table 3.4. Matrices of priors used to calibrate 'integrated Bayesian' models for February and August, with the equation y=b0 + b1*depth + b2*f + b3*temperature + b4*temperature^2 + b5*f^2. Tau is the inverse of the variance (1/Standard deviation^2), arbitrarily set at 0.01 (i.e. variance = 100) when the parameter is considered to be a vague prior.

Parameter		Source	Mean	St. deviation	Tau
b0	intercept	simple SDM	-18.37 (Feb.) 0.32 (Aug.)	11.00 (Feb.) 3.98 (Aug.)	0.01 (Feb.) 0.01 (Aug.)
b1	depth	simple SDM	0.18 (Feb.) 0.16 (Aug.)	0.01 (Feb.) 0.01 (Aug.)	0.01 (Feb.) 0.01 (Aug.)
b2	*f*	Physiological submodel	-0.89 (Feb.) 8.43 (Aug.)	0.09 (Feb.) 0.06 (Aug.)	125.16 (Feb.) 260.22 (Aug.)
b3	temperature	simple SDM	1.03 (Feb.) -0.19 (Aug.)	2.99 (Feb.) 2.60 (Aug.)	0.01 (Feb.) 0.01 (Aug.)
b4	temperature^2	simple SDM	0.19 (Feb.) -0.09 (Aug.)	0.22 (Feb.) 0.47 (Aug.)	0.01 (Feb.) 0.01 (Aug.)
b5	f^2	Physiological submodel	27.78 (Feb.) 11.38 (Aug.)	0.15 (Feb.) 0.20 (Aug.)	41.64 (Feb.) 25.88 (Aug.)

In 'integrated Bayesian' models, distribution probabilities vary within a large range, between 0.1 and 1, a sharp difference with low probability values (< 0.5) obtained with the 'simple SDM' approach (Fig. 3.23, Fig. 3.25A,B). The 'integrated Bayesian' approach also predicts differences between the two seasons, but not as important as the 'integrated SDM-DEB' model results (Fig. 3.24). Overall, the study area is predicted as less suitable to *A. cordatus* in August than in February, when food availability and temperatures are higher (Fig. 3.20B,E). More precisely, in August, suitable areas are mainly restricted to shallow waters and nearshore zones, especially in the west. In February, habitat suitability is more extended but remains mainly located close to the coasts (Fig. 3.25A). Standard deviation scores (Fig. 3.25C,D) are within the range of values obtained for the two other models (0.2-0.4) and values are similar between the two seasons, although high values (around 0.45) cover a broader area in August. Compared to February, some patchy areas nearby coasts present low values in August (Fig. 3.25D).

In February, most of the areas for which standard deviation are the highest for the 'integrated Bayesian' model (Fig. 3.25C) correspond to the extrapolation areas of the 'simple SDM' maps (Fig. 3.23C). This is less clear for the August scenario (Fig. 3.25D, Fig. 3.23F).

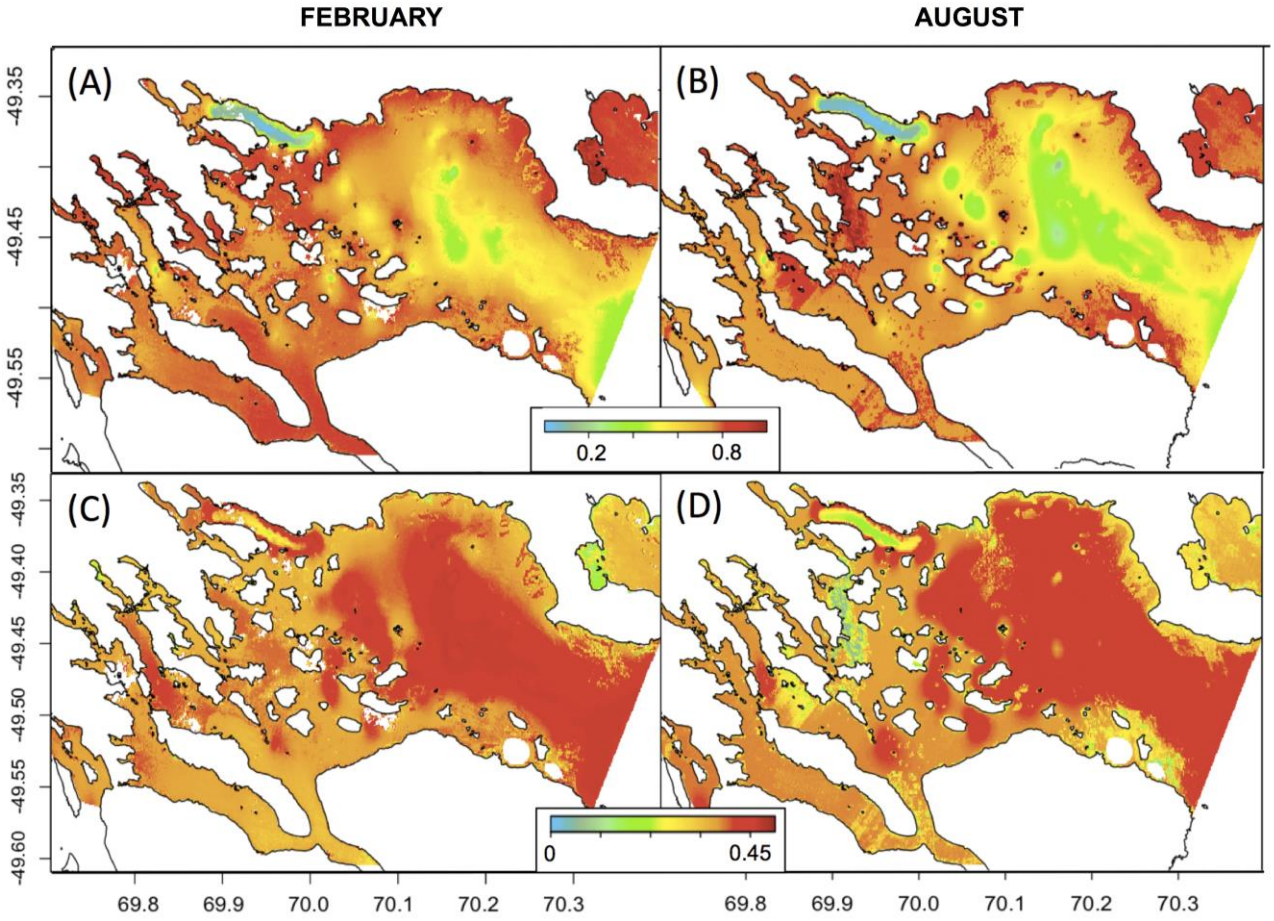


Figure 3.25. Spatial projections of the 'integrated Bayesian' models for February (A,C) and August (B,D), averaged of 50 model replicates. Average distributions (A,B) and associated standard deviations (C,D).

3.5. Contribution of predictors and model performance

'Model performance (Table 3.5) is good for all approaches except for the 'spatial DEB' approach, for which the percentage of correctly predicted presence data is very low in August (38.5%). Among the three other approaches, model performance is very similar between the two seasons in the 'integrated Bayesian' approach. AUC scores are significantly the highest (t test with p-values < 0.001), with values reaching a minimal score of 0.76 in August with the lowest variability. The percentage of correctly classified presence data are good (> 81.7 %) for February, significantly higher than in the two other approaches (compare to 77.8 and 67.3%), but a bit lower for August (88.8% compared to 94.8 and 94.4% for the 'simple SDM' and integrated 'SDM-DEB' approaches, respectively).

Table 3.5. Comparison of model performances (percentage of presence data correctly classified and Area Under the Curve, AUC, metric) for the two seasons. Average and standard deviation of 50 model replicates.

	Spatial DEB	Simple SDM	Integrated SDM-DEB	Integrated Bayesian
% presence data correctly classified	96.15% (Feb.) 38.5% (Aug.)	77.8 ± 12.8 (Feb.) 94.8 ± 1.9 (Aug.)	67.3 ± 18.1 (Feb.) 94.4 ± 6.1 (Aug.)	81.7 ± 12.1 (Feb.) 88.8 ± 7.1 (Aug.)
AUC		0.71 ± 0.03 (Feb.) 0.72 ± 0.03 (Aug.)	0.60 ± 0.12 (Feb.) 0.75 ± 0.04 (Aug.)	0.80 ± 0.02 (Feb.) 0.76 ± 0.02 (Aug.)

Partial dependence plots (Fig. 3.26) were generated to evaluate the influence of each environmental predictor (depth, food availability, and temperature) on model predictions. Overall, comparison between models show that integrated modelling approaches ('integrated SDM-DEB and 'integrated Bayesian) provide more contrasting response curves for all three predictors compared to the 'simple SDM' approach, both for February and August (Fig. 3.26).
The 'integrated Bayesian' model results (Fig. 3.26) suggest a more substantial influence of environmental values on predicted probabilities, with higher temperatures, higher food availability and lower depths associated with higher predicted habitat suitability. This highlights a more important sensitivity of the seasonal effect on model predictions. These results are more in line with the ecology of *A. cordatus* and are confirmed by the higher performance metrics observed for the 'integrated Bayesian' approach, noteworthy in February (Table 3.5).

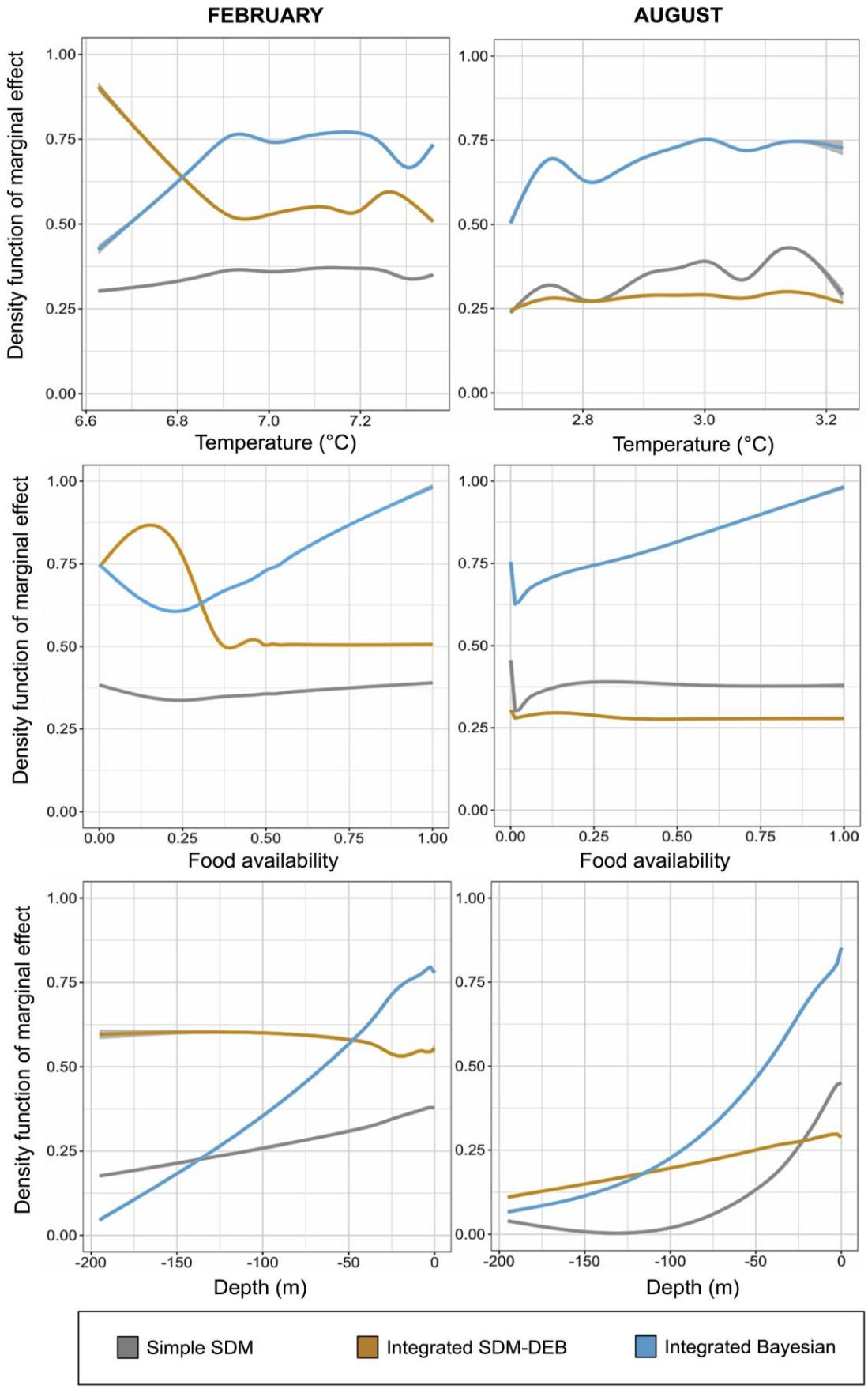


Figure 3.26. Partial dependence plots, representing model predictions (y axis, probabilities between 0 and 1) aligned with the environmental values (x axis). Grey solid line: simple SDM, yellow solid line: integrated SDM-DEB model; blue solid line: integrated Bayesian model. Average prediction values of 50 model replicates.

4. DISCUSSION

4.1. Potential and main limitations to the different modelling approaches

Correlative approaches ('simple SDMs') are aimed at describing the correlation between species occurrence records and environmental conditions. SDM outputs can provide knowledge on the main environmental factors that drive species distribution (Elith et al. 2006, Peterson 2011). Because presence records are used as input data, SDMs also indirectly integrate the influence of other factors such as the effect of biotic interactions (either competition, exclusion or facilitation between species) and the biogeographic context (barriers or dispersal vectors) on species distribution, thereby simply and explicitly assessing the species realised niche (Soberón 2010). However, the relevance of niche estimation often constitutes the main limitation to 'simple SDMs', because their predictive performance strongly relies on sampling completeness (Loehle and Leblanc 1996, Vaughan and Ormerod 2003, Araújo et al. 2005, Randin et al. 2006, Broennimann et al. 2007, Holt 2009). The heterogeneity of presence sampling induce statistical artefacts that can bias model predictions (Bahn and McGill 2007, Currie 2007), a substantial limitation that has already been stressed in former works on the Southern Ocean (Guillaumot et al. 2018a - Appendix, Guillaumot et al. 2020b - Chapter 2, Guillaumot et al. in press - Chapter 2).
Compared to SDMs, mechanistic models require more data (and require a good knowledge of species ecology or physiology) for parameter estimation and model implementation (Kearney and Porter 2009). However, if the model can be built, the approach is powerful to evaluate the survival capacity of individuals in given environmental conditions (Arnould-Pétré et al. 2020 - Chapter 1, Fabri-Ruiz et al. in press - Chapter 3) and can estimate the species fundamental niche (Kearney and Porter 2009).
Combining the merits of both correlative and mechanistic approaches to fine-tune the estimation of the species realised niche can provide important benefits (Dormann et al. 2012a), as prior information on the influence of the environment on species metabolism, given by physiological models, can be used to improve correlative models (Feng et al. 2020). This combined approach is also valuable to assess the effect of fast changing environmental conditions (e.g. seasonality or future predictions), which generate non-equilibrium states (Kearney et al. 2008, Keith et al. 2008) that cannot be accurately modelled by static, correlative approaches (Loehle and Leblanc 1996, Schouten et al. 2020, Fabri-Ruiz et al. in press - Chapter 3).

In the present study, the comparison of 'simple SDM', the most commonly used approach in ecological studies of Southern Ocean species, with 'integrated' approaches, was performed. All approaches have good performance statistics (Table 3.5), except for the 'spatial DEB' model. Spatial projections of the 'spatial DEB' approach are strongly driven by food availability (strong similarities between Fig. 3.20D,E and Fig. 3.22), and provide a biased representation of species distribution for August (Table 3.5), as "low food" areas are simply and systematically predicted as unsuitable to the species survival, with no consideration for the influence of the other environmental drivers. However, the model is interesting because it stresses the link between energetic costs and one major environmental driver (Appendix 3.19), a good complement to physiological submodels, and interesting to assess the environmental conditions that drive species distribution.

'The simple SDM' is characterised by good validation scores (AUC > 0.71 and percentage of correctly classified presence data > 77.8%) (Table 3.5) but distribution probabilities are contrasting for August compared to February (Fig. 3.23), when food concentration is high and evenly distributed in the all bay area (Fig. 3.20D). As a consequence, the contribution of this variable to model predictions is low (Fig. 3.26), an unrealistic prediction that contrasts with results obtained with the integrated approaches ('integrated SDM-DEB' and 'integrated Bayesian') (Fig. 3.26).

Using a physiological submodel to inform a SDM has been applied in recent works by directly adding a physiological layer to the SDM (Elith et al. 2010, Buckley et al. 2011, Mathewson et al. 2017, Rodríguez et al. 2019) or by generating absence data from the modelled physiological information (Elith et al. 2010, Feng and Papes 2017). Model outputs are easy to interpret but the approach requires the combination of several models, as in any hybrid approach, and implies a

risk inherent in the addition of biased estimations of each individual models (Feng and Papes 2017). In the present work, predictions of the 'integrated SDM-DEB' model are similar to results obtained with the 'spatial DEB' projections. This was expected, especially for the August model with corresponding low-food concentration conditions (Fig. 3.20), consequent low survival capacities (Fig. 3.22) leading to predicted low species distribution probabilities for the entire area (Fig. 3.24). The DEB layer contributes to the model as do environmental predictors (Elith et al. 2010) resulting in some inconsistencies, as shown by the lower model performances obtained for February (percentage of correctly classified presence data= 67.3% and AUC= 0.60) (Table 3.5), whereas predictions are the highest in areas where survival and reproduction are impossible (i.e. where reserve $\dot{pC}$ is lower than the energy required for overall maintenance ($\dot{pM}$ - $\dot{pJ}$)) (Fig. 3.24A). This statistical artefact is due to the spatial correlation between the occurrence of a high number of presence records in areas where $\dot{pC}$ - ($\dot{pM}$+$\dot{pJ}$) values are low (i.e. energy available into the reserves $\dot{pC}$ is barely sufficient to pay for maintenance costs). This is shown in Figure 3.26, where the highest distribution probabilities are associated to low food values. The integration of the 'spatial DEB' layer into the 'integrated SDM-DEB' model led to over-estimating the influence of food availability on the prediction of species occurrences.
Another noticeable drawback of the 'integrated SDM-DEB' method is that important variations are obtained between model outputs depending on the DEB layer that is added into the SDM (Mathewson et al. 2017) (Fig. 3.24, Appendix 3.20A). The choice of the DEB layer to be used also influences model extrapolation (Appendix 3.20B) (Rodríguez et al. 2019), which must be taken into consideration when interpreting model results (Elith et al. 2010, Buckley et al. 2011), and increases the complexity of model calibration. Therefore, real benefits of adding modelled physiological information to SDMs are case dependent and the improvement of modelling performances are not certain (Buckley et al. 2011, Rodríguez et al. 2019). However, the method can prove helpful for future predictions and analyses of non-equilibrium states, which constitute the main limitation to the SDM approach (Elith et al. 2010, Buckley et al. 2011, Martínez et al. 2015, Mathewson et al. 2017). When there is few data available and the causal relationship between organism physiology and environment drivers difficult to model in a robust way, using the 'integrated SDM-DEB' approach can be problematic and model outputs must be interpreted with caution.

Bayesian methods are increasingly used in marine sciences (Colloca et al. 2009, Muñoz et al. 2013, Pennino et al. 2014, Roos et al. 2015, Gamliel et al. 2020). They were proved to have several advantages compared to other methods, including (1) a more accurate and realistic estimation of uncertainty as observations and model parameters are both used as random variables in model predictions (Robert 2007) and (2) the possibility to integrate information from different sources, scales or nature (Peters et al. 2004, Hobbs and Ogle 2011, Hartig et al. 2012).
In the present work, the highest AUC scores and correctly classified presence data were obtained with the 'integrated Bayesian' approach. Models performed well in representing uncertain areas, compared to other approaches (Fig. 3.23,3.24), as the areas predicted with the highest standard deviation scores by the 'integrated Bayesian' approach (Fig. 3.25) strongly overlap with the extrapolation areas estimated for the 'simple SDMs' (Fig. 3.23). The influence of environmental variations on model predictions are more marked (Fig. 3.26), with a better fit of the species response to environmental variations, and prediction performances show less contrast in evaluation scores between February and August (Table 3.5). This suggests that the 'integrative Bayesian' approach is the best among the three tested approaches, at estimating the realised niche of *A. cordatus*.

4.2. Seasonality and predicted distribution of Abatus cordatus

In all model predictions, distribution probabilities are the highest in coastal areas, where populations of *A. cordatus* were known to be the most abundant (Poulin and Féral 1995, Poulin 1996). Interestingly, with some rare exceptions, sediment granulometry and hydrodynamics were shown to be important drivers of population densities in *A. cordatus* (Poulin and Féral 1995). These two key factors were not included in our models but suitable areas to the species perfectly match with conditions of high food availability and high temperature that prevail in coastal areas.

Important contrasts, however, were obtained in model predictions between February and August, suggesting that seasonal variations significantly affect the metabolism of *A. cordatus* as organisms face different conditions in terms of food availability and temperature. According to the physiological model ('spatial DEB', Fig. 3.22, Appendix 3.19), maintenance costs are higher in winter (August) than in summer (February) due to lower temperatures that increase the demand of energy to maintain the metabolism (Kooijman 2010). Besides, there is less energy available in the reserve compartment to compensate for the increased maintenance costs as food availability is low in winter too (Appendix 3.19).
These results are strongly dependant on the assumption that metabolism performance (and therefore requested energy) follows Arrhenius laws as determined with summer acclimated individuals (Motreuil et al. 2018). For some Antarctic sea urchins, such as *Sterechinus neumayeri* (Meissner, 1900) it was reported a sharp metabolic switch during winter conditions. During this hypothesised non-feeding period, metabolic rates are decreasing with lower recorded oxygen consumption and slow or absent somatic growth (Brockington et al. 2001, Brockington and Peck 2001, Brockington et al. 2007). Such seasonal metabolic changes has never been observed nor studied for *A. cordatus*, but, if existing, it could bias the estimation of the Arrhenius curve implemented in the model and change some of the metabolic estimations.

Model outputs are in line with our knowledge of the reproduction cycle of *A. cordatus* and its timing. In most places of the *Golfe du Morbihan*, individuals invest energy in the growth of gonads in March, when food is the most abundant (Magniez 1983). Once fertilized, the eggs are brooded in the female incubating chambers for almost nine months (a period of low-food availability and low temperature) before the young are released and settle on the seabed (Schatt and Féral 1996) or live sheltered between holdfasts of the giant kelp *Macrocystis pyrifera* (Poulin 1996). The reproduction cycle of *A. cordatus* is constant across years for a given place (Magniez 1983). However, it was observed that the reproduction period can shift from a few months between sites (Schatt and Féral 1991, Mespoulhé 1992, Poulin 1996), which was explained by spatial and temporal variations in food availability and sediment enrichment in nutrients (Schatt and Féral 1991).

Modelling with such details the influence of environmental variations on the species metabolic performance and distribution bring valuable insights to interpret model predictions and assess the species realised niche. Integrating the effect of seasonal variations in niche modelling, herein assessed as differences between February and August, has long been suggested in SDMs (Elith and Leathwick 2009, Franklin 2010a) but it is seldom achieved due to limited data availability (Guillaumot et al. 2018b). Conversely, ignoring the effect of seasonality in ecological niche estimation has been recently shown to reduce prediction performance (Smeraldo et al. 2018). Seasonality is a fundamental feature of environmental systems. It is particularly critical to life in temperate and high latitudes, and one key phenomenon to consider for studying both species distribution (Morelle and Lejeune 2015, Zuckerberg et al. 2016) and metabolism (Bahlburg et al. 2021).

4.3. Study improvements

To generate accurate models, this study focused on a well-documented echinoid species, *A. cordatus*, which had long been studied in the favorable context of a long-term observing system of coastal marine life, in the *Golfe du Morbihan*, the most visited area of the highly remote archipelago of the Kerguelen Islands. However, some limitations were highlighted by our results. (1) The first limitation is the absence of a precise evaluation of food availability for *A. cordatus* in the total area of the *Golfe du Morbihan*. Estimates of chlorophyll-a concentration were used as a proxy of food abundance and availability but this constitutes a strong assumption that can impact model outputs. Chlorophyll-a concentration in sea surface waters is a partial surrogate to the measurement of food availability for a benthic species like *A. cordatus* as the abundance of nutrients on the sea bottom depends on the processes of organic matter consumption, degradation and transfer from the water column to the sea bottom (Laurenceau-Cornec et al. 2015). Food Availability Models could be developed (Jansen et al. 2018) to estimate the proportion of organic matter that reach the seafloor based on the knowledge of water currents. It could be also

interesting to have some information about benthic detritic organic matter that the sea urchins could consume (Pascal et al. in press). These data were however not available for the study area but such models offer promising perspectives. (2) Detailed information on the link between temperatures and physiological performances are still missing, as we only have and use here the results of a survival experiment performed at different temperatures in 2018 (Motreuil et al. 2018). DEB modelling has the potential to include five Arrhenius parameters to precisely characterise the link between temperature and metabolism (Kooijman 2010, Thomas and Bacher 2018), but available experimental data on *A. cordatus* do not permit measuring them with precision. More data are still needed for our case study to reach this precision and improve the performance of the DEB model. (3) Finally, there is a lack of presence data to correctly calibrate the model and to validate it. Generating ecological models with small datasets was indeed shown to reduce modelling performances (Stockwell and Peterson 2002, Liu et al. 2019) as it truncates predicted distribution and niche definition (Hortal et al. 2008, El-Gabbas and Dormann 2018), and may lead to a reduction in model accuracy because presence and background datasets would not differ markedly (Luoto et al. 2005) and constrains the evaluation process (Pearson et al. 2007) (reviewed in Guillaumot et al. in press - Chapter 2). Therefore, common validation approaches such as the cross-validation method (that uses a part of the dataset to train the model and another part to test it independently, Hijmans 2012, Guillaumot et al. 2019 - Chapter 2) could not have been used for our study, which limited the power of our model evaluation.

4.4. Conclusions

Our results suggest good performances of 'integrated Bayesian' approaches to estimate species realised niche, compared to single correlative approaches or 'integrated SDM-DEB' approaches that might be biased by the subjective choice of the DEB layer used as an input into the SDM. More data are still necessary to better evaluate the model, to more accurately establish the relationship between the environmental conditions and the species physiology and to better represent the whole environment, but this study showed the possibility to apply the method for a data-poor case study, which opens perspectives for future applications to a broad panel of natural examples.

APPENDIX 3.17. Satellite-derived measurements

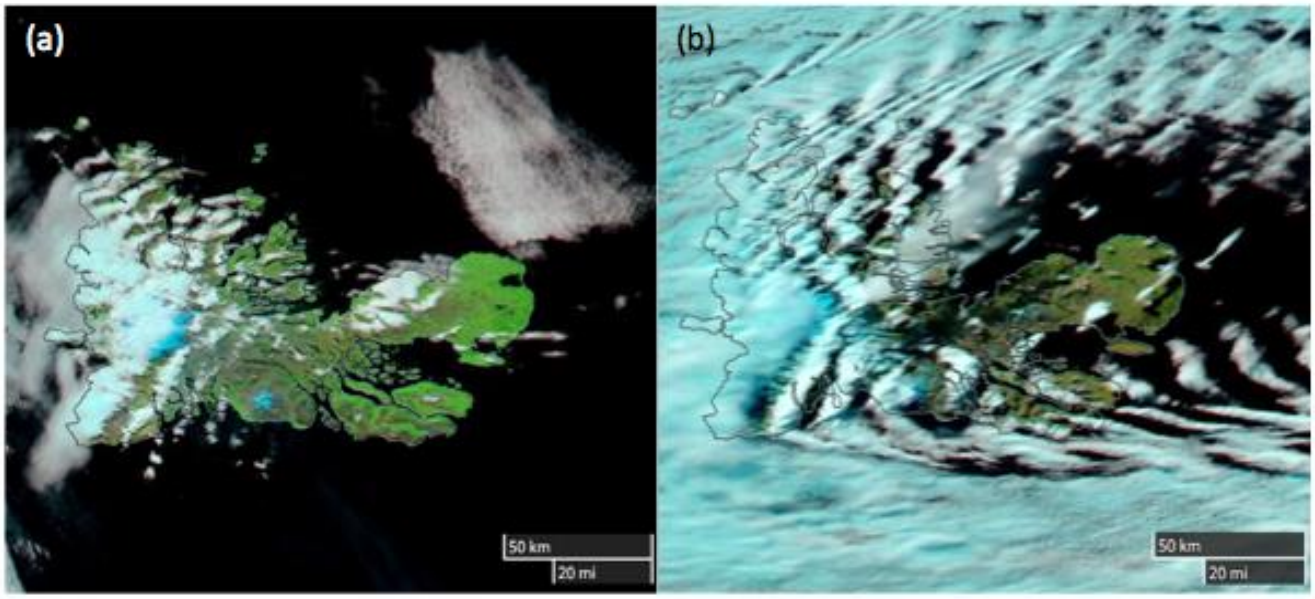


Figure S3.17 Overview of images captured by the Landsat 8 satellite for the selected dates (a) 2017/02/09 and (b) 2017/08/20. Some clouds are present in August over the *Golfe du Morbihan* that should be considered for interpretation. Source: https://worldview.earthdata.nasa.gov, accessed May 2020.

Table S3.17.A. Details of SNAP parameterization for chlorophyll-a measurement, processing parameters.

Processing parameters	Value
Salinity	35 PSU
Temperature	4°C
Ozone	330 DU
Air Pressure at sea level	1000hPa
Elevation	0 m
TSM factor bpart	1.72
TSM factor bwit	3.1
CHL exponent	1.04
CHL factor	21
Threshold rtosa OOS	0.05
Threshold AC reflectances OOS	0.1
Threshold for cloud flag on transmittance down @865	0.955 (default)
Atmospheric aux data path	-
Alternative NN Path	-
Set of neuronal nets	C2RCC-Nets

Table S3.17.B. Comparison between daily *in situ* temperatures (°C) recorded by the PROTEKER program at defined stations within the *Golfe du Morbihan* (Fig. 3.19) and satellite-derived sea surface temperatures from the MUR dataset. Temperatures at *Port-aux-Français* are measured at a 2-minutes frequency by a tide gauge installed on the dock, at shallow depth. It is highly influenced by warm air temperatures.

Station (Latitude ; Longitude)	**February 9th, 2017**		**August 20th, 2017**	
	In situ T° (°C)	MUR (°C)	In situ T° (°C)	MUR (°C)
Ile Haute (-49.3875 ; 69.9415)	6.87	7.31	3.03	2.71
Ile Longue (-49.5387 ; 69.8838)	7.77	6.98	3.13	3.07
Port aux Français (-49.352 ; 70.221)	9.37	7.31	3.21	2.95
Ile Suhm (-49.493 ; 70.1613)	7.34	7.24	3.04	2.87
Ilot Channer (-49.3826 ; 70.1857)	8.36	7.33	3.1	2.91

APPENDIX 3.18. Model forcing by food availability

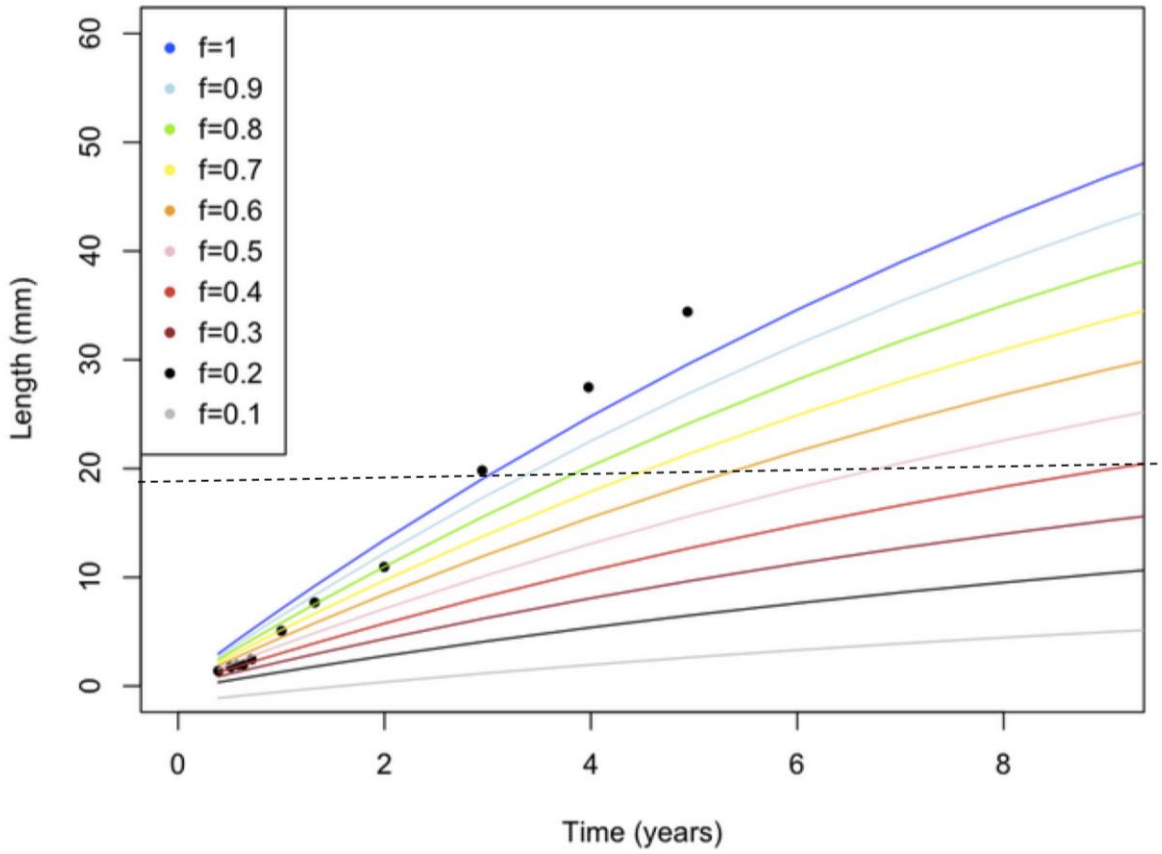


Figure S3.18. Simulated growth rates by the DEB model for different levels of food availability (*f* values). Results show that below *f* = 0.45, the sea urchin never reaches the size at which it should reproduce (1.9 cm, dotted horizontal line) within its average life expectancy (9 year-old). We therefore hypothesized that a background value f = 0.45 should be at least available in the sediment for population survival.

APPENDIX 3.19. Spatial DEB

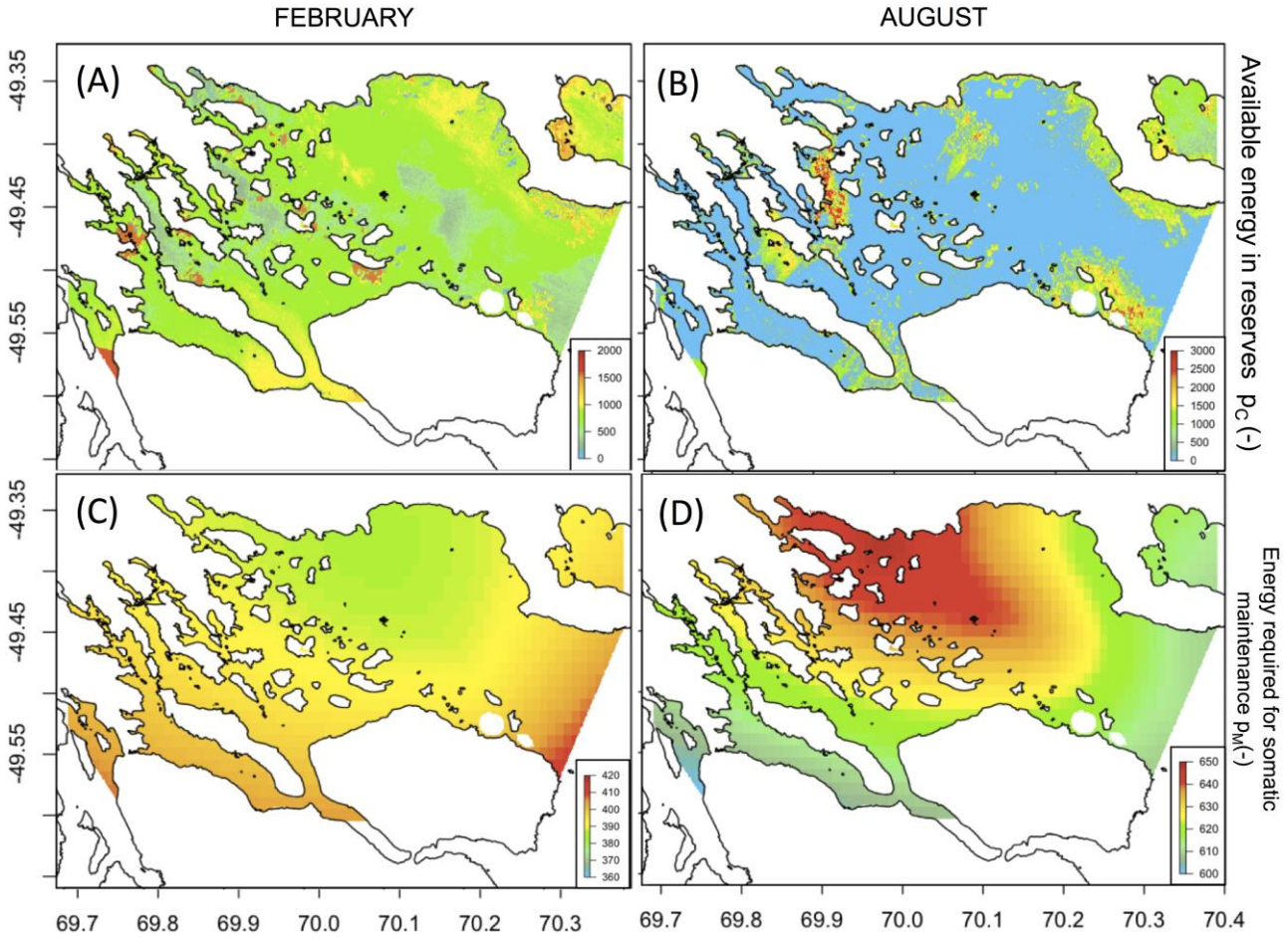


Figure S3.19. Results of the spatial projection of the DEB model for February (A,C) and August (B,D). Evaluation of the available energy in the reserve compartment $\dot{p_C}$ (A,B). Energy required for somatic maintenance $\dot{p_M}$ (C,D).

For February, the model predicts around four times more energy contained in the reserve compartment compared to August (Fig. S3.19A,B). Probably due to decreased temperatures in August, the energy required for somatic maintenance is up to three-fold higher for August compared to February (Fig S3.19D,C).

APPENDIX 3.20. Integrated SDM-DEB models

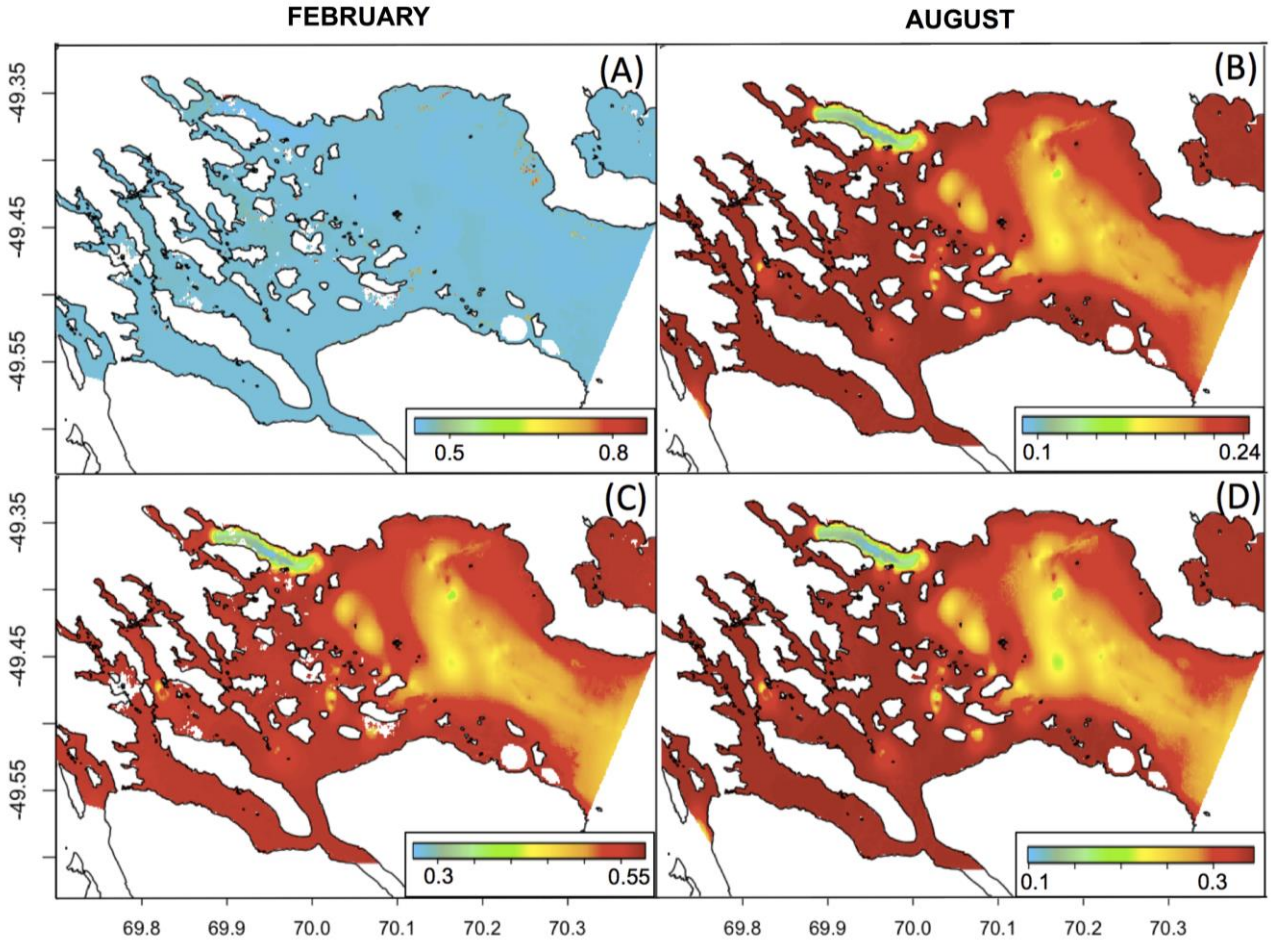


Figure S3.20.A. Distribution probabilities predicted for integrated SDM-DEB models. Average predictions of 50 model replicates for February (A,C) and August (B,D). (A,B) DEB layer being the amount of energy contained in the reserve compartment $\dot{p_C}$; (C,D) DEB layer being the amount of energy required for somatic maintenance $\dot{p_M}$.

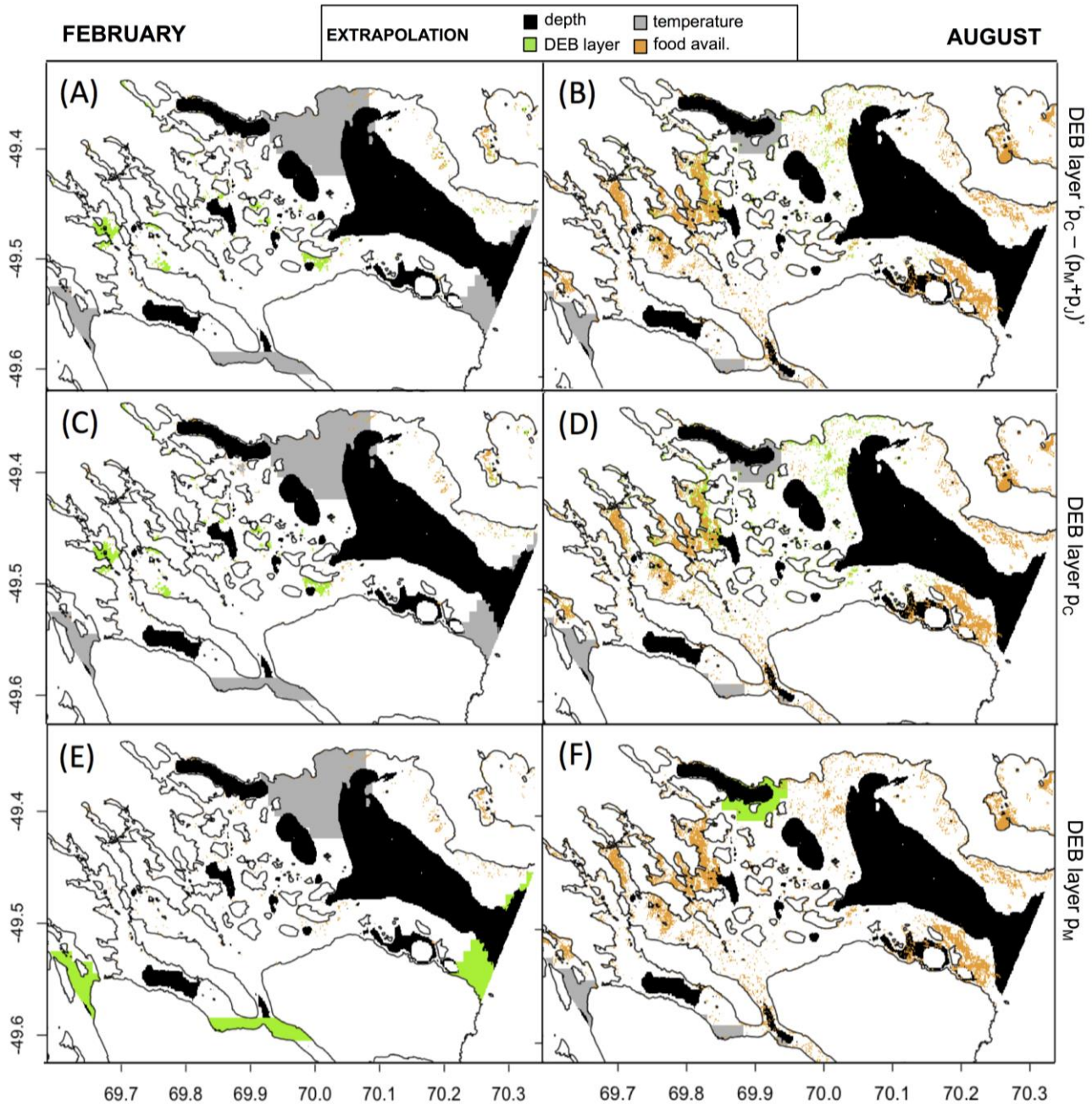


Figure S3.20.B Extrapolation areas (MESS) associated with the descriptor responsible for the extrapolation (depth, temperature, food availability or DEB layer), for the 'integrated SDM-DEB' approach. Application with three different DEB layers: '$\dot{p_C}$ - ($\dot{p_M}$+$\dot{p_J}$)', $\dot{p_C}$ or $\dot{p_M}$.

CHAPTER 4

DISPERSAL MODELS: LAGRANGIAN APPROACH

Chapter 4 finally focusses on another driver of the species ecological niche and BAM diagram: dispersal capacities. Among modelling approaches that exist to model the movement of propagules in water masses are lagrangian models. This approach was exemplified in this chapter.

•The first study (presented in the appendix section) used a lagrangian model to evaluate the capacity of the Patagonian crab *Halicarcinus planatus* (Fabricius, 1775) to reach Antarctic coasts. The model was developed at the scale of the Southern Ocean. Propagules were launched from South America southern tip and from some sub-Antarctic Islands, were the crab was observed. The model simulated particle drift in 3D and according to several climatic scenarios (negative or positive Southern Annular Mode index). Results highlighted an eastward drift and the impossibility of propagules to reach Antarctic coasts in any scenario: they are blocked by the Antarctic Circumpolar Current.

•The second study focussed on the Western Antarctic Peninsula region. From the previous study, it was hypothesized that the female individual of *Halicarcinus planatus* found in Deception Island in 2010, should have been brought by ballast waters released closeby Antarctic coasts. The lagrangian model was used to simulate the passive drift of virtual propagules departing from ballast waters released at contrasting distances from the nearest coasts: 200 (international guidelines), 50 or 11 nautical miles. Results showed that releasing ballast waters at 200 nautical miles considerably reduces the arrival of propagules in proposed marine protected areas. Simulations suggested that the crab could have reached Deception Island only if the international guidelines have been violated, with ballast water exchanged at 50 nautical miles or less from the coasts.

•The last study used a lagrangian approach, combined with SDM simulations and genetic analyses to assess the spatial connectivity of the overexploited marbled rockcod *Notothenia rossii* Richardson, 1844. The study was conducted at the scale of the Southern Ocean. Results highlighted a lack of genetic differentiation between Southern Ocean populations, influenced by an important connectivity between sub-Antarctic Islands, induced by the stepping-stone transport and promoted by the strong water flow of the Antarctic Circumpolar Current.

•[*Appendix section*] López-Farrán Z, Frugone MJ, Gerard K, Vargas-Chacoff L, Poulin E, **Guillaumot C** and Dulière V (in preparation). Can the Patagonian crab *Halicarcinus planatus* reach the Antarctic Peninsula ? Study of the dispersal potential of its larvae using a lagrangian approach.

•Dulière V / **Guillaumot C** (co-first authorship), López-Farrán Z, Lacroix G, Saucède T, Danis B and Baetens K (submitted). Potential impact of ballast water exchanges on the introduction of invasive species in Marine Protected Areas of the Western Antarctic Peninsula. *Diversity and Distributions.*

•Christiansen H, Van de Putte A, **Guillaumot C**, Barrera-Oro E, Volckaert FAM and Young EF (in preparation). Integrated assessment reveals large scale connectivity of a historically overexploited fish in the Southern Ocean.

Dispersal model alert on the risk of alien species introduction by ballast waters in protected areas from the Western Antarctic Peninsula

[Dulière Valérie[1]/ Charlène Guillaumot[2,3]], Lacroix Geneviève[1], Saucède Thomas[3], López-Farrán Zambra[4,5,6], Danis Bruno[2], Schön Isa[1], Baetens Katrijn[1]

[1] Royal Belgian Institute Natural Sciences, Brussels, Belgium
[2] Laboratoire de Biologie marine CP160/15, Université Libre de Bruxelles, Bruxelles, Belgium.
[3] Biogéosciences, UMR 6282 CNRS, Université Bourgogne Franche-Comté, Dijon, France.
[4] LEM-Laboratorio de Ecología Molecular, Instituto de Ecología y Biodiversidad, Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile.
[5] Research Center Dynamics of High Latitude Marine, Ecosystem (Fondap-IDEAL), Universidad Austral de Chile, Valdivia, Chile
[6] LEMAS-Laboratorio de Ecología de Macroalgas Antárticas y Sub antárticas, Universidad de Magallanes, Punta Arenas, Chile

Diversity and Distribution, submitted.

Abstract

*Location***.** The Western Antarctic Peninsula (WAP) is challenged by climate change and increasing maritime traffic that together facilitate the introduction of marine non-native species from warmer regions neighboring the Southern Ocean (SO). Ballast water release has been frequently reported as an introduction vector.

*Aim***.** This study uses a Lagrangian approach to model the passive drift of virtual propagules departing from ballast waters hypothetic release zones, at contrasting distances from the WAP coasts.

*Methods***.** Virtual propagules were released over the 2008-2016 period and at three distances from the nearest coasts: 200 (international guidelines), 50 or 11 nautical miles (NM).

*Results***.** Results show that releasing at 200 NM considerably reduces the arrival of propagules in proposed marine protected areas (MPAs) of the western WAP. On the eastern part, propagules can reach north-eastern MPAs within a few days due to strong currents for all tested scenarios. Seasonal and yearly variations indicate that exceptional climate events could influence the trajectory of particles in the region. Ballast water should be released at least 200 NM offshore on the western side of the WAP and avoided on the eastern side to limit particle arrival in proposed MPAs. Focussing on Deception Island, our results suggested that ballast water could be at the origin of the observation in 2010 of the Patagonian crab (*Halicarcinus planatus*) in case of water release at 50 NM or less from the coast.

*Main conclusions***.** This study suggests that existing guidelines are not sufficient to limit the risk of non-native species introduction. Managing ballast water release of ships visiting the SO and joining such recommendations to future MPA proposals to reduce the risk of non-native species introduction is highly recommended, especially in the context of a more touristic and warmer SO.

Keywords

ballast water, Southern Ocean, Antarctic tourism, invasive species, Marine Protected Areas, dispersal modelling, maritime traffic

ACKNOWLEDGEMENTS

This is contribution no. 49 of the vERSO and no. 26 of the RECTO projects (https://rectoversoprojects.be/), both funded by the Belgian Science Policy Office (BELSPO). This work was also supported by a “Fonds pour la formation à la Recherche dans l’Industrie et l’Agriculture” (FRIA) and “Bourse Fondation de la Mer” grants to C. Guillaumot. We are thankful to Argentina and Chile delegations for QGIS maps of the proposal for a conservation measure establishing a MPA in Domain 1, to Sebastian Cisneros for its contribution to gathering the model forcing and to Emma Young and Léo Barbut for the fruitful discussions on Lagrangian models. Thanks to Ben Wallis for helpful comments during the ideas development.

AUTHORS’ CONTRIBUTION

Valérie Dulière, Charlène Guillaumot and Katrijn Baetens co-conceived the ideas, designed the scenarios, analyzed and interpreted the results and wrote the manuscript. Zambra López-Farran contributed to the development of the ideas, Valérie Dulière developed the model, carried out the model simulations and contributed to the post-processing of the model results and figure production. Charlène Guillaumot post-processed the model results and produced the figures. All authors provided critical feedback and contributed to the final version manuscript.

1. INTRODUCTION

Antarctic marine life is characterised by high levels of endemism (Griffiths 2010) as a result of the long climatic, geodynamic and oceanographic histories of the Southern Ocean (SO) (Crame 1999, Pfuhl and McCave 2005, Aronson et al. 2007, Clarke and Crame 2010). The SO, here, is defined as water masses bounded by the Antarctic continent to the south and the Polar Front (PF) to the north (Rintoul 2009); with the PF being the most significant of a series of circumpolar marine fronts associated with the eastward-flowing jets of the Antarctic Circumpolar Current (ACC) (Orsi et al. 1995). Both the PF and the ACC form physical barriers preventing Antarctic surface water exchanges between the SO and northern ocean expanses (Aronson et al. 2007, Griffiths 2010, Sanches et al. 2016), hence blocking the dispersal of most marine organisms (Peck et al. 2014, Convey and Peck 2019). As a result from the prevalence of such important marine fronts, combined with strong currents and the remoteness from other land masses, a unique SO marine diversity has been shaped (Lawver et al. 1992, Crame 1999, Clarke et al. 2005, Barnes and Clarke 2011).

Polar regions are currently challenged by the multiple effects of climate change at a fast pace (Ansorge et al. 2014, Henley et al. 2019). Antarctic coastal marine ecosystems are notoriously sensitive because many shallow-water species have limited resilience abilities and limited southward migration capacities, towards more suitable areas (Stenni et al. 2017, Cárdenas et al. 2018, Gutt et al. 2018). Direct and indirect impacts of climate change are expected to alter the structure and functions of these marine ecosystems leading to species distribution shifts, local extinctions, and favorable condi-tions for colonization by introduced non-native species (Hughes and Convey 2010, Bender et al. 2016). Anthropogenic impacts induced by fisheries, tourism and research activities have been shown to facilitate the transport and introduction of alien organisms through ship hull fouling and ballast water exchanges (Lewis et al. 2005, Lee and Chown 2009c, Hugues and Ashton 2017).

The Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) was created in 1982 to regulate trade and exploitation of Antarctic marine living resources, and set up marine protected areas (MPAs) (https://www.ccamlr.org/en/organisation/home-page, accessed October 2020). The Scientific Committee and Commission of CCAMLR yearly review and rule on new MPA projects proposed by national experts. To date, two Antarctic MPAs have been initiated and include waters off the South Orkney Islands (in 2009) and within the Ross Sea region (in 2016). Antarctic mineral and core resources are not exploited yet (Westermeyer et al. 2020) but commercial fishing and tourism are on the rise, in particular along the west coast of the WAP (Lee and Chown 2009a, Bender et al. 2016, McCarthy et al. 2019). During the last five austral summers (2014 to 2019), the yearly number of tourists visiting Antarctica has increased from 36,700 to 55,400 (IAATO 2019).

As a consequence of increasing human pressure, the Antarctic region has progressively become less isolated and more affected by human footprint (Chu et al. 2019, Joblin et al. 2020). Among others, the rising maritime traffic has resulted in an increasing risk of introducing non-native species to the SO (Hughes and Convey 2010, Bender et al. 2016), as already reported for terrestrial (*e.g*. the bluegrass *Poa annua*, the brachypterous chironomid *Eretmoptera murphyi* or the enchytraeid worm *Christensenidrilus blocki,* Hughes and Convey 2010, Chown et al. 2012, Chwedorzewska et al. 2015) and marine environments (*e.g*. the seaweed *Ulva intestinalis*, the crab *Hyas araneus*, the mussel *Mytilus platensis* or the tunicate *Ciona intestinalis*, see Hughes et al. 2019 and McCarthy et al. 2019 for a review). Introduction of non-native species have almost exclusively been reported in the vicinity of research stations and visitor landing sites (Lee and Chown 2009b, Volonterio et al. 2013, Hughes et al. 2015).

Ship hull fouling and ballast water release are major vectors of alien species dispersal and introduction to Antarctic coastal waters (Lavoie 1999, Barnes 2002, Lewis et al. 2003, Lewis et al. 2005, Chan et al. 2015, Hughes et al. 2019). Ballast water tanks are filled at ships' home ports in South America to safely navigate across the Drake passage to Antarctica. Fishing vessels progressively discharge most of their ballast waters as it is replaced by their catch. Cruise ships

typically discharge ballast waters to travel faster and regularly take up new water to replace the volume left by fuel consumption (Hughes et al. 2019).

The text of the Antarctic Treaty Consultative Meeting (ATCM, 2006) provides practical guidelines for ballast water release in the SO to mitigate the risk of introducing non-native species in coastal areas. Point 5 of the text says “For vessels needing to discharge ballast water within the Antarctic Treaty area, ballast water should [...] (be released) at least 200 nautical miles from the nearest land [...] if this is not possible for operational reasons then such exchanges should be undertaken in waters at least 50 nautical miles from the nearest land”. Complementary, the International Maritime Organization (IMO) adopted the International Convention for the Control and Management of Ships' Ballast Water and Sediments (BWM) in September 2017 (http://www.imo.org/en/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Control-and-Management-of-Ships'-Ballast-Water-and-Sediments-(BWM).aspx-, accessed October 2020). This convention, ratified by 83 countries so far, establishes standards and procedures for the management and control of ship ballast water and sediments to avoid the unintentional dispersal of alien species. The main document also states that ballast water exchange should be done at least 200 nautical miles from the nearest land and in water at least 200m deep, but “in cases where the ship is unable to conduct ballast water exchange [in these conditions], this should be as far from the nearest land as possible, and in all cases at least 50 nautical miles from the nearest land”.

In the present study, a Lagrangian model was developed to simulate the drift of virtual particles as they are transported along the water masses. The model calculates particle trajectories (identified here as potential propagules) according to different point locations where ballast water is discharged. This was exemplified by the Patagonian crab, *Halicarcinus planatus* (Fabricius, 1775), reported in Deception Island (Western Antarctic Peninsula) in 2010. The potential impact of ballast water release on the introduction of alien species in Antarctic coastal waters was analyzed through pluriannual and multi-seasonal time scales. A map of recommended areas for ballast water release is proposed as a tool to support good practices for ballast water discharge and for conservation purposes

2. MATERIAL AND METHODS

2.1. Study area

The study area is enclosed by the strong eastward flowing ACC (Appendix 4.1) and includes the Scotia Arc region, the Antarctic Peninsula, and the Weddell Sea, as they concentrate most of the maritime traffic between Antarctica and southern South America and therefore, the highest risk of alien species introduction (Lynch et al. 2010, McCarthy et al. 2019).

2.2. Lagrangian model principle and hydrodynamic settings

In this study, we used a Lagrangian particle model, which combines oceanographic information (*e.g*. bathymetry, current direction and speed) forced by atmospheric factors (temperatures, winds, atmospheric pressure) (Huthnance 1991, Robinson and Golnaraghi 1994) with biological features (*e.g.* organisms' size, development rate, buoyancy, Van Sebille et al. 2018). The model used in this study is based on the model described in Dulière et al. (2013) and made available as a module of the free and open-source aquatic modelling system COHERENS v2 (Luyten 2011). This system has already been used to study, among others, oil spill dispersal (Legrand and Dulière 2014), jellyfish drift (Dulière et al. 2014) and the movement of harbor porpoises (Haelters et al. 2015). Particles are transported under advective and diffusive processes. The classical fourth-order Runge-Kutta method is used to estimate horizontal transport. The diffusive velocities are obtained from random walk theory with constant horizontal and vertical diffusion coefficients of 10 and 0.0001$m^2.s^{-1}$, respectively. The same diffusion coefficient values are used as in Young et al. (2014) and are equivalent to values observed in the SO (empirical values or commonly accepted by modelers; Sheen et al. 2013, Watson et al. 2013). A bouncing condition is used for particles

reaching the sea surface or seabed, and particles that leave the model domain through the ocean open boundary are assumed to have left the region. Stranding is not allowed, so when a particle reaches a dry cell, its position is set to its previous position at sea. The Lagrangian module is used off-line with a computation time step of 5 minutes. To ensure the general purpose of this study, the model has been set up with no specific organismal behaviour (*i.e.* swimming or tidal or diurnal vertical migration) and particles are assumed to move along with water masses (*i.e.* no buoyancy effect).

The hydrodynamic conditions used to force the Lagrangian model are based on the 2008-2016 PHY_001_024 datasets produced by the high-resolution global analysis and forecasting system, provided by Mercator Ocean (Law Shune et al. 2019). These products contain daily mean fields of sea surface elevation and horizontal ocean currents. In addition, they also contain sea ice information (*i.e*. concentration, thickness and velocity), sea water potential temperature, sea water salinity and ocean mixed layer thickness. These datasets have been generated with NEMO 3.1 and LIM2 EVP models forced with 3-hourly atmospheric forcing from ECMWF (European Centre for Medium-Range Weather Forecasts, https://www.ecmwf.int/). Daily averaged model products are made available after interpolation from the native model grid to a global standard Arakawa C grid of 1/12° horizontal resolution and 50 fixed vertical levels (from 0 to 5,000 m). The quality of the Global high-resolution products has been assessed in Lellouche et al. (2019). 3D vertical ocean currents are estimated from the divergence in the horizontal velocity from the PHY_001_024 forcing fields, assuming null surface and bottom vertical velocity.

The model grid was built from a sub-sample of the global grid of the hydrodynamic forcing field from latitude 45°S down to the South Pole. The horizontal resolution of 1/12° (~8km) was kept and the 50 vertical levels have been adapted to 50 sigma levels for the COHERENS system. The Lagrangian particle model has been previously validated in Dulière et al. (2013), Legrand and Dulière (2014) and a quality analysis of the hydrodynamic forcing is provided in Lellouche et al. (2019).

2.3. Particle release scenarios

Three scenarios were defined for simulating drift trajectories of organisms potentially released during ballast water discharge along the Antarctic coasts, according to ATCM (2006). The first scenario considers that ballast waters are released 190 to 210 Nautical Miles (NM) away from the nearest coasts (‘200 NM scenario’), which complies with the ATCM ratified guidelines (Fig. 4.1A). The two other scenarios represent cases of technical issues preventing from carrying out ballast water release at 200 NM: the second scenario simulates release from 40 to 60 NM from the nearest coasts (‘50 NM scenario’, Fig. 4.1B) and the third scenario hypothesizes a major transgression of the guidelines, with release from 2 to 20 NM from the closest coasts (‘11 NM scenario’, Fig. 4.1C). Six release zones (Fig. 4.1) were defined: Western Antarctic Peninsula (Rz.1), Eastern Antarctic Peninsula (Rz.2), East Weddell Sea (Rz.3), South Orkney Islands (Rz.4), South Georgia Islands (Rz.5) and South Sandwich Islands (Rz.6).

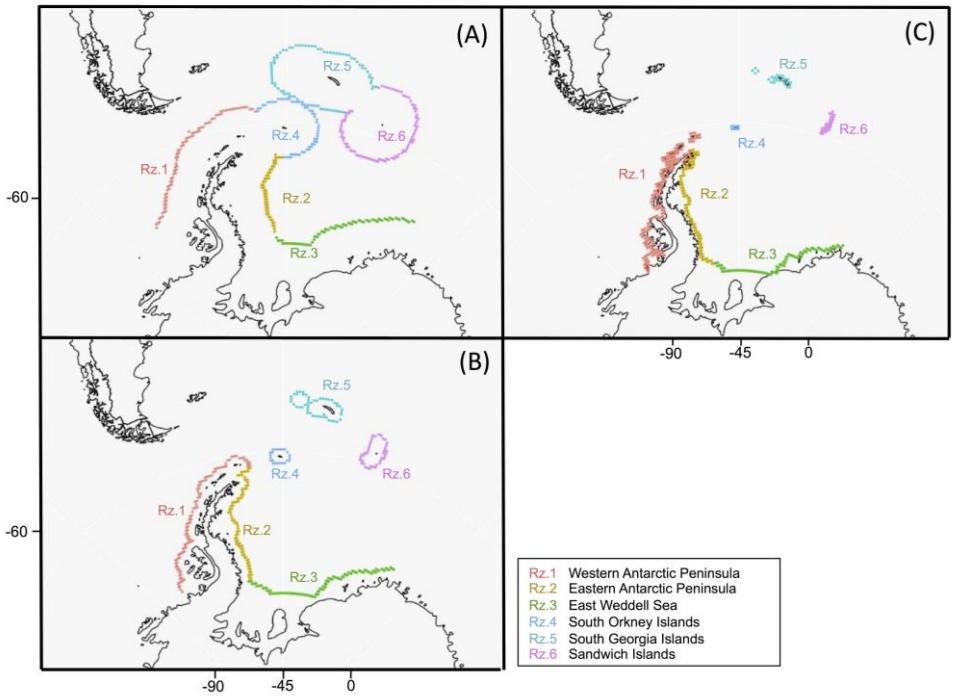


Figure 4.1. Locations of the six particle release zones for the 200 NM (A), 50 NM (B) and 11 NM (C) scenarios.

For each release location, particles were released daily over a 9-year period (from 2008 to 2016) to account for seasonal and interannual variabilities. For technical reasons, particles have been released every six grid-cell pixel in latitude (every 1/2°) and every four grid-cell pixel in longitude (every 1/3°), with respect to 200 m depth. The number of release locations ranges from 8 (for Rz.5 in scenario 11 NM) to 224 (for Rz.1 in scenario 11 NM) release points among the release zones and scenarios. All together, it is more than 4.5 million particle trajectories that have been studied in the model. A 6-month duration of the drift was chosen to match the duration of larval development periods of many SO organisms, which can drift in the water column for sometimes up to six months before starting metamorphosis and settling down on the seabed (White 1998, Stanwell-Smith and Clarke 1998, Stanwell-Smith et al. 1999). Particles initiate their drift at 10 m depth.

2.4. Particles' trajectory and age: statistical comparisons

Model simulations have produced large datasets with model estimations of the daily age and positions (latitude, longitude and depth) of virtual particles. Model results have been post-processed for different years and seasons (January-February-March; April-May-June; July-August-September; October-November-December) and for each ballast water release scenario, to generate maps of dispersal patterns. Results for averaged years and seasons are first provided to describe the overall dispersal patterns of particles drift. Then, interannual and seasonal variabilities are described. Due to the different number of released particles among release areas and scenarios, a scaling correction has been applied for statistical analyses (giving a 'weighted number of particles').

2.5. Proposed Marine Protected Areas

The likely consequences of ballast water release on the potential introduction of alien organisms in MPAs of the WAP, was assessed by analyzing particle entry into proposed MPAs. Proposed MPAs for this region are the interest of the Chilean and Argentinian delegations at CCAMLR. The SC-CAMLR-38/BG/03 report (CCAMLR report SC-CAMLR-38/BG/03 2019), proposes seven regulated areas, selected according to multiple arguments, including the spatial distribution of the benthos to top predators, oceanographic processes, climate change and fishing activities (Fig. 4.2). CCAMLR will rule on this proposal at the next international meeting. Among these proposed regulated areas, CCAMLR distinguishes (1) General Protection Zones (GPZ) that aim at protecting habitats,

bioregions and species in an attempt to mitigate or eliminate specifically identified ecosystem threats from fishing; and support existing and future scientific research and monitoring and (2) Krill Fishery Zones (KFZ) that include fishing areas in addition to protecting benthic habitats (CCAMLR report SC-CAMLR-38/BG/03 2019). (3) The established MPA of the South Orkney Islands (May 2010) belongs to this conservation proposal and already prohibits any fishing activity, any transshipment activities and any discharge or dumping. All activities occurring in the area should be declared according to the CCAMLR delegation (CCAMLR report 91-03 2009, Trathan and Grant 2020).

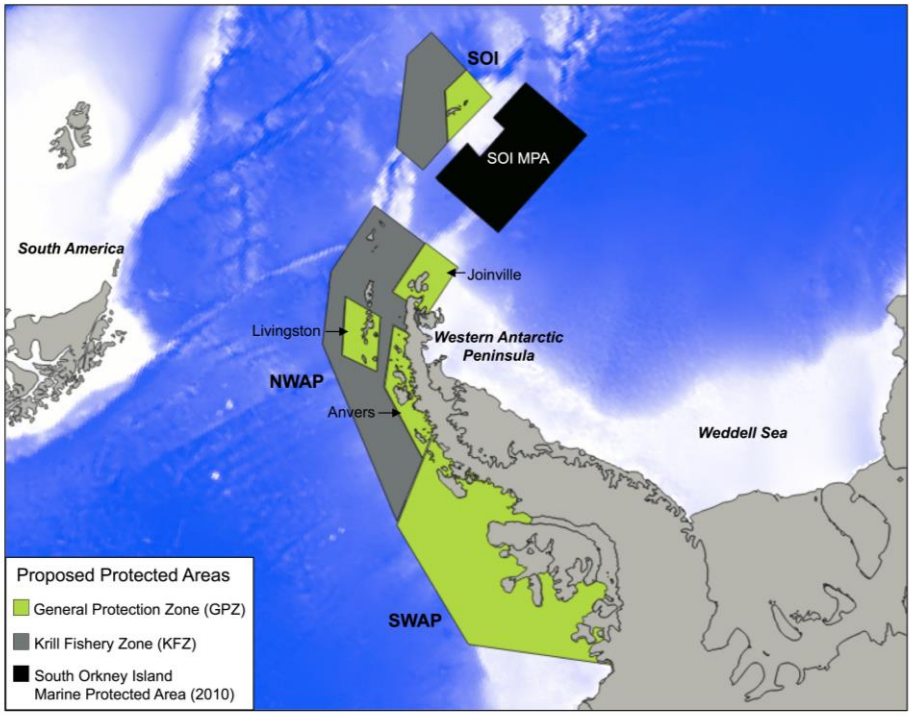


Figure 4.2. Map of proposed marine protected areas. Modified from SC-CAMLR-38/BG/03 report. NWAP and SWAP stand for Northern and Southern Western Antarctic Peninsula, respectively. SOI and SOI MPA stand for South Orkney Islands and South Orkney Islands Marine Protected Area, respectively.

2.6. Focus on Deception Island

In February 2010, a living and mature female of the brachyuran crab *Halicarcinus planatus* (Fabricius, 1775) was reported in shallow, subtidal waters of Deception Island (WAP) (Aronson et al. 2014) (Fig. 4.3). *Halicarcinus planatus* is usually distributed in shallow water areas of southern South America and along coastal areas of some sub-Antarctic Islands (Falkland, Marion, Crozet, Kerguelen and Macquary islands) (Boschi et al. 1969, Richer de Forges 1977, Aronson et al. 2014, Varisco et al. 2016). This little crab (shell diameter from 15 to 20 mm, Fig. 4.3) is an opportunistic feeder (Boschi et al. 1969) that is commonly found sheltered below intertidal and subtidal rocks (Richer de Forges 1977, Vinuesa and Ferrari 2008).

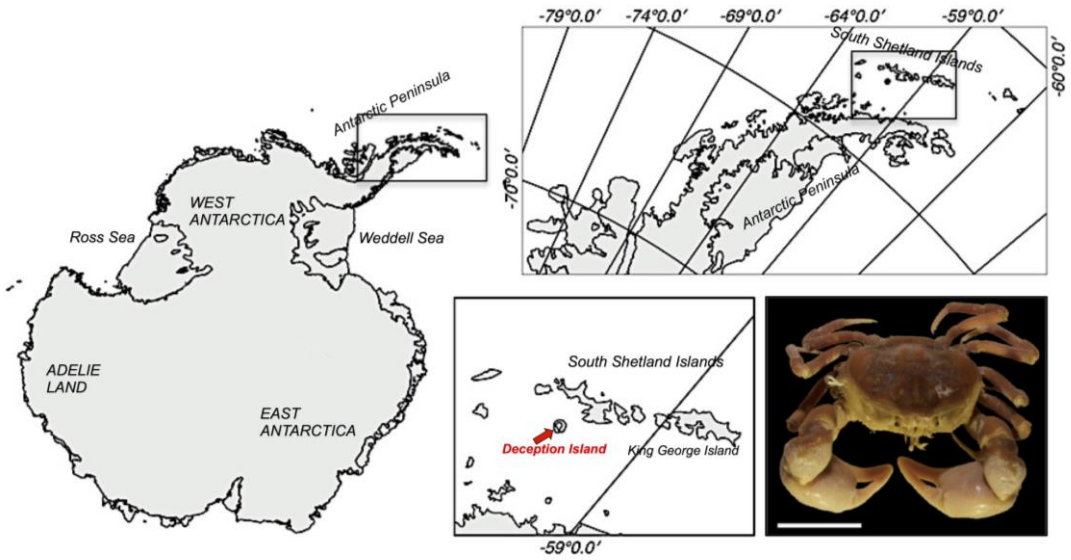


Figure 4.3. Location of Deception Island in the Western Antarctic Peninsula and representative male individual of the crab *Halicarcinus planatus*, scale= 1cm © Karin Gérard.

Halicarcinus planatus has a high dispersal potential, with the release of planktonic larvae in the water column that can drift for 45 to 60 days at temperatures of 11-13°C and 8°C respectively, before settling on the seabed and triggering metamorphosis (Boschi et al. 1969, Richer de Forges 1977, Diez and Lovrich 2010). Recent studies suggest that larvae of *H. planatus* are not capable of drifting across the SO when transported by the eastward flow of the ACC (López-Farrán et al. in prep. - Appendix). If introduced into shallow waters of the WAP (*e.g*. by ballast waters), climate warming is predicted to favour the species' survival (López-Farrán/Guillaumot et al. in press - Chapter 3). The hypothesis of a potential introduction of *H. planatus* in Deception Island through ballast water release was tested by subsetting the model results to a maximum drift period of 2 months, following the known maximal drifting time of the crab larvae (Boschi et al. 1969, Diez and Lovrich 2010). The hypothesis has been tested for all three release scenarios.

3. RESULTS

3.1. Dispersal patterns according to the different release scenarios

General dispersal patterns were different among the three release scenarios (Fig. 4.4). The 200 NM scenario leads to the widest dispersal pattern that expands further eastward across the PF to the sub-Antarctic area. The 50 and 11 NM scenarios comparatively lead to narrower and less extended dispersal patterns (Fig. 4.4). Geographical and oceanographic features (such as the Weddell Sea gyre, the PF, the ACC, the Scotia Ridge) clearly delineate the shape of the dispersal of particles in the corresponding areas.

Release scenarios also show differences in the weighted number of particles (Fig. 4.4B), with 15-fold more particles reaching the coastlines for the 11 NM scenario compared to the other scenarios, mainly closeby the coasts of the WAP.

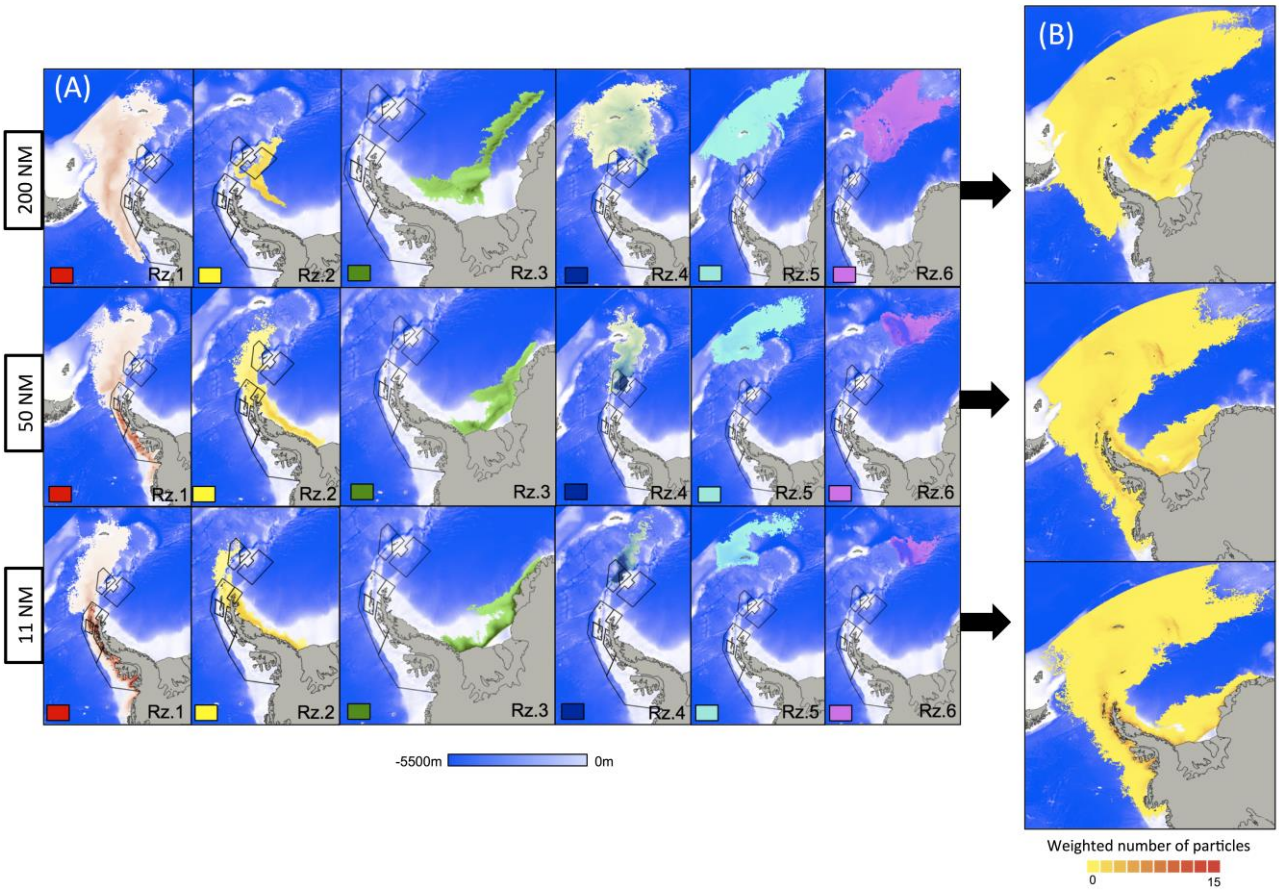


Figure 4.4. Model estimated dispersal patterns for the three release scenarios: 200 NM, 50 NM and 11 NM. Left panels (A): dispersal patterns per release zones (presented in Fig. 4.1); right panels (B) dispersal patterns for all release zones combined. These dispersal patterns are averaged trajectories for the nine-year simulation period, all seasons included. The weighted number of particles is the scaled number of particles, relative to the number of release locations which differs between scenarios. Blue background: bathymetry chart.

3.2. Recurrent arrivals in MPAs during austral summer periods

Gridded weighted numbers of particles have been summed up in each proposed MPA. Results show for the austral summer period contrast between scenarios (Fig. 4.5). These sums (size of circles) are smaller in the 200 NM scenario compared to the 11 NM and 50 NM scenarios for all proposed MPAs but the SOI-MPA. These sums decrease by ~99% for GPZ-SWAP, ~30% for KFZ-SWAP, 100% for Anvers, >96% for Livingston and 100% for Joinville, for the 200 NM scenario in comparison to the two other scenarios. Particles that reach MPAs in scenario 200 NM mostly originate from the western or eastern coasts of the WAP (Rz.1 and Rz.2).

In the 11 NM and 50 NM scenarios, the proposed MPAs (SOI-MPA, KFZ-SOI and GPZ-SOI) in the north-east are mainly affected by particles released close to South Orkney (Rz.4). Contrastingly, when particles are released 200 NM away from the coasts of Rz.4, the number of particles reaching these three proposed MPAs is predicted to strongly decrease (-98.9 and -98.5% for 50 and 11 NM scenarios, respectively). When released from 200 NM offshore, particles coming from the Eastern Antarctic Peninsula zone (Rz.2) drift north-eastward and reach the SOI-MPA in large numbers. The eastward flow of the ACC prevents particles released from South Georgia and Sandwich Islands areas (Rz.5 and Rz.6) to reach the proposed MPAs although they might impact other areas located further east. Releasing particles from the Weddell Sea area (Rz.3) also never impacts the proposed MPAs.

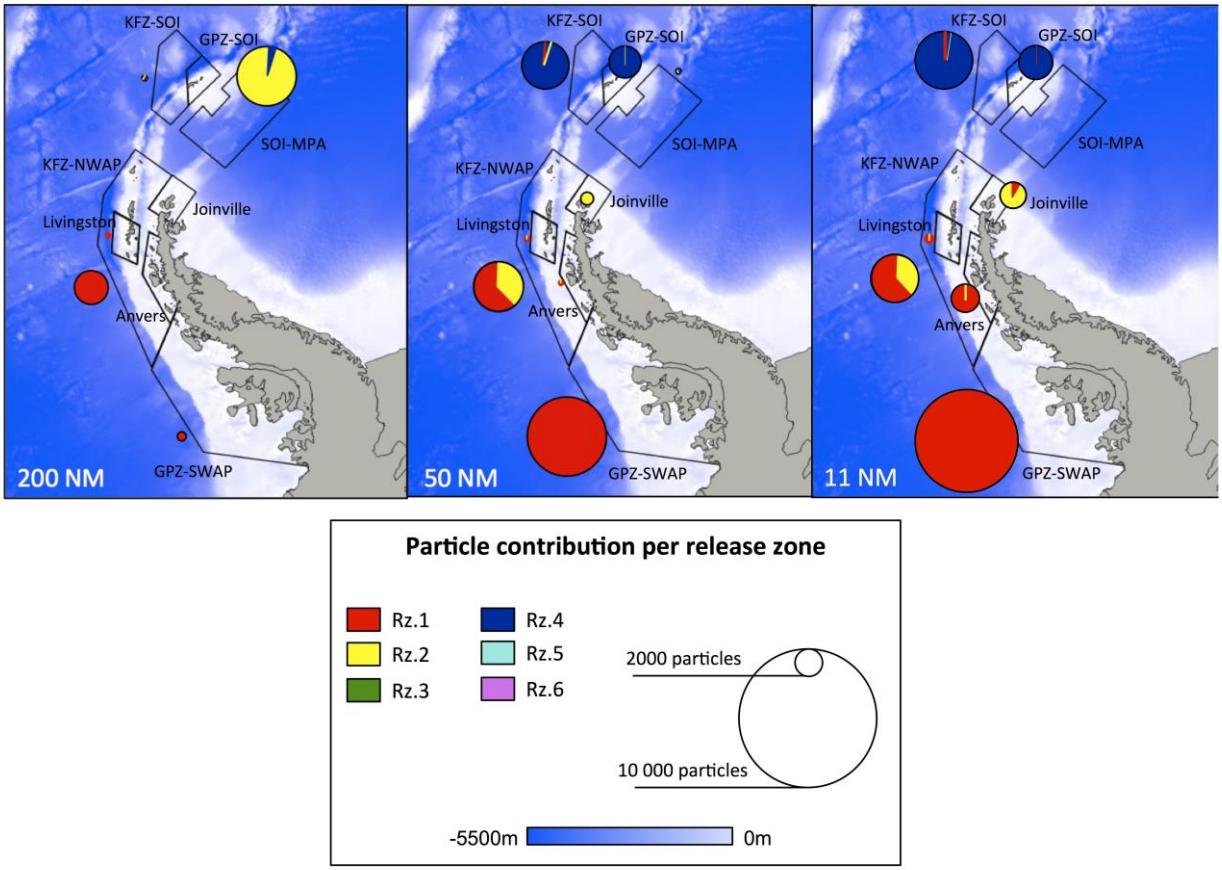


Figure 4.5. Sums of the weighted numbers of particles reaching the proposed marine protected areas (MPAs) during the January-February-March season (austral summer, being the season with the largest number of ships entering the Southern Ocean) over the 9-year period (2008-2016) and for each release scenario (200 NM, 50 NM and 11 NM). Release zone positions are shown in Fig. 4.1. Details about proposed marine protected areas are given in Fig. 4.2.

3.3. Particle age upon arrival in MPAs

The average age of particles reaching the proposed MPAs varies from 93 to 165 days, 74 to 131 days, and 59 to 136 days for the 200, 50 and 11 NM scenarios, respectively (Fig. 4.6). For the 11 NM scenario, the first particles generally reach the proposed MPAs in less than 10 days (except for SOI-MPA) and for the 50 NM scenario, in less than 20 days (except for Anvers and SOI-MPA). For the 200 NM scenario, it generally takes longer (over 25 days, except for GPZ-SWAP and SOI-MPA) for the particles to reach the MPAs. Particles reaching GPZ-SWAP, KFZ-NWAP, KFZ-SOI, GPZ-SOI and SOI-MPA are older when released at a distance of 11 NM from the coast than when released at a distance of 50 NM.

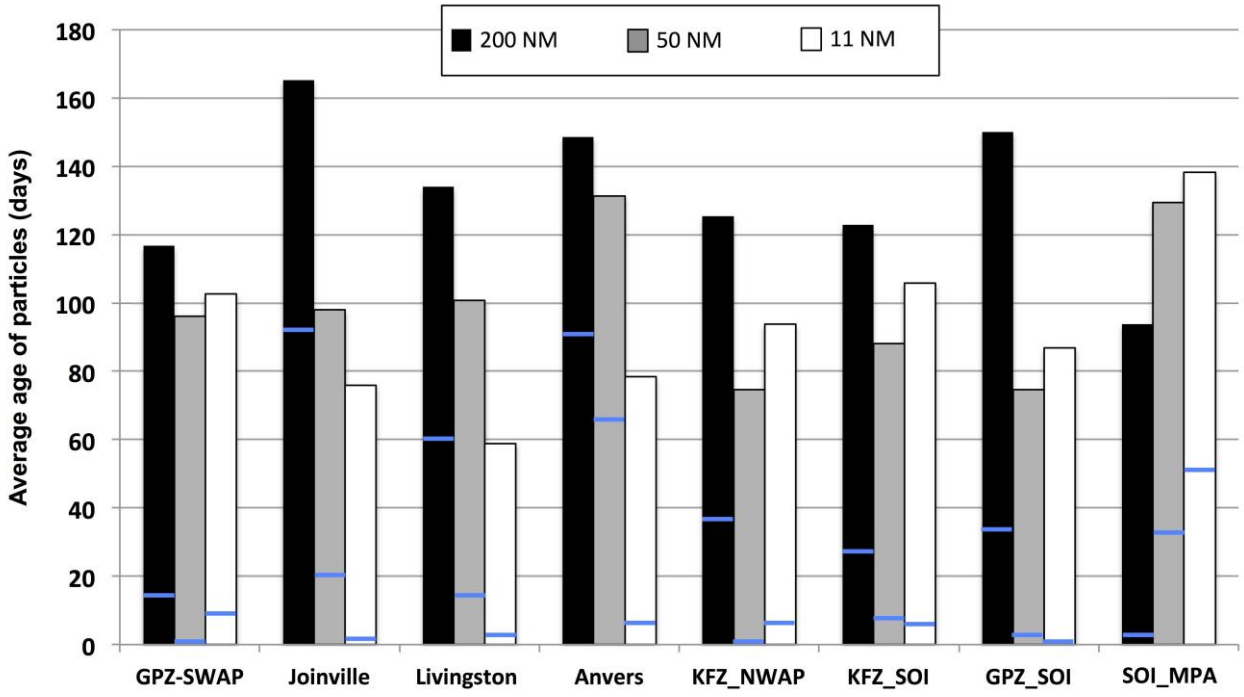


Figure 4.6. Age of particles (in days) reaching the proposed marine protected areas under the 200 NM scenario (black), 50 NM scenario (grey) and 11 NM scenario (white) for the January-February-March season. Values are averaged over the 9 years (2008-2016), blue solid horizontal lines indicate the average year minimal value recorded within the period (2008-2016).

3.4. Intra- and inter-annual variabilities

Comparison of dispersal patterns among the nine simulated years show inter-annual variations in the extent of dispersal areas: such variation is mainly noticeable in the sub-Antarctic region and in the East Weddell Sea. Interannual variation is more obvious in the 11 NM scenario relative to the total extent of the dispersal pattern (Fig. 4.7; right panel). Interestingly, the dispersal area is broader in years 2008 and 2009, more extended to the east in 2014 and 2015 and conversely, more contracted in 2011 and 2012 (results not shown). Inter-seasonal variation is comparatively less marked than inter-annual variation (Appendix 4.2), and main shifts in particle distributions are concentrated to the north-east and south of particle overall distribution, following the same pattern of inter-annual variation (Fig. 4.7, Appendix 4.2).

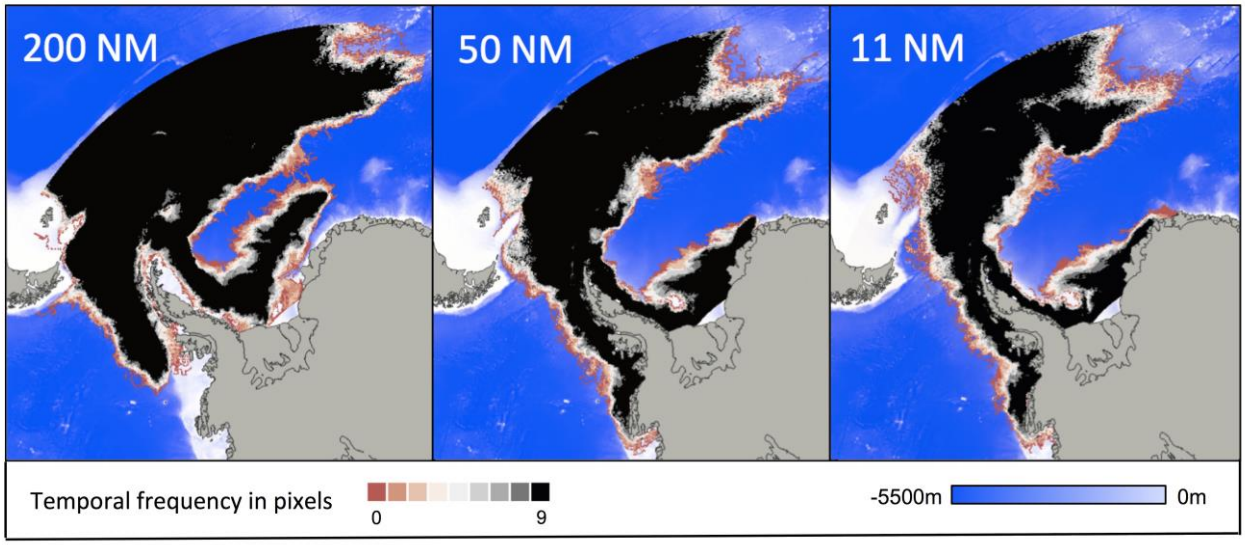


Figure 4.7. Model estimated dispersal patterns assuming the release scenarios: 200, 50 and 11 NM, for particles released from all release zones at the same time. Colors represent the frequency of occurrence among the nine years (2008-2016) with a maximal score of 9 for pixels that receive particles every year. Blue background: bathymetry chart.

Inter-seasonal and inter-annual variations in the origin of particles (release zones) that reach the proposed MPAs highlighted a comparable influence of years and seasons on dispersal contrasts (Fig. 4.8). The origin of particles reaching the MPAs located along the WAP (GPZ-SWAP, KFZ-NWAP, Livingston, Anvers, Joinville), especially the one located in the south (GPZ-SWAP) is less variable. In these MPAs, particles mainly originate from the WAP and the Eastern Antarctic Peninsula zones (Rz.1, Rz.2).

The variability in the origin of particles reaching the GPZ-SOI, KFZ-SOI and SOI-MPA areas, located further north-eastward, is much higher and strongly varies according to the release scenario, season and year, with particles originating either from the WAP, the Eastern Antarctic Peninsula or the South Orkney zones (Rz.1, Rz.2 and Rz.4, respectively).

For the 200 NM scenario, some particles (less than 100 particles, *i.e.* less than 2%) from the South Georgia zone (Rz.5) appear for the first time in the statistics but logically does not present important proportions given the eastward flowing ACC (Fig. 4.8).

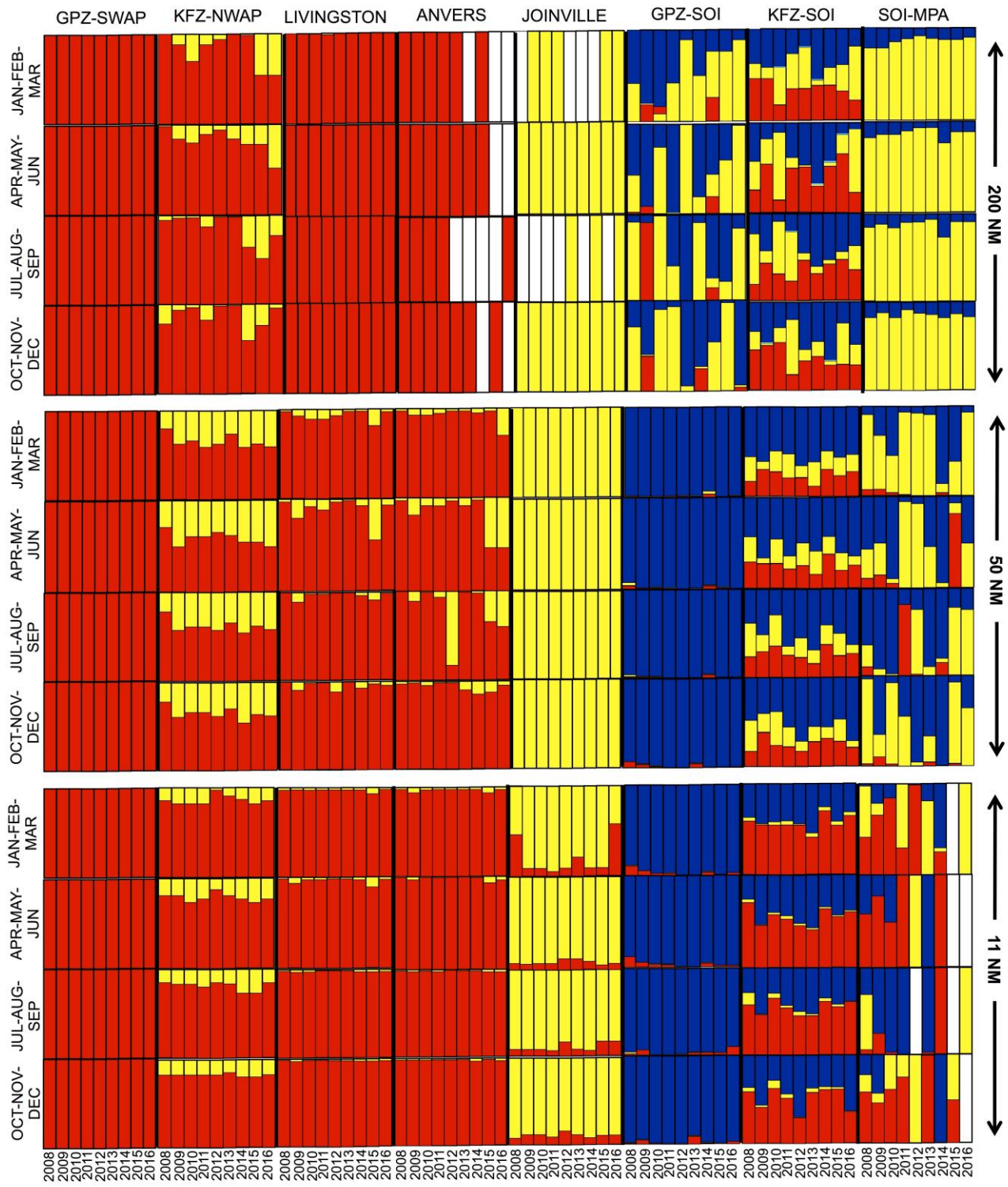


Figure 4.8. Intra- (vertical sub-panels) and interannual (horizontal sub-panels) variations in the origin (release zone) of particles reaching the proposed marine protected areas according to the 200, 50 or 11 NM scenarios. Red: Western Antarctic Peninsula (Rz.1), yellow: Eastern Antarctic Peninsula (Rz.2), dark blue: South Orkney Islands (Rz.4) and turquoise : South Georgia (Rz.5) (very small proportions for KFZ-SOI for 200 NM scenario).

3.5. Invasion risks

Previous results were summarized in a synthesis map (Fig. 4.9) that indicates the release zones and the simulated risk of particle introduction into proposed MPAs. We defined a 'high risk', when models simulate the arrival of particles every year and every season in all neighboring MPAs. The 'no risk' release zones correspond to zones where released particles never reach any proposed

MPAs. A 'moderate risk' category was added for zones where particles may not reach neighboring MPAs during some years and some seasons (according to climatic events) and/or in significantly lower densities.

Results clearly show that releasing ballast waters on the western and eastern sides of the WAP and nearby Scotia Islands generally leads to a high to moderate risk to introduce particles into proposed MPAs, even if released at 200 NM from the nearest coast. In the case of ballast water released in the East Weddell Sea and around South Georgia and Sandwich Islands (Rz.3, Rz.5 and Rz.6, respectively) particles never reach proposed MPAs.

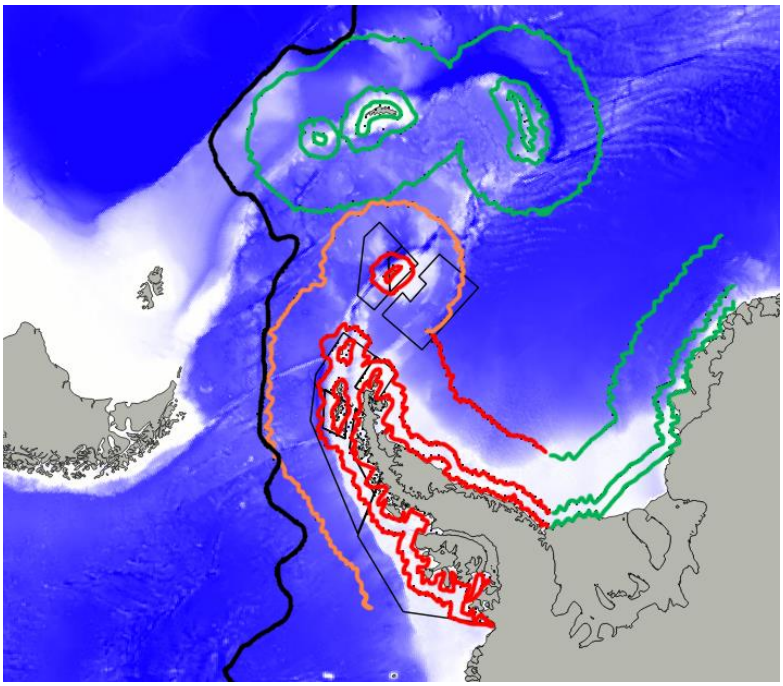


Figure 4.9. Ballast water release zones and their associated simulated risk (green: 'no risk'; orange: 'moderate risk'; red: 'high risk') for particles to reach proposed marine protected areas (MPAs). The black solid line represents the Polar Front yearly mean position. The risk was estimated for proposed MPAs of the considered region only. Other areas that might also be at risk were not included in this study.

3.6. A focus on Deception Island

When ballast waters are released from distances exceeding 200 NM from the nearest coasts, the Lagrangian model predicts that no particle reaches the coasts of the WAP, nor the Gerlache Strait where Deception Island is located (Fig. 4.10A). In contrast, particles reach Deception Island and the Gerlache Strait within 2 months drift under the 50 NM and 11 NM scenarios (Fig. 4.10B-C). The weighted number of particles reaching the South Shetland Islands, entering fjords and drifting along the coasts is up to 15 times larger in the 11 NM scenario than in the 50 NM one; a very constant result across years (Appendix 4.3) and seasons (results not shown). Results are less uniform in the 50 NM scenario, where inter-annual variations cause particles to reach the Gerlache Strait and Deception Island either completely (2009, 2010, 2011, 2012) or partly (2008, 2013, 2014, 2015, 2016).

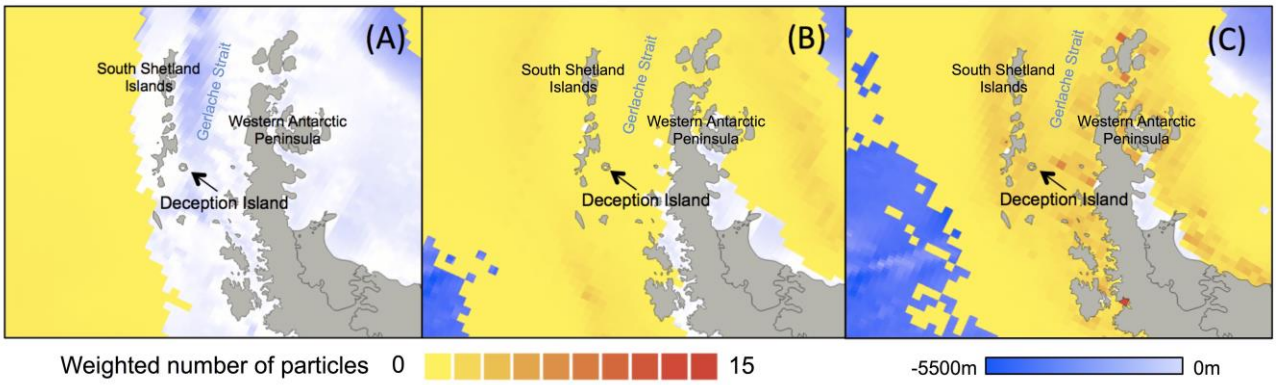


Figure 4.10. Model estimated dispersal patterns, averaged for the nine-year period (2008-2016), for the January-February-March season (southern summer). Particle drift was simulated during two months. The weighted number of particles was obtained by scaling the number of particles by the number of release locations (which differ among scenarios). Results are presented for the three different release scenarios: 200 NM (A), 50 NM (B) and 11 NM (C). Particles are released in all areas at the same time. Blue background: bathymetry chart.

4. DISCUSSION

Particle dispersal

Lagrangian models have been widely used over the last decades for defining and delineating MPAs in many different regions and oceans (Gaines et al. 2003, Berglund et al. 2012, Burgess et al. 2014, Thomas et al. 2014), and study the spread dynamics of invasive species (Brandt et al. 2008, Brickman 2014). In the SO, Lagrangian models have already been used to simulate dispersal abilities and the distribution of fish species or top predators, to understand the main key drivers of population connectivity and assess the position of the main foraging areas, in the aim of determining an effective management of natural resources (Young et al. 2012, 2014, 2015, Della Penna et al. 2017).

In the present work, daily variations of the environment were simulated over a 9-year period and model outputs were analyzed to test the significance of dispersal patterns with regards to inter-seasonal and inter-annual variations. In simulating a large number of particles, Lagrangian models integrate natural variability of hydrodynamic systems (Van Sebille et al. 2018). However, the model should rely on assumptions (parameterization of the general environment, of the properties of the simulated particle), which may not be trivial considering the broad spatial scale of the analysis, the overall system complexity, and the unknown propagule pressure state (*i.e.*, occurrence and density of non-native species in ship ballast waters). Furthermore, some propagule traits such as buoyancy, physiology, survival rate and tidal behaviour were hypothesized; the actual traits of invasive species could potentially have a substantial impact on model outputs (Stanwell-Smith et al. 1999, North et al. 2008, Young et al. 2012, Miller et al. 2013, Barbut et al. 2019).

Model simulations show that in six months, particles can drift along the coasts of the WAP up to South Georgia, driven by the power of the ACC and highlighting the importance of connectivity between Antarctic coasts and the Scotia Sea region (Appendix 4.1, Rintoul 2009, Caccavo et al. 2018, Moffat and Meredith 2018). Other oceanographical features such as the Weddell Sea gyre and major marine fronts off the Scotia Sea (PF)(Della Penna et al. 2017), along with geomorphological features such as the South Orkney Ridge or the South Georgia shelf also clearly influence dispersal patterns (Fig. 4.4, Young et al. 2012, 2015, Vernet et al. 2019), and play a crucial role in the connectivity among sub-Antarctic islands (Young et al. 2012).

Comparisons of the three different scenarios indicate that release distance from the nearest coasts has significant impacts on particle trajectories, on the frequency and weighted number of particles reaching the Antarctic coasts (Fig. 4.4-5-7). Overall, particles are less likely to reach Antarctic coastal areas when ballast waters are released at least 200 NM away from the nearest coasts (Fig. 4.4-5).

Inter-seasonal and inter-annual variability were also shown to have significant effects on modelled dispersal patterns (Appendix 4.2, Fig. 4.7-8). This was expected here, given that SO hydrodynamics are highly controlled by the variability of atmospheric and climate regimes at both high and low frequencies (Henley et al. 2019). Some of our results indicate that in years 2008, 2009, 2014 and 2015 particles were spread furthest while dispersal was the lowest in 2011 and 2012. These results are linked to specific climate events and in particular, to regimes of westerly winds and to the intensity of the Southern Annular Mode (Limpasuvan and Hartmann 1999), which has been shown to be strongly and linearly teleconnected to the phase of El Niño Southern Oscillation, explaining roughly 25% of the SAM interannual variance during the austral summer season (Carvalho et al. 2005, L'Heureux and Thompson 2006, Ciasto and Thompson 2008). Years 2009, 2014 and 2015 were characterised by a negative Southern Oscillation Index with strong El Niño episodes (warmer temperatures and stronger westerly winds); in contrast, years 2010-2011 were characterised by strong positive Southern Ocean Indexes and with La Niña episodes (weaker westerly winds, dryer and colder atmosphere) (Nicolas et al. 2017).

Results' overview in the general context

The WAP is among the regions on Earth that experience climate warming at the fastest pace, where rising temperatures also directly or indirectly drive other environmental shifts (*i.e*. glacier melting, phytoplankton community shifts, changes in sea ice duration and extent) (Convey et al. 2009, Bers 2013, Schram et al. 2015, Schofield et al. 2017, Convey and Peck 2019). This makes the WAP one of the most sensitive regions to potential invasions by introduced species in Antarctica (Meredith and King 2005, Hellman et al. 2008, McGeoch et al. 2015, Hughes et al. 2019) because increased temperatures and related environmental shifts may favour the acclimation of alien species introduced from warmer climates over cold-adapted native taxa (Hellmann et al. 2008, Galera et al. 2018).

For a few decades, maritime traffic has also steadily increased in the SO and in the WAP in particular, due to its relative proximity to harbors of southern South America (McCarthy et al. 2019). This increasing traffic has been cited as the main cause for alien species introduction in coastal waters of the WAP (Tavares and De Melo 2004, Lee and Chown 2007, Diez and Lovrich 2010). Many observations of non-native species have been reported in the last years including decapods, algae, bivalves (Thatje and Fuentes 2003, Fraser et al. 2018, McCarthy et al. 2019, Avila et al. 2020, Cárdenas et al. 2020), as well as the Patagonian crab *H. planatus*, which was found in Deception Island in summer 2010 (Aronson et al. 2014).

Our study is strongly embedded within this context, by evaluating the impact of ship circulation on marine environments, and in MPAs in particular. Results highlight the importance of the distance of ballast water release from coasts to control the frequency and density of particles reaching Antarctic coasts. Focusing more specifically on Deception Island, our simulations indicate that no particle reach the Gerlache Strait when ballast waters are released at 200 NM from the coasts, suggesting that the non-native crab *H. planatus* could not have been introduced to Deception Island due to ballast water release if the ATCM guidelines had been respected. If the introduction to Deception Island indeed occurred through ballast water of cruise ships sailing southwards from ports of southern South America, which we consider to be a likely scenario, the crab must have been released at a distance equal or less than 50 NM from the Antarctic coasts (Fig. 4.10, Appendix 4.3). These results could be generalized to other species, with the ensuing consequences of species introductions (Walsh et al. 2016, David et al. 2017, Britton et al. 2018).

Our results also highlight that the variability in climate regimes has a strong effect on dispersal patterns meaning that in certain years, particles may drift further and reach areas that are on

average not considered to be potentially impacted by ballast water release and the risk of alien species introduction, as already discussed by Fraser et al. (2018) and Waters et al. (2018) for kelp rafting. Other authors stressed the significance of transient events in long-distance dispersal (Leese et al. 2010, Saucède et al. 2014). Such events may become more frequent in future decades, owing to ongoing climate change, since climate projections for the Southern Hemisphere for the 21st century predict a further southward shift and intensification of storm tracks (Perlwitz 2011) and therefore hypothesize an increasing threat for potential species introductions (Hugues et al. 2020).

Future management of MPAs

The ATCM (2006) guidelines recommend that ballast water should be discharged north of the PF before entering Antarctic waters and "preferably north of either the Antarctic Polar Frontal Zone or 60°S, whichever is the furthest north". In practice, the position of the Polar Front is usually noticed after passing it and exchanging ballast waters in these regions is not realistic considering the weather and sea conditions (Wallis B., person. comm.). Consequently, the 200 NM guidelines, provided both by IMO and ATCM (2006) texts, should be as widely as possibly applied by ships that sail across the PF. Our results however suggest that releasing ballast waters at 200 NM around the WAP (Rz.1) may still lead to particles reaching the Antarctic coasts including, the eastern and northernmost proposed MPAs of the region (notably KFZ-NWAP, GPZ-SWAP and Livingston, Fig. 4.5, Fig. 4.8). The particle numbers reaching the Antarctic coasts are considerably reduced when released at 200 NM from the nearest coasts than when released at 50 NM or 11 NM. Although the origin of particles arriving in MPAs varies among years and seasons, the model indicates that ballast water release should best be conducted further away than 200 NM or wherever possibly, avoided altogether on the western side of the Antarctic Peninsula (Fig. 4.9).

When ballast waters are released on the eastern side of the WAP (Rz.2 and Rz.4), particles are predicted to drift north-eastward in the sub-Antarctic region, reaching the KFZ-SOI, GPZ-SOI or SOI-MPA areas within a few days at the earliest, and within 3 months on average (Fig. 4.5-7-8). Regardless of the release scenario, our simulations indicate that it is not possible to prevent particles from reaching the aforementioned MPAs when ballast water is discharged on the eastern side of the WAP (Rz.2; Fig. 4.5 and Fig. 4.8). Avoiding this region for ballast water release is therefore recommended (Fig. 4.9). Discharging ballast water in the East Weddell Sea and around South Georgia and Sandwich Islands (Rz.3, Rz.5 and Rz.6, respectively) results in the absence of particles reaching the MPAs. East Weddell Sea (Rz.3) is however not suitable for ballast water release due to practical reasons, because this region is ice-covered all year long (Vernet et al. 2019). Results also show that particles released in the South Georgia and Sandwich Islands region will not reach the proposed MPAs of the WAP, but the impact on islands located further east was not investigated in the present work.

Given the dense and increasing maritime traffic along the Antarctic Peninsula and Scotia Sea regions, the present model could be improved with more detailed data on ship routes, ballast water discharge events (McCarthy et al. 2019, Hugues et al. 2020) and propagule pressure (Lee and Chown 2009a). Such adapted models could then be used to generate maps delineating recommended zones for ballast water release with a higher precision. Our results strongly stress the necessity to further strengthen existing conservation measures for visitors and ships approaching Antarctic coasts and complementing studies that highlighted the urgency of protecting Antarctica from species introduction (Lennox et al. 2015, McGeoch et al. 2015, Hugues et al. 2020). Such conservation measures should be joined to MPAs proposals.

Awaiting the definition and acceptance of recommended ballast water release zones, countries that ratified the International Convention for the Control and Management of Ships' Ballast Water and Sediments (BWM), could meanwhile already make ballast water treatments compulsory. Infrastructures for ballast water treatment are technologically improving (Aravossis and Pavlopoulou 2013, Chaplin 2019), although this approach brings additional practical and financial issues being the responsibility of ship owners (Aravossis and Pavlopoulou 2013).

5. CONCLUSIONS

This study provides insights on how ballast water release can contribute to the arrival of potentially invasive species in current and proposed MPAs of the WAP, being one of the most vulnerable Antarctic regions to biological invasions.

The existing ballast water release guidelines produced by the IMO and Antarctic Treaty (Antarctic Treaty 1959, ATCM 2006) are not sufficient to prevent the introduction of non-native species in these MPAs, although respecting ballast water discharges at 200 NM away from the nearest coasts lowers the risk of introduction. This is especially true for ballast water being discharged in the areas of the western and eastern WAP. Because of the expected future increase in maritime traffic and the correlated risk of alien species introduction and invasions potentially increasing due to global warming, we here advocate for delineating ballast water discharge zones, so that propagules released within ballast waters would not reach the most fragile Antarctic ecosystems. These discharge zones could be further fine-tuned with more data about maritime traffic and accounting for climatic variability.

We also recommend increasing the ratified distance of ballast water release over 200 NM in the WAP and avoiding discharges in the Eastern Antarctic Peninsula, two recommendations that could be included in future MPA proposals. This study shows that ballast water release at 50 NM or closer to the coasts pose a dangerous threat, as these results in drifting propagules reaching Antarctic coasts. This is in particular exemplified by the case study of the introduction of the Patagonian crab *H. planatus* in Deception Island. Our results indicate that, if the crab was indeed brought after ballast water discharge, the ballast water would have likely been discharged at 50 NM or closer to the Antarctic coast.

APPENDIX 4.1. Main currents along the Western Antarctic Peninsula region

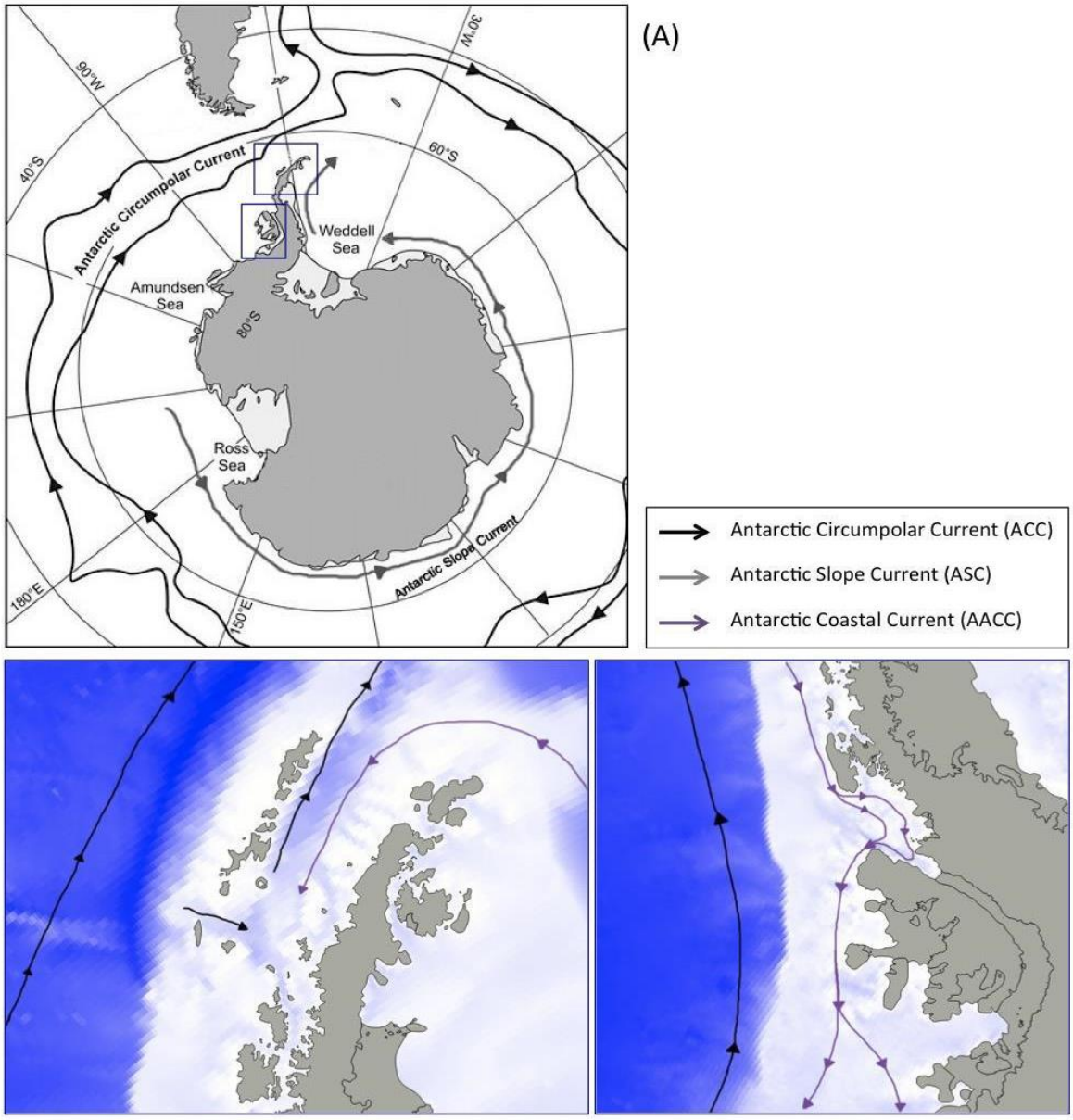


Figure S4.1. (A) Main currents in the Southern Ocean region and focus on the Western Antarctic Peninsula. Modified from Caccavo et al. (2018) and Moffat and Meredith (2018).

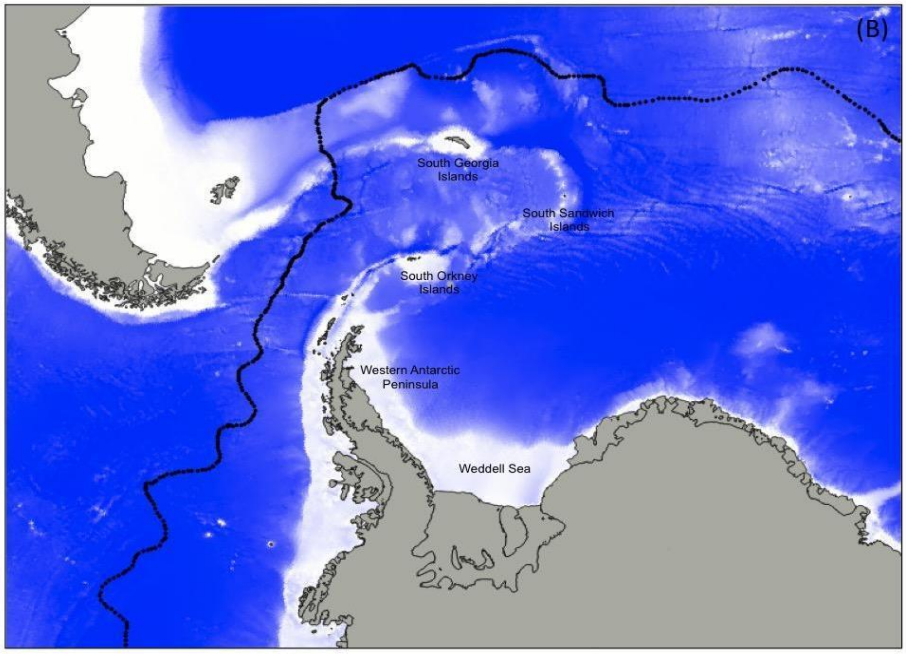


Figure S4.1. (B) Position of the Polar Front (black dotted line), from Orsi and Harris (2019).

APPENDIX 4.2. Seasonal variability

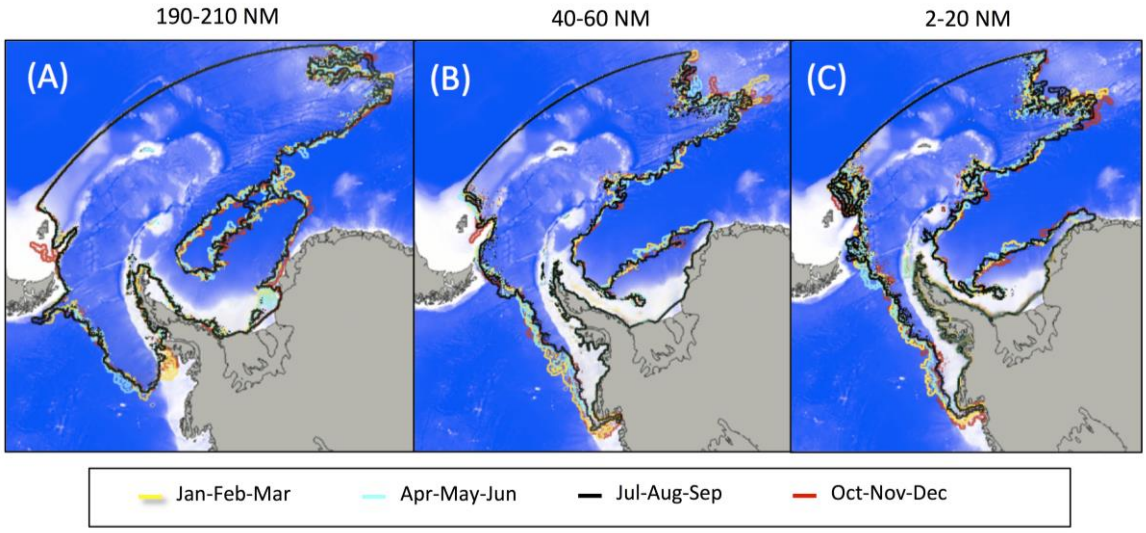


Figure S4.2. Dispersal patterns according to the different scenarios of ballast water release (A : 200 NM, B : 50 NM, C : 11 NM) for contrasting seasons. Release in all areas at the same time. Average values of the 9-year simulations (2008-2016). Blue background: bathymetric chart.

APPENDIX 4.3. Inter-annual variability of dispersal patterns (focus on the Gerlache Strait region)

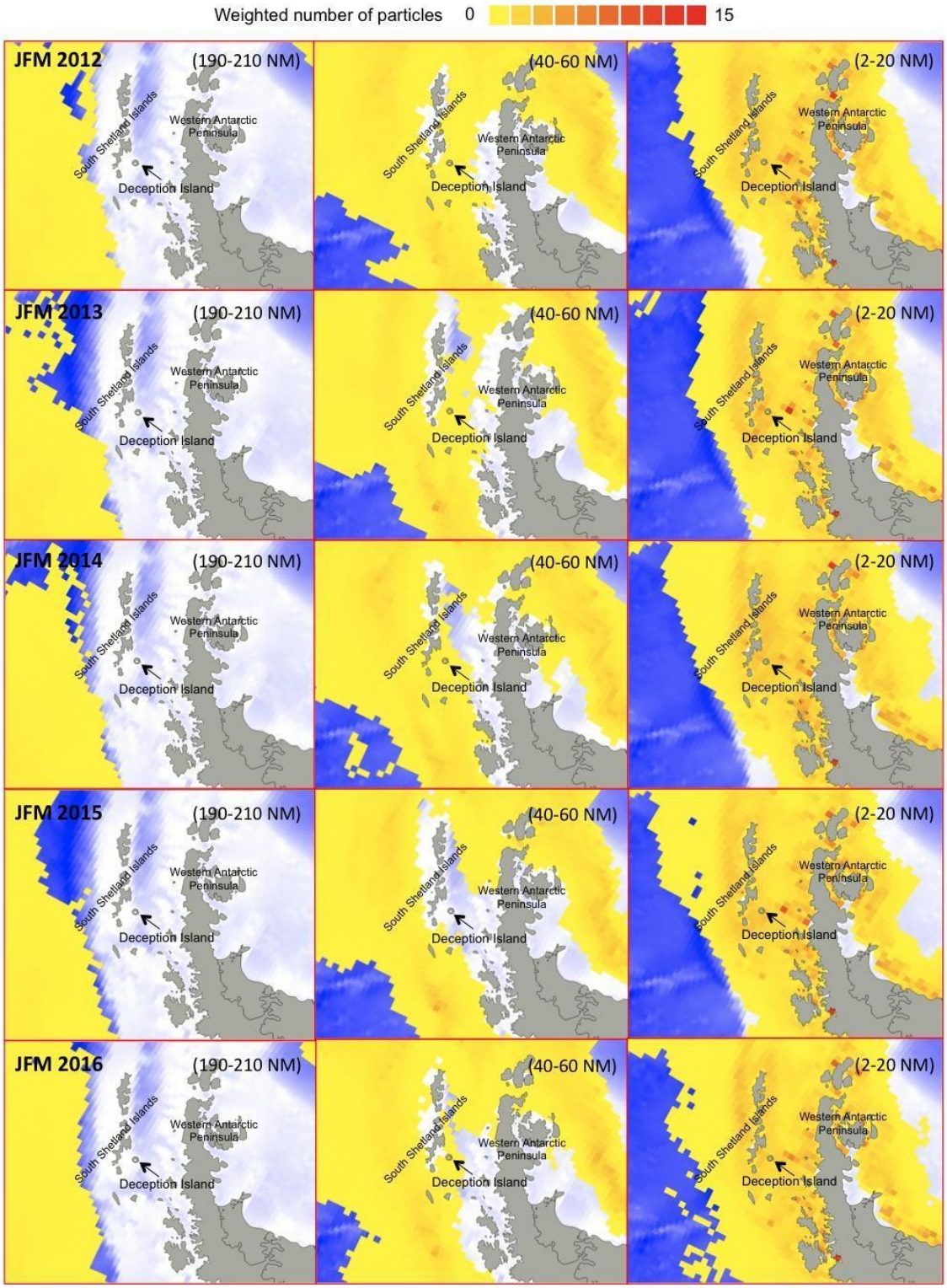


Figure S4.3. Dispersal patterns and weighted number of particles according to the different scenarios of ballast water release (200 NM, 50 NM and 11 NM) for the January-February-March period (summer) and for different years (2012 to 2016). Release in all areas at the same time. Blue background: bathymetric chart.

Integrated assessment of large-scale connectivity in a historically overexploited fish population in the Southern Ocean

Christiansen Henrik[1], Van de Putte Anton[1,2], Guillaumot Charlène[3,4], Barrera-Oro Esteban[5], Volckaert Filip A.[2], Young Emma F.[6]

[1] KU Leuven, Laboratory of Biodiversity and Evolutionary Genomics (LBEG), Charles Deberiotstraat 32, B-3000 Leuven, Belgium.
[2] Royal Belgian Institute of Natural Sciences, OD Nature, Vautiersstraat 29, B-1000 Brussels, Belgium.
[3] Laboratoire de Biologie marine CP160/15, Université Libre de Bruxelles, Bruxelles, Belgium.
[4] Biogéosciences, UMR 6282 CNRS, Université Bourgogne Franche-Comté, Dijon, France.
[5] Instituto Antártico Argentino, 25 de Mayo 1143, San Martín, Provincia de Buenos Aires, Argentina.
[6] British Antarctic Survey, High Cross, Madingley Rd, Cambridge, CB3 0ET, UK.

Draft in preparation

Abstract

Natural populations are often heterogeneously distributed in space, potentially leading to spatial genetic structure and patterns of local adaptation. Ecological traits, the environment, and evolutionary forces determine the connectivity between population patches. In the Southern Ocean, demersal fish disperse widely between the continental shelf, and oceanic island plateaus and seamounts. The marbled rockcod *Notothenia rossii* Richardson, 1844 was historically overharvested and only recently shows signs of recovery. We applied an integrated multidisciplinary approach to determine connectivity of this ecologically important species over contemporary and evolutionary time scales. Thousands of population genomic markers reveal high levels of gene flow and a lack of genetic differentiation over vast distances. Individual-based modelling, however, suggests that large-scale connectivity can only be achieved via stepping-stone transport. In conjunction with species distribution modelling, these results highlight how genetic data alone may overestimate the extent of connectivity. Limited ecological connectivity, that is reduced exchange of larvae or juveniles within one season, and reduced effective population size may have contributed to the long recovery time of the marbled rockcod. Current conservation plans, that aim to create a network of marine protected areas in the Southern Ocean, can benefit from multi-method assessments as presented here, especially in view of global change.

Keywords

Connectivity, fish, genotyping by sequencing, local adaptation, marine protected area, Notothenioidei, seascape genomics, Southern Ocean, species distribution modelling.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contribution of those colleagues providing samples for this study. In particular we thank the British Antarctic Survey, the crew and captain of RV James Clark Ross, the Alfred Wegener Institute, captain and crew of RV Polarstern, the National Natural History Museum in Paris, France and the help of G. Carvalho, G. Duhamel, M. Hautecoeur, E. Moreira, J. Rock, A. North, L. Zane, M. Belchier, N. Tysklind, and M. De Bruyn. Simulated currents from NEMO were provided by the National Oceanography Centre, Southampton. We thank F. Heindler, B. Hellemans, I. Coscia, A. McDevitt, S.R. Maier, F. Calboli and C. Moreau for laboratory help and/or advice. This work is contribution no. 38 to the vERSO project, funded by the Belgian Science Policy Office (BELSPO, Contract no. BR/132/A1/vERSO) and contribution no. 11 to the RECTO project (BELSPO, Contract no. BR/154/A1/RECTO).The modelling work was supported by Natural Environment Research Council (NERC, UK) grant NE/H0230238/1. HC was supported by an individual PhD grant from Flanders Innovation & Entrepreneurship (VLAIO, grant no. 141328).

DATA ARCHIVING STATEMENT

Demultiplexed, but otherwise raw sequencing data has been deposited on NCBI's Sequence Read Archive (SRA). Metadata for each individual using the same identification codes is available and cross-linked at GeOMe (Deck et al. 2017) and data.biodiversity.aq. In addition, vcf and genepop files, R scripts and additional input files for analysis are available at https://doi.org/10.5281/zenodo.3552609.

1. INTRODUCTION

Connectivity and spatial genetic structure in marine organisms is determined through the interplay of ecological traits, such as dispersal mode, duration and behavior, and the physical setting, that is environmental conditions including hydrodynamics, and evolutionary forces, such as selection and genetic drift (Hidalgo et al. 2017, Hoey and Pinsky 2018, Xuereb et al. 2018). This multitude of factors can lead to complex patterns and makes the relative importance of each factor difficult to discriminate (Moon et al. 2017, Miller et al. 2018, Milligan et al. 2018). However, spatial population structure and its temporal dynamics are crucial information for sound biodiversity management and protection (Funk et al. 2012, Momigliano et al. 2019). Management of marine organisms aims to protect the biodiversity of species, populations and ecosystems under competing influences of various anthropogenic disturbances (Everson 2017, Ropert-Coudert et al. 2019).

Taxonomic ranks at the species level are readily available, at least for macro-organisms, but recent research has shown that it is imperative to also consider intraspecific variation (Mee et al. 2015, Carvalho et al. 2017, Des Roches et al. 2018, Paz-Vinas et al. 2018). In order to assess intraspecific variation in the ocean, putative subpopulations are characterised with respect to their ecology and evolution. These two aspects are intertwined, leading to eco-evolutionary dynamics that determine the long-term fate of a species. Similarly, a species may persist in the form of a metapopulation comprised of subpopulations that are linked through ecological and evolutionary connectivity (Cowen and Sponaugle 2009, Pinsky et al. 2017). Ecological connectivity is the contemporary exchange of individuals between fragmented habitats. Evolutionary connectivity additionally considers the long-term degree of connection between a given number of separate (sub)populations through genetic exchange and drift (Waples and Gaggiotti 2006). All these aspects gain relevance in an exploitation context. Fisheries in particular can have drastic and immediate effects on reproductive output and thus ecological connectivity. In addition, fisheries can reduce genetic diversity and may have evolutionary consequences by imposing artificial selection on a species (Pinsky and Palumbi 2014, Heino et al. 2015). Recognizing and mitigating these consequences is a global challenge for fisheries and ocean management, especially in areas beyond national jurisdiction (Ortuño Crespo et al. 2019).

The Southern Ocean provides an example of ecosystem- and consensus-based fisheries management in an area that is not governed by a single nation (Kock et al. 2007, Constable 2011, Everson 2017). As such the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR), can be considered progressive (Constable et al. 2000, Nilsson et al. 2016, Hofman 2019), albeit constantly facing challenges (Ainley and Pauly 2013, Brooks 2013, Abrams et al. 2016). Before CCAMLR came into force in 1982, a number of fish populations were severely overfished. Most strikingly, the endemic marbled rockcod *Notothenia rossii* Richardson, 1844 (Nototheniidae, Perciformes) was an early target species with a cumulative reported catch of 501,262 t in the first two fishing seasons (1969/70 and 70/71) around South Georgia (Kock 1992) and near 150,000 t around Kerguelen in 1971 (Duhamel 1982). Thereafter, large trawlers were still active throughout the Southern Ocean, but with considerably lower catches, until the fishery was closed by CCAMLR in 1986/87 (Kock 1992). Since the inception of CCAMLR, conservation measures have been adopted progressively in order to assist the recovery of several notothenioid species by banning directed fisheries and establishing stringent by-catch limits in many Antarctic zones (CCAMLR 2019a). Recovery of the *N. rossii* stocks took more than 35 years with the species only in the past decade showing clear signs of increasing abundance (Barrera-Oro and Marschoff 2007, Marschoff et al. 2012, Barrera-Oro et al. 2017, Duhamel et al. 2019).

At present, new initiatives are underway to enhance biodiversity protection in the Southern Ocean through a network of Marine Protected Areas (MPAs). The South Orkney Islands Southern Shelf (SOISS) and large parts of the Ross Sea are designated MPAs since 2009 and 2016, respectively (Fig. 4.11). The waters around South Georgia and the South Sandwich Islands are also widely protected against overexploitation (Trathan et al. 2014). Additional MPAs were proposed in the Atlantic, Indian and Pacific sectors of the Southern Ocean to create a MPA network (CCAMLR

2019b). Much research has been devoted to provide the scientific information needed for the establishment of an appropriate MPA network (Teschke et al. 2015, 2019, Constable et al. 2016, Hill et al. 2017, Brasier et al. 2019, Parker et al. 2019). In addition, several large-scale research initiatives have increased our knowledge of biodiversity and bioregions in the Southern Ocean over the past decades (Schiaparelli et al. 2013, De Broyer et al. 2014). Nevertheless, Antarctica and the Southern Ocean remain a data-poor environment in the global context. Data availability is biased, due to high sampling effort around research bases (Griffiths 2010) and difficult accessibility of the remote and often ice-covered Southern Ocean (Convey and Peck 2019). A combined methodological approach, including genetic analyses, environmental measurements and modelling techniques, could help eliminate knowledge gaps concerning population structure and connectivity of key species in the Southern Ocean ecosystems (Gutt et al. 2018).

The genomic revolution has enabled the relatively fast and economic characterisation of thousands of genetic markers in non-model organisms (Elshire et al. 2011, Andrews et al. 2016). High resolution genetic data promises to yield new insights into speciation, differentiation and adaptation patterns and is thus a valuable tool to fully use the potential of precious Antarctic samples (Christiansen 2020). Species distribution modelling (SDM), sometimes referred to as ecological niche modelling, is a powerful technique to correlate environmental and occurrence data and subsequently predict the occurrence probability of a given species in other habitats or under changing environmental conditions (Elith and Leathwick 2009). Such techniques can be useful in data-limited situations, albeit care must be taken to ensure appropriate parameterization (Guillaumot et al. 2018a - Appendix, 2019 - Chapter 2). Lastly, individual-based modelling (IBM) can be used to simulate dispersal and thus obtain a spatially explicit prediction of connectivity between habitats (Cowen and Sponaugle 2009). Combining genomic data and modelled connectivity estimates is most useful in the marine realm, where direct observations of migration and dispersal are often virtually impossible (Pinsky et al. 2017, Xuereb et al. 2018). We use the methods mentioned above complimentarily to advance our understanding of large-scale connectivity in *N. rossii*, a fish that is both valuable as a living resource and vulnerable to overfishing.

The marbled rockcod grows to more than 50 cm in length, can form dense shoals in sub-Antarctic and Antarctic fjords and shelf waters, and occurs widely in the Southern Ocean (DeWitt et al. 1990, Duhamel et al. 2014). Its life cycle has been well described for the population at South Georgia by Olsen (1954) and Burchett (1983) and for the population at Kerguelen Islands by Duhamel (1982). Spawning takes place between April and June on the bottom of continental shelf areas at about 200-360 m depth, where ripe adults migrate during fall. Hatching occurs between September and October in the water column, where larval and young pelagic blue-phase fingerling stages remain before they migrate inshore approximately in January-February (DeWitt et al. 1990, Kock and Kellermann 1991, Kock and Jones 2000, North 2001). The fingerlings then change morphologically to the brown-phase fingerling stage and become demersal, settling in the algae beds. At about 5-7 years of age and 41-45 cm of length, *N. rossii* reaches sexual maturity and migrates to the offshore shelf feeding area joining the adult population. These offshore-inshore phases in the life cycle of the marbled rockcod are assumed to be similar in the geographical areas of its range (Kock 1992). The ecological habits of *N. rossii* as a benthic-benthopelagic species constitute an important trophic link between lower trophic levels (macroalgae, benthic invertebrates, small fish) and Antarctic top predators, such as seals and birds (Barrera-Oro 2002, McInnes et al. 2017, Bertolin and Casaux 2018).

Previous studies have reasoned that this extended pelagic period contributes to the widespread distribution and low or absent genetic structure of *N. rossii* (DeWitt et al. 1990, Young et al. 2015). The Antarctic Circumpolar Current (ACC), the world's largest ocean current system, is a prime candidate to facilitate eastward advection and thus connectivity (Orsi et al. 1995, Matschiner et al. 2009). The ACC is comprised of a series of approximately zonal fronts, where there are rapid changes in water mass properties and associated geostrophic currents. These current jets are not fixed in time and space, rather they show a high degree of meso-scale variability, with frequent

splitting and merging. However, long-term trends in frontal positions show very little variability (Chapman 2017), in part due to the strong steering of the ACC by the seafloor topography. Broadly speaking, the ACC has a pervasive eastward flow. However, within the Scotia Sea this flow assumes a more northward component as the current system recovers from its most southerly excursion in Drake Passage. Population genetic investigations of *N. rossii* have focused on this area, using microsatellite loci and connectivity models (Young et al. 2015). However, the connection between *N. rossii* in the Scotia Sea and the Kerguelen Plateau, initially described as two subspecies (DeWitt et al. 1990), has not been investigated with modern genomic methods or numerical modelling.

Here, we hypothesize that despite the fragmented distribution only subtle genetic structure and adaptive divergence are present in *N. rossii* due to the long pelagic phase. Species distribution modelling is used to determine which localities are likely important habitats for *N. rossii*, while genome-wide polymorphisms are used to test for genetic population structure. Individual-based modelling is employed to quantify dispersal between sites. Combining population genomics with distribution and dispersal modelling enables us to infer gene flow and environmental specialization in a spatially explicit framework. Finally, we evaluate large-scale distribution and spatial connectivity patterns in the Southern Ocean in light of current fisheries management and conservation actions.

2. MATERIAL AND METHODS

2.1. Species occurrence and sampling

Publicly available occurrence data for marbled rockcod *N. rossii* were mined from the Ocean Biogeographic Information System (OBIS) and the Global Biodiversity Information Facility (GBIF) (both accessed September 2019) using R softwage with the packages 'robis' v2.1.8 (Provoost and Bosch 2019), 'rgbif' v1.3.0 (Chamberlain et al. 2019) and 'SOmap' for plotting (Maschette et al. 2019). Duplicate entries (identical coordinates) were removed. A total of six occurrences that appear likely to be misidentified (based on these occurrences being drastically outside the generally accepted species distribution) were removed, in an attempt to provide exclusively highly reliable input data for species distribution modelling. For genetic analyses, adult fish were caught during many expeditions throughout the Southern Ocean (Fig. 4.11, Table 4.1). Fin, muscle, or liver biopsies were taken and stored in 90% ethanol or frozen until further processing. The samples from South Georgia, the South Orkney Islands, Elephant Island, and King George Island/Isla 25 de Mayo (collected in 2006) were previously analyzed using microsatellites (Young et al. 2015). In addition, samples from trammel nets were taken in 2016 on King George Island/Isla 25 de Mayo (Barrera-Oro et al. 2019) and from research trawling in 2016 on Skiff Bank (Leclaire Rise) and on the Northeast part of the Kerguelen Islands shelf. The occurrence data from these samples were added to the OBIS/GBIF occurrence data set for species distribution modelling if not already included. All available metadata per sampled individual can be found on data.biodiversity.aq.

Table 4.1. Sampling details (location, location code, latitude (Lat) and longitude (Lon), sample size (N) and year) and genetic diversity of *Notothenia rossii* from the Southern Ocean. Geographical coordinates are listed in decimal degrees; note that values are approximate for most locality samples; all available metadata per individual can be found on data.biodiversity.aq. Expected (H_E) and observed heterozygosity (H_O) were calculated for filtered genotypes from *de novo* and reference-based bioinformatics.

Location	Code	Lat	Lon	N	Year	H_O *de novo*	H_O reference	H_E *de novo*	H_E reference
South Shetlands, Deception Island	SSD-06	-62.95	-60.65	34	2006	0.246 ± 0.004	0.212 ± 0.007	0.253 ± 0.005	0.219 ± 0.006
South Shetlands, King George Island	SSK-06	-62.23	-58.68	35	2006	0.230 ± 0.004	0.217 ± 0.006	0.235 ± 0.004	0.227 ± 0.006
South Shetlands, King George Island	SSK-15-16	-62.23	-58.68	40	2015/16	0.244 ± 0.004	0.213 ± 0.006	0.249 ± 0.004	0.223 ± 0.006
Elephant Island	EI-02	-61.24	-55.62	33	2002	0.255 ± 0.004	0.215 ± 0.006	0.261 ± 0.004	0.226 ± 0.006
Elephant Island	EI-06-07	-61.24	-55.62	31	2006/7	0.253 ± 0.005	0.211 ± 0.006	0.260 ± 0.005	0.219 ± 0.006
South Orkney Islands	SO-06	-60.70	-45.57	22	2006	0.251 ± 0.005	0.224 ± 0.008	0.258 ± 0.005	0.236 ± 0.008
South Georgia	SG-02-03	-55.24; -53.70	-35.6; -37.51	35	2002/3	0.247 ± 0.004	0.216 ± 0.006	0.252 ± 0.004	0.224 ± 0.006
South Georgia	SG-05	-53.70	-37.51	45	2005	0.244 ± 0.004	0.210 ± 0.005	0.250 ± 0.004	0.220 ± 0.005
Kerguelen Islands Shelf	KI-15	-47.41; -48.67	69.7; 70.98	39	2016	0.238 ± 0.004	0.204 ± 0.007	0.243 ± 0.004	0.213 ± 0.007
Skiff Bank, Kerguelen Islands	SB-15	-49.8; -50.01	64.8; 65.64	40	2016	0.236 ± 0.004	0.217 ± 0.006	0.242 ± 0.004	0.228 ± 0.006

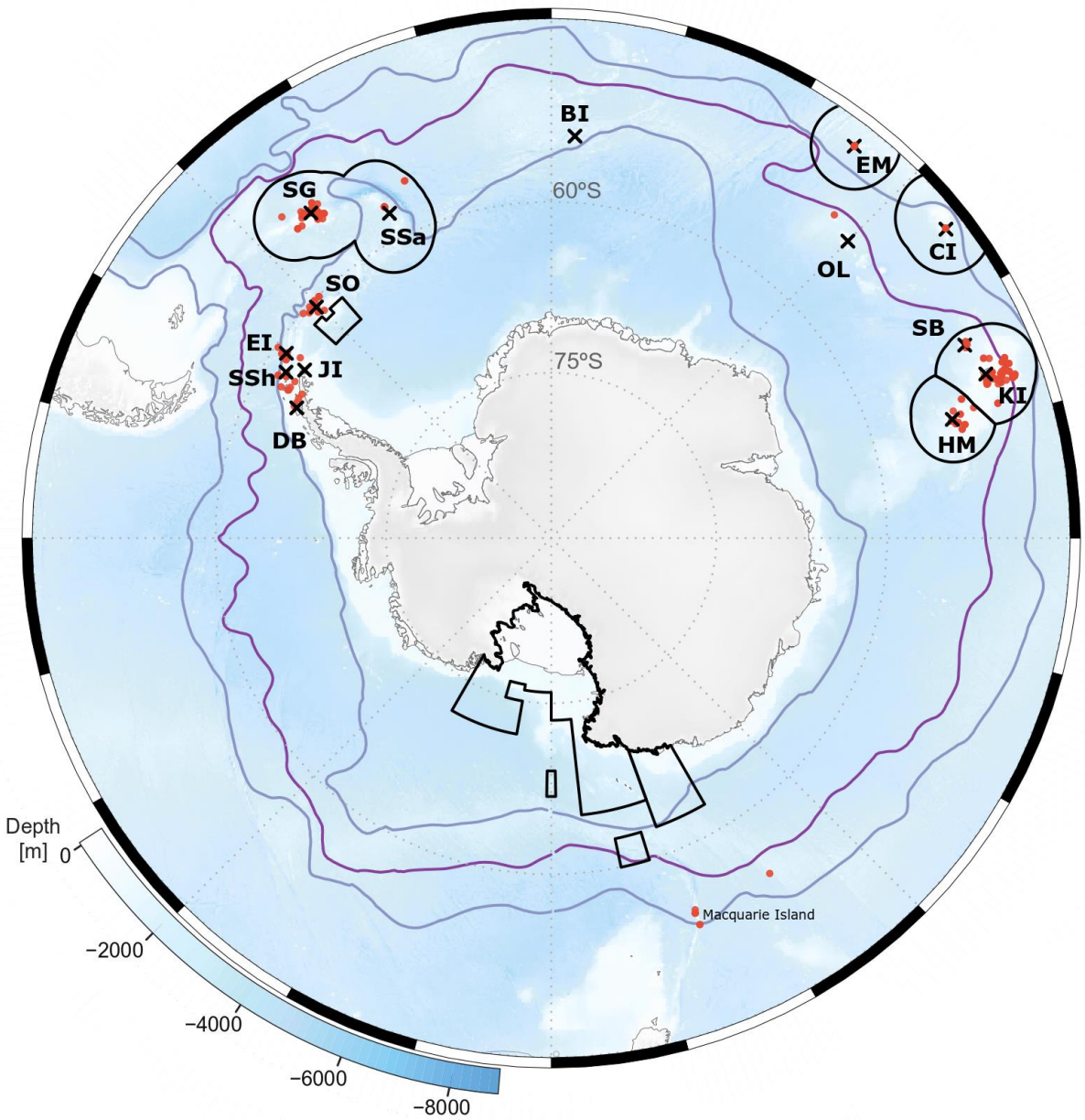


Fig. 4.11. Species occurrence of *Notothenia rossii* in the Southern Ocean (red dots) and localities used for individual-based hydrodynamic connectivity modelling (black crosses). Ocean fronts after Orsi et al. (1995) indicated from north to south: sub-Antarctic Front, Polar Front, Southern Boundary of the Antarctic Circumpolar Current. Current marine protected areas (MPAs), i.e. the South Orkney Islands Southern Shelf and Ross Sea MPAs, are shown as black rectangles and exclusive economic zones (EEZs) as black circles. Background shading (white-blue) reflects ocean depth. Modelling sites from west to east: Dallman Bay (DB), South Shetland Islands (SSh), Joinville Island (JI), Elephant Island (EI), South Orkney Islands (SO), South Georgia (SG), South Sandwich Islands (SSa), Bouvet Island (BI), Edward and Marion Islands (EM), Ob and Lena Banks (OL), Crozet Island (CI), Skiff Bank (SB), Kerguelen Islands (KI), Heard and McDonald Islands (HM). Samples for genetics were available from SSh, EI, SO, SG, SB, and KI (see Table 4.1).

2.2. Species distribution modelling

The assembled occurrence data (Fig. 4.11) and a subset of environmental variables describing the habitat of benthic Antarctic organisms at 0.1° resolution (Guillaumot et al. 2018b) were used for predictive species distribution modelling (SDM) with boosted regression trees (BRT, Elith et al. 2008). Twelve environmental variables were included after selecting the most informative variables based on biological knowledge, and the Variation Inflation Factor (VIF) stepwise procedure was

used to prune highly correlated descriptors (Naimi et al. 2014). Retained variables were: depth, geomorphology, sediments, slope, seafloor current speed, maximum ice cover, maximum ice thickness, mixed layer depth, minimum and maximum particulate organic carbon (POC), and minimum seafloor salinity and temperature. POC, temperature and salinity are an average of interpolated values of the time period 2005-2012. Models were run using only unique occurrences per 0.1° grid-cell and environmental layers were masked to areas shallower than 1000 m depth, which encompasses the observed depth range of adult *N. rossii* (the upper 550 m of the water column, DeWitt et al. 1990). The models were spatially limited to -100 and 100° longitude and -45 and -70° latitude. This longitudinal selection excludes most of the Pacific, and parts of the Indian Southern Ocean sectors. *Notothenia rossii* is documented to occur around Macquarie Island (Fig. 4.11). However, no samples from this comparatively isolated site (only known occurrence in the Pacific sector) were available. Hence, modelling focused on the area from the western Antarctic Peninsula to the Kerguelen Plateau. The latitudinal selection covers all assembled occurrences and therefore the area in which the species can likely biologically occur. Optimal BRT parameters, which in combination reduce modelling error while avoiding overfitting to the occurrence data set, were calibrated following Elith et al. (2008). The selected combination used for the final SDMs were: tree complexity 4, bag fraction 0.8 and learning rate 0.02. The number of background data used to characterise the environmental conditions was set at 500 and a four-fold 'CLOCK' method was applied to spatially segregate the proportion of occurrence records used to train the model (75%) and test the model (25%) (Guillaumot et al. 2019 - Chapter 2). In addition, a kernel density estimation layer (Phillips et al. 2009) and a multivariate environmental similarity surface index (Elith et al. 2010) were estimated and applied as described in detail in Guillaumot et al. (2019 - Chapter 2). These corrections were applied to correct for the influence of autocorrelation within occurrence records and model extrapolation, respectively. A final total of 240 replicate SDMs were run and the mean probabilities of occurrence were used for plotting. Model performance was assessed by measuring the Area Under the receiver operating Curve (AUC, Fielding and Bell 1997), and assessing the number of presence test data correctly classified as suitable areas by the model predictions. Analyses were conducted in R using the packages 'ncdf4' v1.16.1 (Pierce 2019), 'raster' v3.0-2 (Hijmans et al. 2019), 'usdm' v1.1-18 (Naimi and Araújo 2016), 'dismo' v1.1-4 (Hijmans et al. 2017), 'MASS' v7.3-51.4 (Venables and Ripley 2002), and 'gbm' v2.15 (Greenwell et al. 2019). Input data and R scripts are available at https://doi.org/10.5281/zenodo.3552609.

2.3. Reduced representation sequencing

Large numbers of single nucleotide polymorphism (SNP) loci were sourced with reduced representation sequencing, a methodology that reproducibly samples the full genome (Andrews et al. 2016). Genomic DNA was extracted using a standard salting out protocol to avoid shearing during column purification. DNA concentration was determined using the Quant-iT PicoGreen dsDNA kit (Thermo Fisher Scientific Inc.) and an Infinite M200 microplate reader (Tecan Group Ltd.) according to the manufacturer's instructions. Extractions were then standardized, checked for signs of degradation on agarose gels and quantified and standardized again to approximately 10 ng μL^{-1}. Four reduced representation sequencing libraries containing 96 individuals each were constructed (Table 4.1). Thirty of these were within- and between-library controls (i.e. DNA replicates from identical individuals). A modified GBS library preparation protocol based on Elshire et al. (2011) but with size selection was used, as described in full detail in Christiansen (2020). The restriction enzyme *ApeKI* and a size selection window of 240-340 bp, using a Pippin Prep (Sage Science), were applied to achieve high marker density. The libraries were paired-end sequenced on four lanes of a HiSeq 2500 with v4 chemistry (Illumina Inc.) at the KU Leuven Center for Human Genetics (GenomicsCore).

2.4. Variant calling and filtering

Sequencing data was checked for general quality using FastQC v0.11.5 (Andrews 2010). The Stacks pipeline v2.4 (Rochette et al. 2019) was used to genotype samples both *de novo* and using

the reference genome of *Notothenia coriiceps* (Shin et al. 2014). First, each library was demultiplexed and quality filtered using process_radtags (options: -c -q -r). Library 2 and 4 were demultiplexed with quality control disabled (without -q), because low Phred scores at the cutsite otherwise lead to discarding of all reverse reads. For the *de novo* approach the sequences were trimmed to 119 bp, to fulfill Stacks' requirement of uniform read length. In addition, forward, reverse and remainder reads were concatenated after demultiplexing to treat both reads as individual loci, because GBS reads cannot be oriented preventing Stacks from building paired-end contigs (Rochette et al. 2019). A parameter test series using a subset of 24 individuals (four from each locality) was conducted as described in Rochette and Catchen (2017). The retained parameters were m=3, M=4, n=4 (see Appendix 4.4) and subsequently applied to the entire data set. A catalog was built using 10 individuals per locality (cstacks module); all individuals were matched against the catalog (sstacks) and data was transposed (tsv2bam). For the reference approach, forward and reverse reads were aligned to the *N. coriiceps* genome after demultiplexing using BWA v0.7.17 and SAMtools v1.9 and default parameters (Li and Durbin 2009, Li et al. 2009). In both cases (reference-based and *de novo*) genotyping of SNPs was conducted using gstacks with default parameters, that is under a Bayesian low coverage framework from Maruki and Lynch (2017).

Stringent quality control and filtering is necessary before downstream processing of GBS data, because high-throughput sequencing data has comparably high error rates (Shafer et al. 2017, O'Leary et al. 2018). In a first filtering step, genotypes of the reference-based and *de novo* data set were pruned using the populations module of Stacks, requiring loci to be present in at least 80 % of the individuals of each population, to have a minor allele frequency > 0.05 and heterozygosity < 0.7 (Rochette and Catchen 2017). Subsequently, genepop files were imported and filtered extensively in R software using the 'radiator' package v1.1.1 (Gosselin 2019). This filtering approach was conducted without and with technical replicates to assess the genotyping error rate before and after filtering. In brief, data was filtered on missing values, heterozygosity, minor allele count, coverage, SNP position, linkage disequilibrium (LD) and departures from Hardy-Weinberg proportions (Appendix 4.5).

2.5. Population genomics

Analyses of the filtered genomic data sets were conducted mostly in R, with code and input data available under https://doi.org/10.5281/zenodo.3552609. Overall and pairwise differentiation measures (F_{ST}, G_{ST}, D), expected and observed heterozygosity and hierarchical analyses of molecular variance (AMOVA) were calculated using 'adegenet' v2.1.1 (Jombart 2008, Jombart and Ahmed 2011), 'hierfstat' v0.04-30 (Goudet and Jombart 2015), 'mmod' v1.3.3 (Winter 2012), and 'pegas' v0.11-12 R packages (Paradis 2010). Additional data filtering was conducted using 'poppr' v2.8.3 (Kamvar et al. 2014) (see Appendix 4.5). Principal component analysis (PCA), non-metric multidimensional scaling and discriminant analyses of principal components (DAPC) were conducted using 'adegenet', 'vegan' v2.5-6 (Oksanen et al. 2018), 'MASS' v7.3-51.4 (Venables and Ripley 2002), and 'factoextra' v1.0.5 (Kassambara and Mundt 2017) and 'ggsci' v2.9 (Xiao 2017) aiding plotting. Following cross-validation to avoid overfitting, 30 (*de novo* data) and 50 (reference data) principal components were used for the DAPC. Migration was estimated and visualized using the divMigrate function (Sundqvist et al. 2016) from the 'diveRsity' package v1.9.90 (Keenan et al. 2013) and the 'qgraph' package v1.6.3 (Epskamp et al. 2012).
In addition, the Bayesian clustering software Structure v2.3.4 (Pritchard et al. 2000) was called from within R using 'ParallelStructure' v1.0 (Besnier and Glover 2013) and a function from Clark (2017) to prepare input files. Structure was run on both data sets with K ranging from one to ten, with five replicates of each run and always using 10,000 repetitions as burn-in and 100,000 subsequent iterations. The likely number of genetic clusters was inferred using ΔK (Evanno et al. 2005) and Structure Harvester (Earl 2012). The contemporary effective population size (N_e) of each sampling location and year was estimated using the LD method (Waples and Do 2008) under a random mating model using a MAF cutoff of 0.05 with bias correction (Waples 2006) and

updates for dealing with missing data (Peel et al. 2013), as implemented in 'NeEstimator' v2.1 (Do et al. 2014). Finally, a simple genome scan for signs of selection was conducted using R packages 'pcadapt' v4.1.0 (Luu et al. 2016) and 'qvalue' v2.16.0 (Storey et al. 2019). Loci with $q < 0.05$ were retained as candidate loci and the contigs that contained these SNPs were matched against the nucleotide (nt) collection of the NCBI database using BLASTN 2.10.0+ (Altschul et al. 1997). Only top hits with an E-value ≤ 1×10^{-6} and at least 70 % similarity were retained (Benestan et al. 2017) and then further investigated using BLASTX 2.9.0+ and the UniProtKB vertebrate database for their putative function (Apweiler et al. 2004).

2.6. Individual-based connectivity modelling

Five-day mean flow fields for the Southern Ocean region from a state-of-the-art oceanographic modelling framework, Nucleus for European Modelling of the Ocean (NEMO), underpinned the numerical modelling simulations. Simulated flows for the period 1996–2001 were provided by the National Oceanography Centre, Southampton (UK), from a global application of NEMO with an eddy-permitting nominal horizontal resolution of 1/4°, and a partial step z-coordinate with 64 levels in the vertical. Full details of the ocean model are available at http://www.nemo-ocean.eu/About-NEMO. NEMO has been widely used over a range of spatial scales and resolutions and has been shown to provide a good representation of the dominant oceanography of the southern Atlantic Ocean region (Renner et al. 2009, 2012).
Mean flows from the circulation model were used to advect Lagrangian particles representing the early life stages of *N. rossii*. The Lagrangian model has been applied previously to the simulation of the dispersal of the eggs and larvae of *N. rossii* around South Georgia (Young et al. 2012) and in the Scotia Sea region (Young et al. 2015, 2018). In summary, particles were advected at each model time step (5 min) according to the imposed three-dimensional velocity field, using a second-order Runge-Kutta method. Additional horizontal and vertical diffusions were included using a random-walk approach (Dyke 2001), to account for unresolved turbulent motion in the ocean model flow fields. Particles representing the early life stages of *N. rossii* were released at known spawning populations (Fig. 4.11) (DeWitt et al. 1990, Barrera-Oro and Casaux 1992, Duhamel et al. 1995, Kock and Belchier 2004). Appropriate spawning areas at each location were identified by a comparison of local model depth with the known spawning depth range, 200-360 m (Kock and Belchier 2004). Modelled particles were released randomly within appropriate grid cells, with one thousand particles released per day at each site for the duration of the observed spawning periods. Dispersal of eggs was simulated for four months (Atlantic Ocean) or three months (Indian Ocean), with subsequent larval dispersal simulated for three months. There are no data to suggest that *N. rossii* larvae perform diel vertical migration and like all notothenioids the species does not possess a swimbladder; thus model eggs and larvae were allowed to move randomly within observed depth ranges: upper 100 m for eggs and upper 50 m for larvae (A. W. North, personal communication).

The potential for successful dispersal between isolated populations was assessed by comparing the position of each model larva with recruitment boxes encompassing each known population location (Fig. 4.11) over a 4-week period centered on the end of the planktonic phase. If a larva was within a recruitment box at any point during the 4-week period, it was considered to potentially recruit successfully to a nursery ground at this site. The percentage of larvae from each spawning site successfully reaching each recruitment box was calculated, and the results were combined into a single connectivity matrix describing the proportion of individuals arriving in a destination population (rows) from a given source population (columns). Such matrices describe potential connectivity; they do not include mortality or inter-annual variability in biological processes such as spawning and development rates. The effect of inter-annual variability in the underlying flow fields on predicted dispersal and connectivity was assessed by repeating the simulations for a five-year period (1996-2000). The results were combined to give a single mean connectivity matrix, and a matrix showing the number of years in which non-zero connectivity occurred (i.e. persistence).

3. RESULTS

3.1. Species distribution probability

Species distribution modelling results showed high modelling relevance with an AUC score of 0.975 and > 90 % of test data correctly classified (Table 4.2). Generally, model predictions are confined to areas of the Southern Ocean that are depth-wise potentially relevant for *N. rossii* ; large deep-sea areas are not included in the model here. The largest contribution to the modelling results was from mixed layer depth (51.6 ± 7.4%), with moderate contribution from maximum ice thickness (15.4 ± 3.7%), geomorphology (6.9 ± 4.1%), and depth (5.3 ± 2.5%), and little contribution from the remaining variables (< 5 % each). Areas with high occurrence probability overlap well with documented occurrence records. South Georgia and the Kerguelen Plateau are the largest areas of high occurrence in the part of the Southern Ocean that was evaluated (Fig. 4.12). In addition, suitable habitat for *N. rossii* is predicted in areas that are not documented in the occurrence data set: around Patagonia (low probability), around Bouvet Island and some seamounts north of that (high to medium probability), and west and east of Prydz Bay off the Antarctic continent (medium to low probability). Interestingly, the occurrence probability at the Ob and Lena banks is comparatively high. In contrast, predicted occurrence is low around the Edward and Marion Islands and Crozet Island.

Table 4.2. Model statistics describing the outcome of species distribution modelling to predict occurrence probability of *Notothenia rossii* in the Atlantic and Indian sectors of the Southern Ocean (mean ± standard deviation). AUC: Area Under the Receiver Operating Curve; COR: Point Biserial Correlation; TSS: True Skill Statistic; maxSSS: maximum sensitivity plus specificity threshold; Correctly classified test data (%): percentage of presence-test and background-test records falling on predicted suitable areas (prediction > maxSSS).

Model statistic	Mean and standard deviation
AUC	0.975 ± 0.016
COR	0.831 ± 0.069
TSS	0.768 ± 0.117
maxSSS	0.469 ± 0.230
Correctly classified test data (%)	92.3 ± 3.0 %

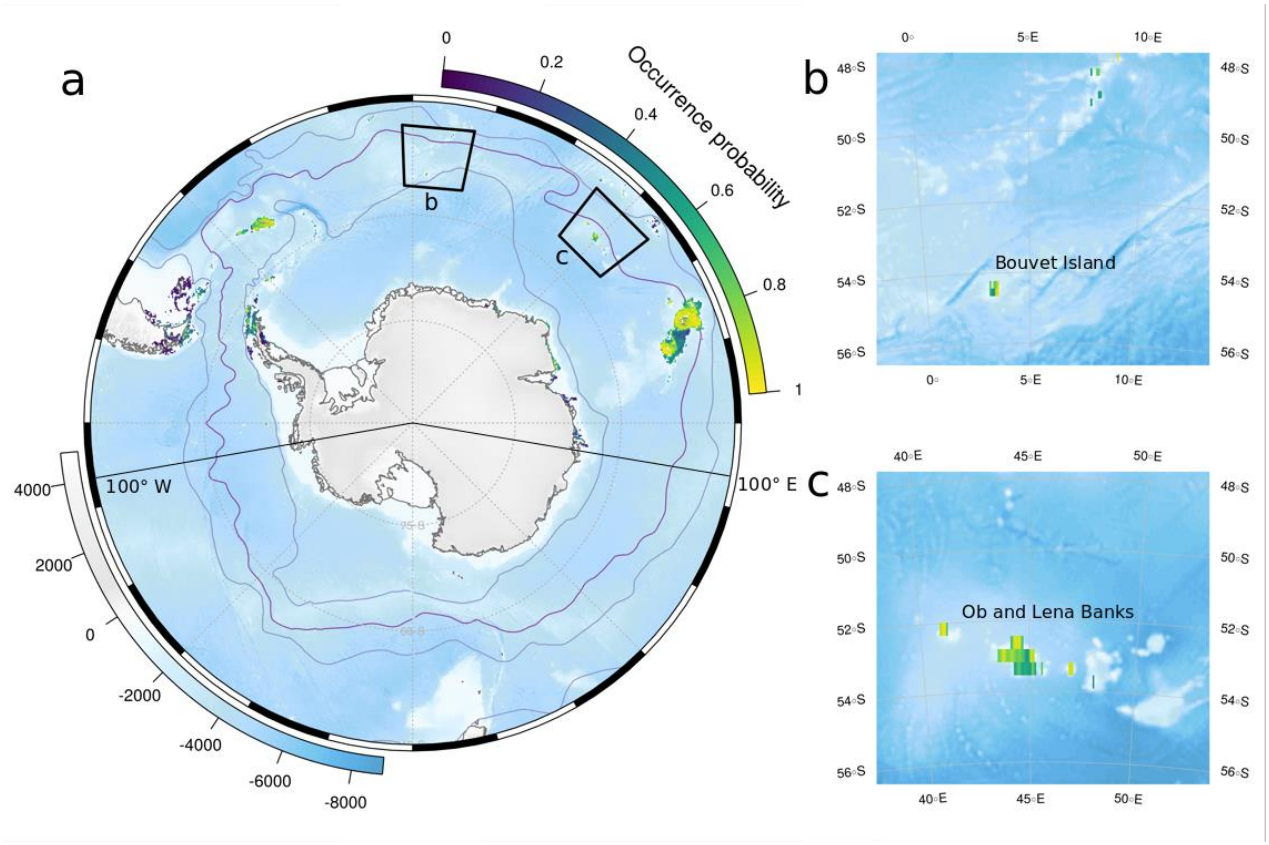


Figure 4.12. Predicted species occurrence probability for *Notothenia rossii* in the Atlantic and Indian sector of the Southern Ocean (prediction only between -100 and 100° longitude) based on mean prediction values from 240 model replicates using boosted regression trees (a). Insets show predictions around Bouvet Island (b) and the Ob and Lena Banks (c). Ocean fronts after Orsi et al. (1995) indicated from north to south: sub-Antarctic Front, Polar Front, southern Boundary of the Antarctic Circumpolar Current. Background shading (white-blue) reflects ocean depth. Only predictions in areas where the model does not extrapolate are shown.

3.2. Sequencing data

No sequencing problems were indicated by FastQC reports. On average each individual received 4.86 ± 2.29 million (M) reads. Four low coverage (< 1 M reads), as well as four high coverage (> 13 M reads) individuals were excluded before bioinformatics. After genotyping using Stacks, global coverage (as the average number of reads per locus per individuals) was at 10.64 ± 11.84. This coverage is at the minimum required for calling heterozygotes reliably (Rochette and Catchen 2017). Therefore, extensive downstream filtering of SNP data sets was conducted with relatively high thresholds (Appendix 4.5). After pre-filtering using the population module of Stacks, 73,554 and 85,980 SNPs were present in the reference-based and *de novo* based data sets, respectively. Global genotyping error rates of these data sets (calculated from technical replicates) were between 1.48 % and 5.30 % in the reference-based data and 0.61 % and 8.80 % in the *de novo* data. Such genotyping error rates are not ideal, but also not uncommon, and likely related to the comparably low coverage (Mastretta-Yanes et al. 2014, Fountain et al. 2016). We circumvented negative impacts as far as possible by applying very extensive downstream filtering in R (Appendix 4.5). In brief, individual genotypes with high amounts of missing data, abnormal heterozygosity patterns or signs of duplicate genomes were filtered. Loci were filtered based on missing data, minor allele count, minor allele frequency, coverage, linkage, SNP position, and departure form Hardy-Weinberg proportions. Even after these filtering steps, a bias related to sequencing library remained detectable in the data, as evidenced by PCA and AMOVA. Loci that contributed to this bias were also excluded. All these steps are described in detail in Appendix 4.5. Eventually, 272

individuals and 3,503 SNPs in the reference data set and 261 individuals and 7,501 SNPs in the *de novo* data set remained and were used for all subsequent analyses.

3.3. Genomic variability

While the number of SNPs is more than twice as high in the final filtered *de novo* data set compared to the reference-based data set, patterns of genetic diversity are largely congruent between the data sets. Observed heterozygosity was minimally lower than expected heterozygosity in both data sets (Table 4.1, Fig. 4.3a) and on average heterozygosity was slightly higher in the *de novo* data compared to the reference data (Table 4.1). Pairwise population estimates of differentiation were generally low with F_{ST} values (Weir and Cockerham 1984) reaching 0.0027 and 0.0018 in the *de novo* and reference-based data sets, respectively (Table 4.3 and 4.4). Few pairwise comparisons were significantly different from zero as indicated from confidence intervals (14 in the *de novo* data, six in the reference data). Similar results were obtained using alternative differentiation metrics (Appendix 4.6). The sample from the South Orkney Islands seems slightly differentiated with five (*de novo*) or three (reference) significant pairwise estimates of F_{ST}. This pattern is also discernible in the NMDS plots based on Hedrick's G_{ST} (Fig. 4.13c). Individual-based clustering approaches including PCA, DAPC (functions find.clusters and snapclust.choose.k) failed to identify meaningful genetic clusters (Fig. 4.13b and 54.14a). Results from STRUCTURE indicated the number of clusters based on maximum ΔK as six (*de novo* data) or five (reference data), but the log likelihood for K did not increase significantly suggesting an absence of genetic structure. Using DAPC with sampling location as priors offsets the population centroids of the South Orkneys, South Georgia and Skiff Bank slightly from the remaining locations (Fig. 4.14a). Genomic data suggests that the effective population size (N_e) of *N. rossii* is relatively large at most localities, with values ranging from 1,449 to 42,299, but often with upper limits of the 95 % confidence intervals as infinite (Table 4.5). The migration analysis revealed very high levels and no asymmetric pattern of gene flow, corresponding to the observed absence of distinct genetic structure (Fig. 4.14b). Genome scans for signs of selection identified 12 (*de novo*) and 37 (reference) candidate loci for further investigation (Fig 4.14c). Matching the flanking regions of these loci against the nucleotide database showed that some of them are highly similar to genomic DNA from other perciform fishes, such as *Chionodraco hamatus*, *Gymnodraco acuticeps*, *Notothenia coriiceps*, *Cottoperca gobio*, and *Dissostichus mawsoni* (Appendix 4.7). The DNA sequence of *D. mawsoni* is part of an antifreeze glycoprotein locus (Nicodemus-Johnson et al. 2011). Several contigs matched to predicted mRNA sequences of *N. coriiceps*, e.g. with putative functions in metal ion binding (Appendix 4.7). The PCA, DAPC, migration and outlier plots are only shown for the *de novo* data set (Fig. 4.13 and 4.14); the reference-based data yielded similar results (Appendix 4.8).

Table 4.3. Pairwise genetic differentiation of *Notothenia rossii* per sampling locality (see Table 4.1 for codes) based on 7,501 SNP genotypes derived from mapping against a *de novo* assembly. F_{ST} following Weir and Cockerham (1984) (also referred to as G_{ST}) below the diagonal (negative values set to zero) and confidence intervals after 1000 bootstraps above the diagonal. F_{ST} values where confidence intervals do not span zero are marked in bold.

	SSD-06	**SSK-06**	**SSK-15-16**	**EI-06-07**	**EI-02**	**SO-06**	**SG-02-03**	**SG-05**	**SB-15**	**KI-15**
SSD-06		-0.0002 – 0.0003	-0.0001 – 0.0003	-0.0002 – 0.0002	-0.0003 – 0.0001	0.0001 – 0.0006	0.0001 – 0.0005	-0.0001 – 0.0003	0.0000 – 0.0004	-0.0004 – 0.0001
SSK-06	0.0000		0.0000 – 0.0004	-0.0002 – 0.0002	-0.0002 – 0.0002	0.0001 – 0.0006	0.0001 – 0.0005	-0.0003 – 0.0000	-0.0001 – 0.0002	-0.0002 – 0.0003
SSK-15-16	0.0005	**0.0004**		-0.0005 – -0.0001	-0.0001 – 0.0002	-0.0002 – 0.0003	-0.0003 – 0.0001	-0.0002 – 0.0001	-0.0001 – 0.0003	-0.0002 – 0.0001
EI-06-07	0.0003	0.0007	0.0000		-0.0003 – 0.0001	0.0000 – 0.0004	-0.0001 – 0.0003	-0.0002 – 0.0002	-0.0003 – 0.0001	-0.0002 – 0.0002
EI-02	0.0000	0.0000	0.0000	0.0000		-0.0002 – 0.0002	0.0000 – 0.0004	-0.0002 – 0.0001	-0.0003 – 0.0001	0.0000 – 0.0004
SO-06	**0.0018**	**0.0011**	0.0002	**0.0012**	0.0003		0.0001 – 0.0005	-0.0004 – 0.0000	0.0002– 0.0006	-0.0001 – 0.0004
SG-02-03	**0.0005**	**0.0006**	0.0000	0.0000	**0.0000**	**0.0006**		-0.0001 – 0.0002	0.0000 – 0.0004	0.0000 – 0.0004
SG-05	0.0005	0.0000	0.0002	0.0007	0.0001	0.0000	0.0000		-0.0003 – 0.0000	0.0000 – 0.0003
SB-15	**0.0006**	0.0002	0.0004	0.0003	0.0000	**0.0007**	**0.0008**	0.0000		-0.0003 – 0.0001
KI-15	0.0000	0.0003	0.0000	0.0002	**0.0003**	0.0004	**0.0005**	**0.0001**	0.0000	

Table 4.4. Pairwise genetic differentiation of *Notothenia rossii* per sampling locality (see Table 4.1 for codes) based on 3,503 SNP genotypes derived from mapping against the reference genome of *N. coriiceps* (Shin et al. 2014). F_{ST} following Weir and Cockerham (1984) (also referred to as G_{ST}) below the diagonal (negative values set to zero) and confidence intervals after 1000 bootstraps above the diagonal. F_{ST} values where confidence intervals do not span zero are marked in bold.

	SSD-06	**SSK-06**	**SSK-15-16**	**EI-06-07**	**EI-02**	**SO-06**	**SG-02-03**	**SG-05**	**SB-15**	**KI-15**
SSD-06		-0.0002 – 0.0003	-0.0005 – 0.0001	-0.0007 – 0.0000	-0.0004 – 0.0001	-0.0003 – 0.0004	-0.0002 – 0.0004	-0.0003 – 0.0002	-0.0004 – 0.0002	-0.0007 – -0.0001
SSK-06	0.0006		0.0001 – 0.0007	-0.0002 – 0.0003	-0.0001 – 0.0004	0.0000 – 0.0006	0.0002 – 0.0008	-0.0001 – 0.0004	-0.0004 – 0.0001	-0.0001 – 0.0005
SSK-15-16	0.0000	**0.0017**		0.0000 – 0.0006	-0.0002 – 0.0002	-0.0005 – 0.0001	-0.0005 – 0.0000	-0.0002 – 0.0002	-0.0001 – 0.0004	-0.0001 – 0.0005
EI-06-07	0.0001	0.0013	**0.0011**		-0.0003 – 0.0003	0.0001 – 0.0007	-0.0003 – 0.0003	-0.0001 – 0.0005	-0.0003– 0.0002	-0.0005 – 0.0004
EI-02	0.0000	0.0007	0.0002	0.0002		-0.0002 – 0.0003	-0.0004 – 0.0001	-0.0003– 0.0001	-0.0003– 0.0002	-0.0001 – 0.0004
SO-06	0.0008	**0.0027**	0.0000	**0.0015**	0.0012		-0.0002 – 0.0005	-0.0001 – 0.0005	0.0001– 0.0007	-0.0003 – 0.0003
SG-02-03	0.0005	**0.0010**	0.0000	0.0005	0.0000	0.0009		-0.0002 – 0.0003	-0.0002 – 0.0003	-0.0002 – 0.0004
SG-05	0.0000	0.0003	0.0002	0.0004	0.0000	0.0007	0.0000		-0.0001 – 0.0004	-0.0002 – 0.0003
SB-15	0.0000	0.0001	0.0010	0.0000	0.0003	**0.0016**	0.0003	0.0002		-0.0001 – 0.0004
KI-15	0.0000	0.0005	0.0004	0.0000	0.0007	0.0005	0.0004	0.0001	0.0000	

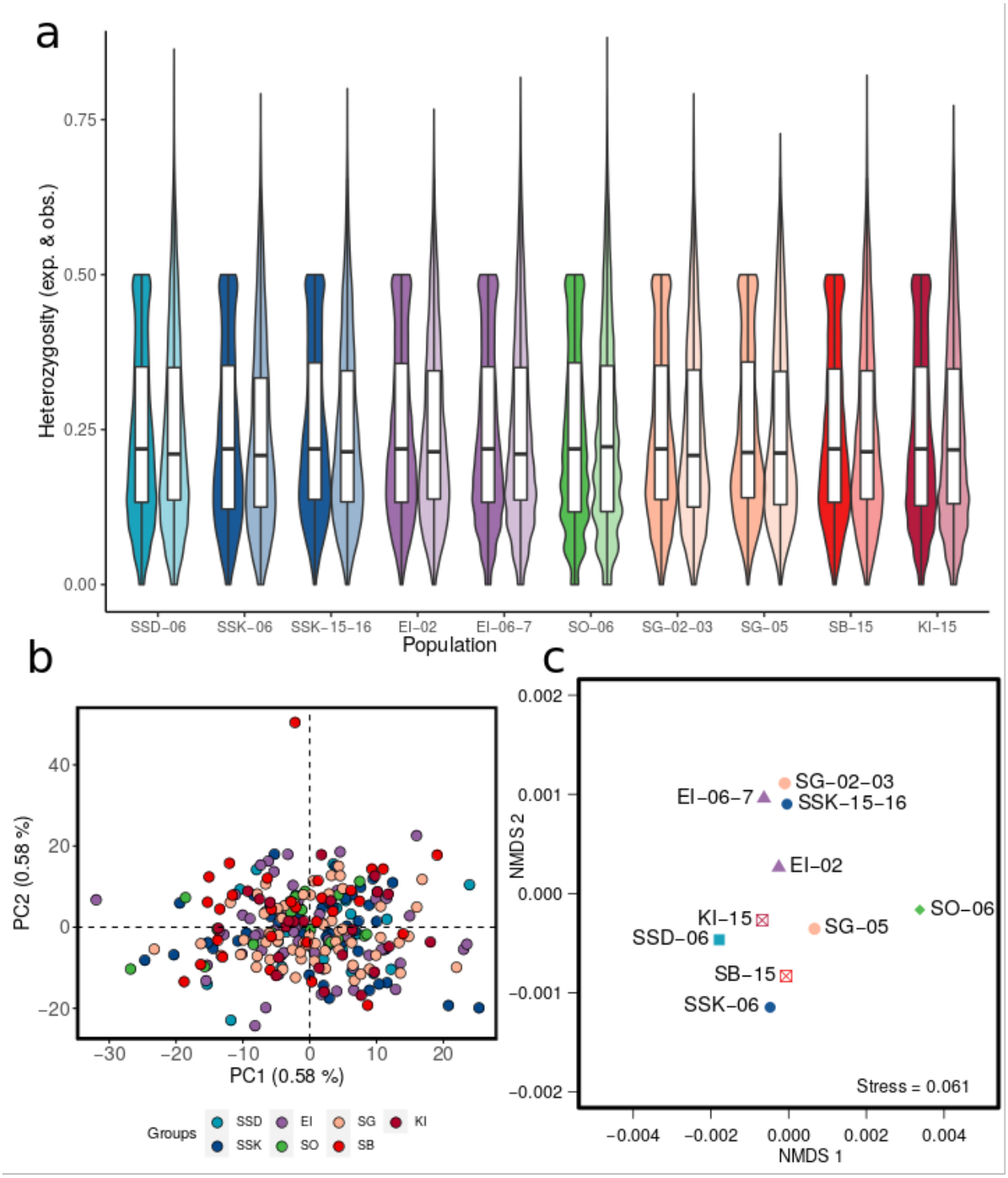


Figure 4.13. Genomic diversity of *Notothenia rossii* in the Southern Ocean based on 7,501 SNP loci. Expected (darker shading, left) and observed (light shading, right) heterozygosity is shown as box and violin plots for each genetically screened population (a). Principal component analysis (PCA) reveals little individual-based differentiation (b), while non-metric multidimensional scaling based on G_{ST} distances shows subtle differences between population samples (c). Sample codes as in Table 4.1; samples from different years but same locality are not shown separately on the PCA.

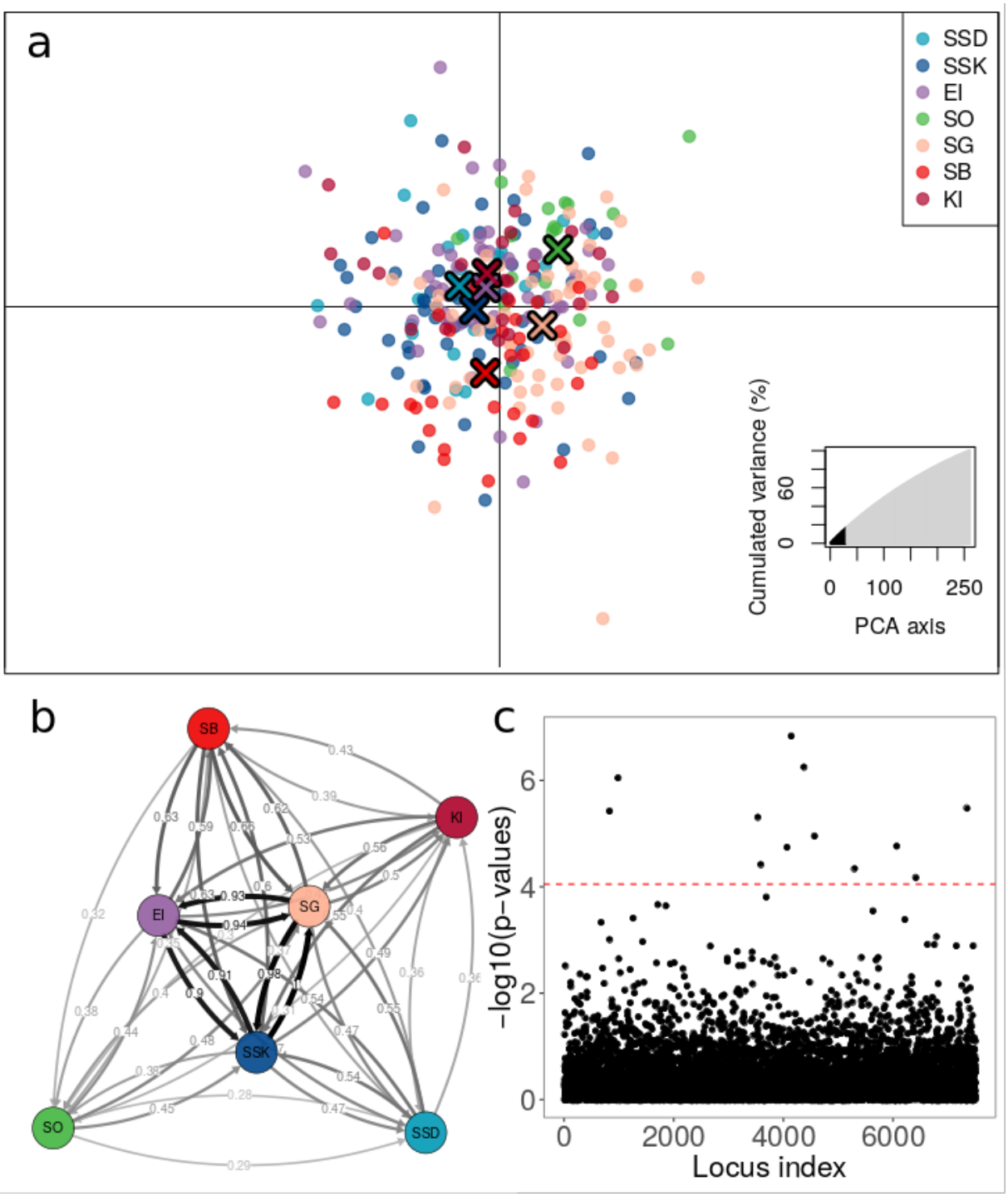


Figure 4.14. Genomic differentiation of *Notothenia rossii* in the Southern Ocean based on 7,501 SNP loci. Geographic clustering as attempted through discriminant analysis of principal components is shown along the first two principal components (a). Relative migration as estimated from Nei's G_{ST} reveals overall high and no asymmetric gene flow (b). Genome scans for loci putatively under influence of selection detected 12 outliers at q > 0.05 (c). Sample codes as in Table 4.1; samples from different years but same locality are combined.

Table 4.5. Effective population size (N_e) of *Notothenia rossii* from various locations in the Southern Ocean. Estimates were calculated using the linkage disequilibrium method for filtered genotypes from *de novo* and reference-based bioinformatics; with 95 % confidence intervals (CI) calculated based on the jackknife method of Waples and Do (2008).

Sample	N_e *de novo*	CI *de novo*	N_e reference	CI reference
SSD	2,089	197 – Infinite	2,677	319 – Infinite
SSK	6,207	1,312 - Infinite	6,846	1,302 - Infinite
EI	4,777	1,036 - Infinite	Infinite	1,422 - Infinite
SO	6,629	182 - Infinite	42,299	228 - Infinite
SG	4,837	1,327 - Infinite	21,601	1,864 – Infinite
SB	2,327	385 - Infinite	1,957	390 – Infinite
KI	1,665	256 - Infinite	1,449	206 - Infinite

3.4. Modelled connectivity

The predicted mean connectivity matrix suggests wide dispersal of *N. rossii* within the Scotia Sea, with high and persistent levels of connectivity between populations around the Antarctic Peninsula (AP) and South Georgia, and lower but persistent connectivity with the South Orkney and South Sandwich Islands (Fig 4.15). There is low but persistent connectivity from the South Sandwich Islands to Bouvet Island, and from Bouvet to populations in the Indian Ocean, in particular Edward and Marion Islands and Crozet Island. Although persistent connectivity pathways are predicted in the Indian Ocean, for example from Ob and Lena Banks to Crozet Island, Skiff Bank and Kerguelen Islands, the magnitude of connectivity is generally weaker than in the Scotia Sea, with the exception of the Kerguelen Plateau region. The patterns of connectivity suggest highly asymmetric dispersal, with a greater occurrence of non-zero values below the diagonal of the connectivity matrix, indicating unidirectional transport to the northeast across the Scotia Sea, and eastward towards and within the Indian Ocean in accordance with the dominant flows of the Antarctic Circumpolar Current. Bidirectional transport is predicted between proximate sites with complex local oceanography, in particular around the Antarctic Peninsula and the Kerguelen Plateau region. The pattern of connectivity suggests that gene flow includes an element of stepping-stone transport. *Notothenia rossii* is widely dispersed within the Scotia Sea, but there is no direct connectivity between sites in the Scotia Sea and those in the Indian Ocean. Gene flow over this larger geographic scale is achieved through stepping-stone transport via Bouvet Island.

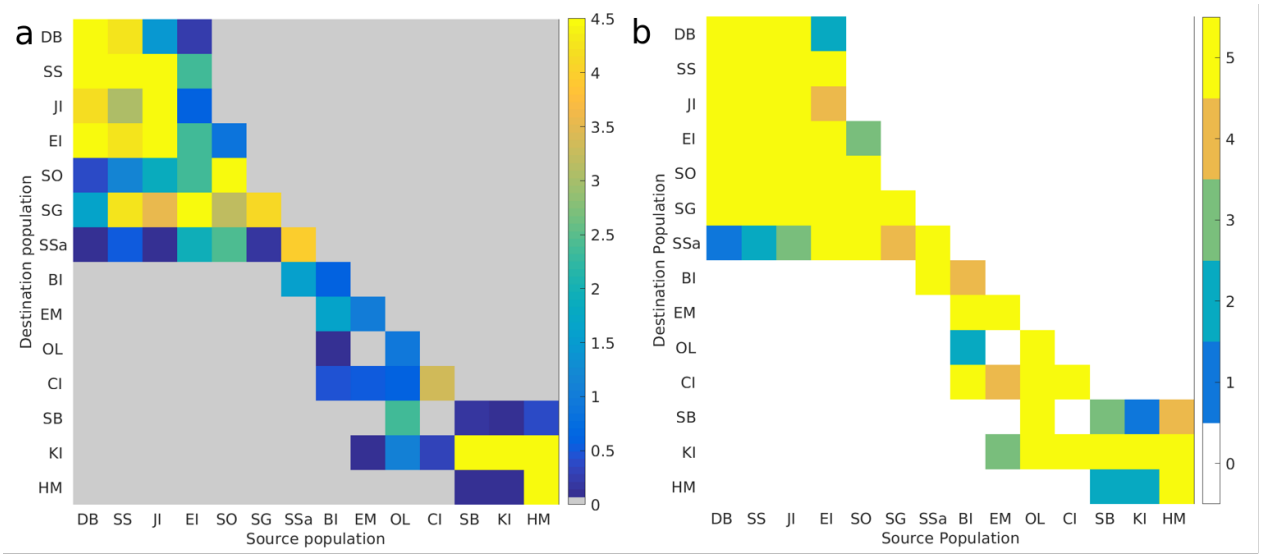


Figure 4.15. Simulated dispersal connectivity of *Notothenia rossii* throughout most of the Southern Ocean: (a) mean connectivity as a percentage of particles from source populations (columns) successfully reaching destination populations (rows), on a transformed log scale [log((10x)+1)], (b) frequency of non-zero connectivity.

4. DISCUSSION

4.1 Distribution, genomic diversity and connectivity of Notothenia rossii

Contrary to our working hypothesis, even thousands of genomic markers screened across many individuals from locations > 5,000 km apart reveal no evidence of genetic differentiation. The historically overexploited Antarctic fish *N. rossii* therefore exhibits no contemporary spatial genetic population structure, although connectivity modelling suggests that exchange between the Scotia Sea and the Kerguelen Plateau regions occurs via stepping-stone transport. This apparent discrepancy may be due to the different temporal scales resolved by the observational and modelling analyses, i.e. ecological vs. evolutionary time. As we will further detail, our results show 1) the detailed distribution of the species in the Atlantic and Indian Ocean sectors of the Southern Ocean, 2) the contemporary genome-wide levels of diversity and 3) the direction and magnitude of dispersal connectivity.

As a shelf dwelling fish that prefers shallower waters to feed and reproduce, the occurrence of *N. rossii* is relatively localized in the Southern Ocean. Juveniles develop in algae beds (Duhamel et al. 1982, Barrera-Oro and Winter 2008), which are only found close to the coast (Wiencke et al. 2014). The cold temperatures of the high-Antarctic are, however, likely detrimental for the species as its blood equilibrium freezing point is comparatively high for a notothenioid (Bilyk and DeVries 2010, Miya et al. 2013). Therefore, the habitat requirements alone induce a fragmented distribution in this species. Species distribution modelling predictions confirm this and highlight many well-documented population hotspots as highly probable habitats (Fig. 4.12). The South Shetland Islands, South Georgia and the Kerguelen Plateau are localities where *N. rossii* is most often caught (DeWitt et al. 1990, Duhamel et al. 2014). However, other areas, some of which are less well-studied, may also be relevant habitats for *N. rossii*. For example, SDM data show high occurrence probability at Bouvet Island, even though the shelf around this island is narrow and the species was never caught there (DeWitt et al. 1990, Jones et al. 2008a, Padilla et al. 2015). However, incidental data (chinstrap penguin stomach content) indicate that *N. rossii* is present at Bouvet Island (Niemandt et al. 2015). This would corroborate our SDM prediction, although the record is based on a single otolith, which could also be misidentified. Ultimately, only more extensive sampling, particularly of the near shore fish fauna, will be able to resolve the matter. It also remains to be investigated whether other new occurrence localities predicted by the SDM, that is around Prydz Bay and Patagonia, are realistic. At least the latter seems very unlikely as the ecological niche in Patagonian waters is filled by different species such as *Paranotothenia magellanica* or *Patagonothen spp*. (Cousseau et al. 2019). *Notothenia rossii* has not been recorded off Prydz Bay so far either despite regular surveys (Hoddell et al. 2000, Van de Putte et al. 2010).

The genetic diversity of *N. rossii* in terms of heterozygosity is similar to that observed in other fishes with SNPs (Fig. 4.13a, Berg et al. 2016, Pérez-Portela et al. 2018, Christiansen 2020). In addition, there are no signs of spatial variation in heterozygosity, despite spatially heterogeneous fishing pressure (Duhamel 1982, Kock 1992). These results are in contrast to the expectation that overharvesting reduces genetic diversity (Pinsky and Palumbi 2014). However, objective comparisons between studies are challenging due to the wide variety of settings employed to generate a “final” SNP data set (Shafer et al. 2017). Fifty years post-exploitation the genetic diversity of *N. rossii* does not seem dramatically reduced across thousands of markers. As there is no baseline for pre-exploitation diversity levels, it remains elusive at this point whether these levels of heterozygosity are representative of the unperturbed state. In fact, even if overall average diversity is high, rare alleles, potentially important for rapid adaptation, may be lost (Pinsky and Palumbi 2014). The genetic data furthermore demonstrate a striking lack of spatial structure with very low F_{ST} values and no genetic clusters discernable (Table 4.3 & 4.4, Fig. 4.13b). Consequently, this implies regular gene flow, at least via stepping stones between all established population patches of *N. rossii* (Fig. 4.14b). This is an important implication to consider in the context of the results of the connectivity modelling exercise. Finally, several candidate loci show indications of putative recent selective pressure, despite the lack of overall population structure. This is not unexpected, given that selection and adaptation can occur in the presence of high gene

flow (Tigano and Friesen 2016, Hoey and Pinsky 2018). Nevertheless, we emphasize that these candidate loci might be false positives or even related to genotyping error. High coverage studies or whole genome sequencing approaches are needed for a more detailed understanding of local or global adaptation in *N. rossii* (Booker et al. 2019).

Dispersal of the early life stages of *N. rossii* is generally high, but the large-scale unidirectional connectivity predicted by the modelling suggests that inter-ocean connectivity is achieved through stepping-stone transport (Fig. 4.15). The South Sandwich Islands and Bouvet Island in particular are predicted to be key links between the abundant *N. rossii* aggregations at the Antarctic Peninsula and Scotia Sea, and the Kerguelen Plateau. The relatively large effective population sizes indicated by the genomic data may secure successful large-scale connectivity over evolutionary time scales. Thus, for example, if the exchange of individuals between the Scotia Sea and the Kerguelen Plateau were to fail in some years, successful recruitment in other years may suffice to maintain gene flow.

4.2 A unifying framework to explain contemporary patterns

Here, we suggest that our results can be collectively explained through a scenario that incorporates the life history, physical setting and exploitation history of *N. rossii*. Three main aspects help to unify the patterns observed through SDM, genomics and dispersal modelling. First, important stepping stones such as the South Sandwich Islands and potentially Bouvet Island may also act as temporal refuges for juvenile fish. Even if small, the local benthic ecosystem at Bouvet Island provides suitable conditions including some macroalgae and a variety of invertebrate prey items for *N. rossii* fingerlings (Arntz et al. 2006, Jacob et al. 2006). The model simulations stopped after seven months, at which point observational data suggest larvae develop into brown -phase fingerlings and recruit to kelp beds (North 2001). However, the recruitment behavior of early juveniles is not well known. For example, blue-phase fingerlings may be able to continue a pelagic life style for an extended period until a suitable recruitment site is reached. Such behavior would increase the dispersal range of the early life stages, and potentially reduce the dependence on small, isolated stepping stones for inter-ocean connectivity. In addition, it is unknown so far, but not inconceivable that, for example, juveniles that reached a stepping stone may later proceed to migrate further toward habitats in the Kerguelen Plateau where they continue to grow, mature and eventually reproduce. In fact, Shcherbich (1975) for South Georgia and Barrera-Oro et al. (2014) for the South Shetland Islands indicated that some juveniles spend at least a full year as blue-phase fingerlings in the water column before settling to a benthic life style. The adults also undertake at least short distance migrations, such as from coastal kelp belts to the outer archipelago (on the Kerguelen Plateau) or to their spawning ground (Duhamel 1982, DeWitt et al. 1990). Tagging studies have been used so far to validate age determination (Moreira et al. 2013), but could be used in the future to verify whether juveniles or adults of *N. rossii* are capable of more extensive migrations. Such behavior is documented in Antarctic toothfish, that undertake long-distance migrations at least occasionally (Hanchet et al. 2010). As the genomic data suggests circumpolar transport, it is possible that other, potentially small stepping-stone population patches exist. An important area for circumpolar connectivity could be the documented occurrence off Macquarie Island in the Pacific sector of the Southern Ocean. The lack of samples and data precluded us from investigating this further. Model simulations, however, suggested the potential for transport from the Kerguelen Plateau to Macquarie Island with an extended dispersal period of a year, although successful dispersal from Macquarie Island to the Antarctic Peninsula was not predicted within this time frame. Therefore, to achieve circumpolar connectivity, the modelling setup would suggest a longer pelagic phase or the existence of undocumented population patches, or both.

The slow recovery of *N. rossii* following severe overfishing may be the result of historically diminished effective population size in conjunction with the stepping-stone nature of large-scale connectivity. Large stocks around South Georgia and Kerguelen were heavily exploited in the 1970s, likely leading to considerably reduced effective population size at these localities. Possibly, the South Georgia stock would have been resupplied through dispersal from the western Antarctic Peninsula in the years following its overexploitation. However, the spawning stock at the Antarctic

Peninsula was also largely removed through fishing in 1979/80 (Kock 1992). In turn, the Kerguelen stock was not supplied sufficiently, because the influx of larvae or fingerlings from South Georgia via stepping stones was interrupted. Thus, the already low levels of long-distance ecological connectivity that we estimated here may be an explanation for the long recovery time. This could mean that the original, unexploited population went through a genetic bottleneck leading to large genetic homogeneity of the extant population. In addition, other species, for example the opportunistically feeding grey notothen *Lepidonotothen squamifrons* at South Georgia (Gregory et al. 2014), may have filled vacant ecological niches in the meantime and further hampered the re-establishment of highly abundant *N. rossii* stocks. Kock and Belchier (2004) rightfully pointed out that the biomass estimation is particularly difficult in *N. rossii* due to its patchy distribution as adults. Yet, the most recent surveys were able to document large catches of *N. rossii* once again, at least in the Kerguelen Plateau (Duhamel et al. 2019). This trend is corroborated in the South Shetland Islands (Barrera-Oro et al. 2017), suggesting that the species is indeed slowly recovering.

Lastly, a comparison with other Southern Ocean fish indicates that relatively high connectivity may be the most common scenario among sub-Antarctic fish. Matschiner et al. (2009) summarized population genetic studies of notothenioids over a period of 15 years and noted that significant differentiation over small scales (< 100 km) was only exceptionally documented in three out of 27 cases. More recently, several studies have uncovered previously unknown genetic differentiation, for example in *Trematomus spp.* (Van de Putte et al. 2012), *Champsocephalus gunnari* and *Chaenocephalus aceratus* (Damerau et al. 2014, Young et al. 2015), *Pleuragramma antarctica* (Agostini et al. 2015, Caccavo et al. 2019), and *Notothenia coriiceps* (Christiansen 2020). However, except for *C. gunnari* and *C. aceratus* these results relate to long-distance differentiation in high-Antarctic species. On the one hand, habitats for non-deep-sea species near the Antarctic continent may be physically closer to each other than offshore habitats of the sub-Antarctic, where vast deep sea basins cause habitat fragmentation. On the other hand, ice cover, iceberg scouring and the advance and retreat of ice during glacial cycles may have driven high-Antarctic species into local refugia, leading to genetic dissimilarities that are still traceable in the genome (Allcock and Strugnell 2012). In the sub-Antarctic the habitat is less affected by ice, but discontinuous for benthic species through the sheer geographical setting, while it is comparatively barrier-free for pelagic species. In order to persist in this habitat and maintain vast distribution ranges, species may have adapted their dispersal capabilities to achieve persistent (even if low) long-distance connectivity as in the case of *N. rossii*, but also kelp, toothfish and crustaceans, for example (see Moon et al. 2017 and references therein).

4.3 Implications for MPA design and fisheries management

The fisheries data shows clearly that *N. rossii* experienced a dramatic overexploitation in the 1970s (Kock 1992, CCAMLR 2019a). The recovery is more difficult to assess due to fewer systematic stock assessment methods (trawling, acoustics) compared to less remote oceans. Nevertheless, several recent ecological and fisheries surveys indicate an ongoing recovery (Marschoff et al. 2012, Barrera-Oro et al. 2017, Duhamel et al. 2019). If a slow recovery process was indeed the result of long generation time, slow growth, reduced effective population size and stepping stone connectivity, as we suggest here, then future management plans should remain very precautionary, which is in accordance with CCAMLR. Importantly, a precautionary approach should not only regard each management area separately, but considers the interconnectivity between these areas. Successful ecological connectivity in at least some years may be an important prerequisite for a stable circum-Antarctic population. In turn, this suggests that it is important to protect areas that act as key stepping stone habitats. The waters around Bouvet Island have been fished in the past (Arntz et al. 2006), but are currently a designated marine reserve to 12 nautical miles from the coast. Controversially though, some krill fisheries permits are granted in this area as well. Our results demonstrate that the Bouvet Island marine ecosystem may be a unique stepping stone of large ecological importance. Endemism levels at Bouvet are very low, further supporting the premise that many species are in fact transported here by advection (Arntz et al. 2006, Gutt et al. 2006, Moreau et al. 2017). Therefore, not only *N. rossii* but also other invertebrate and vertebrate species, particularly those present in both the Scotia Sea and Kerguelen Plateau regions, may rely on this comparatively tiny ecological hotspot, which should

receive adequate protection. Long-distance connectivity has clear benefits for the effectiveness of MPA networks, although it has previously rarely been quantified (Manel et al. 2019).
A second important conservation implication concerns the Antarctic Peninsula. A recent MPA proposal that was presented to CCAMLR concerns this area and was put forward by Argentina and Chile during 2018, but so far not adopted (CCAMLR 2019b). The modelled oceanographic connectivity strongly suggests that *N. rossii* populations at the Antarctic Peninsula are sources for the fish assemblages off South Georgia and, to a lesser extent, the South Sandwich and South Orkney Islands. They are therefore important for re-establishing and maintaining a large population throughout the Scotia Sea. In addition, if the population of *N. rossii* is to be managed throughout its range in a precautionary approach in the future, conservation and monitoring of the northwest region of the Antarctic Peninsula marine ecosystem, including Bransfield Strait, will be important. This region experiences drastic climate change effects, including increases in temperature and reduction of ice cover, with consequences for the entire ecosystem (McClintock et al. 2008b, Ducklow et al. 2013). It can therefore be a natural laboratory to detect the effects of global warming, for example, on the high-Antarctic and sub-Antarctic Notothenia species (*N. coriiceps* and *N. rossii*) that occur here in sympatry. Adaptive genetic variance in these species may bear the potential to mitigate climate change effects as shown theoretically in terrestrial species (Razgour et al. 2019). Continued multidisciplinary investigations as presented here could help achieve adequate monitoring and prevent unsustainable loss of biodiversity.

4.4 Methodological considerations and future research perspectives

Applying species distribution models at large scales and in data-poor environments is challenging. Particular problems include spatially aggregated data, presence-only data, and extensive gaps in environmental data, which can be partly circumvented with appropriate calibration and validation methods (Guillaumot et al. 2018a - Appendix, Guillaumot et al. 2019 - Chapter 2). In conjunction with other methods, SDM data can be used successfully to fill specific knowledge gaps, such as the case of Bouvet Island, presented here. Sometimes, interpolating from data collected elsewhere is a valuable alternative to costly or near impossible direct observation (Gutt et al. 2012). In addition to collecting more data, a future improvement of SDM approaches would be the separation of the model by life stages. For *Notothenia rossii* in particular it could be highly informative to generate a robust SDM for larval stages, provided that sufficient ecological information can be gathered.
The genomic data created here have limitations especially with regard to sequencing coverage, which causes further downstream issues such as potential genotyping error and low quality genotypes that necessitated very extensive filtering (Mastretta-Yanes et al. 2014). The available data were thus dramatically reduced, but the overall pattern of little genetic structure and comparatively normal diversity levels is likely accurate. In contrast, determining with certainty which loci or genes are important for local adaptation would require further efforts. The reason for our low coverage is likely an underestimation of the true genome size of *N. rossii* or of the number of fragments that the restriction enzyme ApeKI produces. More extensive *a priori* testing could help alleviate such issues (Christiansen 2020) as well as a high quality reference genome (Fountain et al. 2016, Shafer et al. 2017).

The oceanographic model used in this study has a relatively coarse resolution, due to the large spatial scale at which it is applied and resultant computation demands. The development of high-resolution oceanographic models that better resolve fine-scale circulation features has the potential to reveal further details on the connectivity of *N. rossii*, for example regarding the extent of local retention. In addition, the skill of the IBM is highly dependent on the accuracy of its biological parameterization. We have used the best available biological knowledge gathered over decades of research, but some uncertainty remains. In particular, there is uncertainty over the total permissible length of the pelagic phase, and the active behavior of larvae, fingerlings, and juveniles. Further knowledge of the behavior of fingerlings and juveniles, for example, would allow the incorporation of these additional life stages into the IBM. Observational evidence for feeding behavior, active swimming and diel vertical migration, would allow further refinement of the IBM, improving its predictive skill. In addition, both SDM and IBM models could be refined by further integrating spatially variable biological traits, when additional such information becomes available. Even in the

absence of genetic differentiation, some spatial differences in life history parameters between *N. rossii* assemblages have been reported over large distances, but recent ecological comparisons are scarce (DeWitt et al. 1990, but see also Calì et al. 2017). Considering intraspecific diversity is difficult but the necessary next step for most accurate biological models (in SDM, IBM and population genomics) with clear conservation benefits (Mee et al. 2015, Marcer et al. 2016, Des Roches et al. 2018, Paz-Vinas et al. 2018).

5. CONCLUSIONS

Multidisciplinary approaches to assess connectivity are extremely useful in data-limited situations as is the case in the vast and remote Southern Ocean. The integration of data from three different sources allowed us to identify areas important for conservation and provide a hypothesis that explains the slow recovery of *Notothenia rossii* stocks. These results are relevant for the ongoing effort to establish a network of MPAs and implement ecosystem based management for the region. Further challenges lay ahead, with climate change potentially altering the suitable habitat and connectivity, which demands continued research and monitoring, and flexible, adaptive management.

APPENDIX 4.4.

Results from parameter optimization with Stacks v2.4 (Rochette et al. 2019) for *de novo* assembly and genotyping of four genotyping-by-sequencing (GBS) libraries of *Notothenia rossii*.

A parameter test series using a subset of 24 individuals (4 from each population) was conducted as described in Rochette and Catchen (2017). The Stacks parameter m was kept constant in two test series (m = 2 and m = 3), while parameters M and n were varied together from 1 to 9. Subsequently, only loci present in 80 % of the samples were retained and for each M=n parameter the number of loci and polymorphic loci was plotted, as well as the proportion of these loci containing 0 to 10 or >10 SNPs. Optimal values were inferred from these results as m=3 and M=n=4.

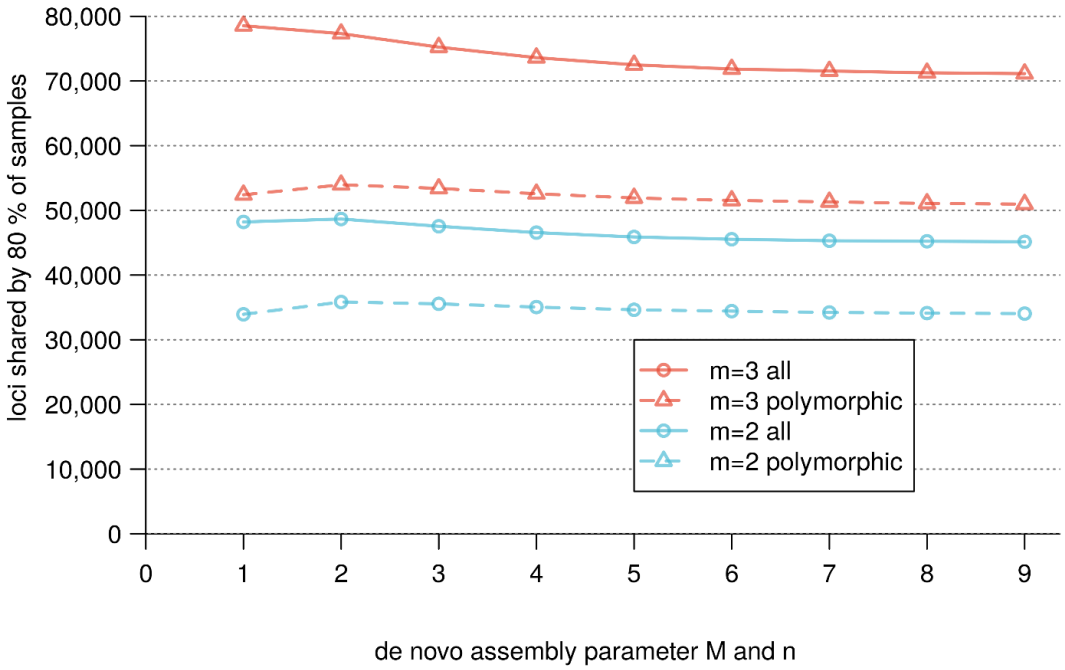


Figure S4.4.A. Number of loci and polymorphic loci shared by 80 % of samples from the *Notothenia rossii* GBS libraries across nine values for parameter M and n and two values for parameter m (2 and 3; blue vs. red) in Stacks v.2.4.

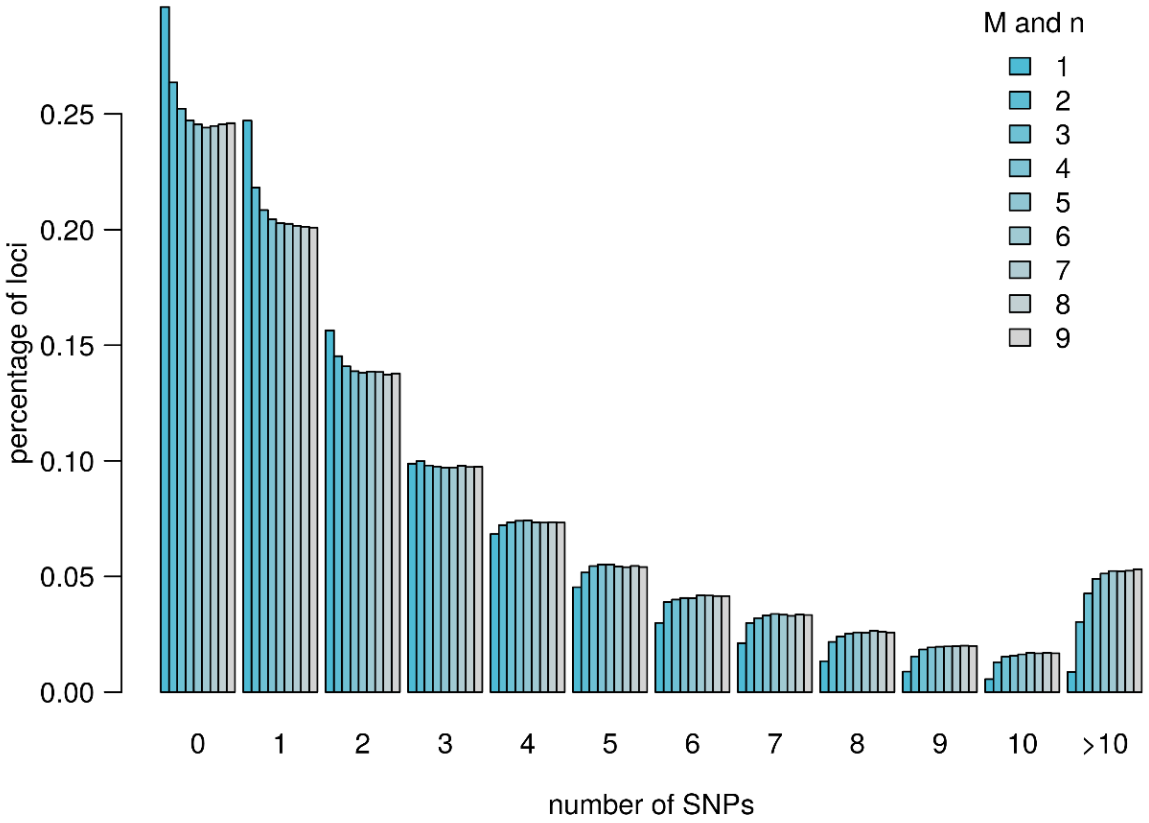


Figure S4.4.B. Number of SNPs per locus shared by 80 % of samples from the *Notothenia rossii* GBS libraries across nine parameters of M and n under constant m = 2 in Stacks v.2.4.

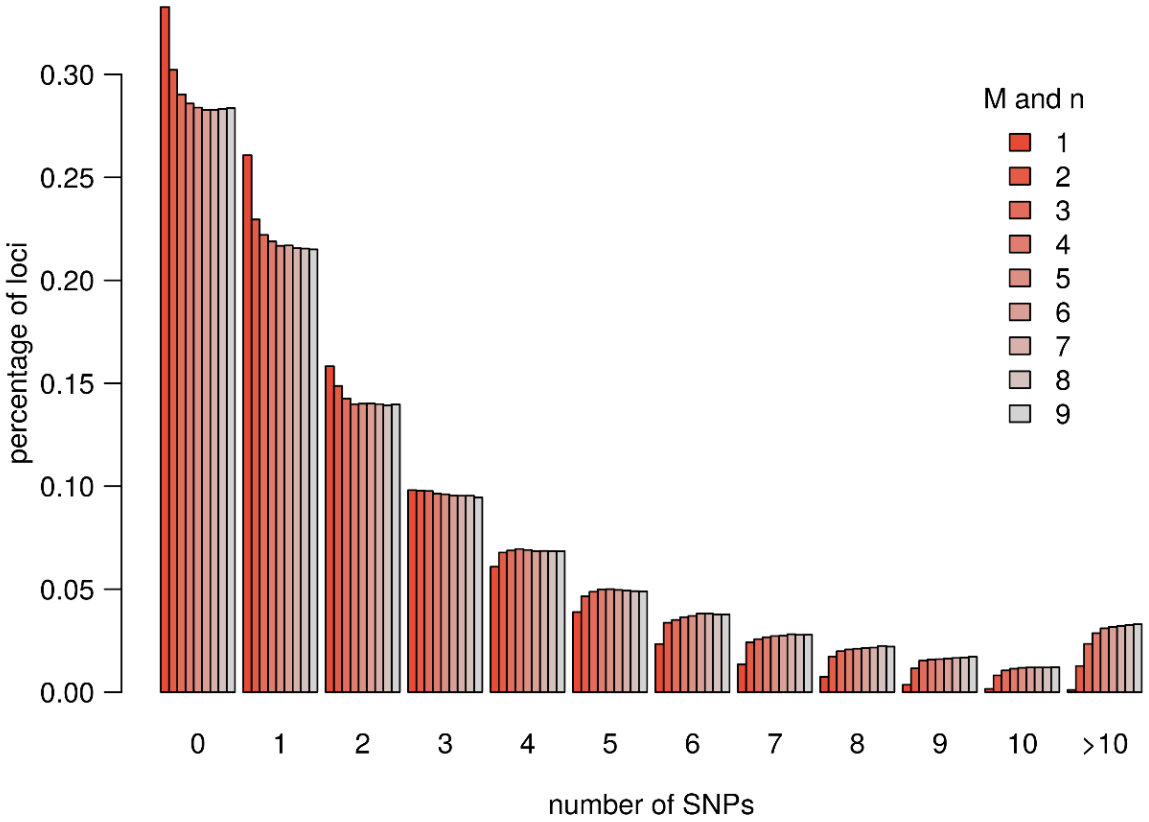


Figure S4.4.C. Number of SNPs per locus shared by 80 % of samples from the *Notothenia rossii* GBS libraries across nine parameters of M and n under constant m = 3 in Stacks v.2.4.

APPENDIX 4.5.

Filtering conducted on raw SNP data sets produced from bioinformatics of four GBS libraries of *Notothenia rossii* from the Southern Ocean.

Genotypes of the reference-based and *de novo* data sets were first pruned using the population module of Stacks v2.4 (Rochette et al. 2019), requiring loci to be present in at least 80 % of the individuals of each population, to have a minor allele frequency > 0.05 and heterozygosity < 0.7 (Rochette and Catchen 2017). Subsequently, genepop files were imported and filtered extensively in R software using the 'radiator' package v1.1.1 (Gosselin 2019).

At the start of the radiator filtering pipeline, the data sets contained genotypes from 349 individuals at 85,980 (*de novo*) and 73,554 (reference-based) SNPs. First, loci that were not shared across all populations were removed, i.e. 65,269 and 54,850 SNPs. Subsequently, individual genotypes were filtered based on an outlier statistic of missing data and of heterozygosity. These steps removed data from 62 (*de novo*) and 45 (reference) individuals. Then, markers were filtered based on minor allele count (mac), requiring a minimum mac of 10, and on coverage, removing all loci with coverage below 10 or above 100. SNPs were also removed when showing signs of unnormal positioning with the RAD fragment and on short linkage disequilibrium. For the latter, only one SNP per fragment was retained, the one with highest mac. Finally, duplicate genomes were detected and removed and loci significantly ($p < 0.01$) departing from Hardy-Weinberg proportions were removing. After these steps, 277 individuals and 12,400 loci remained in the *de novo* data set and 294 individuals and 4,505 loci in the reference data set. As a last filtering steps, these data were filtered on minor allele frequency (maf, threshold: 0.05), leaving 9,806 and 4,079 SNPs.

Because a bias related to sequencing library was still detectable using principal component analysis in the above data sets, further loci, contributing to this bias, were removed. The final data sets then contain 261 individual genotypes at 7,501 SNPs in the *de novo* data and 272 individual genotypes at 3,503 SNPs in the reference-based data.

APPENDIX 4.6.

Pairwise genetic differentiation of *Notothenia rossii* in the Southern Ocean using alternative differentiation metrics.

Table S4.6.A. Pairwise genetic differentiation of *Notothenia rossii* per sampling locality (see Table 4.1 for codes) based on 7,501 SNP genotypes derived from mapping against a *de novo* assembly. Jost's D (2008) below the diagonal and Hedrick's GST (2005) above the diagonal, as calculated with R package 'mmod' (Winter 2012).

	SSD-06	SSK-06	SSK-15-16	EI-06-07	EI-02	SO-06	SG-02-03	SG-05	SB-15	KI-15
SSD-06		0.0008	0.0015	0.0010	0.0002	0.0030	0.0013	0.0013	0.0015	0.0000
SSK-06	0.0002		0.0015	0.0017	0.0006	0.0024	0.0015	0.0003	0.0010	0.0013
SSK-15-16	0.0004	0.0004		0.0002	0.0004	0.0013	0.0000	0.0009	0.0012	0.0008
EI-06-07	0.0003	0.0004	0.0001		0.0004	0.0023	0.0004	0.0015	0.0010	0.0009
EI-02	0.0000	0.0001	0.0001	0.0001		0.0013	0.0006	0.0006	0.0006	0.0012
SO-06	0.0008	0.0006	0.0003	0.0006	0.0003		0.0016	0.0005	0.0018	0.0014
SG-02-03	0.0003	0.0004	0.0000	0.0001	0.0001	0.0004		0.0006	0.0018	0.0014
SG-05	0.0003	0.0001	0.0002	0.0004	0.0002	0.0001	0.0002		0.0005	0.0007
SB-15	0.0004	0.0003	0.0003	0.0003	0.0001	0.0004	0.0004	0.0001		0.0006
KI-15	0.0000	0.0003	0.0002	0.0002	0.0003	0.0003	0.0004	0.0002	0.0002	

Table 4.6.B. Pairwise genetic differentiation of *Notothenia rossii* per sampling locality (see Table 4.1 for codes) based on 3,503 SNP genotypes derived from mapping against the reference genome of *N. coriiceps* (Shin *et al.* 2014). Jost's D (2008) below the diagonal and Hedrick's GST (2005) above the diagonal, as calculated with R package 'mmod' (Winter 2012).

	SSD-06	SSK-06	SSK-15-16	EI-06-07	EI-02	SO-06	SG-02-03	SG-05	SB-15	KI-15
SSD-06		0.0020	0.0017	0.0015	0.0006	0.0029	0.0017	0.0012	0.0008	0.0000
SSK-06	0.0004		0.0036	0.0025	0.0019	0.0048	0.0021	0.0013	0.0010	0.0016
SSK-15-16	0.0004	0.0008		0.0031	0.0016	0.0018	0.0010	0.0015	0.0026	0.0022
EI-06-07	0.0003	0.0006	0.0007		0.0012	0.0035	0.0015	0.0014	0.0009	0.0009
EI-02	0.0001	0.0004	0.0004	0.0003		0.0031	0.0006	0.0003	0.0012	0.0019
SO-06	0.0006	0.0011	0.0004	0.0008	0.0007		0.0025	0.0024	0.0033	0.0022
SG-02-03	0.0004	0.0005	0.0002	0.0003	0.0001	0.0005		0.0005	0.0012	0.0015
SG-05	0.0003	0.0003	0.0003	0.0003	0.0001	0.0005	0.0001		0.0011	0.0012
SB-15	0.0002	0.0002	0.0006	0.0002	0.0003	0.0007	0.0003	0.0003		0.0009
KI-15	0.0000	0.0004	0.0005	0.0002	0.0004	0.0005	0.0003	0.0003	0.0002	

APPENDIX 4.7

Annotation of loci identified through genome scans for selection in GBS data of *Notothenia rossii*. Candidate outlier SNPs were matched against the NCBI database.

Table S4.7.A. BLAST results from 12 candidate SNPs from the *de novo* data set.

Nr	Name	Top Hit	Accession Number	Percent Identity	E value
1	>CLocus_130488	*Cottoperca gobio* genome assembly, chromosome: 19	LR131926.1	82.353	4.72E-23
2	>CLocus_138804	*Cottoperca gobio* genome assembly, chromosome: 12	LR131919.1	81.818	4.73E-04
3	>CLocus_221130	*Cottoperca gobio* genome assembly, chromosome: 12	LR597562.1	89.474	4.42E-17
4	>CLocus_223132	PREDICTED: *Notothenia coriiceps* ubiquitin specific peptidase 38 (usp38), mRNA	XM_010793758.1	100	3.88E-24
5	>CLocus_237675	*Sparus aurata* genome assembly, chromosome: 6	LR537126.1	91.892	2.00E-03
6	>CLocus_240253	*Gossypioides kirkii* chromosome KI_01	CP032244.1	90.909	2.40E-01
7	>CLocus_251284	*Myripristis murdjan* genome assembly, chromosome: 22	LR597571.1	88.889	2.29E-33
8	>CLocus_263048	*Centromochlus existimatus* isolate S1A2_08 ATPase subunit 8 (ATPase 8) gene, complete cds; and ATPase subunit 6 (ATPase 6) gene, partial cds; mitochondrial	JX910183.1	88.889	8.50E-01
9	>CLocus_38304	*Apteryx australis* mantelli genome assembly AptMant0, scaffold scaffold1406	LK066414.1	88.095	7.00E-02
10	>CLocus_60053	*Cottoperca gobio* genome assembly, chromosome: 3	LR131933.1	82.353	9.74E-13
11	>CLocus_6959	*Lateolabrax maculatus* linkage group 21 sequence	CP032596.1	84.507	4.14E-11
12	>CLocus_96107	*Cottoperca gobio* genome assembly, chromosome: 12	LR131919.1	87.288	1.19E-30

Table S4.7.B. BLAST results from 37 candidate SNPs from the reference data set.

Nr	Name	Top Hit	Accession Number	Percent Identity	E value
1	>CLocus_207926 [AZAD01004947.1, 742, +]	*Chionodraco hamatus* Cu/Zn superoxide dismutase (SOD1) mRNA, partial cds	AY736281.1	90.244	1.38E-18
2	>CLocus_417728 [AZAD01011137.1, 181, +]	*Gymnodraco acuticeps* zona pellucida protein ZPC5 isoform 1 (ZPC5) mRNA, complete cds	KU522427.1	89.655	1.68E-55
3	>CLocus_664045 [AZAD01019142.1, 1235, -]	*Cottoperca gobio* genome assembly, chromosome: 10	LR131917.1	83.092	3.03E-52
4	>CLocus_1150506 [AZAD01034986.1, 5571, -]	PREDICTED: *Notothenia coriiceps* transcription initiation factor IIB-like (LOC104957872), partial mRNA	XM_010785544.1	87.903	2.49E-34
5	>CLocus_1462038 [AZAD01044248.1, 240, +]	*Dissostichus mawsoni* haplotype 1 AFGP/TLP gene locus, partial sequence	HQ447059.1	95.96	3.45E-83
6	>CLocus_1491991 [AZAD01045372.1, 5082, +]	PREDICTED: *Notothenia coriiceps* transcription initiation factor IIB-like (LOC104957872), partial mRNA	XM_010785544.1	85.484	5.49E-30
7	>CLocus_2032174 [AZAD01062243.1, 5021, +]	PREDICTED: *Notothenia coriiceps* transcription initiation factor IIB-like (LOC104957872), partial mRNA	XM_010785544.1	85.484	5.49E-30

8	>CLocus_2457922 [AZAD01071921.1, 3456, +]	*Cottoperca gobio* genome assembly, chromosome: 10	LR131917.1	73.973	6.69E-10
9	>CLocus_51011 [KL662357.1, 5270500, -]	*Cottoperca gobio* genome assembly, chromosome: 21	LR131929.1	76.136	6.26E-23
10	>CLocus_139511 [KL662384.1, 326949, -]	*Cottoperca gobio* genome assembly, chromosome: 10	LR131917.1	74.658	1.29E-12
11	>CLocus_198071 [KL662552.1, 248239, +]	*Cottoperca gobio* genome assembly, chromosome: 1	LR131916.1	83.333	3.46E-26
12	>CLocus_210679 [KL662597.1, 16029, -]	PREDICTED: *Notothenia coriiceps* uncharacterized LOC104965236 (LOC104965236), mRNA	XM_010794195.1	87.5	3.95E-19
13	>CLocus_210838 [KL662599.1, 5111, +]	*Cottoperca gobio* genome assembly, chromosome: 14	LR131921.1	83.571	2.33E-28
14	>CLocus_261220 [KL662789.1, 79206, +]	*Dissostichus mawsoni* haplotype 2 AFGP/TLP gene locus, partial sequence	HQ447060.1	92.405	1.38E-56
15	>CLocus_292518 [KL662880.1, 318427, +]	*Sparus aurata* genome assembly, chromosome: 17	LR537137.1	76.829	7.63E-22
16	>CLocus_313316 [KL662933.1, 117387, -]	*Cottoperca gobio* genome assembly, chromosome: 9	LR131939.1	76.531	6.69E-29
17	>CLocus_512277 [KL663578.1, 176442, +]	*Cottoperca gobio* genome assembly, chromosome: 6	LR131936.1	85.87	5.48E-49
18	>CLocus_550684 [KL663710.1, 313820, +]	PREDICTED: *Notothenia coriiceps* symplekin-like (LOC104960710), mRNA	XM_010788863.1	82.707	8.14E-28
19	>CLocus_612378 [KL663896.1, 102980, +]	*Cottoperca gobio* genome assembly, chromosome: 10	LR131917.1	77.397	5.86E-17
20	>CLocus_676688 [KL664078.1, 28095, -]	*Lateolabrax maculatus* chromosome Lm22	CP027283.1	76.301	6.69E-29
21	>CLocus_984478 [KL665099.1, 640019, +]	*Thalassophryne amazonica* genome assembly, chromosome: 13	LR722978.1	80	8.00E-03
22	>CLocus_1083361 [KL665382.1, 29521, +]	*Cicer arietinum* chromosome Ca2	CP039332.1	85.714	3.20E-01
23	>CLocus_1090835 [KL665412.1, 447335, +]	*Cottoperca gobio* genome assembly, chromosome: 1	LR131916.1	84.277	9.28E-40
24	>CLocus_1092384 [KL665412.1, 1044395, +]	PREDICTED: *Aplysia californica* calmodulin-like protein 3 (LOC106012422), mRNA	XM_013085361.1	82.222	4.00E+00
25	>CLocus_1179376 [KL665586.1, 242786, -]	*Gouania willdenowi* genome assembly, chromosome: 8	LR131991.1	75	1.21E-06
26	>CLocus_1224872 [KL665708.1, 68516, -]	*Lateolabrax maculatus* linkage group 21 sequence	CP032596.1	72.358	9.93E-08
27	>CLocus_1425216 [KL666295.1, 8396428, +]	*Cottoperca gobio* genome assembly, chromosome: 4	LR131934.1	89.103	5.48E-49
28	>CLocus_1534498 [KL666587.1, 97216, +]	*Sparus aurata* genome assembly, chromosome: 9	LR537129.1	89.899	6.67E-67
29	>CLocus_1551285 [KL666590.1, 947841, +]	*Chionodraco hamatus* transposon helitron polyprotein (HeliNoto) gene, complete cds	GU014476.2	77.236	7.14E-16

30	>CLocus_1552830 [KL666590.1, 1381214, -]	PREDICTED: *Notothenia coriiceps* symplekin-like (LOC104960710), mRNA	XM_010788863.1	86.667	2.49E-34
31	>CLocus_1630310 [KL666849.1, 6404, +]	*Cottoperca gobio* genome assembly, chromosome: 5	LR131935.1	87.879	1.57E-11
32	>CLocus_1954638 [KL667808.1, 350977, -]	PREDICTED: *Notothenia coriiceps* ectonucleotide pyrophosphatase/phosphodiesterase family member 2-like (LOC104962317), mRNA	XM_010790748.1	99.167	3.69E-51
33	>CLocus_2033344 [KL668045.1, 3250, +]	PREDICTED: *Notothenia coriiceps* uncharacterized LOC104955179 (LOC104955179), ncRNA	XR_799431.1	90.625	2.84E-27
34	>CLocus_2067957 [KL668140.1, 732310, +]	*Cottoperca gobio* genome assembly, chromosome: 6	LR131936.1	75	5.14E-05
35	>CLocus_2206060 [KL668296.1, 21625648, +]	PREDICTED: *Arabidopsis lyrata* subsp. lyrata protein SUPPRESSOR OF NIM1 1 (LOC9320118), mRNA	XM_021032701.1	89.474	9.30E-02
36	>CLocus_2245286 [KL668297.1, 4382515, +]	*Dicentrarchus labrax* chromosome sequence corresponding to linkage group 1, top part, complete sequence	FQ310506.3	78.616	1.47E-24
37	>CLocus_2284480 [KL668297.1, 14524502, -]	*Cottoperca gobio* genome assembly, chromosome: 10	LR131917.1	76.712	2.49E-15

APPENDIX 4.8

Results of the population genomic analysis of *Notothenia rossii* in the Southern Ocean using the reference genome aligned SNP data set.

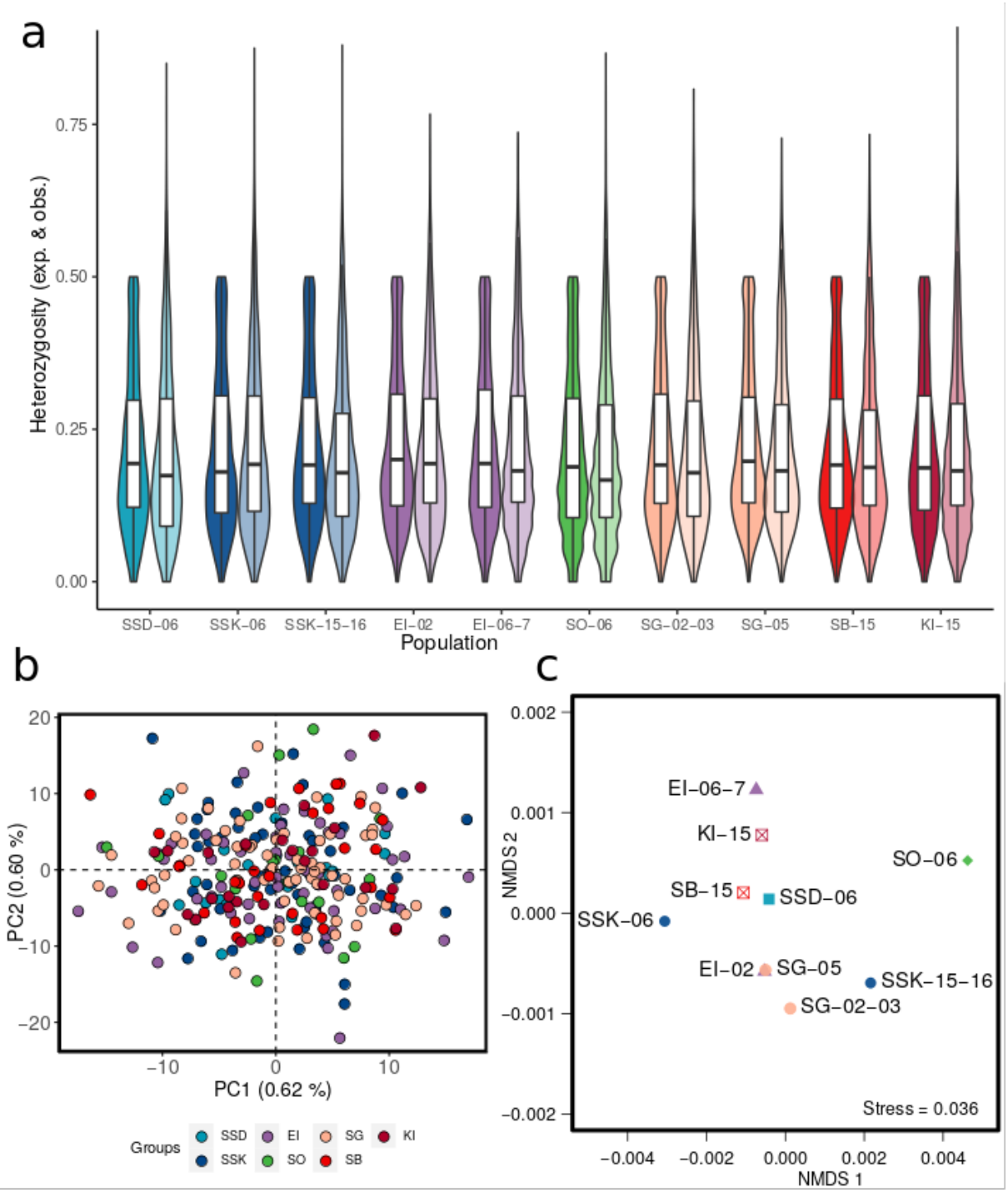


Figure S4.8A. Genomic diversity of *Notothenia rossii* in the Southern Ocean based on 3,503 SNP loci from reference-based variant calling. Expected (darker shading, left) and observed (light shading, right) heterozygosity is shown as box and violin plots for each genetically screened population (a). Principal component analysis (PCA) reveals little individual-based differentiation (b), while non-metric multidimensional scaling based on GST distances shows subtle differences between population samples (c). Sample codes as in Table 4.1; samples from different years but same locality are not shown separately on the PCA.

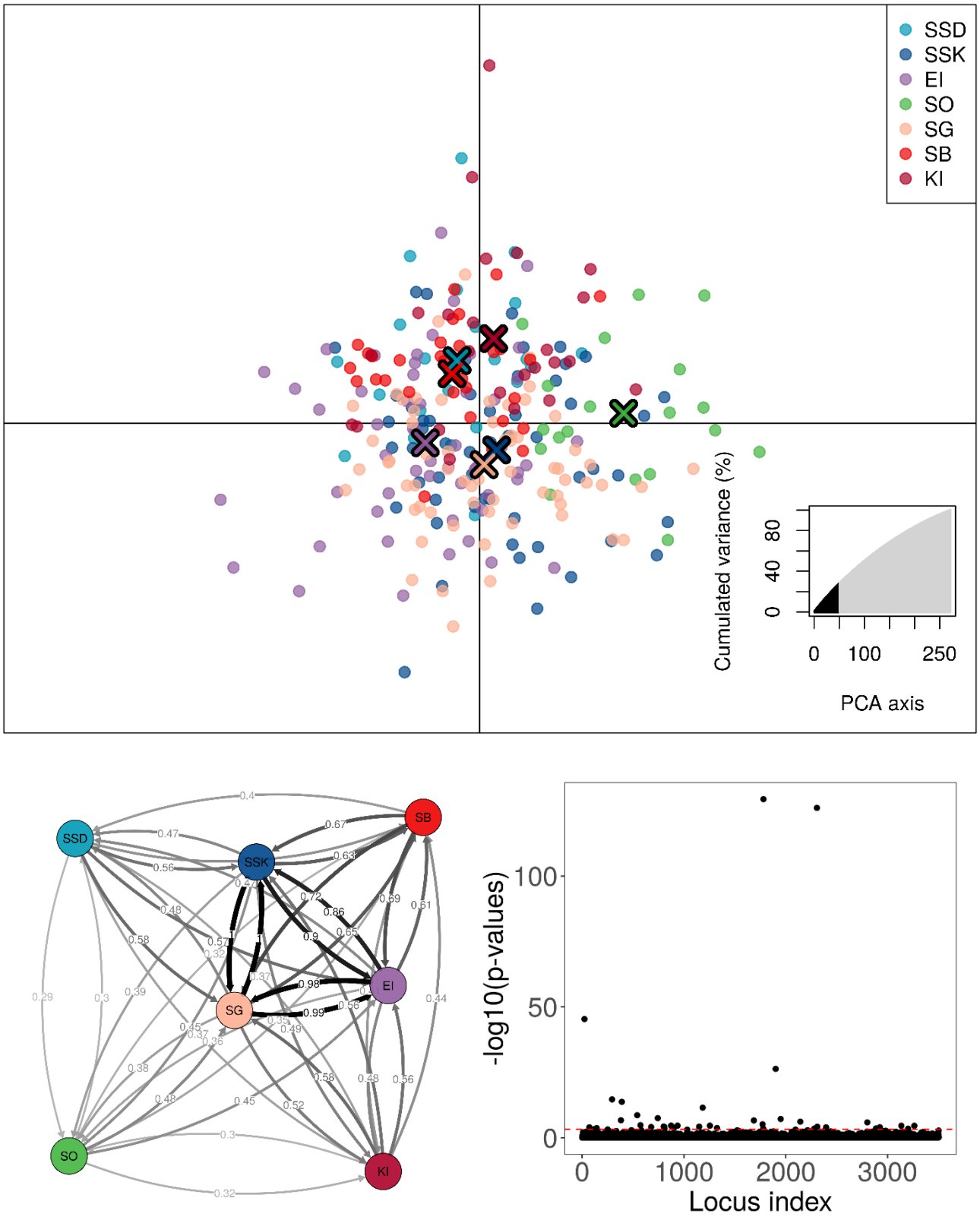


Figure S4.8.B. Genomic differentiation of *Notothenia rossii* in the Southern Ocean based on 3,503 SNP loci from reference-based variant calling. Geographic clustering as attempted through discriminant analysis of principal components is shown along the first two principal components (a). Relative migration as estimated from Nei's GST reveals overall high and no asymmetric gene flow (b). Genome scans for loci putatively under influence of selection detected 37 outliers at q > 0.05 (c). Sample codes as in Table 4.1; samples from different years but same locality are combined.

1. Modelling the Southern Ocean marine life: challenges and model performance

As partly reviewed in Guillaumot et al. (in press - Chapter 2), ecological modelling applied to Southern Ocean (SO) organisms faces many challenges given limitations and peculiarities of available datasets. In my PhD thesis, I aimed at highlighting the different biases that alter modelling performance and I developed several methods to correct for such issues. After all these steps, are settled corrections sufficient to generate accurate models ? And which are the remaining limits?

The quality of biological data: limitations and spatial aggregation.

My PhD results show that the limited number of occurrence records available for modelling is a recurrent issue in SO studies. The number of available occurrence data is not a problem (Stockwell and Peterson 2002, van Proosdij et al. 2016) as far as it respects the assumption that the sampled data cover the species full ecological range (Sánchez-Fernández et al. 2011, Raes 2012). However, this is not always the case in broad-scale SO SDMs (Guillaumot et al. 2020b - Chapter 2). The species prevalence, being the ratio between the species occupied space (represented by presence record locations) and the total surface of the study area (McPherson et al. 2004), can also influence model predictions. The model performance increases with decreasing prevalence (Barbet-Massin et al. 2012, Tessarolo et al. 2014, van Proosdij et al. 2016), meaning that for a similar projection area, a model based on more occurrence records would theoretically perform better. This explains the low performance obtained in SDMs generated at the scale of the SO. In Guillaumot et al. (2020b - Chapter 2), the modelled distribution of six sea star species barely showed any contrast between species ecological niches, although the six species are known to have contrasting feeding diets, reproductive strategies and trophic positions. In Guillaumot et al. (2018b), evaluation scores of SDMs predicting the distribution of species with wide ecological niches, were much weaker than those obtained for species with narrow niches.

Increasing the number of occurrence data available for modelling was shown to improve model predictions (Guillaumot et al. 2018a - Appendix, Fabri-Ruiz et al. 2019), with a small increase in a very small dataset resulting in a large benefit in model performance (van Proosdij et al. 2016). Using historical data to improve datasets does not impact model relevance as far as the species niche, or distribution, has not changed too much during the time period of reference, and as far as datasets are thoroughly checked for georeferencing errors and taxonomic inconsistencies (Newbold 2010, Guillaumot et al. 2018a - Appendix, Guillaumot et al. in press - Chapter 2). Adding absence records (even patchily distributed) to datasets, when available, is also a good option to improve the accuracy of presence-only model predictions (Peel et al. 2019) but the method was developed at the regional scale and never tested at a broader scale.

A limited number of data also limits the efficiency of model evaluation. For SDMs, a limited number of occurrence records was shown to imply a reduction in the number of training data available to generate the model (i.e. reducing modelling performance, Guillaumot et al. 2020b - Chapter 2) or to induce considerable limits when testing the model (Guillaumot et al. submitted - Chapter 3). In Guillaumot et al. (submitted - Chapter 3), solely 26 presence-only records were available to characterise the abiotic environment occupied by the sea urchin species *Abatus cordatus* in a bay of the Kerguelen Islands. This was sufficient to run the model, but methodologically difficult to subset a part of this dataset as independent test data. The model was therefore tested alternatively with presence records and evaluation metrics (Area Under the Curve) but more relevant evaluation strategies, using independent datasets are required to improve the quality of studies.

Regarding other modelling approaches, such as DEB and IBM models, it was also difficult to access supplementary information to evaluate models based on independent observations, as it is usually done (Marn et al. 2017, Haberle et al. 2020). Alternative approaches were used such as in Fabri-Ruiz et al. (in press - Chapter 3), where the profile method developed by Marques et al. (2019) and used by Stavrakidis-Zachou et al. (2019) helped calculate the marginal confidence intervals of the estimated DEB parameters and hence provided the uncertainty related to parameter estimation. In the population model developed in Arnould-Pétré et al. (2020) - Chapter 1, no time series on population densities nor structure were available to test model performance. A sensitivity analysis was applied, aiming at evaluating model stability regarding changes in parameter values (± 30% of the initial value). Such a method is referred to as the "robustness analysis". It evaluates model performance by a systematic deconstruction of the model, by forcefully changing model parameters, structure, and representation of processes (Grimm and Berger 2016, Railsback and Grimm 2019). Awaiting for more data for DEB model evaluation, another alternative, commonly used in our case studies was to compare DEB estimation mean relative errors (MRE, comparison between input observations and model predictions) with the average score of the DEB collection (MRE < 0.1, Marques et al. 2018). So far, DEB models built for benthic SO marine species showed good performances (*Abatus cordatus*: MRE= 0.121; *Adamussium colbecki* MRE: 0.08; *Nacella concinna* MRE: 0.203; *Laternula elliptica* MRE: 0.1; *Odontaster validus* MRE: 0.123; *Sterechinus neumayeri* MRE: 0.136; http://www.bio.vu.nl/thb/deb/deblab/add_my_pet/species_list.html, Agüera et al. 2015, 2017).

Regarding the influence of the spatial aggregation of occurrence records on model predictions, I showed that the targeted background data approach (Dormann 2007, Phillips et al. 2009, Syfert et al. 2013) can be applied and is efficient to correct modelling biases (Guillaumot et al. 2018a - Appendix). Data aggregation was also shown to have some negative effect on the performance of model evaluation (Guillaumot et al. 2019 - Chapter 2). I showed that it was possible to correct such a bias using spatial cross-validation procedures that spatially separate training and test data (Guillaumot et al. 2019 - Chapter 2), as previously stated in other works (Hijman 2012, Muscarella et al. 2014, Roberts et al. 2017). I strongly recommend to apply this method in future studies to improve modelling performance (application codes are available in Guillaumot et al. 2021 – Thesis material).

The poor quality of abiotic environment datasets.

The access to abiotic data with good temporal and spatial resolutions at the scale of the entire SO was challenging, a common issue in broad-scale oceanographic studies (Robinson et al. 2017). This implied to work with average values in time and large grid-cell pixels that do not accurately take into account environmental variability and complexity (Galante et al. 2018, Guillaumot et al. 2018a - Appendix), and biases the reliability of the occurrence-environment relationship, a strong prerequisite in correlative approaches (Morales et al. 2017, Araújo et al. 2019).

Given averaged values available in abiotic datasets, the large panel of ecoregions identified in the SO (Fabri-Ruiz et al. 2020) and the spatial aggregation of occurrence datasets, SDMs were shown to extrapolate, as training occurrence data may not encompass the entire range of environmental values existing in the study area (Guillaumot et al. 2020c - Chapter 2) or the whole possibilities of combinations between all environmental descriptors (Mesgaran et al. 2014). My results highlighted that areas where models extrapolate can cover high proportions (up to 75%) of the projection area (Guillaumot et al. 2020c - Chapter 2). It is therefore necessary to identify extrapolation areas and provide them along with prediction maps to facilitate model interpretation (Guillaumot et al. 2020c - Chapter 2). Extrapolation areas can be reduced by restraining model projection areas based on

species ecological or physiological tolerance thresholds, using experimental data or field observations (Guillaumot et al. 2020c - Chapter 2), which suggests that a good prior knowledge of the studied species is necessary before modelling.

Taxonomic uncertainties.

Dispersal through the Antarctic Circumpolar Current (ACC) and the ability of planktotrophic larvae to drift during several months in the water column both result in complex phylogeographic patterns, questioning the established taxonomy of many marine species in the SO (Hunter and Halanych 2010, Gonzalez-Wevar et al. 2011, Moreau et al. 2017, 2020). When generating a model at the scale of the SO, the uncertain taxonomic status of the studied species also questions the uniqueness of ecological niches such as species physiological performances and responses to environmental forcing factors. This may lead models to alternatively over- or under-estimate the predicted distribution of species suitable environments (Cacciapaglia and van Woesik 2017, Thyrring et al. 2017). To address this issue, I did my best to choose study species with little taxonomic uncertainties, based on the literature or discussions with experts (Guillaumot et al. 2020a - Chapter 1, Guillaumot et al. submitted - Chapter 3, Fabri-Ruiz et al. in press - Chapter 3). However, on-going taxonomic works inevitably question some results of my PhD. For example, recent findings on the sea star *Bathybiaster loripes* Sladen 1889, showed that it is a species complex structured according to depth (Moreau et al. 2020), although I formerly modelled its distribution as a single species (Guillaumot et al. 2020b - Chapter 2).

Even if not applied to SO species yet, it is possible to model the distribution of a species complex and consider local, potential adaptations within the SDM. Breiner et al. (2018) performed small ensemble SDMs to build distribution models based on spatial aggregates defined on genetic differentiation measures (F_{st}). The small, individual SDMs, are projected on sub-areas and the ensemble of results is then averaged to generate the total prediction of the entire species potential distribution (see also Breiner et al. 2015). Other approaches, when implementing the model, subsample the environment (Cacciapaglia and van Woesik 2017) or presence records (Gotelli and Stanton-Geddes 2015, Ikeda et al. 2017) according to the distribution of genetic entities. In other works, alternatives consist in introducing into the SDM a predictor that describes the species genotype (Banta et al. 2012).

Adapting models to the physiology and ecology of SO species.

Particular environmental conditions of the SO (e.g. extreme cold temperatures, seasonal food and light availability, ice dynamics and impact on shallow benthic communities; Vernet et al. 2008, Barnes et al. 2014, Cummings et al. 2018) necessitate some adaptations of physiological models formely developed for other conditions and organisms (Arnould-Pétré et al. 2020, Guillaumot et al. 2020a - Chapter 1). When creating a species model, the standard DEB approach calls pseudo-data, that are extra data coming from close taxa that help calibrate the model estimation similarly to a prior element (Lika et al. 2011a). Parameter estimation is thus performed at a reference temperature of +20°C (Kooijman 2010). This implies that results must be transformed to correctly interpret estimated parameters for polar species and some caution is necessary when comparing model outputs with those obtained for their counterparts from other regions of the world (Guillaumot et al. 2020a - Chapter 1).

SO species also present behavioural, morphological and physiological peculiarities that necessitate the development of new modules in existing physiological models (e.g. adaptation to the cold, Cziko et al. 2014; slower larval development and growth rates, Peck et al. 2007, Peck 2016; low protein production, Pörtner et al. 2007; direct development and brooding of the young, Moreau et al. 2017). In Arnould-Pétré et al. (2020) - Chapter 1, energy fluxes that determine the

development of the sea urchin species *Abatus cordatus* (Verrill, 1876) were constrained in the model due to the fact that juveniles remain isolated in brood pouches for eight months before they become autonomous to feed upon the sea bottom (Magniez 1983). Still in this study, the DEB model was complemented with estimated Arrhenius parameters to threshold metabolic performances within the species temperature tolerance (Kooijman 2010, Thomas and Bacher 2018), as it was already done for the SO species *Odontaster validus* (Agüera et al. 2015) and *Laternula elliptica* (Agüera et al. 2017). To do so, results of an experimental analysis that assesses species metabolic rates according to temperature variation are necessary, and up to five Arrhenius parameters, characterising the upper and down slopes of the Arrhenius curve can be implemented into the DEB model (Agüera et al. 2015, Lavaud et al. 2020). The method is powerful, but unfortunately, this kind of experimental data are not always available for SO marine species (Fabri-Ruiz et al. in press - Chapter 3) or datasets not complete enough to draw the ascending and descending parts of the Arrhenius curve (Arnould-Pétré et al. 2021 - Chapter 1, Guillaumot et al. submitted - Chapter 3), which limits the potential of DEB predictions, above all for temperature change simulations (Fabri-Ruiz et al. in press - Chapter 3).

The influence of data gaps on DEB physiological models.

The performance of DEB models is highly dependent on the completeness of available datasets. Lower significant predictive abilities are shown for models implemented with data gaps, both at the individual and population levels (Accolla et al. 2020). This impacts in turn the capacity of models to address the initial research objectives. For example, in this PhD, we applied the augmented-loss function method (Guillaumot et al. 2020a - Chapter 1) recently developed by Lika et al. (2020). In Lika et al. (2020), they obtained impressive detailed contrasts between their four catfish species (contrasts in predicted weights and sizes, in energy allocation or reproduction performance). However, in our case study, we only found minor differences in model predictions between populations of the Antarctic limpet (intertidal vs. subtidal morphotypes of *Nacella concinna*), although their respective morphology and physiology were proved to differ in field experiments and observations (Butlin et al. 2008, Hoffman et al. 2010). These results show that it was possible to build a DEB model for these two limpet morphotypes by extensively recycling data from the literature. However some data (above all) on larval and juvenile stages were missing and would have been necessary in the model to catch physiological and morphological differences that do exist between populations.

The biology of many SO marine benthic species is still poorly known to generate precise DEB models as it is difficult to study their life cycle *in situ* or *ex situ*. In the field, winter conditions make access to research stations and local investigations difficult, which complicates the settlement of experiments or observations. The lack of knowledge of species larval stage remains the main limitation (i.e. size and weight of eggs and larvae, feeding ability and ecology, precise knowledge of drifting duration in the water column, behaviour in the water column, metamorphosis event, Guillaumot et al. 2020a - Chapter 1, Christiansen et al. - Chapter 4). This is a common issue in the marine realm (Thorrold et al. 2002, Jones et al. 2008b), that is particularly important in the SO given the strong impact of marine currents (Sanches et al. 2016), the substantial duration of larval stages and drift (Stanwell-Smith et al. 1999), harsh weather conditions that constrain offshore sampling and difficulties encountered to bring and raise larvae in aquaria.

Integrating in SO benthic species DEB models an exhaustive overview of environmental conditions that influence species physiology is another limitation. Standard DEB models rely on two main forcing parameters: food availability and temperature (Kooijman 2010), but it is sometimes very

difficult to describe food availability for benthic species (Guillaumot et al. submitted - Chapter 3, Fabri-Ruiz et al. in press - Chapter 3). Generally, sea surface chlorophyll-a concentrations are used as a proxy of food availability (Arnould-Pétré et al. 2020 - Chapter 1, Guillaumot et al. submitted - Chapter 3), with few time series available to make the link between chlorophyll-a concentrations and organisms' nutrition (Agüera et al. 2015, 2017, Fabri-Ruiz et al. in press - Chapter 3). Even when the diet of some benthic species has been accurately studied and described (Dearborn et al. 1991, Calizza et al. 2018, Pascal et al. 2021), most of the time environmental data are not available at a precise spatial or temporal scale to describe in details food availability.

Hydrodynamics, sediment properties, sea ice coverage and dynamics, pH variation or light availability are also poorly documented in coastal areas. These factors strongly influence species distribution, survival and fitness (Grange and Smith 2013, Cummings et al. 2018) but cannot be included into models due to the lack of data (Arnould-Pétré et al. 2020, Guillaumot et al. 2020a - Chapter 1). This also explains why we did not manage to predict the morphological and physiological contrasts between the two morphotypes of *Nacella concinna* in Guillaumot et al. (2020a - Chapter 1). Morphological contrasts are linked to the stronger impact of waves in the intertidal zone compared to the subtidal and the more frequent time spent by individuals out of water (Beaumont and Wei 1991, Hoffman et al. 2010). However, only food and temperature could be used as environmental drivers in the model (Guillaumot et al. 2020a - Chapter 1). Including other environmental data in physiological models is therefore necessary to provide more precise and relevant outputs for SO species.

2. Using models to extrapolate: future simulations and invasive species

Climate change scenarios.

As in other regions on Earth, the SO is exposed to strong environmental changes (Henley et al. 2019) that have a cascading effect on marine species (Convey and Peck 2019). Modelling climate changes and their impact on species distribution is however challenging, as it is embedded in an extrapolation context, given that some future climatic conditions may not have modern analogs (Miller et al. 2004, Fitzpatrick and Hargrove 2009). SDM projections based on future conditions should be therefore considered with care. Furthermore, one of the main assumptions made by SDMs is that occurrence records and environmental conditions are at the equilibrium (Newbold 2010, Elith and Leathwick 2009, Václavík and Meentemeyer 2012). When not at the equilibrium (i.e. non-analog climate conditions), the species predicted distribution can be misinterpreted, as the model projects results for a new environment and does not integrate species interactions, population growth rates nor changes in dispersal abilities (Williams and Jackson 2007, Fitzpatrick and Hargrove 2009, Zurell et al. 2009, Woodin et al. 2013, Tingley et al. 2014).

The lack of future predictions for some environmental descriptors constitutes another issue. Water pH, primary productivity, current speed, are often not available which necessitates to assume that conditions are similar to present-day conditions (Guillaumot et al. 2018b, Fabri-Ruiz et al. in press - Chapter 3), which strongly limits the potential of simulations (Guillaumot et al. 2018b, Guillaumot et al. in press - Chapter 2, López-Farrán/Guillaumot et al. in press - Chapter 3). The poor spatial resolution of future environmental descriptors (1° ≅ 100 km on https://www.bio-oracle.org/) also considerably alters the precision and relevance of model future predictions for marine benthic species (Guillaumot et al. 2018b). Validating future predictions is not possible, which also constitutes a serious issue to address (Guillaumot et al. 2018b). In Fabri-Ruiz et al. (in press - Chapter 3), the distribution of the sea urchin *Sterechinus neumayeri* was modelled with both SDM

and a spatial projection of the DEB model. Contrasting results were obtained in predicted suitable areas between the two methods but I could not select the true prediction, as no validation could be performed. Similarly, in Arnould-Pétré et al. (2020) - Chapter 1, the population model of the sea urchin *Abatus cordatus* predicted population dynamics under future conditions could not be compared with time series of population densities. These future simulations are interesting to draw some preliminary hypotheses on the influence of some environmental changes on species distribution but results should really be considered with care as some important ecological information are not integrated into models.

As for physiological models, the lack of experimental data to correctly link temperature to physiological performance is a strong constraint to generate accurate future simulations (Arnould-Pétré et al. 2020 - Chapter 1, Fabri-Ruiz et al. in press - Chapter 3). However, when available, this can prove powerful to improve models, as it was done in Muller and Nisbet (2014) for simulating growth and calcification rates in the phytoplankton species *Emiliania huxleyi* under ocean acidification and changing ocean carbonate system or in Lavaud et al. (2021) that evaluated the influence of warmer and fresher waters on oyster growth, reproduction and mortality.
Regional or local-scale studies, with abundant experimental information that accurately describe species physiological performances in link with environmental conditions would provide more relevant projections for future simulations. In many cases, future scenarios are simulated with an ensemble of scenarios (range of potential temperature shifts, changes in one or several conditions, Mangano et al. 2019, Arnould-Pétré et al. 2020 - Chapter 1), in order to understand the relative influence of environmental shifts rather than providing a single approximated prediction.
Combining SDM predictions with experimental analyses was also shown to improve the relevance of future model predictions (López-Farrán/Guillaumot et al. in press - Chapter 3). This supplementary physiological information is indeed an interesting complement to evaluate the reliability of model predictions (Buckley et al. 2011, Greiser et al. 2020, López-Farrán/Guillaumot et al. in press - Chapter 3). The knowledge of species physiological tolerance for defined environmental conditions was also shown helpful to restrain the projection area and consequently limit extrapolation uncertainty (Feng et al. 2020, Guillaumot et al. 2020c - Chapter 2).
As a perspective, the Stack-SDM approach (S-SDM, Mateo et al. 2012), which has not been applied to SO species yet, integrates the information brought by several species-level models using a knowledge of species interactions (competition or predation) (Guisan and Rahbek 2011, Hortal et al. 2012b). The performance of SDMs generated for future scenarios is clearly improved when biotic interactions are integrated. This was exemplified in Davis et al. (2021), where the predicted distribution of a sea urchin facilitated the identification of potential kelp refugia in future environmental conditions. However this necessitates to study the influence of environmental conditions on each species and the consequence of future changes on species interactions (Freitas et al. 2007, Freitas 2011), which again necessitates more field observations.

Invasive species modelling.

Along with climate change, the SO experiences a consequent development of tourism, associated with an increasing number of cruise ships and visitors reaching Antarctic coasts (McCarthy et al. 2019). The risk of non-native species introduction and their survival therefore constitutes one critical threat to SO marine communities in the future (Hughes et al. 2020). SDMs tend to be used to address the issues of species introduction, but as for climate change scenarios, models predict in an extrapolation context (Robinson et al. 2010), with difficulties for SDMs to infer species potential distribution under novel environment settings compared to the species native range (Venette et al. 2010, Kumar et al. 2015). Such simulations also ignore ecological processes (e.g.

dispersal and biotic interactions) and thus provide an incomplete picture of invasion risks (Srivastava et al. 2019). Building a SDM for mapping species invasion risk has been done once in the SO for the case study of the Arctic sea star *Asterias amurensis* (Byrne et al. 2016), but it is a meticulous exercise that requires many tests, a thorough knowledge of the species ecology and environmental tolerance, and ideally to complete SDM predictions with information related to the species physiology, dispersal abilities or biotic interactions (Araújo and Guisan 2006, Srivastava et al. 2019). Some of these central points are not considered in Byrne et al. (2016).
In this PhD, the potential of the Patagonian crab *Halicarcinus planatus* to invade SO coastal areas was studied by associating SDM with physiological information from experiments on adult and larvae tolerance to contrasting salinity and temperature values (López-Farrán/Guillaumot et al. in press - Chapter 3) and completed with a study on the species dispersal potential using a Lagrangian approach (López-Farrán et al. in prep. - Appendix). These complementary analyses provided a complete overview of the species invasion risk.
This approach could be extended to the use of physiological models (such as DEB models) to complete our physiological knowledge of non-native species. DEB models could be used to investigate the bioclimatic envelope of non-native species (including adult and larval life stages) in order to evaluate their survival capacity under present or future conditions. This was already applied in Monaco and McQuaid (2018) to assess the ability of two South African bivalve species to survive across a steep intertidal environmental gradient. In Monaco et al. (2019), they also studied the acclimation abilities of the South African mussel in the Mediterranean Sea and Lavaud et al. (2021) predicted the growth and reproductive potential of the eastern oyster *Crassostrea virginica* in future scenarios.

DEB models could also be used to characterise biotic interactions between species of a community in which non-native species could potentially settle. The respective species physiology could be compared (Marn et al. 2019), and the influence of environmental conditions on each species life traits studied (Cardoso 2007, Marn et al. 2017). These models can be used to compare the species' fundamental niches and to provide an evaluation of the non-native species capacity to survive in the community (Fig.D). Occurrence records are positioned in space according to the environmental conditions recorded at different sites (x axis: food availability, y axis: temperature) and metabolic performances are calculated for each area, according to these environmental conditions. This helps indicate areas where species are well adapted to local conditions and in contrast, the limits of the species fundamental niche beyond which metabolism is less performant (Fig.D).

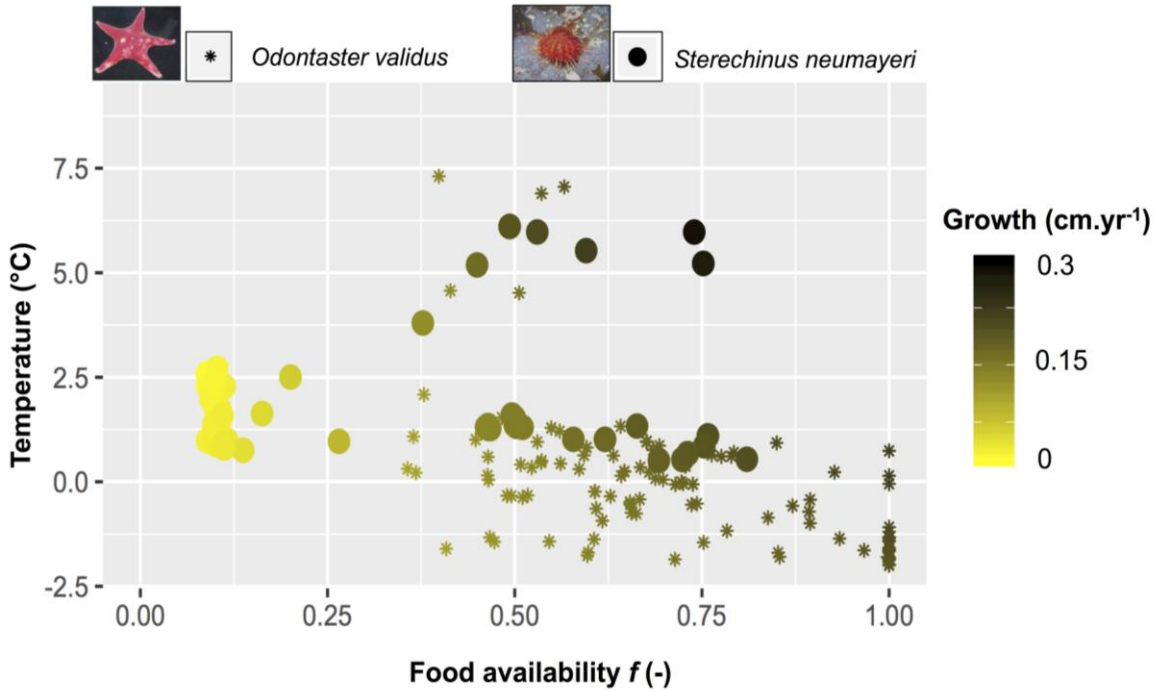


Figure D. Simple representation of species adaptation to local environmental conditions using a DEB model approach. X axis represents food conditions at each occurrence location, y axis temperature values. Environmental conditions are averaged values for summers 2002-2005. Star symbol corresponds to the sea star species *Odontaster validus* and filled circles to the sea urchin species *Sterechinus neumayeri.* The color bar indicates species metabolic performances, represented by growth rates (cm/year). The figure highlights areas where local environmental conditions are very favourable to the species development and in contrasts, the limits of their niche, where metabolim is less performant. Figure presented during the 2019 DEB Symposium.

Finally, associating DEB and spatial dynamic models such as Lagrangian models into an integrative approach enables to analyse the movement of species in complement to metabolic changes that are iteratively updated during the particle journey and following spatio-temporal variations of the environment (Bahlburg et al. 2021). This approach was already used in many marine studies to describe larval development along drifting periods under present (Ayata et al. 2010) or future environmental conditions (Lett et al. 2010, Lacroix et al. 2018, van de Wolfshaar et al. 2021), to evaluate species-specific responses to environmental changes (Falcini et al. 2020) or population connectivity between habitats (La Mesa et al. 2015, Thomas et al. 2020). The method could be applied to SO case studies by combining hydrodynamic models, available so far for the entire SO or smaller regions, with DEB models implemented for larvae or propagules. It constitutes a powerful analysis when studying the potential of an alien species to reach and survive along Antarctic coasts.

3. What did models tell us about the ecology of SO species ? How useful are the generated models ?

Models are a simple representation of a complex reality. After the many corrections introduced to take into account all methodological biases presented above, what did we finally learn about the ecology of SO species? In other words, how useful were the generated models even if they are all fundamentally wrong in the words of Box (1979)?

The generated SDMs were helpful to interpolate the distribution of species from hundreds of observations, including the distribution of the sea urchin *Abatus cordatus,* an endemic species of the Kerguelen Plateau (that belongs to the French marine protected reserve) (Guillaumot et al. 2018a,b, Guillaumot et al. submitted - Chapter 3) and the sea urchin *Ctenocidaris nutrix* (Guillaumot et al. 2018b), considered as a VME species (Vulnerable Marine Ecosystem) by CCAMLR (https://www.ccamlr.org/en/compliance/vulnerable-marine-ecosystems-vmes). Models showed good predictive performance. These models could be interesting for conservation and nature management, such as SDMs generated by Hibberd (2016) that used Random Forests to predict the distribution of several benthic species off Heard and McDonald Islands in the Australian Exclusive Economic Zone. Hill et al. (2017) mapped and quantified the distribution of demersal fish assemblages to gain ecological and management insights, using a multi-species model-based approach (Regions of Common Profile). The CCAMLR Scientific Committee composed Marine Protected Area proposals (CCAMLR report SC-CAMLR-38/BG/03 2019) by integrating outputs of distribution models, dynamic trophic models (Dahood et al. 2019) and population connectivity models (Piñones et al. 2013) of krill to identify sea mammals foraging areas and define fishery zones in the Western Antarctic Peninsula region.

My studies were also helpful to investigate the quality of occurrence datasets (i.e. number, spatial and temporal coverage). In Guillaumot et al. (2018a - Appendix), among four sea urchin studied species of the Kerguelen Plateau area, only one presented a dataset that fulfilled all methodological requirements to produce a reliable distribution model. In Guillaumot et al. (2020b - Chapter 2), I showed that occurrence datasets of six sea stars species were not complete enough to highlight contrasts between species modelled niches at the scale of the SO. Such results are however useful to identify knowledge gaps and guide future sampling plans (Guisan et al. 2006).

When enough data are available, model predictions were proved interesting to delineate species occupied environmental subspaces and thus helped describe environmental conditions preferentially occupied by the species along with the main abiotic descriptors driving the distribution (Guillaumot et al. 2018b, López-Farrán/Guillaumot et al. in press - Chapter 3, Guillaumot et al. submitted - Chapter 3). This was proved particularly relevant and powerful when distribution models are combined with experimental data (López-Farrán/Guillaumot et al. in press - Chapter 3). Such results can be used to interpret the potential response of species to changes in some environmental factors (Guillaumot et al. 2018b, López-Farrán/Guillaumot et al. in press - Chapter 3) and guide conservation strategies accordingly (Guillaumot et al. 2018b).
Similarly, DEB physiological models were interesting to disentangle the respective importance of food availability and temperature on species metabolic performances (Fabri-Ruiz et al. in press - Chapter 3), to describe the species life cycle (Agüera et al. 2015) or to determine species feeding histories according to observed gonadal cycles (Agüera et al. 2017 -Appendix). They also helped highlight the negative effect of increasing temperatures on metabolic costs (*Abatus cordatus,* Guillaumot et al. submitted - Chapter 3; *Sterechinus neumayeri,* Fabri-Ruiz et al. in press - Chapter 3) and population mortality (*Abatus cordatus*, Arnould-Pétré et al. 2020 - Chapter 1).
Applying some new modules of DEB modelling could help increase models ability to describe the relationship between environmental conditions and the organism's metabolism. The standard DEB model only considers food resources and temperatures as environmental drivers of metabolism performances (Kooijman 2010) but complementary model-based modules have already been developed to include the effect of tide cycles influence (Monaco and McQuaid 2018), pH/CO_2

concentrations (Troost et al. 2010, Wijsman and Smaal 2011, Klok et al. 2014a), salinity (Lavaud et al. 2017), O_2 concentration (Lavaud et al. 2019) or exposure to suspended sediment particle loads (La Peyre et al. 2020). Applying this modelling approach to SO case studies is feasible, as long as enough observations are available.

I showed that food availability always has a low influence on species physiological performances compared to temperature (Arnould-Pétré et al. 2020 - Chapter 1, Fabri-Ruiz et al. in press - Chapter 3). This may be consistent with our knowledge of physiological peculiarities of SO organisms, namely metabolisms adapted to low temperatures (Peck 2016, Peck et al. 2018), with poor abilities to acclimate to warmer temperatures (Peck et al. 2014) and a noticeable adaptation to low and seasonal food inputs (Lawrence and McClintock 1994). However, this may also be partly explained by an erroneous representation of food availability and a misleading evaluation of the relationship between food availability and energetic performances as it is implemented in the DEB model (Arnould-Pétré et al. 2020 - Chapter 1). First, the scaled functional response *f*, comprised between 0 and 1, is commonly used in DEB models to represent food availability (Kooijman 2010) but it is too general to accurately characterise food availability for a benthic species. Further, sea surface chlorophyll-a concentration was often used as a relative poor proxy of food availability, as previously stated. Considering the approach of Jansen et al. (2018), that assessed the redistribution of surface productivity at the seafloor using a pelagic-benthic coupling approach, could be an interesting perspective. In that study, a regional ocean model with remotely sensed sea-surface chlorophyll-a maps was combined with data on diatom abundances obtained from sediment grabs and with a particle tracking approach to infer the potential density of food available on the seabed. Detailed information on the species surrounding habitat is necessary to apply this method, but it could substantially improve the representation of food availability for application of physiological models to benthic species.
Another modelling perspective consists in considering several food resources to represent energy supply to organisms in DEB models (Galasso et al. 2020, Reid et al. 2020). This existing modelling framework is based on empirical results on growth performance according to food quality (Galasso et al. 2020). Following these works, prospective studies will link DEB theory with stable isotope dynamics ("Dynamic Isotope Budget", DIB, talk given by Lefebvre et al. at DEB Symposium 2019). The method that describes stable isotope fluxes within organisms already exists (Pecquerie et al. 2010) and Lefebvre et al. (DEB Symposium 2019) objectives are to put the analysis in the framework of trophic network studies and relate organism diet with energetics. Trophic network analyses have been developed for the SO region (Ducklow et al. 2006, 2007, 2013, Western Antarctic Peninsula; Murphy et al. 2007, Scotia sea pelagos; Hill et al. 2012, South Georgia pelagos; Pinkerton et al. 2013, 2014, Ross Sea pelagos; Ballerini et al. 2014, Western Antarctic Peninsula; Ortiz et al. 2016, Marina et al. 2018 and Zenteno et al. 2019, King George Island) and could constitute a good basis to apply this method.

Finally, dispersal models were proved helpful to highlight the role of the ACC as a barrier to the dispersal of propagules arriving from neighbour continents (López-Farrán et al. in prep. - Appendix) and also connect the SO regions between each other (Dulière/Guillaumot et al. submitted - Chapter 4, Christiansen et al. in prep.- Chapter 4). Such information on species dispersal abilities are really interesting to further understand species realised niche (Christiansen et al. in prep. - Chapter 4) and test dispersal scenarios (López-Farrán et al. in prep. - Appendix). Combined together, the three studies of this PhD performed on the Patagonian crab *Halicarcinus planatus* (López-Farrán et al. in prep. - Appendix, López-Farrán/Guillaumot et al. in press - Chapter 3, Dulière/Guillaumot et al. submitted - Chapter 4) suggested that the crab found in Deception Island

in 2010 could not have naturally crossed the Drake Passage to reach the Western Antarctic Peninsula coasts, as the ACC targeted its drift eastward along the 60°S latitude line (López-Farrán et al. in prep. - Appendix). A natural arrival could be by rafting over buoyant kelps (Fraser et al. 2018), but it constitutes a questionable alternative, which is possible during a stormy weather event, and suggests the survival of the adult crab during its long way, attached on macroalgae holdfasts (López-Farrán/Guillaumot et al. in press - Chapter 3). The hypothesis of an accidental anthropogenic introduction was therefore suggested, either by transport on ship hulls or release of ballast waters. Results from Dulière/Guillaumot et al. (submitted - Chapter 4) highlighted a higher probability of its arrival in Deception Island, whether ballast waters were exchanged at 50 nautical miles from the coasts or closer. The respect of the Antarctic Treaty guidelines, promoting ballast water release at least at 200 nautical miles away from the coasts, prevented the crab propagules from reaching the western coasts of the Western Antarctic Peninsula (Dulière/Guillaumot et al. submitted - Chapter 4). Once introduced, physiological experiments showed that it is difficult for the crab to survive in present Antarctic winter conditions (López-Farrán/Guillaumot et al. in press - Chapter 3), but suggested a potential settlement of this non-native species under future conditions, when warmer temperatures will allow its physiological acclimation (López-Farrán/Guillaumot et al. in press - Chapter 3).

4. Personal feedbacks on DEB and SDM applications.

Mechanistic models: DEB theory.

Applying DEB models to SO species was not so easy and took several months from initiation to model finalisation. A lot of concepts have to be understood (van der Meer 2006, Kooijman 2010) to generate a model and to be able to interpret the link between DEB parameters and species' physiological traits (van der Meer 2006). Hopefully, an increasing number of available tutorials (see the DEBwiki page https://en.wikipedia.org/wiki/Dynamic_energy_budget_theory) and Matlab codes help implement DEB models (https://www.bio.vu.nl/thb/deb/) including a newly developed module that automatically fills Matlab codes (*AmPeps*, http://www.bio.vu.nl/thb/deb/deblab/add_my_pet/AmPeps.html). Moreover, every two years, DEB schools are organised and include training courses on DEB model applications (https://deb2021.sciencesconf.org/). They are completed with a 3-weeks MOOC to learn DEB theory principles. These events have been organised for several years and considerably help increase the number of DEB applications (https://www.zotero.org/groups/500643/deb_library/library).

Alternatives to the DEB theory exist, as well as internal alternatives of the standard DEB model that rely on different assumptions regarding energy allocation priorities or new parameters to describe individual metabolism (comments on DEB3 Kooijman 2010, Lika and Kooijman 2011). The DEBkiss model ('kiss' for 'keep it simple stupid') proposes a simple version of the standard DEB (no reserve compartment) and is widely used for routine applications such as for toxicity analyses (Jager et al. 2013, Jager 2016). However, according to Lika and Kooijman (2011), this model also presents weaknesses compared to the standard DEB procedure and cannot be used as an alternative but as a first step before more complex implementations are carried out. Several other conceptual models of individual metabolism also exist, including the WBE West, Brown and Enquist theory (Brown et al. 2004) according to which energy supply is determined by branching networks that change with body size (West et al. 2002, Kearney and White 2012), or the Scope for Growth approaches that specifically focus on a precise metabolic process (Filgueira et al. 2011) and can integrate empirical information (Table 11.1 in Kooijman 2010). Among all these approaches, the DEB theory is the most applied.

Implementing DEB models requires detailed information on species physiology, morphology or energetics, according to environmental conditions and for different life stages (Jusup et al. 2017, Guillaumot et al. 2020a - Chapter 1). When available in the literature, these data can be sufficient enough to parameterize a DEB model for a SO species (Guillaumot 2019a). Unfortunately, this is rarely the case, these data are missing for most species depending on model objectives and on the expected details (Guillaumot et al. 2020a - Chapter 1). Upscaling the DEB approach at the population level (DEB-IBM) was proved time and data consuming (Arnould-Pétré et al. 2020 - Chapter 1). It required a thorough knowledge of the species ecology, physiology, and of the surrounding environment to model processes that drive population dynamics (i.e. mortality, interspecific relationships, reproduction) (Beaudouin et al. 2015, Groeneveld et al. 2015, Arnould-Pétré et al. 2020 - Chapter 1, Groeneveld et al. 2020). However, once applied, DEB models are extremely interesting and provide a large panel of outputs that can be applied to all species development states (e.g. larval development, growth rate, energy allocation description, reproduction performance, survival potential in given food and temperature conditions) and consequently allow a thorough understanding of species physiology and its link to the environment (Aguëra et al. 2015, Marn et al. 2017, Arnould Pétré et al. 2020 - Chapter 1, Haberle et al. 2020, Guillaumot et al. submitted - Chapter 3). The use of DEB models for SO marine species is therefore totally adapted. I also showed it is really interesting to combine DEB estimation of the species fundamental niche along with other modelling approaches that focus on other parts of the ecological niche, such as correlative approaches (i.e. the realised niche) and dispersal models.

DEB modelling is a powerful approach that could open to other applications for SO studies such as ecotoxicology analyses. The amount of works studying the influence of pollutants on SO marine species has been increasing for a few years (Ansari et al. 2004, Poulsen et al. 2012, Majer et al. 2014, Furtado et al. 2019), including microplastics (Jovanović 2017, Cappello et al. 2021). DEB theory was proved efficient in ecotoxicology, at the individual or population scale (Martin et al. 2013), to describe the sublethal effects of toxicants (Muller et al. 2010, Sherborne and Galic 2020), to model the uptake, elimination and (metabolic) transformation of the toxic compounds (Kooijman et al. 2009, Pousse et al. 2019) or using the principle of Synthesizing Units to represent the impact of inhibitors and damaging agents on enzyme kinetics (Muller et al. 2019). Only some experiments and specific information regarding the pollution and its effect on metabolic processes would be necessary before such modelling analyses can be used for SO studies. Some works have already been led in the Arctic to assess the impact of petroleum substances on copepods (Klok et al. 2012a), fish (Klok et al. 2014b) or food chains (Klok et al. 2012b).
To put it in a nutshell, the development of DEB modelling can be really an interesting perspective for future SO modelling works.

Correlative models: SDMs

I found SDMs easy to implement and general principles easy to understand. Some tutorials are available to quickly learn how to generate these models (Naimi and Araújo 2016, Hijmans and Elith 2017, Oliver 2018, Barbosa 2020). To build a SDM, the compilation of occurrence data is generally the most time-consuming part as it often requires to implement databases (Guillaumot et al. 2016, Fabri-Ruiz et al. 2017a, Moreau et al. 2018), to check for data georeferencing and taxonomic accuracy, or to delete duplicates generated during the automatisation of online and free-access catalogs (Moreau et al. 2018). Similarly, to compile the set of environmental descriptors, it is often necessary to get familiarized with Geographic Information Systems (GIS) and homogeneize and stack the ensemble of raster layers that will be necessary to generate the SDM. The choice of

environmental descriptors also requires some ecological knowledge of the studied species, and consequently to go through the literature (Guillaumot et al. 2018a - Appendix). Some statistical analyses should also be performed to study potential correlations between these descriptors, in order to limit model complexity (Guillaumot et al. 2020b - Chapter 2) and resulting biases in model predictions (Harisena et al. 2021). Several syntheses were published on the subject, describing the different steps of dataset preparation (Elith et al. 2006, Peterson et al. 2011, Dormann et al. 2012b), including my recent review focussed on SO case studies (Guillaumot et al. in press - Chapter 2).
However, despite the fact that SDM principles are easy to understand, the choice of the algorithm may generate additional complexity to analyse the link between occurrence records and the environment (Olden et al. 2008, Elith and Leathwick 2009). A good trade-off must be found between generalist algorithms that are insufficient to describe the relationship between the environment and species occurrences, and complex ones that overfit based on training data and cannot efficiently transfer in space or time (Syfert et al. 2013, Merow et al. 2014). For SO studies, BRT (Boosted Regression Trees, Elith et al. 2008) and Random Forests (RF, Breiman 2001) were compared to other algorithms and were proved performant to deal with missing environmental data and presence-only records. They can also easily integrate supplementary modules to correct for spatial aggregation for example (Guillaumot et al. 2018a - Appendix, Fabri-Ruiz et al. 2019). However these two machine learning algorithms are greedy in calculation time and generally require the use of a public cluster to be launched. They also may overfit to datasets, leading to limited transferability performances (Heikkinen et al. 2012, Wenger and Olden 2012, Crimmins et al. 2013). Interpretation should therefore be always done with caution.

My PhD results suggest that modellers should consider with caution SDMs when occurrence datasets are not complete enough to correctly cover the species full ecological range. This is notably the case for SDMs performed at the scale of the SO, but it could also be the case at a regional scale (see section 1). Results should also be treated with care when models extrapolate under future climate scenarios or to study the potential of alien species to invade new areas (see section 2). Apart from these particular applications, SDMs can prove really useful to interpolate species distribution and thus are interesting to fill knowledge gaps on SO species biogeography and ecology. I developed many corrections in a synthesis paper (Guillaumot et al. in press - Chapter 2) and the SDMPlay R package (Guillaumot et al. 2021 – Thesis material) can help users implement models for SO studies. I also recommend modellers, whenever possible, to provide maps that represent model uncertainties along with model predictions (Rocchini et al. 2011, Guillaumot et al. 2020c - Chapter 2) in order to give realistic estimates of confidence intervals around model predictions (Beale and Lennon 2012) and ensure accurate interpretations (Beale and Lennon 2012, Addison et al. 2013, Guisan et al. 2013).

Finally, complementing SDMs with additional physiological information was proved powerful to improve the description of the species realised niche (Guillaumot et al. submitted, López-Farrán/Guillaumot et al. in press - Chapter 3). This could also be done with the addition of information on biotic interactions (Araújo and Luoto 2007, Heikkinen et al. 2007, Anderson 2013, Dormann et al. 2018), one key feature in the BAM diagram model (Soberón 2007, Peterson et al. 2011, Sillero 2011, Saupe et al. 2012). Integration of biotic data can be done by restricting the predicted distribution of a given species by the occurrence of another one (Schweiger et al. 2012), or by using occurrence (or abundance) data of other species as predictors within the abiotic descriptor dataset (Leathwick and Austin 2001, Leathwick 2002, Meier et al. 2010, Pellissier et al. 2010, Bebber and Gurr 2019). This could constitute a strong complement in understanding species

distribution and community structure (Hellmann et al. 2012, Wisz et al. 2013, Alexandridis et al. 2017). However, this targets much more complex modelling approaches. Ecological networks vary along environmental gradients (Pellissier et al. 2017) and some species present strong feeding plasticity (Michel et al. 2016, 2019) which implies that a lot of observations must be performed at local scale to accurately characterise species interactions within communities.

5. Concluding remarks

Ecological modelling has considerably been used during the past two decades to study SO marine species. Modelling is a way to synthetise different information gathered by several scientific teams and during several campaigns. Model maps are handy to interpret, and thus help bridge between scientists and politicians during decision-making processes. Models are interesting to quickly identify knowledge gaps and set research priorities. They are attractive, as they can integrate different types of data and information and can help recycle left-behind historical data. They are helpful to simplify complex processes, therefore offering the possibility of making preliminary assumptions before further researches are conducted. Finally, within a few months of reading and training, anyone can now generate a model to predict species distribution or model an organism's metabolism. Modelling is not only accessible to people having a mathematical background. The impressive collection of articles, R packages or tutorials that guide people to create a model, using open source available scripts. All these points can explain the growing popularity of the approach, as experienced in the community of Antarctic biologists.

However, several issues should be stressed, as final conclusions to this PhD work.

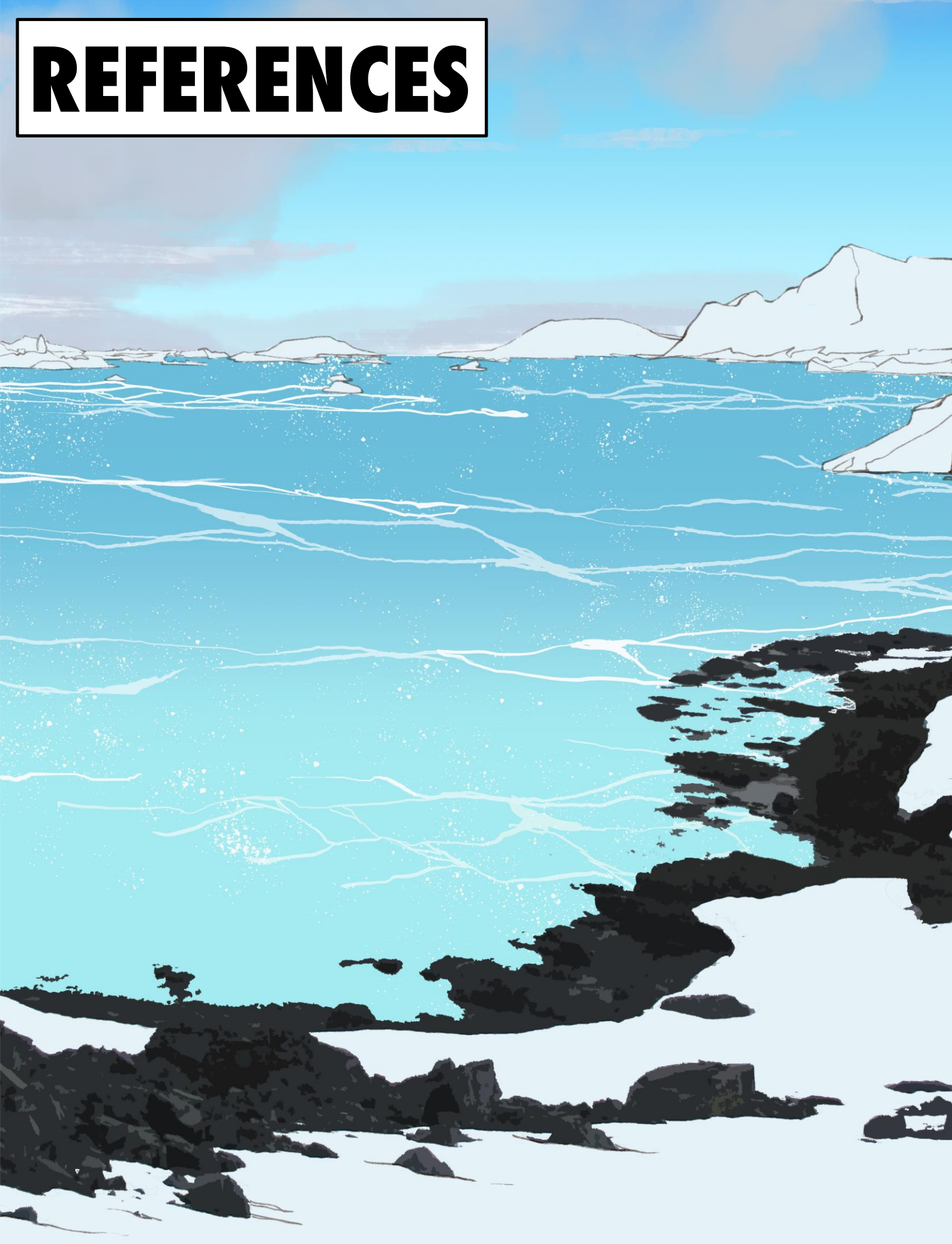
(i) It is essential to anticipate the possibility of evaluating model's predictive performance before creating it (Grimm and Berger 2016, Railsback and Grimm 2019). Model evaluation is indeed not always considered with enough importance in SO modelling applications. The evaluation method is often not adapted to the dataset (Guillaumot et al. 2019 - Chapter 2) or the model not evaluated at all (Griffiths et al. 2017).

(ii) What is the point of generating future climate change simulations without accurate input data? IPCC scenarios are global average scenarios, most of them are not adapted to study species ecology (Cavanagh et al. 2017) and methodological difficulties prevent from incorporating climate scenario uncertainties into model predictions (Freer et al. 2018). Climate scenarios represent the environment with a coarse spatial resolution (100 km) and most environmental conditions are not available. Therefore, doubtful assumptions are usually made for future environmental conditions with no possibility to evaluate final model predictions. Therefore, I would suggest to consider ecological simulations based on future climate scenarios with a very critical eye.

(iii) Results of my PhD suggest that we are not ready to generate distribution models at the scale of the entire SO, as the quality of environmental and occurrence datasets is not sufficient to precisely distinguish contrasts between species (Guillaumot et al. 2020b - Chapter 2). I would recommend to run models at the regional scale and when data are abundant enough to describe environmental conditions, and when the relationship between species and their environment (i.e. abiotic conditions, biotic interactions and dispersal abilities) has been the subject of former studies that provide a detailed overview of species ecological preferences and interactions. Providing uncertainty maps along with model results is also strongly encouraged.

The SO is not as poorly known as often stated (Griffiths 2010). A lot of oceanographic campaigns were undertaken and many datasets are available to modellers. The development of underwater imagery also considerably helps collecting new data. It is just a matter of time before the completeness and quality of datasets can be improved and more accurate models generated. Open-access databases, at the regional scale, could be really helpful for improving modelling studies. Such ideas are progressively being developed (J. Stark MEASO congress, April 2018; special issue https://www.mdpi.com/journal/diversity/special_issues/Ross_Sea_Marine). This would allow an efficient international collaborative sampling and would be also really interesting to support conservation decisions. To conclude, results of my PhD thesis will be, I hope, useful for future modelling works applied to SO species and for conservation purposes. I hope the provided guidelines will be helpful to adjust model predictions, improve their accuracy, relevance and facilitate the interpretation of model outputs.

REFERENCES



A

Aanensen, D.M., Huntley, D.M., Feil, E.J. & Spratt, B.G. (2009). *EpiCollect*: linking smartphones to web applications for epidemiology, ecology and community data collection. *PloS One*, *4*(9), e6968.

Abrams, P.A., Ainley, D.G., Blight, L.K., Dayton, P.K., Eastman, J.T. & Jacquet, J.L. (2016). Necessary elements of precautionary management: implications for the Antarctic toothfish. *Fish and Fisheries*, *17*(4), 1152-1174.

Accolla, C., Vaugeois, M., Rueda-Cediel, P., Moore, A., Marques, G.M., Marella, P. & Forbes, V.E. (2020). DEB-tox and Data Gaps: Consequences for individual-level outputs. *Ecological Modelling*, *431*, 109107.

Addison, P.F., Rumpff, L., Bau, S.S., Carey, J.M., Chee, Y.E., Jarrad, F.C., ... & Burgman, M.A. (2013). Practical solutions for making models indispensable in conservation decision-making. *Diversity and Distributions*, *19*(5-6), 490-502.

Agassiz, A. (1881). Report on the Echinoidea dredged by H.M.S. Challenger during the years 1873-1876. Report on the scientific results of the voyage of H.M.S. Challenger during the years 1873-1876, Zoology, 3, 1-321.

Agostini, C., Patarnello, T., Ashford, J.R., Torres, J.J., Zane, L. & Papetti, C. (2015). Genetic differentiation in the ice-dependent fish *Pleuragramma antarctica* along the Antarctic Peninsula. *Journal of Biogeography*, *42*(6), 1103-1113.

Aguiar, L.M., da Rosa, R.O., Jones, G. & Machado, R.B. (2015). Effect of chronological addition of records to species distribution maps: The case of *Tonatia saurophila maresi* (Chiroptera, Phyllostomidae) in South America. *Austral Ecology*, *40*(7), 836-844.

Aguirre-Gutiérrez, J., Carvalheiro, L.G., Polce, C., van Loon, E.E., Raes, N., Reemer, M. & Biesmeijer, J.C. (2013). Fit-for-purpose: species distribution model performance depends on evaluation criteria–Dutch hoverflies as a case study. *PloS One*, *8*(5), e63708.

Agüera, A., Collard, M., Jossart, Q., Moreau, C. & Danis B. (2015) Parameter estimations of Dynamic Energy Budget (DEB) model over the life history of a key Antarctic species: the Antarctic sea star *Odontaster validus* Koehler, 1906. *PloS One*, 10(10), e0140078.

Agüera, A., Ahn, I-Y., Guillaumot, C. & Danis, B. (2017). A Dynamic Energy Budget (DEB) model to describe *Laternula elliptica* (King, 1832) seasonal feeding and metabolism. *PloS one*, *12*(8), e0183848.

Agusto, M.R., Caselli, A.T. & dos Santos Afonso, M. (2004). Manifestaciones de piritas framboidales en fumarolas de la Isla Decepción (Antártida): Implicancias genéticas. *Revista de La Asociacion Geologica Argentina*, 59(1), 152–157.

Ahn, I-Y., Surh, J., Park, Y.G., Kwon, H., Choi, K.S., Kang, S.H., ... & Chung, H. (2003). Growth and seasonal energetics of the Antarctic bivalve *Laternula elliptica* from King George Island, Antarctica. *Marine Ecology Progress Series*, *257*, 99-110.

Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B. & Anderson, R.P. (2015). *spThin*: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5), 541-545.

Ainley, D.G. & Pauly, D. (2014). Fishing down the food web of the Antarctic continental shelf and slope. *The Polar Record*, *50*(1), 92.

Albers, C.S., Kattner, G. & Hagen, W. (1996). The compositions of wax esters, triacylglycerols and phospholipids in Arctic and Antarctic copepods: evidence of energetic adaptations. *Marine Chemistry*, *55*(3-4), 347-358.

Alexander, F.J., King, C.K., Reichelt-Brushett, A.J. & Harrison, P.L. (2017). Fuel oil and dispersant toxicity to the Antarctic sea urchin (*Sterechinus neumayeri*). *Environmental Toxicology and Chemistry*, *36*(6), 1563-1571.

Alexandridis, N., Dambacher, J.M., Jean, F., Desroy, N. & Bacher, C. (2017). Qualitative modelling of functional relationships in marine benthic communities. *Ecological Modelling*, *360*, 300-312.

Allan, M.G. (2014). Remote sensing, numerical modelling and ground truthing for analysis of lake water quality and temperature (Doctoral dissertation, University of Waikato).

Allcock, A.L. & Strugnell, J.M. (2012). Southern Ocean diversity: new paradigms from molecular ecology. *Trends in Ecology & Evolution*, *27*(9), 520-528.

Allouche, O., Tsoar, A. & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223-1232.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997).

Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389-3402.

Alvarado, J.L. & Castilla, J.C. (1996). Tridimensional matrices of mussels *Perumytilus purpuratus* on intertidal platforms with varying wave forces in central Chile. *Marine Ecology Progress Series*, 133(1–3), 135–141.

Améziane, N., Eléaume, M., Hemery, L., Monniot, F., Hemery, A., Hautecoeur, M. & Dettai, A. (2011). Biodiversity of the Benthos off Kerguelen Islands: Overview and Perspectives, in: Duhamel, G., Welsford, D. (Eds), The Kerguelen Plateau, marine ecosystem and fisheries, Société Française d'Ichtyologie, pp. 157-167.

Amsler, C.D., McClintock, J.B. & Baker, B.J. (1999). An Antarctic feeding triangle: defensive interactions between macroalgae, sea urchins, and sea anemones. *Marine Ecology Progress Series*, *183*, 105-114.

Anderson, B.J., Akçakaya, H.R., Araújo, M.B., Fordham, D.A., Martinez-Meyer, E., Thuiller, W. & Brook, B.W. (2009). Dynamics of range margins for metapopulations under climate change. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1661), 1415-1420.

Anderson, R.P. & Raza, A. (2010). The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. *Journal of Biogeography*, 37(7), 1378-1393.

Anderson, R.P. & Gonzalez Jr, I. (2011). Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecological Modelling*, 222(15), 2796-2811.

Anderson, R.P. (2013). A framework for using niche models to estimate impacts of climate change on species distributions. *Annals of the New York Academy of Sciences*, 1297(1), 8-28.

Anderson, O.F., Guinotte, J.M., Rowden, A.A., Tracey, D.M., Mackay, K.A., & Clark, M.R. (2016). Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand. *Deep Sea Research Part I: Oceanographic Research Papers*, 115, 265-292.

Anderson, R.P. (2017). When and how should biotic interactions be considered in models of species niches and distributions? *Journal of Biogeography*, 44(1), 8-17.

Anderson, C.B. (2018). Biodiversity monitoring, earth observations and the ecology of scale. *Ecology Letters*, *21*(10), 1572-1585.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequencing data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. & Hohenlohe, P.A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81.

Ansaldo, M., Sacristán, H. & Wider, E. (2007). Does starvation influence the antioxidant status of the digestive gland of *Nacella concinna* in experimental conditions? *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 146(1-2), 118-123.

Ansari, T.M., Marr, I.L. & Tariq, N. (2004). Heavy metals in marine pollution perspective-a mini review. *Journal of Applied Sciences*, *4*(1), 1-20.

Ansorge, I.J., Durgadoo, J.V. & Treasure, A.M. (2014). Sentinels to climate change. The need for monitoring at South Africa's Subantarctic laboratory. *South African Journal of Science*, *110*(1-2), 1-4.

Antarctic Treaty (1959). signed in Washington. DC on December, 1, 402. Pdf available at https://documents.ats.aq/keydocs/vol_1/vol1_2_AT_Antarctic_Treaty_e.pdf, accessed November 2019.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., ... & Yeh, L.S. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, *32*(1), D115-D119.

Arango, C.P., Soler-Membrives, A. & Miller, K.J. (2011). Genetic differentiation in the circum-Antarctic sea spider *Nymphon australe* (Pycnogonida; Nymphonidae). *Deep Sea Research Part II: Topical Studies in Oceanography*, 58(1-2), 212-219.

Aranzamendi, M.C., Sahade, R., Tatián, M. & Chiappero, M.B. (2008). Genetic differentiation between morphotypes in the Antarctic limpet *Nacella concinna* as revealed by inter-simple sequence repeat markers. *Marine Biology*, 154(5), 875-885.

Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005). Validation of species-climate impact models under climate change. *Global Change Biology*, *11*(9), 1504-1513.

Araújo, M.B. & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33(10), 1677-1688.

Araújo, M.B. & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, *16*(6), 743-753.

Araújo, M.B. & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, *22*(1), 42-47.

Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., ... & O'Hara, R.B. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), eaat4858.

Aravossis, K. & Pavlopoulou, Y. (2013). Creating shared value with eco-efficient and green chemical systems in ship operations and in ballast water management. *Fresenius Environmental Bulletin*, *22*(12a), 1-9.

Ardron, J.A., Possingham, H.P. & Klein, C.J. (2008). Marxan good practices handbook. *Pacific Marine Analysis and Research Association, Vancouver*, 149.

Arnaud, P.M. (1974). Contribution a la bionomie benthique antarctique et subantarctique. Ph. D. Dissertation, Station marine d'Endoume, Marseille.

Arnould-Pétré, M., Guillaumot, C., Danis, B., Féral, J-P. & Saucède, T. (2020) Individual-based model of population dynamics in a sea urchin of the Kerguelen Plateau (Southern Ocean), *Abatus cordatus*, under changing environmental conditions. *Ecological Modelling*, 440, 109352.

Arntz, W.E. & Gallardo, V.A. (1994). Antarctic benthos: present position and future prospects. In *Antarctic Science* (pp. 243-277). Springer, Berlin, Heidelberg.

Arntz, W.E., Gutt, J. & Klages, M. (1997). Antarctic marine biodiversity an overview. Antarctic communities: Proc 6th SCAR Biology Symposium, Venice 1994 (B Battaglia, J Valencia, D Walton) Cambridge Univ Pr, Cambridge.

Arntz, W.E., Thatje, S., Linse, K., Avila, C., Ballesteros, M., Barnes, D.K., ... & Teixido, N. (2006). Missing link in the Southern Ocean: sampling the marine benthic fauna of remote Bouvet Island. *Polar Biology*, 29(2), 83-96.

Aronson, R.B. & Blake, D.B. (2001). Global climate change and the origin of modern benthic communities in Antarctica. *American Zoologist*, *41*(1), 27-39.

Aronson, R.B., Thatje, S., Clarke, A., Peck, L.S., Blake, D.B., Wilga, C.D. & Seibel, B.A. (2007). Climate change and invasibility of the Antarctic benthos. *Annual Review of Ecology, Evolution and Systematics, 38*, 129-154.

Aronson, R.B., Frederich, M., Price, R. & Thatje, S. (2014). Prospects for the return of shell-crushing crabs to Antarctica. *Journal of Biogeography, 42*(1), 1–7.

Aronson, R.B., Smith, K.E., Vos, S.C., McClintock, J.B., Amsler, M.O., Moksnes, P.O. ... & Schiferl, J.C. (2015). No barrier to emergence of bathyal king crabs on the Antarctic shelf. *Proceedings of the National Academy of Sciences*, *112*(42), 12997-13002.

Arthur, B., Hindell, M., Bester, M., De Bruyn, P.N., Trathan, P., Goebel, M. & Lea, M.A. (2017). Winter habitat predictions of a key Southern Ocean predator, the Antarctic fur seal (*Arctocephalus gazella*). *Deep Sea Research Part II: Topical Studies in Oceanography, 140*, 171-181.

Arthur, B., Hindell, M., Bester, M., De Bruyn, P.N., Goebel, M.E., Trathan, P., & Lea, M.A. (2018). Managing for change: Using vertebrate at sea habitat use to direct management efforts. *Ecological Indicators*, *91*, 338-349.

Artois, J. (2019). Pathogens and parasites, species unlike others: The spatial distribution of avian influenzas in poultry. PhD dissertation, Université Libre de Bruxelles, 164pp.

Ashauer, R., Albert, C., Augustine, S., Cedergreen, N., Charles, S., Ducrot, V. ... & Jager, T. (2016). Modelling survival: exposure pattern, species sensitivity and uncertainty. *Scientific Reports*, *6*(1), 1-11.

Ashford, J., La Mesa, M., Fach, B. A., Jones, C. & Everson, I. (2010). Testing early life connectivity using otolith chemistry and particle-tracking simulations. *Canadian Journal of Fisheries and Aquatic Sciences*, *67*(8), 1303-1315.

Ashford, J., Dinniman, M., Brooks, C., Andrews, A. H., Hofmann, E., Cailliet, G., ... & Ramanna, N. (2012). Does large-scale ocean circulation structure life history connectivity in Antarctic toothfish (*Dissostichus mawsoni*)? *Canadian Journal of Fisheries and Aquatic Sciences*, *69*(12), 1903-1919.

Ashford, J., Zane, L., Torres, J.J., La Mesa, M. & Simms, A.R. (2017). Population structure and life history connectivity of Antarctic silverfish (*Pleuragramma antarctica*) in the Southern Ocean ecosystem. In *The Antarctic silverfish: A keystone*

species in a changing ecosystem (pp. 193-234). Springer, Cham.

Ashton, G.V., Morley, S.A., Barnes, D.K., Clark, M.S. & Peck, L.S. (2017). Warming by 1°C drives species and assemblage level responses in Antarctica's marine shallows. *Current Biology*, *27*(17), 2698-2705.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E.A. & De Clerck, O. (2018). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, 27(3), 277–284.

ATCM (2006). Practical Guidelines for Ballast Water Exchange in the Antarctic Treaty Area. Annex to Resolution 3. https://documents.ats.aq/recatt/Att345_e.pdf, Accessed January 2020.

Aubry, K.B., Raley, C.M. & McKelvey, K.S. (2017). The importance of data quality for generating reliable distribution models for rare, elusive, and cryptic species. *PloS One*, *12*(6), e0179152.

Augustine, S., Rosa, S., Kooijman, S., Carlotti, F. & Poggiale, J.C. (2014). Modeling the eco-physiology of the purple mauve stinger, *Pelagia noctiluca* using Dynamic Energy Budget theory. *Journal of Sea Research*, *94*, 52-64.

Austin, M. (2007). Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, 200(1-2), 1-19.

Austin, M.P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, *157*(2-3), 101-118.

Austin, M.P. & Van Niel, K.P. (2011). Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*, 38(1), 1-8.

Avila, C., Angulo-Preckler, C., Martín-Martín, R.P., Figuerola, B., Griffiths, H.J. & Waller, C.L. (2020). Invasive marine species discovered on non–native kelp rafts in the warmest Antarctic island. *Scientific Reports*, *10*(1), 1–9.

Ayata, S.D., Lazure, P. & Thiébaut, É. (2010). How does the connectivity between populations mediate range limits of marine invertebrates? A case study of larval dispersal between the Bay of Biscay and the English Channel (North-East Atlantic). *Progress in Oceanography*, *87*(1-4), 18-36.

B

Baas, J. & Kooijman, S.A. (2015). Sensitivity of animals to chemical compounds links to metabolic rate. *Ecotoxicology*, *24*(3), 657-663.

Bahlburg, D., Meyer, B. & Berger, U. (2021). The impact of seasonal regulation of metabolism on the life history of Antarctic krill. *Ecological Modelling*, 442, 109427.

Bahn, V. & McGill, B.J. (2007). Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, *16*(6), 733-742.

Bahn, V. & McGill, B.J. (2013). Testing the predictive performance of distribution models. *Oikos, 122(3)*, 321-331.

Baird, S.J. & Mormède, S. (2014). *Assessing the environmental preferences of seabirds and spatial distribution of seabirds and marine mammals in the Southern Ocean Ross Sea region in late summer*. Ministry for Primary Industries.

Ball, I.R., Possingham, H.P. & Watts, M. (2009). Marxan and relatives: software for spatial conservation prioritisation. *Spatial conservation prioritisation: Quantitative methods and computational tools*, 185-195.

Ballard, G., Jongsomjit, D., Veloz, S.D. & Ainley, D.G. (2012). Coexistence of mesopredators in an intact polar ocean ecosystem: the basis for defining a Ross Sea marine protected area. *Biological Conservation*, 156, 72-82.

Ballerini, T., Hofmann, E.E., Ainley, D.G., Daly, K., Marrari, M., Ribic, C.A., ... & Steele, J.H. (2014). Productivity and linkages of the food web of the southern region of the western Antarctic Peninsula continental shelf. *Progress in Oceanography*, *122*, 10-29.

Banta, J.A., Ehrenreich, I.M., Gerard, S., Chou, L., Wilczek, A., Schmitt, J., ... & Purugganan, M.D. (2012). Climate envelope modelling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in *Arabidopsis thaliana*. *Ecology Letters*, *15*(8), 769-777.

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2), 327-338.

Barbosa, M. (2020). A quick guide to species distribution modelling with fuzzySim. http://fuzzysim.r-forge.r-project.org/fuzzySim-modelling-tutorial.html

Barbut, L., Grego, C., Delerue-Ricard, S., Vandamme, S., Volckaert, F. & Lacroix, G. (2019). How larval traits of six flatfish species impact connectivity: Modeling connectivity of larval flatfish. *Limnology and Oceanography. 64*, 10.1002/lno.11104.

Barnes, D.K. (2002). Invasions by marine life on plastic debris. *Nature, 416*(6883), 808-809.

Barnes, D.K. & Peck, L.S. (2008). Vulnerability of Antarctic shelf biodiversity to predicted regional warming. *Climate Research*, 37(2-3), 149-163.

Barnes, D.K. & Griffiths, H.J. (2008). Biodiversity and biogeography of southern temperate and polar bryozoans. *Global Ecology and Biogeography*, *17*(1), 84-99.

Barnes, D.K. & Clarke, A. (2011). Antarctic marine biology. *Current Biology, 21*(12), R451-R457.

Barnes, D.K. & Souster, T. (2011). Reduced survival of Antarctic benthos linked to climate-induced iceberg scouring. *Nature Climate Change*, 1(7), 365-368.

Barnes, D.K., Fenton, M. & Cordingley, A. (2014). Climate-linked iceberg activity massively reduces spatial competition in Antarctic shallow waters. *Current Biology*, *24*(12), R553-R554.

Barnes, D.K., Fleming, A., Sands, C.J., Quartino, M.L. & Deregibus, D. (2018). Icebergs, sea ice, blue carbon and Antarctic climate feedbacks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2122), 20170176.

Barrera-Oro, E.R. & Casaux, R.J. (1992). Age estimation for juvenile *Notothenia rossii* from Potter Cove, South Shetland Islands. *Antarctic Science*, *4*(2), 131-136.

Barrera-Oro, E. (2002). The role of fish in the Antarctic marine food web: differences between inshore and offshore waters in the southern Scotia Arc and west Antarctic Peninsula. *Antarctic Science*, *14*(4), 293.

Barrera-Oro, E.R. & Marschoff, E. (2007). Information on the status of fjord *Notothenia rossii*, *Gobionotothen gibberifrons* and *Notothenia coriiceps* in the lower South Shetland Islands, derived from the 2000–2006 monitoring program at Potter Cove. *CCAMLR Science*, *14*, 83-87.

Barrera-Oro, E.R. & Winter, D.J. (2008). Age composition and feeding ecology of early juvenile *Notothenia rossii* (Pisces, Nototheniidae) at Potter Cove, South Shetland Islands, Antarctica. *Antarctic Science*, *20*(4), 339.

Barrera-Oro, E., La Mesa, M. & Moreira, E. (2014). Early life history timings in marbled rockcod (*Notothenia rossii*) fingerlings from the South Shetland Islands as revealed by otolith microincrement. *Polar Biology*, *37*(8), 1099-1109.

Barrera-Oro, E., Marschoff, E. & Ainley, D. (2017). Changing status of three notothenioid fish at the South Shetland Islands (1983-2016) after impacts of the 1970-80s commercial fishery. *Polar Biology*, *40*(10), 2047-2054.

Barrera-Oro, E., Moreira, E., Seefeldt, M.A., Francione, M.V. & Quartino, M.L. (2019). The importance of macroalgae and associated amphipods in the selective benthic feeding of sister rockcod species *Notothenia rossii* and *N. coriiceps* (Nototheniidae) in West Antarctica. *Polar Biology*, *42*(2), 317-334.

Barry, S. & Elith, J. (2006). Error and uncertainty in habitat models. *Journal of Applied Ecology*, 43(3), 413-423.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., ... & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, *222*(11), 1810-1819.

Basher, Z., & Costello, M.J. (2016). The past, present and future distribution of a deep-sea shrimp in the Southern Ocean. *PeerJ*, 4, e1713.

Beale, C.M. & Lennon, J.J. (2012). Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transaction of the Royal Society B*, 367(1586), 247-258.

Beaman, R.J. & O'Brien, P.E. (2011). Kerguelen Plateau Bathymetric Grid, November 2010. Record 2011/22, Geoscience Australia, Canberra, Australia, pp. 18. Available at: http://www.deepreef.org/publications/reports/99-kergdem.html

Bean, W.T., Stafford, R. & Brashares, J.S. (2012). The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography*, 35(3), 250-258.

Beaudouin, R., Goussen, B., Piccini, B., Augustine, S., Devillers, J., Brion, F. & Péry, A.R. (2015). An individual-based model of zebrafish population dynamics accounting for energy dynamics. PloS One, 10(5), e0125841.

Beaumont, A.R. & Wei, J.H. (1991). Morphological and genetic variation in the Antarctic limpet *Nacella concinna* (Strebel, 1908). *Journal of Molluscan Studies*, *57*(4), 443-450.

Beaumont, L.J., Hughes, L., & Poulsen, M. (2005). Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. *Ecological Modelling*, 186(2), 251-270.

Beaumont, L.J., Hughes, L. & Pitman, A.J. (2008). Why is the choice of future climate scenarios for species distribution modelling important? *Ecology Letters, 11*(11), 1135-1146.

Beaumont, L.J., Graham, E., Duursma, D.E., Wilson, P.D., Cabrelli, A., Baumgartner, J.B., .. & Laffan, S.W. (2016). Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges? *Ecological Modelling*, 342, 135-146.

Bebber, D. & Gurr, S. (2019). Biotic interactions and climate in species distribution modelling. *bioRxiv*, 520320.

Becquevort, S. & Smith Jr, W.O. (2001). Aggregation, sedimentation and biodegradability of phytoplankton-derived material during spring in the Ross Sea, Antarctica. *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(19-20), 4155-4178.

Bender, N.A., Crosbie, K. & Lynch, H.J. (2016). Patterns of tourism in the Antarctic Peninsula region: a 20-year analysis. *Antarctic Science, 28*(3), 194.

Benestan, L., Moore, J.S., Sutherland, B.J., Le Luyer, J., Maaroufi, H., Rougeux, C., ... & Bernatchez, L. (2017). Sex matters in massive parallel sequencing: Evidence for biases in genetic parameter estimation and investigation of sex determination systems. *Molecular Ecology*, *26*(24), 6767-6783.

Benito Garzón, M., Robson, T.M. & Hampe, A. (2019). ΔTrait SDMs: species distribution models that account for local adaptation and phenotypic plasticity. *New Phytologist*, 222(4), 1757-1765.

Bennett, A. (2006). *Lagrangian fluid dynamics*. Cambridge University Press.

Bennett, J.M., Calosi, P., Clusella-Trullas, S., Martínez, B., Sunday, J., Algar, A.C., ... & Morales-Castilla, I. (2018). GlobTherm, a global database on thermal tolerances for aquatic and terrestrial organisms. *Scientific Data*, *5*(1), 1-7.

Benson, B.B. & Krause Jr, D. (1984). The concentration and isotopic fractionation of oxygen dissolved in freshwater and seawater in equilibrium with the atmosphere. *Limnology and Oceanography*, *29*(3), 620-632.

Berg, P.R., Star, B., Pampoulie, C., Sodeland, M., Barth, J.M., Knutsen, H., ... & Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports*, *6*(1), 1-12.

Berglund, M., Jacobi, M.N. & Jonsson, P.R. (2012). Optimal selection of marine protected areas based on connectivity and habitat quality. *Ecological Modelling, 240*, 105-112.

Berline, L., Zakardjian, B., Molcard, A., Ourmieres, Y. & Guihou, K. (2013). Modeling jellyfish *Pelagia noctiluca* transport and stranding in the Ligurian Sea. *Marine Pollution Bulletin*, *70*(1-2), 90-99.

Berry, P.M., Dawson, T.P., Harrison, P.A. & Pearson, R.G. (2002). Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Global Ecology and Biogeography*, 11, 453-462.

Bers, A.V., Momo, F., Schloss, I.R. & Abele, D. (2013). Analysis of trends and sudden changes in long-term environmental data from King George Island (Antarctica): relationships between global climatic oscillations and local system response. *Climatic Change*, *116*(3-4), 789-803.

Bertolin, M.L. & Casaux, R. (2019). Diet overlap among top predators at the South Orkney Islands, Antarctica. *Polar Biology*, *42*(2), 371-383.

Bertrand, A., Habasque, J., Hattab, T., Hintzen, N.T., Oliveros-Ramos, R., Gutiérrez, M., ... & Gerlotto, F. (2016). 3-D habitat suitability of jack mackerel *Trachurus murphyi* in the Southeastern Pacific, a comprehensive study. *Progress in Oceanography*, *146*, 199-211.

Besnier, F. & Glover, K.A. (2013). ParallelStructure: a R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *PLoS One*, *8*(7), e70651.

Bevelhimer, M.S., Stein, R.A. & Carline, R.F. (1985). Assessing significance of physiological differences among three esocids with a bioenergefics model. *Canadian Journal of Fisheries and Aquatic Sciences*, 42(1), 57-69.

Bevins, S.N. (2019). Parasitism, host behavior, and invasive species. pgs 273-278. In: J.C. Choe, editor. Encyclopedia of Animal Behavior, 2nd edition, vol 1. Elsevier, Academic Press. 3048 pp.

Bilyk, K.T. & DeVries, A.L. (2010). Freezing avoidance of the Antarctic icefishes (*Channichthyidae*) across thermal gradients in the Southern Ocean. *Polar Biology*, *33*(2), 203-213.

Bilyk, K.T. & DeVries, A.L. (2011). Heat tolerance and its plasticity in Antarctic fishes. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *158*(4), 382-390.

Blackburn, T.M. & Gaston, K.J. (2002). Scale in macroecology. *Global Ecology and Biogeography*, 11(3), 185-189.

Blanchard, J.L., Heneghan, R.F., Everett, J.D., Trebilco, R. & Richardson, A.J. (2017). From bacteria to whales: using functional size spectra to model marine ecosystems. *Trends in Ecology & Evolution*, *32*(3), 174-186.

Bloom, T.D., Flower, A. & DeChaine, E.G. (2018). Why georeferencing matters: introducing a practical protocol to prepare species occurrence records for spatial analysis. *Ecology and Evolution*, *8*(1), 765-777.

Bobashev, G.V., Goedecke, D.M., Yu, F. & Epstein, J.M. (2007). A hybrid epidemic model: combining the advantages of agent-based and equation-based approaches. In 2007 Winter Simulation Conference (pp. 1532-1537). IEEE.

Bodiguel, X., Maury, O., Mellon-Duval, C., Roupsard, F., Le Guellec, A.M. & Loizeau, V. (2009). A dynamic and mechanistic model of PCB bioaccumulation in the European hake (*Merluccius merluccius*). *Journal of Sea Research*, *62*(2-3), 124-134.

Bombosch, A., Zitterbart, D.P., Van Opzeeland, I., Frickenhaus, S., Burkhardt, E., Wisz, M.S. & Boebel, O. (2014). Predictive habitat modelling of humpback (*Megaptera novaeangliae*) and Antarctic minke (*Balaenoptera bonaerensis*) whales in the Southern Ocean as a planning tool for seismic surveys. *Deep Sea Research Part I: Oceanographic Research Papers*, *91*, 101-114.

Böning, C.W., Dispert, A., Visbeck, M., Rintoul, S.R. & Schwarzkopf, F.U. (2008). The response of the Antarctic Circumpolar Current to recent climate change. *Nature Geoscience*, *1*(12), 864-869.

Booker, T.R., Yeaman, S. & Whitlock, M.C. (2019). Global adaptation confounds the search for local adaptation. *bioRxiv*, 742247.

Borgman, C.L., Wallis, J.C. & Enyedy, N. (2007). Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, *7*(1), 17-30.

Boria, R.A., Olson, L.E., Goodman, S.M. & Anderson, R.P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73-77.

Bosch, I., Beauchamp, K.A., Steele, M.E. & Pearse, J.S. (1987). Development, metamorphosis, and seasonal abundance of embryos and larvae of the Antarctic sea urchin *Sterechinus neumayeri*. *The Biological Bulletin*, *173*(1), 126-135.

Bosch, I. (1989). Contrasting modes of reproduction in two Antarctic asteroids of the genus Porania, with a description of unusual feeding and non-feeding larval types. *The Biological Bulletin*, 177(1), 77-82.

Bosch, I. & Pearse, J.S. (1990). Developmental types of shallow-water asteroids of McMurdo Sound, Antarctica. *Marine Biology*, 104(1), 41-46.

Boschi, E.E. (1964). Los crustáceos decápodos Brachyura del litoral bonaerense. *Boletín Instituto de Biología Marina*, 6, 1–99.

Boschi, E.E., Scelzo, M.A. & Goldstein, B. (1969). Desarrollo larval del cangrejo, *Halicarcinus planatus* (Fabricius) (Crustacea, Decapoda, Hymenosomidae), en el laboratorio, con observaciones sobre la distribución de la especie. *Bulletin of Marine Science, 19*(1), 225-242.

Boulanger, Y., Parisien, M.A. & Wang, X. (2018). Model-specification uncertainty in future area burned by wildfires in Canada. *International Journal of wildland fire*, *27*(3), 164-175.

Bourg, N.A., McShea, W.J. & Gill, D.E. (2005). Putting a CART before the search: successful habitat prediction for a rare forest herb. *Ecology*, *86*(10), 2793-2804.

Bouzaza, Z. & Mezali, K. (2013). Etude systématique, phylogénétique et phylogéographique de quelques espèces de Patelles. *Rapport de Communications Internationales de la Mer Méditerranée*. 40: 859.

Bouzaza, Z. & Mezali, K. (2018). Discriminant-based study of the shell morphometric relationships of *Patella caerulea* (Gastropoda: Prosobranchia) of the western Mediterranean Sea. *Turkish Journal of Zoology*, 42(5), 513-522.

Bowden, D.A. (2005). Seasonality of recruitment in Antarctic sessile marine benthos. *Marine Ecology Progress Series*, *297*, 101-118.

Box, G.E. (1979). Robustness in the strategy of scientific model building. In *Robustness in statistics* (pp. 201-236). Academic Press.

Box, E.O. (1981). Macroclimate and plant forms: An introduction to predictive modeling in phytogeography. Tasks for Vegetation Science-1. Dr. W. Junk, The Hague, 258 pp.

Boyd, P.W., Collins, S., Dupont, S., Fabricius, K., Gattuso, J.P., Havenhand, J., ... & Biswas, H. (2018). Experimental strategies to assess the biological ramifications of multiple drivers of global ocean change, a review. *Global Change Biology*, *24*(6), 2239-2261.

Bracegirdle, T.J., Connolley, W.M. & Turner, J. (2008). Antarctic climate change over the twenty first century. *Journal of Geophysical Research: Atmospheres*, *113*(D3).

Bracegirdle, T.J. & Stephenson, D.B. (2012). Higher precision estimates of regional polar warming by ensemble regression of climate model projections. *Climate Dynamics*, *39*(12), 2805-2821.

Bradhurst, R.A., Roche, S.E., East, I.J., Kwan, P. & Garner, M.G. (2015). A hybrid modeling approach to simulating foot-and-mouth disease outbreaks in Australian livestock. *Frontiers in Environmental Science*, 3, 17.

Bradie, J. & Leung, B. (2017). A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *Journal of Biogeography*, *44*(6), 1344-1361.

Braeckman, U., Pasotti, F., Hoffmann, R., Vázquez, S., Wulff, A., Schloss, I.R., ... & Vanreusel, A. (2021). Glacial melt disturbance shifts community metabolism of an Antarctic seafloor ecosystem from net autotrophy to heterotrophy. *Communications Biology*, *4*(1), 1-11.

Branch, G.M. (1975). Ecology of Patella species from the Cape Peninsula, South Africa. IV. Desiccation. *Marine biology*, 32(2), 179-188.

Branch, G.M. (1981). The ecology of limpets: physical factors, energy flow, and ecological interactions. *Oceanography and Marine Biology: Annual Revies*, 19, 235-380.

Branch, G.M. (1985). Limpets: evolution and adaptation. In The mollusca. Trueman ER, Clarke MR eds, Orlando Academic Press 187–220 pp.

Brandt, A., Gooday, A.J., Brandao, S.N., Brix, S., Brökeland, W., Cedhagen, T., ... & Diaz, R.J. (2007). First insights into the biodiversity and biogeography of the Southern Ocean deep sea. *Nature*, *447*(7142), 307-311.

Brandt, G., Wehrmann, A. & Wirtz, K.W. (2008). Rapid invasion of *Crassostrea gigas* into the German Wadden Sea dominated by larval supply. *Journal of Sea Research, 59*(4), 279-296.

Brandt, A. & Gutt, J. (2011). Biodiversity of a unique environment: the Southern Ocean benthos shaped and threatened by climate change. In *Biodiversity Hotspots* (pp. 503-526). Springer, Berlin, Heidelberg.

Brandt, A., Van de Putte, A. & Griffiths, H.J. (2014) Southern Ocean benthic deep-sea biodiversity and biogeography. Biogeographic atlas of the Southern Ocean (ed. by C. De Broyer, P. Koubbi, H.J. Griffiths, B. Raymond, C. d'Udekem d'Acoz, Van dePutte A., B. Danis, B. David, S. Grant,J. Gutt, C. Held, G. Hosie, F. Huettmann, A. Post and Y. Ropert-Coudert), pp. 510. SCAR, Cambridge.

Brasier, M.J., Constable, A., Melbourne-Thomas, J., Trebilco, R., Griffiths, H., Van de Putte, A. & Sumner, M. (2019). Observations and models to support the first Marine Ecosystem Assessment for the Southern Ocean (MEASO). *Journal of Marine Systems*, *197*, 103182.

Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H., & Bollmann, K. (2013). Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography*, 36(9), 971-983.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984) Classification and regression Trees. Wadsworth International Group, Belmont, CA, USA.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Breiner, F.T., Guisan, A., Bergamini, A. & Nobis, M.P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, 6(10), 1210-1218.

Breiner, F.T., Nobis, M.P., Bergamini, A. & Guisan, A. (2018). Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods in Ecology and Evolution*, *9*(4), 802-808.

Breitzke, M. (2014). Overview of seismic research activities in the Southern Ocean-quantifying the environmental impact. *Antarctic Science, 26*(1), 80.

Brenning, A. (2005). Spatial prediction models for landslide hazards: review, comparison and evaluation. *Natural Hazards and Earth System Science*, 5(6), 853-862.

Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package *sperrorest*. In *Geoscience and Remote Sensing Symposium* (IGARSS), 2012 IEEE International (pp. 5372-5375).

Brêthes, J.C., Ferreyra, G. & De la Vega, S. (1994). Distribution, growth and reproduction of the limpet *Nacella* (*Patinigera*) *concinna* (Strebel, 1908) in relation to potential food availability, in Esperanza Bay (Antarctic Peninsula). *Polar Biology*, 14(3), 161-170.

Brewer, M.J., O'Hara, R.B., Anderson, B.J. & Ohlemüller, R. (2016). *Plateau*: a new method for ecologically plausible climate envelopes for species distribution modelling. *Methods in Ecology and Evolution*, *7*(12), 1489-1502.

Brey, T. (1991). Population dynamics of *Sterechinus antarcticus* (Echinodermata: Echinoidea) on the Weddell Sea shelf and slope, Antarctica. *Antarctic Science*, *3*(3), 251-256.

Brey, T. & Clarke, A. (1993). Population dynamics of marine benthic invertebrates in Antarctic and subantarctic environments: are there unique adaptations? *Antarctic Science*, *5*, 253-253.

Brey, T., Pearse, J., Basch, L., McClintock, J. & Slattery, M. (1995). Growth and production of *Sterechinus neumayeri* (Echinoidea: Echinodermata) in McMurdo sound, Antarctica. *Marine Biology*, *124*(2), 279-292.

Brickman, D. (2014). Could ocean currents be responsible for the west to east spread of aquatic invasive species in Maritime Canadian waters? *Marine Pollution Bulletin, 85*(1), 235-243.

Briscoe, N.J., Kearney, M.R., Taylor, C.A. & Wintle, B.A. (2016). Unpacking the mechanisms captured by a correlative species distribution model to improve predictions of climate refugia. *Global Change Biology*, *22*(7), 2425-2439.

Britton, J.R., Ruiz-Navarro, A., Verreycken, H. & Amat-Trigo, F. (2018). Trophic consequences of introduced species: Comparative impacts of increased interspecific versus intraspecific competitive interactions. *Functional Ecology, 32*(2), 486–495.

Brockington, S., Clarke, A. & Chapman, A. (2001). Seasonality of feeding and nutritional status during the austral winter in the Antarctic sea urchin *Sterechinus neumayeri. Marine Biology,* 139(1), 127-138.

Brockington, S. & Peck, L.S. (2001). Seasonality of respiration and ammonium excretion in the Antarctic echinoid *Sterechinus neumayeri. Marine Ecology Progress Series*, 219, 159-168.

Brockington, S., Peck, L.S. & Tyler, P.A. (2007). Gametogenesis and gonad mass cycles in the common circumpolar Antarctic echinoid *Sterechinus neumayeri. Marine Ecology Progress Series*, 330, 139-147.

Brockmann, C., Doerffer, R., Peters, M., Stelzer, K., Embacher, S., Ruescas, A. (2016). Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. In Proceedings of the Living Planet Symposium 2016, Prague, Czech Republic, 9–13 May 2016; pp. 9–13.

Broennimann, O., Treier, U.A., Müller‐Schärer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. *Ecology Letters*, *10*(8), 701-709.

Broennimann, O., Mráz, P., Petitpierre, B., Guisan, A. & Müller-Schärer, H. (2014). Contrasting spatio-temporal climatic niche dynamics during the eastern and western invasions of spotted knapweed in North America. *Journal of Biogeography*, 41(6), 1126-1136.

Brooks, C.M. (2013). Competing values on the Antarctic high seas: CCAMLR and the challenge of marine-protected areas. *The Polar Journal*, *3*(2), 277-300.

Brooks, C.M., Crowder, L.B., Österblom, H. & Strong, A.L. (2020). Reaching consensus for conserving the global commons: The case of the Ross Sea, Antarctica. *Conservation Letters*, *13*(1), e12676.

Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27(4), 437-448.

Brotons, L., De Cáceres, M., Fall, A. & Fortin, M.J. (2012). Modeling bird species distribution change in fire prone Mediterranean landscapes: incorporating species dispersal and landscape dynamics. *Ecography*, 35(5), 458-467.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004). Toward a metabolic theory of ecology. *Ecology*, *85*(7), 1771-1789.

Brown, J.L., Bennett, J.R. & French, C.M. (2017). SDMtoolbox 2.0: the next generation Python-

based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, 5, e4095.

Brueggeman, P. (1998). Underwater Field Guide to Ross Island & McMurdo Sound, Antarctica. The National Science Foundation's Office of Polar Progams sponsored Norbert Wu.–Univ. California, San Diego.

Bryson, M., Ferrari, R., Figueira, W., Pizarro, O., Madin, J., Williams, S. & Byrne, M. (2017). Characterization of measurement errors using structure-from-motion and photogrammetry to measure marine habitat structural complexity. *Ecology and Evolution*, *7*(15), 5669-5681.

Buckley, L.B., Urban, M.C., Angilletta, M.J., Crozier, L.G., Rissler, L.J. & Sears, M.W. (2010). Can mechanism inform species' distribution models? *Ecology Letters*, *13*(8), 1041-1054.

Buckley, L.B., Waaser, S.A., MacLean, H.J. & Fox, R. (2011). Does including physiology improve species distribution model predictions of responses to recent climate change? *Ecology*, *92*(12), 2214-2221.

Bucklin, D.N., Basille, M., Benscoter, A.M., Brandt, L.A., Mazzotti, F.J., Romanach, S.S., ... & Watling, J.I. (2015). Comparing species distribution models constructed with different subsets of environmental predictors. *Diversity and Distributions*, 21(1), 23-35.

Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, 16(4), 1145-1157.

Burgess, S.C., Nickols, K.J., Griesemer, C.D., Barnett, L.A., Dedrick, A.G., Satterthwaite, E.V. ... & Botsford, L.W. (2014). Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected-area design. *Ecological Applications, 24*(2), 257-270.

Burchett, M.S. (1983) The life cycle of *Notothenia rossii* from South Georgia. *British Antarctic Survey Bulletin*, 61, 71-73.

Butlin, R.K., Galindo, J. & Grahame, J.W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1506), 2997-3007.

Byrne, M., Gall, M., Wolfe, K. & Agüera, A. (2016). From pole to pole: the potential for the Arctic seastar *Asterias amurensis* to invade a warming Southern Ocean. *Global Change Biology*, *22*(12), 3874-3887.

C

Cabral, J.S. & Kreft, H. (2012). Linking ecological niche, community ecology and biogeography: insights from a mechanistic niche model. *Journal of Biogeography*, *39*(12), 2212-2224.

Caccavo, J.A., Papetti, C., Wetjen, M., Knust, R., Ashford, J.R. & Zane, L. (2018). Along-shelf connectivity and circumpolar gene flow in Antarctic silverfish (*Pleuragramma antarctica*). *Scientific Reports*, *8*(1), 1-16.

Caccavo, J.A., Ashford, J.R., Ryan, S., Papetti, C., Schröder, M. & Zane, L. (2019). Spatial structuring and life history connectivity of Antarctic silverfish along the southern continental shelf of the Weddell Sea. *Marine Ecology Progress Series*, *624*, 195-212.

Cacciapaglia, C. & van Woesik, R. (2017). Marine species distribution modelling and the effects of genetic isolation under climate change. *Journal of Biogeography*, *45*(1), 154-163.

Cadée, G.C. (1999). Shell damage and shell repair in the Antarctic limpet *Nacella concinna* from King George Island. *Journal of Sea Research*, 41(1-2), 149-161.

Calì, F., Riginella, E., La Mesa, M. & Mazzoldi, C. (2017). Life history traits of *Notothenia rossii* and *N. coriiceps* along the southern Scotia Arc. *Polar Biology*, *40*(7), 1409-1423.

Calizza, E., Careddu, G., Sporta Caputi, S., Rossi, L. & Costantini, M.L. (2018). Time-and depth-wise trophic niche shifts in Antarctic benthos. *PloS One*, *13*(3), e0194796.

Candy, S.G. (2004). Modelling catch and effort data using generalised linear models, the Tweedie distribution, random vessel effects and random stratum-by-year effects. *CCAMLR Science*, *11*, 59-80.

Cappello, S., Caruso, G., Bergami, E., Macrì, A., Venuti, V., Majolino, D. & Corsi, I. (2021). New insights into the structure and function of the prokaryotic communities colonizing plastic debris collected in King George Island (Antarctica): preliminary observations from two plastic fragments. *Journal of Hazardous Materials*, 125586.

Cardador, L., Carrete, M., Gallardo, B. & Tella, J.L. (2016). Combining trade data and niche modelling improves predictions of the origin and distribution of non-native European populations of a globally invasive species. *Journal of Biogeography*, 43(5), 967-978.

Cárdenas, C.A., González-Aravena, M. & Santibañez, P.A. (2018). The importance of local settings: within-year variability in seawater temperature at South Bay, Western Antarctic Peninsula. *PeerJ*, *6*, e4289.

Cárdenas, L., Leclerc, J.C., Bruning, P., Garrido, I., Détrée, C., Figueroa, A., … & Pardo, L. (2020). First mussel settlement observed in Antarctica reveals the potential for future invasions. *Scientific Reports*, *10*(1), 1–8.

Cardoso, J.F. (2007). Growth and reproduction in Bivalves. *An energy budget approach. Groningen: University of Groningen*.

Carvalho, L.M., Jones, C. & Ambrizzi, T. (2005) Opposite phases of the Antarctic oscillation and relationships with intraseasonal to interannual activity in the tropics during the austral summer. *Journal of Climate,* 18, 702–718.

Carvalho, S.B., Brito, J.C., Crespo, E.G., Watts, M.E. & Possingham, H.P. (2011). Conservation planning under climate change: toward accounting for uncertainty in predicted species distributions to increase confidence in conservation investments in space and time. *Biological Conservation*, 144(7), 2020-2030.

Carvalho, B.M., Rangel, E.F., Ready, P.D. & Vale, M.M. (2015). Ecological niche modelling predicts southward expansion of *Lutzomyia (Nyssomyia) flaviscutellata* (Diptera: Psychodidae: Phlebotominae), vector of *Leishmania (Leishmania) amazonensis* in South America, under climate change. *PLoS One*, *10*(11), e0143282.

Carvalho, S.B., Velo-Anton, G., Tarroso, P., Portela, A.P., Barata, M., Carranza, S., ... & Possingham, H.P. (2017). Spatial conservation prioritization of biodiversity spanning the evolutionary continuum. *Nature Ecology & Evolution*, *1*(6), 1-8.

Cavanagh, R.D., Murphy, E.J., Bracegirdle, T.J., Turner, J., Knowland, C.A., Corney, S.P. ... & Constable, A.J. (2017). A synergistic approach for evaluating climate model output for ecological applications. *Frontiers in Marine Science*, *4*, 308.

CCAMLR (2009). Commission for the Conservation of Antarctic Marine Living Resources. Conservation Measure 91-03. Protection of the South Orkney Islands southern shelf. 2pp. https://www.ccamlr.org/sites/drupal.ccamlr.org/files//91-03.pdf. Accessed 1/08/2018

CCAMLR (2016). Commission for the Conservation of Antarctic Marine Living Resources. Conservation Measure 91-05. Ross Sea region marine protected area. 17pp. https://www.ccamlr.org/sites/drupal.ccamlr.org/files//91-05_2.pdf. Accessed 1/08/2018

CCAMLR report 91-03 (2009). "Conservation Measure 91-03. Protection of the South Orkney Islands southern shelf", available online at https://www.ccamlr.org/sites/default/files/91-03_9.pdf

CCAMLR report SC-CAMLR-38/BG/03 (2019) "Domain 1 MPA Proposal CM 91-XXrev1: Rationale of the changes for the Proposal for the Establishment of a Marine Protected Area in the Western Antarctic Peninsula- South Scotia Arc" Delegations of Argentina and Chile. Revised proposal for a conservation measure establishing a Marine Protected Area in Domain 1 (Western Antarctic Peninsula and South Scotia Arc) CCAMLR 38/25 rev 1. Hobart, Australia 2019.

CCAMLR (2019a) *Schedule of Conservation Measures in Force 2019/20*. https://www.ccamlr.org/en/system/files/e-schedule2019-20_1.pdf

CCAMLR (2019b) *Report of the Thirty-eighth meeting of the Commission*. Hobart, Australia. https://www.ccamlr.org/en/system/files/e-cc-38-prelim-v2.pdf.

Ceia-Hasse, A., Sinervo, B., Vicente, L. & Pereira, H.M. (2014). Integrating ecophysiological models into species distribution projections of European reptile range shifts in response to climate change. *Ecography*, *37*(7), 679-688.

Chan, F.T., MacIsaac, H.J. & Bailey, S.A. (2015). Relative importance of vessel hull fouling and ballast water as transport vectors of nonindigenous species to the Canadian Arctic. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(8), 1230-1242.

Chamberlain, S., Barve, V., Mcglinn, D. *et al.* (2019) rgbif: Interface to the Global Biodiversity Information Facility API. https://cran.r-project.org/web/packages/rgbif/index.html

Chaplin, B.P. (2019). The prospect of electrochemical technologies advancing worldwide water treatment. *Accounts of Chemical Research*, *52*(3), 596-604.

Chapman, C.C. (2017). New perspectives on frontal variability in the Southern Ocean. *Journal of Physical Oceanography*, *47*(5), 1151-1168.

Chase, J.M. & Leibold, M.A. (2003). *Ecological niches: linking classical and contemporary approaches*. University of Chicago Press.

Chefaoui, R.M. & Lobo, J.M. (2008). Assessing the effects of pseudo-absences on predictive distribution

model performance. *Ecological Modelling*, *210*(4), 478-486.

Chelton, D.B. (1994). Physical oceanography: a brief overview for statisticians. *Statistical Science*, *9*(2), 150-166.

Chenillat, F., Blanke, B., Grima, N., Franks, P.J., Capet, X. & Rivière, P. (2015). Quantifying tracer dynamics in moving fluids: a combined Eulerian-Lagrangian approach. *Frontiers in Environmental Science*, 3, 43.

Cheung, W.W., Lam, V.W., & Pauly, D. (2008). Modelling present and climate-shifted distribution of marine fishes and invertebrates. Fisheries Center Research Reponses. 16:1-76.

Chevin, L.M., Lande, R. & Mace, G.M. (2010). Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biology*, *8*(4).

Chiantore, M., Cattaneo-Vietti, R., Elia, L., Guidetti, M. & Antonini, M. (2002) Reproduction and condition of the scallop *Adamussium colbecki* (Smith, 1902), the sea-urchin *Sterechinus neumayeri* (Meissner, 1900) and the sea-star *Odontaster validus* (Koehler, 1911) at Terra Nova Bay (Ross Sea): different strategies related to inter-annual variations in food availability. *Polar Biology*, 25(4), 251-255.

Chown, S.L., Lee, J.E., Hughes, K.A., Barnes, J., Barrett, P.J., Bergstrom, D.M., Convey, P., *et al*. (2012). Challenges to the future conservation of the Antarctic. *Science, 337,* 158-159.

Christiansen, H. (2020). PhD thesis. Connectivity and evolution of fishes in the Southern Ocean - from species to populations. KULeuven.

Chu, W.L., Dang, N.L., Kok, Y.Y., Yap, K.S., Phang, S.M. & Convey, P. (2019). Heavy metal pollution in Antarctica and its potential impacts on algae. *Polar Science*, *20*, 75-83.

Chuang, C.T. & Ng, P.K. (1994). The ecology and biology of Southeast Asian false spider crabs (Crustacea: Decapoda: Brachyura: Hymenosomatidae). *Hydrobiologia*, 285(1–3), 85–92.

Chwedorzewska, K., Korczak, M., Bednarek, P. & Markowska-Potocka, M. (2010). Low genetic differentiation between two morphotypes of the gastropod *Nacella concinna* from Admiralty Bay, Antarctica. *Polish Polar Research*, 31(2), 195-200.

Chwedorzewska, K.J., Giełwanowska, I., Olech, M., Molina-Montenegro, M.A., Wódkiewicz, M., Galera, H. (2015). *Poa annua* in the maritime Antarctic: an overview. *Polar Record, 51(6),* 637-643.

Ciasto, L.M. & Thompson, D. (2008). Observations of large-scale ocean–atmosphere interaction in the Southern Hemisphere. *Journal of Climate*, 21, 1244–1259.

Clark, J.A. & May, R.M. (2002). Taxonomic bias in conservation research. *Science*, *297*(5579), 191-193.

Clark, M.S., Fraser, K.P. & Peck, L.S. (2008). Antarctic marine molluscs do have an HSP70 heat shock response. *Cell Stress and Chaperones*, 13(1), 39-49.

Clark, M.S. & Peck, L.S. (2009). HSP70 heat shock proteins and environmental stress in Antarctic marine organisms: a mini-review. *Marine Genomics*, 1(2), 11-18.

Clark, L. (2017). lvclark/R_genetics_conv: R_genetics_conv 1.1.

Clark, M.S., Villota Nieva, L., Hoffman, J.I., Davies, A.J., Trivedi, U.H., Turner, F. … Peck, L.S. (2019). Lack of long-term acclimation in Antarctic encrusting species suggests vulnerability to warming. *Nature Communications*, 10(1), 1–10.

Clarke, A. (1980). A reappraisal of the concept of metabolic cold adaptation in polar marine invertebrates. *Biological Journal of the Linnean Society*, *14*(1), 77-92.

Clarke, A. (1989). Faecal production and an estimate of food intake in the wild of the Antarctic limpet *Nacella concinna* (Strebel). *Journal of Molluscan Studies*, 55(2), 261-262.

Clarke, A. & Johnston, N.M. (2003). Antarctic marine benthic diversity. In *Oceanography and Marine Biology, An Annual Review*, Volume 41 (pp. 55-57). CRC Press.

Clarke, A., Aronson, R.B., Crame, J.A., Gili, J.M. & Blake, D.B. (2004). Evolution and diversity of the benthic fauna of the Southern Ocean continental shelf. *Antarctic Science*, *16*(4), 559-568.

Clarke, A., Barnes, D.K. & Hodgson, D.A. (2005). How isolated is Antarctica? *Trends in Ecology and Evolution, 20*, 1-3.

Clarke, A., Griffiths, H.J., Linse, K., Barnes, D.K. & Crame, J.A. (2007). How well do we know the Antarctic marine fauna? A preliminary study of macroecological and biogeographical patterns in Southern Ocean gastropod and bivalve molluscs. *Diversity and Distributions*, *13*(5), 620-632.

Clarke, A., Meredith, M.P., Wallace, M.I., Brandon, M.A. & Thomas, D.N. (2008). Seasonal and interannual variability in temperature, chlorophyll

and macronutrients in northern Marguerite Bay, Antarctica. *Deep Sea Research Part II: Topical Studies in Oceanography*, 55(18-19), 1988-2006.

Clarke, A. & Crame, J. (2010). Evolutionary dynamics at high latitudes: speciation and extinction in polar marine faunas. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*, 3655-3666.

Cliff, A. & Ord, J.K. (1981). Spatial Processes Models and Applications. Pion Ltd.

Colloca, F., Bartolino, V., Lasinio, G.J., Maiorano, L., Sartor, P. & Ardizzone, G. (2009). Identifying fish nurseries using density and persistence measures. *Marine Ecology Progress Series*, *381*, 287-296.

Comiso, J.C. & Nishio, F. (2008). Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data. *Journal of Geophysical Research: Oceans*, *113*(C2).

Conlisk, E., Syphard, A.D., Franklin, J., Flint, L., Flint, A. & Regan, H. (2013). Uncertainty in assessing the impacts of global change with coupled dynamic species distribution and population models. *Global Change Biology*, *19*(3), 858-869.

Connor, T., Hull, V., Viña, A., Shortridge, A., Tang, Y., Zhang, J. ... & Liu, J. (2018). Effects of grain size and niche breadth on species distribution modeling. *Ecography, 41*(8), 1270-1282.

Constable, A.J., de la Mare, W.K., Agnew, D.J., Everson, I. & Miller, D. (2000). Managing fisheries to conserve the Antarctic marine ecosystem: practical implementation of the Convention on the Conservation of Antarctic Marine Living Resources (CCAMLR). *ICES Journal of marine Science*, *57*(3), 778-791.

Constable, A.J. (2011). Lessons from CCAMLR on the implementation of the ecosystem approach to managing fisheries. *Fish and Fisheries*, *12*(2), 138-151.

Constable, A., Costa, D., Murphy, E., Hofmann, E., Schofield, O., Press, A., Johnson, N.M., Newman, L. (2014a). Chapter 9.3. Assessing status and change in Southern Ocean Ecosystems. In: De Broyer C., Koubbi P., Griffiths H.J., Raymond B., Udekem d'Acoz C. d', et al. (eds.). Biogeographic Atlas of the Southern Ocean. Scientific Committee on Antarctic Research, Cambridge, pp. 404-407.

Constable, A.J., Melbourne-Thomas, J., Corney, S.P., Arrigo, K.R., Barbraud, C., Barnes, D.K., Bindoff, N.L., Boyd, P.W., Brandt, A., Costa, D.P. et al. (2014b). Climate change and Southern Ocean ecosystems I: How changes in physical habitats directly affect marine biota. *Global Change Biology*, 20(10), 3004-3025.

Constable, A.J., Costa, D.P., Schofield, O., Newman, L., Urban Jr, E., Fulton, E.A., ... & Willis, Z. (2016). Developing priority variables ("ecosystem Essential Ocean Variables"—eEOVs) for observing dynamics and change in Southern Ocean ecosystems. *Journal of Marine Systems*, *161*, 26-41.

Convey, P., Bindschadler, R., Di Prisco, G., Fahrbach, E., Gutt, J., Hodgson, D.A., ... & ACCE Consortium. (2009). Antarctic climate change and the environment. *Antarctic Science*, 21(6), 541-563.

Convey, P. & Peck, L.S. (2019). Antarctic environmental change and biological responses. *Science Advances*, 5(11), eaaz0888.

Cook, A.J., Holland, P.R., Meredith, M.P., Murray, T., Luckman, A. & Vaughan, D.G. (2016). Ocean forcing of glacier retreat in the western Antarctic Peninsula. *Science*, *353*(6296), 283-286.

Cooke, S.J. & O'Connor, C.M. (2010). Making conservation physiology relevant to policy makers and conservation practitioners. *Conservation Letters*, *3*(3), 159-166.

Cooke, S.J., Hinch, S.G., Donaldson, M.R., Clark, T.D., Eliason, E.J., Crossin, G.T., ... & Farrell, A.P. (2012). Conservation physiology in practice: how physiological knowledge has improved our ability to sustainably manage Pacific salmon during up-river migration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1596), 1757-1769.

Cooke, S.J., Blumstein, D.T., Buchholz, R., Caro, T., Fernandez-Juricic, E., Franklin, C.E., ... & Wikelski, M. (2014). Physiology, behavior, and conservation. *Physiological and Biochemical Zoology*, *87*(1), 1-14.

Costa, G.C., Nogueira, C., Machado, R.B. & Colli, G.R. (2010). Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodiversity and Conservation*, 19(3), 883-899.

Costa, H., Foody, G.M., Jiménez, S. & Silva, L. (2015). Impacts of species misidentification on species distribution modeling with presence-only data. *ISPRS International Journal of Geo-Information*, *4*(4), 2496-2518.

Cousseau, M.B., Pequeño, G., Mabragaña, E., Lucifora, L.O., Martínez, P. & Giussi, A. (2019). The Magellanic Province and its fish fauna (South America): Several provinces or one? *Journal of Biogeography*, *47*(1), 220-234.

Cowen, R.K. & Sponaugle, S. (2009) Larval dispersal

and marine population connectivity. *Annual Review of Marine Science*, 1, 443-466.

Crame, J.A. (1999). An evolutionary perspective on marine faunal connections between southernmost South America and Antarctica. *Scientia Marina*, *63*(S1), 1–14.

Crame, J.A. (2000). Evolution of taxonomic diversity gradients in the marine realm: evidence from the composition of Recent bivalve faunas. *Paleobiology*, *26*(2), 188-214.

Crame, J.A. (2018). Key stages in the evolution of the Antarctic marine fauna. *Journal of Biogeography*, *45*(5), 986–994.

Crampton, J.S., Cody, R.D., Levy, R., Harwood, D., McKay, R. & Naish, T.R. (2016). Southern Ocean phytoplankton turnover in response to stepwise Antarctic cooling over the past 15 million years. *Proceedings of the National Academy of Sciences of the United States of America*, 113(25), 6868–6873.

Crase, B., Liedloff, A.C. & Wintle, B.A. (2012) A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography*, 35(10), 879-888.

Crimmins, S.M., Dobrowski, S.Z. & Mynsberge, A.R. (2013). Evaluating ensemble forecasts of plant species distributions under climate change. *Ecological Modelling*, 266, 126-130.

Cryer, M. (2015). Progress on predicting the distribution of Vulnerable Marine Ecosystems and options for designing spatial management areas for bottom fisheries within the SPRFMO Convention Area. In *3rd Meeting of the SPRFMO Scientific Committee*.

Cuba-Díaz, M., Troncoso, J.M., Cordero, C., Finot, V.L. & Rondanelli-Reyes, M. (2013). *Juncus bufonius*, a new non-native vascular plant in King George Island, South Shetland Islands. *Antarctic Science*, *25*(3), 385.

Culumber, Z.W. & Tobler, M. (2016). Ecological divergence and conservatism: spatiotemporal patterns of niche evolution in a genus of livebearing fishes (Poeciliidae: Xiphophorus). *BMC Evolutionary Biology*, *16*(1), 44.

Cummings, V.J., Hewitt, J.E., Thrush, S.F., Marriott, P.M., Halliday, N.J. & Norkko, A. (2018). Linking Ross Sea coastal benthic communities to environmental conditions: documenting baselines in a spatially variable and changing world. *Frontiers in Marine Science*, *5*, 232.

Currie, D.J. (2007). Disentangling the roles of environment and space in ecology. *Journal of Biogeography*, *34*(12), 2009-2011.

Cziko, P.A., DeVries, A.L., Evans, C.W. & Cheng, C.H. (2014). Antifreeze protein-induced superheating of ice inside Antarctic notothenioid fishes inhibits melting during summer warming. *Proceedings of the National Academy of Sciences*, *111*(40), 14583-14588.

D

Dahood, A., Watters, G.M. & de Mutsert, K. (2019). Using sea-ice to calibrate a dynamic trophic model for the Western Antarctic Peninsula. *PloS One*, *14*(4).

Danis, B., Griffiths, H.J. & Jangoux, M. (2014) Asteroidea. Biogeographic atlas of the Southern Ocean (ed. by C. De Broyer, P. Koubbi, H.J. Griffiths, B. Raymond, C. d'Udekem d'Acoz, Van de Putte A., B. Danis, B. David, S. Grant, J. Gutt, C. Held, G. Hosie, F. Huettmann, A. Post and Y. Ropert-Coudert), pp. 510. SCAR, Cambridge.

Damerau, M., Matschiner, M., Salzburger, W. & Hanel, R. (2014). Population divergences despite long pelagic larval stages: lessons from crocodile icefishes (Channichthyidae). *Molecular Ecology*, *23*(2), 284-299.

Danis, B., Christiansen, H., Guillaumot, C., Heindler, F.M., Houston, R., Jossart, Q., Lucas, K., Moreau, C., Pasotti, F., Robert, H., Wallis, B. & Saucède, T. (2019). Report of the Belgica121 expedition to the West Antarctic Peninsula. 96pp.

David, B. & Néraudeau, D. (1989). Tubercle loss in Spatangoids (Echinodermata, Echinoides): Original skeletal structures and underlying processes. *ZooMorphology*, 109, 39–53.

David, B., Choné, T., Mooi, R. & De Ridder, C. (2005). Antarctic Echinoidea. Synopses of the Antarctic benthos. Koeltz Scientific Books, Königstein: 273 pp.

David, B. & Saucède, T. (2015). *Biodiversity of the Southern Ocean*. Elsevier.

David, P., Thébault, E., Anneville, O., Duyck, P.F., Chapuis, E. & Loeuille, N. (2017). Impacts of Invasive Species on Food Webs: A Review of Empirical Data. *Advances in Ecological Research*, *56*, 1–60.

David, V., Joachim, S., Tebby, C., Porcher, J.M. & Beaudouin, R. (2019). Modelling population dynamics in mesocosms using an individual-based model coupled to a bioenergetics model, *Ecological Modelling*, 398(C), 55-66.

Davies, A.J., Wisshak, M., Orr, J.C. & Roberts, J.M. (2008). Predicting suitable habitat for the cold-water coral *Lophelia pertusa* (Scleractinia). *Deep Sea Research Part I: Oceanographic Research Papers*, 55(8), 1048-1062.

Davis, L.B., Hofmann, E.E., Klinck, J.M., Piñones, A. & Dinniman, M.S. (2017). Distributions of krill and Antarctic silverfish and correlations with environmental variables in the western Ross Sea, Antarctica. *Marine Ecology Progress Series*, 584, 45-65.

Davis, T.R., Champion, C. & Coleman, M.A. (2021). Climate refugia for kelp within an ocean warming hotspot revealed by stacked species distribution modelling. *Marine Environmental Research*, 105267.

Dayton, P.K. (1990) Polar benthos. In: Smith WO (ed) Polar Oceanography, Part B: Chemistry, Biology and Geology. Academic Press, San Diego, pp 631-685.

DeAngelis, D.L., Godbout, L. & Shuter, B.J. (1991). An individual-based approach for predicting density-dependent dynamics in smallmouth bass populations. *Ecological Modelling*, 57, 91-115.

DeAngelis, D.L., Rose, K.A. & Huston, M.A. (1994). Individual-oriented approaches to modeling ecological populations and communities. In Frontiers in mathematical biology (pp. 390-410). Springer, Berlin, Heidelberg.

DeAngelis, D.L. & Mooij, W.M. (2005). Individual-based modeling of ecological and evolutionary processes. *Annual Review of Ecology, Evolution and Systematics*, 36,147-168.

Dearborn, J.H. (1977). Foods and feeding characteristics of Antarctic asteroids and ophiuroids. *Adaptations within Antarctic ecosystems*, 293-326.

Dearborn, J.H., Edwards, K.C. & Fratt, D.B. (1991). Diet, feeding behavior, and surface morphology of the multi-armed Antarctic sea star *Labidiaster annulatus* (Echinodermata: Asteroidea). *Marine Ecology Progress Series. Oldendorf*, 77(1), 65-84.

De'ath, G. & Fabricius, K.E. (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81, 3178–3192.

De Broyer, C. & Danis, B. (2011). How many species in the Southern Ocean? Towards a dynamic inventory of the Antarctic marine species. *Deep sea research Part II: Topical studies in oceanography*, 58(1-2), 5-17.

De Broyer, C., Koubbi, P., Griffiths, H.J., Raymond, B., Udekem d'Acoz, C. d', Van de Putte, A.P., Danis, B., David, B., Grant, S., Gutt, J., Held, C., Hosie, G., Huettmann, F., Post, A., Ropert-Coudert, Y. (eds.): 510 pp., (2014). Cambridge, SCAR. ISBN 978-0-948277-28-3

De Broyer, C., Koubbi, P. (2014) The biogeography of the Southern Ocean. In: De Broyer C, Koubbi P, Griffiths HJ, et al. (eds) Biogeographic Atlas of the Southern Ocean. Scientific Committee on Antarctic Research, Cambridge UK, pp 2–9.

De Candolle, A. (1855). Géographie botanique raisonnée ou exposition des faits principaux et des lois concernant la distribution géographique des plantes de l'époque actuelle (Vol. 2). V. Masson.

Deck, J., Gaither, M.R., Ewing, R., Bird, C.E., Davies, N., Meyer, C., ... & Crandall, E.D. (2017). The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples. *PLoS Biology*, *15*(8), e2002925.

DeConto, R.M. & Pollard, D. (2016). Contribution of Antarctica to past and future sea-level rise. *Nature*, *531*(7596), 591-597.

Decoursey, D.G. (1992). Developing models with more detail: do more algorithms give more truth? *Weed Technology*, *6*(3), 709-715.

de la Hoz, C.F., Ramos, E., Acevedo, A., Puente, A., Losada, Í.J. & Juanes, J.A. (2018). OCLE: A European open access database on climate change effects on littoral and oceanic ecosystems. *Progress in Oceanography*, *168*, 222-231.

de la Hoz, C.F., Ramos, E., Puente, A. & Juanes, J.A. (2019). Temporal transferability of marine distribution models: The role of algorithm selection. *Ecological Indicators*, *106*, 105499.

Deleersnijder, E., Van Ypersele, J.P. & Campin, J.M. (1993). An orthogonal curvilinear coordinate system for a world ocean model. Ocean Modelling (SCOR WG Newsletter).

Delille, D., Gadel, F. & Cahet, G. (1979). La matière organique dans les dépots de l'archipel des Kerguelen. Distribution spatiale et saisonnière. *Oceanologica Acta*, 2(2), 181-193.

Delille, D. & Bouvy, M. (1989). Bacterial responses to natural organic inputs in a marine sub-Antarctic area. *Hydrobiologia*, 182(3), 225-238.

Delille, D., Fiala, M. & Razouls, S. (1996). Seasonal changes in bacterial and phytoplankton biomass in a subantarctic coastal area (Kerguelen Islands). *Hydrobiologia*, *330*(2), 143-150.

Della Penna, A., Koubbi, P., Cotté, C., Bon, C., Bost, C.A. & d'Ovidio, F. (2017). Lagrangian analysis of multi-satellite data in support of open ocean Marine

Protected Area design. *Deep Sea Research Part II: Topical Studies in Oceanography*, *140*, 212-221.

De Magalhaes, J.P. & Costa, J. (2009). A database of vertebrate longevity records and their relation to other life-history traits. *Journal of Evolutionary Biology*, *22*(8), 1770-1774.

Denny, M.W. & Blanchette, C.A. (2000). Hydrodynamics, shell shape, behavior and survivorship in the owl limpet *Lottia gigantea*. *Journal of Experimental Biology*, 203(17), 2623-2639.

Deppeler, S.L. & Davidson, A.T. (2017). Southern Ocean phytoplankton in a changing climate. *Frontiers in Marine Science*, *4*, 40.

De Ridder, C. & Lawrence, J.M. (1982). Food and feeding mechanisms: Echinoidea, in: Jangoux, M., Mawrence, J.M. (Eds), Echinoderm nutrition. A.A. Balkema Publisheres, Rotterdam, pp. 57-115.

De Ridder, C., David, B. & Larrain, A. (1992). Antarctic and Subantarctic echinoids from "Marion Dufresne" expeditions MD03, MD04, MD08 and from the "Polarstern" expedition Epos III. Bulletin du Muséum national d'Histoire naturelle. Section A, Zoologie, biologie et écologie animales, 14(2), 405-441.

Des Roches, S., Post, D.M., Turley, N.E., Bailey, J.K., Hendry, A.P., Kinnison, M.T., ... & Palkovacs, E.P. (2018). The ecological importance of intraspecific variation. *Nature Ecology & Evolution*, *2*(1), 57-64.

De Villiers, M., Hattingh, V. & Kriticos, D.J. (2013). Combining field phenological observations with distribution data to model the potential distribution of the fruit fly *Ceratitis rosa* Karsch (Diptera: Tephritidae). *Bulletin of Entomological Research*, 103(1), 60-73.

de Siqueira, M.F., Durigan, G., de Marco Júnior, P. & Peterson, A.T. (2009). Something from nothing: using landscape similarity and ecological niche modeling to find rare plant species. *Journal for Nature Conservation*, *17*(1), 25-32.

DeWitt, H.H., Heemstra, P.C. & Gon, O. (1990) Nototheniidae. In: *Fishes of the Southern Ocean* (eds Gon O, Heemstra PC), pp. 279–331. J.L.B. Smith Institute of Ichthyology, Grahamstown, South Africa.

Dhingra, M.S., Artois, J., Robinson, T.P., Linard, C., Chaiban, C., Xenarios, I. ... & Von Dobschuetz, S. (2016). Global mapping of highly pathogenic avian influenza H5N1 and H5Nx clade 2.3. 4.4 viruses with spatial cross-validation. *elife*, 5, e19571.

Díaz, A., Féral, J-P., David, B., Saucède, T. & Poulin, E. (2011). Evolutionary pathways among shallow and deep-sea echinoids of the genus Sterechinus in the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, *58*(1-2), 205-211.

Díaz, A., Gerard, K., González-Wevar, C., Maturana, C., Féral, J-P., David, B., ... & Poulin, E. (2018). Genetic structure and demographic inference of the regular sea urchin *Sterechinus neumayeri* (Meissner, 1900) in the Southern Ocean: The role of the last glaciation. *PloS One*, *13*(6), e0197611.

Dierssen, H.M., Smith, R.C. & Vernet, M. (2002). Glacial meltwater dynamics in coastal waters west of the Antarctic peninsula. *Proceedings of the National Academy of Sciences*, *99*(4), 1790-1795.

Diez, M.J. & Lovrich, G.A. (2010). Reproductive biology of the crab *Halicarcinus planatus* (Brachyura, Hymenosomatidae) in sub-Antarctic waters. *Polar Biology*, *33*(3), 389-401.

Diez, M.J., Florentín, O. & Lovrich, G.A. (2011). Distribution and population structure of the crab *Halicarcinus planatus* (Brachyura, Hymenosomatidae) in the Beagle channel, Tierra del Fuego. *Revista de Biologia Marina y Oceanografia*, 46(2), 141–155

Diniz-Filho, J.A., Bini, L.M. & Hawkins, B.A. (2003). Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, *12*(1), 53-64.

Diniz-Filho, J.A., Mauricio Bini, L., Fernando Rangel, T., Loyola, R.D., Hof, C., Nogués-Bravo, D. & Araújo, M.B. (2009). Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography*, *32*(6), 897-906.

Distler, T., Schuetz, J.G., Velásquez-Tibatá, J. & Langham, G.M. (2015). Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *Journal of Biogeography*, *42*(5), 976-988.

Dixon, G.B., Davies, S.W., Aglyamova, G.V., Meyer, E., Bay, L.K. & Matz, M.V. (2015). Genomic determinants of coral heat tolerance across latitudes. *Science*, *348*(6242), 1460-1462.

Doney, S.C. & Hecht, M.W. (2002). Antarctic bottom water formation and deep-water chlorofluorocarbon distributions in a global ocean climate model. *Journal of Physical Oceanography*, *32*(6), 1642-1666.

Döös, K., Nilsson, J., Nycander, J., Brodeau, L. & Ballarotta, M. (2012). The world ocean thermohaline circulation. *Journal of Physical Oceanography*, *42*(9), 1445-1460.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J. & Ovenden, J.R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209-214.

Dormann, C.F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16(2),129-138.

Dormann, C., McPherson, J., Araújo, M., Bivand, R., Bolliger, J., Carl, G., ... & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, *30*(5), 609-628.

Dormann, C.F., Purschke, O., Márquez, J.R.G., Lautenbach, S., & Schröder, B. (2008). Components of uncertainty in species distribution analysis: a case study of the great grey shrike. *Ecology*, 89(12), 3371-3386.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., ... & Singer, A. (2012a). Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography*, *39*(12), 2119-2131.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G. ... & Münkemüller, T. (2012b). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46.

Dormann, C.F., Bobrowski, M., Dehling, D.M., Harris, D.J., Hartig, F., Lischke, H., ... & Schmidt, S. I. (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, *27*(9), 1004-1016.

Douglass, L.L., Turner, J., Grantham, H.S., Kaiser, S., Constable, A., Nicoll, R., ... & Beaver, D. (2014). A hierarchical classification of benthic biodiversity and assessment of protected areas in the Southern Ocean. PloS One, 9(7), e100551.

Drogoul, A., Huynh, N.Q. & Truong, Q.C. (2016). Coupling environmental, social and economic models to understand land-use change dynamics in the Mekong Delta. *Frontiers in Environmental Science*, 4, 19.

Duan, R.Y., Kong, X.Q., Huang, M.Y., Fan, W.Y. & Wang, Z.G. (2014). The predictive performance and stability of six species distribution models. *PLoS One*, 9(11), e112764.

Ducklow, H.W., Fraser, W., Karl, D.M., Quetin, L.B., Ross, R.M., Smith, R.C., ... & Daniels, R.M. (2006). Water-column processes in the West Antarctic Peninsula and the Ross Sea: interannual variations and foodweb structure. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(8-10), 834-852.

Ducklow, H.W., Baker, K., Martinson, D.G., Quetin, L.B., Ross, R.M., Smith, R.C., ... & Fraser, W. (2007). Marine pelagic ecosystems: the west Antarctic Peninsula. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1477), 67-94.

Ducklow, H.W., Fraser, W.R., Meredith, M.P., Stammerjohn, S.E., Doney, S.C., Martinson, D.G., ... & Amsler, C.D. (2013). West Antarctic Peninsula: an ice-dependent coastal marine ecosystem in transition. *Oceanography*, *26*(3), 190-203.

Duffy, G.A. & Chown, S.L. (2017). Explicitly integrating a third dimension in marine species distribution modelling. *Marine Ecology Progress Series*, *564*, 1-8.

Duhamel, G. (1982). Biology and Population Dynamics of *Notothenia rossii* from the Kerguelen Islands (Indian Sector of Southern Ocean). *Polar Biology*, 1, 141–151.

Duhamel, G., Hureau. J-C. & Ozouf-Costaz, C. (1982) Ecological Survey of the notothenioid fishes in the Southern Ocean from Bouvet to Kerguelen Islands. In: *Proceedings of the BIOMASS Colloquium in 1982* (eds Nemoto T, Matsuda T). National Institute of Polar Research, Tokyo.

Duhamel, G., Ozouf-Costaz, C., Cattaneo-Berrebi, G. & Berrebi, P. (1995). Interpopulation relationships in two species of Antarctic fish *Notothenia rossii* and *Champsocephalus gunnari* from the Kerguelen Islands: an allozyme study. *Oceanographic Literature Review*, *7*(43), 698.

Duhamel, G., Hulley, P-A, Causse, R., *et al.* (2014) Biogeographic patterns of fish. In: Biogeographic atlas of the Southern Ocean. Published by The Scientific Committee on Antarctic Research, Scott Polar research Institute, Cambridge. pp. 328-362.

Duhamel, G., Péron, C., Sinègre, R., Chazeau, C. & Gasco, N. (2019). Important readjustments in the biomass and distribution of groundfish species in the northern part of the Kerguelen plateau and skiff bank. In *The Kerguelen Plateau: Marine Ecosystem and Fisheries. Proceedings of the Second Symposium, Proceedings of the Second*

Symposium, eds D. Welsford, J. Dell, and G. Duhamel (Hobart, TAS) (pp. 135-184).

Dulière, V., Ovidio F. & Legrand S. (2013) Development of an Integrated Software for Forecasting the Impacts of Accidental Oil Pollution-OSERIT. Final Report. Brussels: Belgian Science Policy, 68 pp. (Research Programme Science for a Sustainable Development) http://www.belspo.be/belspo/SSD/science/Reports/OSERIT_FinRep_AD.pdf

Dulière, V., Kerckhof, F. & Lacroix G. (2014): Where is my jelly? *De Strandvlo*, 34(2), pp. 48–56.

Duque-Lazo, J., Van Gils, H.A., Groen, T.A. & Navarro-Cerrillo, R.M. (2016). Transferability of species distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. *Ecological Modeling*, 320, 62-70.

Dyke, P.P. (2001) *Coastal and Shelf Sea Modelling*. Kluwer Academic, Boston, MA.

E

Earl, D.A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359-361.

Easterling, D.R., Meehl, G.A., Parmesan, C., Changnon, S.A., Karl, T.R. & Mearns, L.O. (2000). Climate extremes: observations, modeling, and impacts. *Science*, 289(5487), 2068-2074.

Ebert, T.A. (1975). Growth and mortality of post-larval echinoids. *Integrative and Comparative Biology*, 15(3), 755-775.

Ebert, T.A. (2013). Growth and survival of post-settlement sea urchins, in Lawrence, J.M. (Ed), Sea Urchins: Biology and Ecology, pp. 83-117. Developments in Aquaculture and Fisheries Science, 38.

Eby, M. & Holloway, G. (1994). Grid transformation for incorporating the Arctic in a global ocean model. *Climate Dynamics*, 10(4-5), 241-247.

El-Gabbas, A. & Dormann, C.F. (2018). Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. *Ecology and Evolution*, 8(4), 2196-2206.

Elith, J., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. & Loiselle, B. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151.

Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.

Elith, J. & Leathwick, J.R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.

Elith, J. & Graham, C.H. (2009). Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, 32(1), 66-77.

Elith, J., Kearney, M. & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330-342.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17, 43–57.

Ellison, A.M. (2004). Bayesian inference in ecology. *Ecology Letters*, *7*(6), 509-520.

El Mahrad, B., Newton, A., Icely, J.D., Kacimi, I., Abalansa, S. & Snoussi, M. (2020). Contribution of remote sensing technologies to a holistic coastal and marine environmental management framework: A review. *Remote Sensing*, *12*(14), 2313.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. & Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, *6*(5), e19379.

Elsawah, S., Guillaume, J.H., Filatova, T., Rook, J. & Jakeman, A.J. (2015). A methodology for eliciting, representing, and analysing stakeholder knowledge for decision making on complex socio-ecological systems: From cognitive maps to agent-based models. *Journal of Environmental Management*, *151*, 500-516.

Elton, C.S. (1927). Animal ecology: London. *Sidgwick and Jackson, Ltd*.

Elzhov, T.V., Mullen, K.M., Spiess, A-N, Bolker, B. (2013) minpack. lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds.

Emrich, S., Breitenecker, F., Zauner, G. & Popper, N. (2008). Simulation of influenza epidemics with a hybrid model-combining cellular automata and agent based features. In ITI 2008-30th International Conference on Information Technology Interfaces (pp. 709-714). IEEE.

Enderling, H., Alfonso, J.C., Moros, E., Caudell, J.J. & Harrison, L.B. (2019). Integrating mathematical modeling into the roadmap for personalized adaptive radiation therapy. *Trends in Cancer*, *5*(8), 467-474.

Engler, R. & Guisan, A. (2009). MigClim: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, 15, 590-601.

Engler, R., Randin, C F., Vittoz, P., Czáka, T., Beniston, M., Zimmermann, N.E. & Guisan, A. (2009). Predicting future distributions of mountain plants under climate change: Does dispersal capacity matter? *Ecography*, 32, 34–45.

Enriquez-Urzelai, U., Kearney, M.R., Nicieza, A.G. & Tingley, R. (2019). Integrating mechanistic and correlative niche models to unravel range-limiting processes in a temperate amphibian. *Global Change Biology*, *25*(8), 2633-2647.

Ensing, D.J., Moffat, C.E. & Pither, J. (2012). Taxonomic identification errors generate misleading ecological niche model predictions of an invasive hawkweed. *Botany*, *91*(3), 137-147.

Epskamp, S., Cramer, A.O., Waldorp, L.J., Schmittmann, V.D. & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1-18.

Eren, M.I., Lycett, S.J., Patten, R.J., Buchanan, B., Pargeter, J. & O'Brien, M.J. (2016). Test, model, and method validation: the role of experimental stone artifact replication in hypothesis-driven archaeology. *Ethnoarchaeology*, *8*(2), 103-136.

Espinosa, F., Rivera-Ingraham, G.A., Fa, D. & García-Gómez, J.C. (2009). Effect of human pressure on population size structures of the endangered ferruginean limpet: toward future management measures. *Journal of Coastal Research*, 857-863.

Escobar, L.E., Ryan, S.J., Stewart-Ibarra, A.M., Finkelstein, J.L., King, C.A., Qiao, H. & Polhemus, M.E. (2015). A global map of suitability for coastal *Vibrio cholerae* under current and future climate conditions. *Acta tropica*, 149, 202-211.

Estoup, A. & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: Why, how and so what? *Molecular Ecology*, 19(19), 4113–4130.

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620.

Evans, T.G., Diamond, S.E. & Kelly, M.W. (2015). Mechanistic species distribution modelling as a link between physiology and conservation. *Conservation Physiology*, *3*(1), cov056.

Everson, I. (2017). Designation and management of large-scale MPAs drawing on the experiences of CCAMLR. *Fish and Fisheries*, *18*(1), 145-159.

F

Fabri-Ruiz, S., Saucède, T., Danis, B. & David, B. (2017a). Southern Ocean Echinoids database–An updated version of Antarctic, Sub-Antarctic and cold temperate echinoid database. *ZooKeys*, (697), 1.

Fabri-Ruiz, S., Danis, B., David, B. & Saucede, T. (2017b) Environmental data of the Southern Ocean, 1955-2012, Ver. 1, *Australian Antarctic Data Centre* - doi:10.4225/15/5a30800470009.

Fabri-Ruiz, S. (2018). Modèles de distribution et changements environnementaux: Application aux faunes d'échinides de l'océan Austral et écorégionalisation (Doctoral dissertation, Université de Bourgogne Franche-Comté).

Fabri-Ruiz, S., Danis, B., David, B., Saucède, T. (2019) Can we generate robust Species Distribution Models at the scale of the Southern Ocean? *Diversity and Distributions.* 25(1), 21-37.

Fabri-Ruiz, S., Danis, B., Navarro, N., Koubbi, P., Laffont, R. & Saucède, T. (2020). Benthic ecoregionalization based on echinoid fauna of the Southern Ocean supports current proposals of Antarctic Marine Protected Areas under IPCC scenarios of climate change. *Global Change Biology*. 26(4), 2161-2180.

Fabri-Ruiz, S., Guillaumot, C., Agüera, A., Danis, B. & Saucède, T. (in press). Using correlative and mechanistic niche models to assess the sensitivity of the Antarctic echinoid *Sterechinus neumayeri* (Meissner, 1900) to climate change. *Polar Biology.*

Fach, B.A., Hofmann, E.E. & Murphy, E.J. (2002). Modeling studies of Antarctic krill *Euphausia superba* survival during transport across the Scotia Sea. *Marine Ecology Progress Series*, *231*, 187-203.

Falcini, F., Corrado, R., Torri, M., Mangano, M.C., Zarrad, R., Di Cintio, A., ... & Lacorata, G. (2020). Seascape connectivity of European anchovy in the Central Mediterranean Sea revealed by weighted Lagrangian backtracking and bio-energetic modelling. *Scientific Reports*, *10*(1), 1-13.

Falk-Petersen, J., Renaud, P. & Anisimova, N. (2011). Establishment and ecosystem effects of the alien invasive red king crab (*Paralithodes camtschaticus*) in the Barents Sea–a review. *ICES Journal of Marine Science*, *68*(3), 479-488.

Feeley, K.J. & Silman, M.R. (2011). Keep collecting: accurate species distribution modelling requires more collections than previously thought. *Diversity and Distributions*, 17(6), 1132-1140.

Fell, H.B. (1962). West-wind-drift dispersal of echinoderms in the southern hemisphere. *Nature*, *193*(4817), 759-761.

Fenaux, L., Malara, G. & Charra, R. (1975). Effets d'un jeûne de courte durée sur les principaux constituants biochimiques de l'oursin *Arbacia lixula*. I. Stade de repos sexuel. *Marine Biology*, 30(3), 239-244.

Feng, X., & Papeş, M. (2017). Can incomplete knowledge of species' physiology facilitate ecological niche modelling? A case study with virtual species. *Diversity and Distributions*, 23(10), 1157-1168.

Feng, X., Liang, Y., Gallardo, B. & Papeş, M. (2020). Physiology in ecological niche modeling: using zebra mussel's upper thermal tolerance to refine model predictions through Bayesian analysis. *Ecography*, 43(2), 270-282.

Féral, J.P. & Magniez, P. (1988). Relationship between rates of oxygen consumption and somatic and gonadal size in the sub-Antarctic echinoid *Abatus cordatus* from Kerguelen. In *6th International Echinoderm Congress* (pp. 581-587).

Féral, J.P., Poulin, E., González-Wevar, C.A., Améziane, N., Guillaumot, C., Develay, E. & Saucède, T. (2019). Long-term monitoring of coastal benthic habitats in the Kerguelen Islands: a legacy of decades of marine biology research. In: Welsford, D., Dell J., Duhamel G. (Eds), The Kerguelen Plateau: marine ecosystem and fisheries. Proceedings of the Second Symposium. Australian Antarctic Division, Kingston, Tasmania, Australia, pp. 383-402.

Ferguson, N., White, C. & Marshall, D. (2013). Competition in benthic marine invertebrates: The unrecognized role of exploitative competition for oxygen. *Ecology*, 94, 126-35.

Ferrari, L., Gil, D.G. & Vinuesa, J.H. (2011). Breeding and fecundity of the sub-Antarctic crab *Halicarcinus planatus* (Crustacea: Hymenosomatidae) in the Deseado River estuary, Argentina. *Journal of the Marine Biological Association of the United Kingdom*, 91(5), 1023–1029.

Ferrari, R., Malcolm, H., Neilson, J., Lucieer, V., Jordan, A., Ingleton, T., ... & Hill, N. (2018). Integrating distribution models and habitat classification maps into marine protected area planning. *Estuarine, Coastal and Shelf Science*, 212, 40-50.

Ficetola, G.F., Thuiller, W. & Miaud, C. (2007). Prediction and validation of the potential global distribution of a problematic alien invasive species - the American bullfrog. *Diversity and Distributions*, 13(4), 476-485.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38-49.

Filgueira, R., Rosland, R. & Grant, J. (2011). A comparison of scope for growth (SFG) and dynamic energy budget (DEB) models applied to the blue mussel (*Mytilus edulis*). *Journal of Sea Research*, *66*(4), 403-410.

Fillinger, L., Janussen, D., Lundälv, T. & Richter, C. (2013). Rapid glass sponge expansion after climate-induced Antarctic ice shelf collapse. *Current Biology*, *23*(14), 1330-1334.

Fisher, R.A., Koven, C.D., Anderegg, W.R., Christoffersen, B.O., Dietze, M.C., Farrior, C.E., ... & Lichstein, J.W. (2018). Vegetation demographics in Earth System Models: A review of progress and priorities. *Global Change Biology*, *24*(1), 35-54.

Fitzpatrick, M.C. & Hargrove, W.W. (2009). The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, 18(8), 2255.

Figuerola, B., Taboada, S., Monleón-Getino, T., Vázquez, J. & Avila, C. (2013). Cytotoxic activity of Antarctic benthic organisms against the common sea urchin *Sterechinus neumayeri*. *Oceanography*, *1*(107), 2.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W. ... & Forest, C. (2014). Evaluation of climate models. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 741-866). Cambridge University Press.

Fois, M., Cuena-Lombraña, A., Fenu, G., & Bacchetta, G. (2018). Using species distribution models at local scale to guide the search of poorly known species: Review, methodological issues and future directions. *Ecological Modelling*, 385, 124-132.

Fordham, D.A., Mellin, C., Russell, B.D., Akçakaya, R.H., Bradshaw, C.J., Aiello-Lammens, M.E., ... & Brook, B.W. (2013). Population dynamics can be more important than physiological limits for determining range shifts under climate change. *Global Change Biology*, 19(10), 3224-3237.

Fountain, E.D., Pauli, J.N., Reid, B.N., Palsbøll, P.J. & Peery, M.Z. (2016). Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Molecular Ecology Resources*, *16*(4), 966-978.

Fourcade, Y., Engler, J.O., Rödder, D. & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PloS One*, *9*(5).

Fourcade, Y., Besnard, A.G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography,* 27(2), 245-256.

Franklin, J. (2010a). Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, 16(3), 321–330.

Franklin, J. (2010b). Mapping species distributions: spatial inference and prediction. Cambridge University Press.

Fraser, K.P., Peck, L.S. & Clarke, A. (2004). Protein synthesis, RNA concentrations, nitrogen excretion, and metabolism vary seasonally in the Antarctic holothurian *Heterocucumis steineni* (Ludwig 1898). *Physiological and Biochemical Zoology*, *77*(4), 556-569.

Fraser, K.P., Clarke, A. & Peck, L.S. (2007). Growth in the slow lane: protein metabolism in the Antarctic limpet *Nacella concinna* (Strebel 1908). *Journal of Experimental Biology*, 210(15), 2691-2699.

Fraser, C.I., Morrison, A.K., Hogg, A.M., Macaya, E.C., van Sebille, E., Ryan, P.G. ... & Waters, J.M. (2018). Antarctica's ecological isolation will be broken by storm-driven dispersal and warming. *Nature Climate Change, 8*(8), 704.

Frederich, M., Sartoris, J. & Pörtner, H.O. (2001). Distribution patterns of decapod crustaceans in polar areas: A result of magnesium regulation? *Polar Biology*, 24(10), 719–723.

Freer, J.J. (2018). *Ecological niches and geographic distributions of lanternfishes* (Doctoral dissertation, University of Bristol).

Freer, J.J., Partridge, J.C., Tarling, G.A., Collins, M.A. & Genner, M.J. (2018). Predicting ecological responses in a changing ocean: the effects of future climate uncertainty. *Marine Biology*, *165*(1), 7.

Freer, J.J., Tarling, G.A., Collins, M.A., Partridge, J.C., & Genner, M.J. (2019). Predicting future distributions of lanternfish, a significant ecological resource within the Southern Ocean. *Diversity and Distributions*. 25: 1259-1272

Freitas, V., Campos, J., Fonds, M. & Van der Veer, H.W. (2007). Potential impact of temperature change on epibenthic predator–bivalve prey interactions in temperate estuaries. *Journal of Thermal Biology*, *32*(6), 328-340.

Freitas, V., Cardoso, J.F., Lika, K., Peck, M.A., Campos, J., Kooijman, S.A. & Van der Veer, H.W. (2010). Temperature tolerance and energetics: a dynamic energy budget-based comparison of North Atlantic marine species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1557), 3553-3565.

Freitas, V. (2011). Climate-induced changes in estuarine predator-prey systems: A DEB approach. PhD dissertation. Vrije Universiteit Amsterdam.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics,* 1189-1232.

Friedman, J., Hastie, T. & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.

Frigg, R. & Hartmann, S. "Models in Science", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2020/entries/models-science/

Frölicher, T.L., Rodgers, K.B., Stock, C.A. & Cheung, W.W. (2016). Sources of uncertainties in 21st century projections of potential ocean ecosystem stressors. *Global Biogeochemical Cycles*, *30*(8), 1224-1243.

Fulton, E.A., Bax, N.J., Bustamante, R.H., Dambacher, J.M., Dichmont, C., Dunstan, P.K., Hayes, K.R., Hobday, A.J., Pitcher R., Plagányi E.E. et al. (2015). Modelling marine protected areas: insights and hurdles. *Philosophical Transactions of the Royal Society B*, 370, 1681.

Funk, W.C., McKay, J.K., Hohenlohe, P.A. & Allendorf, F.W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489-496.

Furtado, R., Pereira, M.E., Granadeiro, J.P. & Catry, P. (2019). Body feather mercury and arsenic concentrations in five species of seabirds from the Falkland Islands. *Marine Pollution Bulletin*, *149*, 110574.

G

Gage, J.D. (2004). Diversity in deep-sea benthic macrofauna: the importance of local ecology, the larger scale, history and the Antarctic. *Deep Sea Research Part II: Topical Studies in Oceanography*, 51(14-16), 1689-1708.

Gaichas, S.K., Seagraves, R.J., Coakley, J.M., DePiper, G.S., Guida, V.G., Hare, J.A., ... & Wilberg, M.J. (2016). A framework for incorporating species, fleet, habitat, and climate interactions into fishery management. *Frontiers in Marine Science*, *3*, 105.

Gaines, S.D., Gaylord, B. & Largier, J.L. (2003). Avoiding current oversights in marine reserve design. *Ecological Applications*, *13*(sp1), 32-46.

Galante, P.J., Alade, B., Muscarella, R., Jansa, S.A., Goodman, S.M. & Anderson, R.P. (2018). The challenge of modeling niches and distributions for data-poor species: a comprehensive approach to model complexity. *Ecography*, *41*(5), 726-736.

Galasso, H.L., Lefebvre, S., Aliaume, C., Sadoul, B. & Callier, M.D. (2020). Using the Dynamic Energy Budget theory to evaluate the bioremediation potential of the polychaete *Hediste diversicolor* in an integrated multi-trophic aquaculture system. *Ecological Modelling*, *437*, 109296.

Galera, H., Chwedorzewska, K.J., Korczak-Abshire, M. & Wódkiewicz, M. (2018). What affects the probability of biological invasions in Antarctica? Using an expanded conceptual framework to anticipate the risk of alien species expansion. *Biodiversity and Conservation*, *27*(8), 1789-1809.

Gallego, R., Dennis, T.E., Basher, Z., Lavery, S. & Sewell, M.A. (2017). On the need to consider multiphasic sensitivity of marine organisms to climate change: A case study of the Antarctic acorn barnacle. *Journal of Biogeography*, 44, 2165-2175.

Gamliel, I., Buba, Y., Guy-Haim, T., Garval, T., Willette, D., Rilov, G. & Belmaker, J. (2020). Incorporating physiology into species distribution models moderates the projected impact of warming on selected Mediterranean marine species. *Ecography*, 43(7), 1-17.

García-Callejas, D. & Araújo, M.B. (2016). The effects of model and data complexity on predictions from species distributions models. *Ecological Modelling*, *326*, 4-12.

Garth, J.S. (1958). Brachyura of the Pacific Coast of America. *Allan Hancock Pacific Expeditions*, 21, i–xii, 1– 804.

Gatti, P., Petitgas, P. & Huret, M. (2017). Comparing biological traits of anchovy and sardine in the Bay of Biscay: A modelling approach with the Dynamic Energy Budget. *Ecological Modelling*, 348, 93-109.

Giglio, D. & Johnson, G.C. (2017). Middepth decadal warming and freshening in the South Atlantic. *Journal of Geophysical Research: Oceans*, *122*(2), 973-979.

Gilbert, C.S., Gentleman, W.C., Johnson, C.L., DiBacco, C., Pringle, J.M. & Chen, C. (2010). Modelling dispersal of sea scallop (*Placopecten magellanicus*) larvae on Georges Bank: The influence of depth-distribution, planktonic duration and spawning seasonality. *Progress in Oceanography*, *87*(1-4), 37-48.

Gille, S.T. (2002). Warming of the Southern Ocean since the 1950s. *Science*, *295*(5558), 1275-1277.

Giovanelli, J.G., de Siqueira, M.F., Haddad, C.F. & Alexandrino, J. (2010). Modeling a spatially restricted distribution in the Neotropics: How the size of calibration area affects the performance of five presence-only methods. *Ecological Modelling*, *221*(2), 215-224.

Girard, P., Levison, J., Parrott, L., Larocque, M., Ouellet, M.A. & Green, D.M. (2015). Modeling cross-scale relationships between climate, hydrology, and individual animals: generating scenarios for stream salamanders. *Frontiers in Environmental Science*, *3*, 51.

Gobeyn, S., Mouton, A.M., Cord, A.F., Kaim, A., Volk, M. & Goethals, P.L. (2019). Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecological Modelling*, 392, 179-195.

Godsoe, W. (2010). I can't define the niche but I know it when I see it: a formal link between statistical theory and the ecological niche. *Oikos*, *119*(1), 53-60.

Goedegebuure, M., Melbourne-Thomas, J., Corney, S.P., McMahon, C.R. & Hindell, M.A. (2018). Modelling southern elephant seals *Mirounga leonina* using an individual-based model coupled with a dynamic energy budget. *PloS One*, 13(3), e0194950.

Gonzalez, A., Cardinale, B.J., Allington, G.R., Byrnes, J., Arthur Endsley, K., Brown, D.G., ... & Loreau, M. (2016). Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. *Ecology*, *97*(8), 1949-1960.

González-Wevar, C.A., David, B. & Poulin, E. (2011). Phylogeography and demographic inference in *Nacella* (*Patinigera*) *concinna* (Strebel, 1908) in the western Antarctic Peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography*, *58*(1-2), 220-229.

González-Wevar, C.A., Díaz, A., Gerard, K., Cañete, J.I. & Poulin, E. (2012). Divergence time estimations and contrasting patterns of genetic diversity between Antarctic and southern South America benthic invertebrates. *Revista Chilena de Historia Natural*, *85*(4), 445-456.

González-Wevar, C. A., Segovia, N. I., Rosenfeld, S., Ojeda, J., Hüne, M., Naretto, J., ... & Poulin, E. (2018a). Unexpected absence of island endemics: Long-distance dispersal in higher latitude sub-Antarctic Siphonaria (Gastropoda: Euthyneura) species. *Journal of Biogeography*, *45*(4), 874-884.

González-Wevar, C.A., Hüne, M., Rosenfeld, S., Nakano, T., Saucède, T., Spencer, H. & Poulin, E. (2018b). Systematic revision of *Nacella* (Patellogastropoda: Nacellidae) based on a complete phylogeny of the genus, with the description of a new species from the southern tip of South America. *Zoological Journal of the Linnean Society*, 186(2), 303-336.

González-Wevar, C.A., Gérard, K., Rosenfeld, S., Saucède, T., Naretto, J., Díaz, A., ... & Poulin, E. (2019). Cryptic speciation in Southern Ocean *Aequiyoldia eightsii* (Jay, 1839): Mio-Pliocene trans-Drake Passage separation and diversification. *Progress in Oceanography*, *174*, 44-54.

Good, R. (1947). The geography of the flowering plants.

Goodman, S.H. & Gotlib, I.H. (1999). Risk for psychopathology in the children of depressed mothers: a developmental model for understanding mechanisms of transmission. *Psychological Review*, *106*(3), 458.

Goodwin, R.A., Nestler, J.M., Anderson, J.J., Weber, L.J. & Loucks, D.P. (2006). Forecasting 3-D fish movement behavior using a Eulerian–Lagrangian–agent method (ELAM). *Ecological Modelling*, 192(1-2), 197-223.

Gordon, A.L. (1971). Antarctic polar front zone. *Antarctic Oceanology I, Antarctic Res. Ser*, *15*, 205-221.

Gosselin, T. (2019) *Radiator*: RADseq Data Exploration, Manipulation and Visualization using R. doi: 10.5281/zenodo.1475182.

Gotelli, N.J., Anderson, M.J., Arita, H.T., Chao, A., Colwell, R.K., Connolly, S.R., ... & Willig, M.R. (2009). Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, *12*(9), 873-886.

Gotelli, N.J. & Stanton-Geddes, J. (2015). Climate change, genetic markers and species distribution modelling. *Journal of Biogeography*, *42*(9), 1577-1585.

Goudet, J. & Jombart, T. (2015) *hierfstat*: estimation and tests of hierarchical F-statistics. http://www.r-project.org, http://github.com/jgx65/hierfstat.

Graham, C.H., Ron, S.R., Santos, J.C., Schneider, C.J. & Moritz, C. (2004). Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, *58*(8), 1781-1793.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, T.A., Loiselle, B.A. & NCEAS Predicting Species Distributions Working Group. (2008). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, *45*(1), 239-247.

Grandfils, R. (1982). Contribución al conocimiento de *Patella ferruginea* (Gmelin, 1789). *Iberus*, 2, 57-69.

Grange, L.J. & Smith, C.R. (2013). Megafaunal communities in rapidly warming fjords along the West Antarctic Peninsula: hotspots of abundance and beta diversity. *PloS One*, *8*(12).

Gray, J.S. (2001). Antarctic marine benthic biodiversity in a world-wide latitudinal context. *Polar Biology*, *24*(9), 633-641.

Gray, D.R. & Hodgson, A.N. (2003). Growth and reproduction in the high-shore South African limpet *Helcion pectunculus* (Mollusca: Patellogastropoda). *African Zoology*, 38(2), 371-386.

Gray, R. & Wotherspoon, S. (2012). Increasing model efficiency by dynamically changing model representations. *Environmental Modelling & Software*, 30, 115-122.

Greenwell, B., Boehmke, B., Cunningham, J. & GBM Developers (2019) R package *gbm.* https://CRAN.R-project.org/package=gbm

Gregory, S., Brown, J. & Belchier, M. (2014). Ecology and distribution of the grey notothen *Lepidonotothen squamifrons* around South Georgia and Shag Rocks, Southern Ocean. *Antarctic Science*, *26*(3), 239-249.

Greiser, C., Hylander, K., Meineri, E., Luoto, M. & Ehrlén, J. (2020). Climate limitation at the cold edge: contrasting perspectives from species distribution modelling and a transplant experiment. *Ecography*, *43*(5), 637-647.

Griffiths, H.J., Barnes, D.K. & Linse, K. (2009). Towards a generalized biogeography of the Southern Ocean benthos. *Journal of Biogeography*, *36*(1), 162-177.

Griffiths, H.J. (2010). Antarctic Marine Biodiversity – What Do We Know About the Distribution of Life in the Southern Ocean? *PLoS One, 5*(8): e11683.

Griffiths, H.J., Whittle, R.J., Roberts, S.J., Belchier, M. & Linse, K. (2013). Antarctic crabs: invasion or endurance? *PLoS One*, *8*(7).

Griffiths, H.J., Van de Putte, A.P. & Danis, B. (2014). CHAPTER 2.2. Data distribution: Patterns and implications. In: De Broyer, C., Koubbi, P., Griffiths, H.J., Raymond, B., Udekem d'Acoz, C. d', et al. (eds.). Biogeographic Atlas of the Southern Ocean. Scientific Committee on Antarctic Research, Cambridge, pp. 16-26.

Griffiths, H.J., Meijers, A.J. & Bracegirdle, T.J. (2017). More losers than winners in a century of future Southern Ocean seafloor warming. *Nature Climate Change*, *7*(10), 749-754.

Grimm, V. (1994). Mathematical models and understanding in ecology. *Ecological Modelling*, *75*, 641-651.

Grimm, V. & Railsback, S.F. (2005). Individual-based modeling and ecology. Princeton university press.

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J., Railsback, S.F. (2010). The ODD protocol: A review and first update. Ecological Modelling, 221, 2760-2768.

Grimm, V., Augusiak, J., Focks, A., Frank, B.M., Gabsi, F., Johnston, A.S. ... & Thorbek, P. (2014). Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. *Ecological Modelling, 280*, 129-139.

Grimm, V. & Berger, U. (2016). Robustness analysis: Deconstructing computational models for ecological theory and applications. *Ecological Modelling*, 326, 162-167.

Grinnell, J. (1904). The origin and distribution of the chest-nut-backed chickadee. *The Auk*, *21*(3), 364-382.

Grinnell, J. (1917). The niche-relationships of the California Thrasher. *Auk*, *34*(4), 427-433.

Groeneveld, J., Johst, K., Kawaguchi, S., Meyer, B., Teschke, M. & Grimm, V. (2015). How biological clocks and changing environmental conditions determine local population growth and species distribution in Antarctic krill (*Euphausia superba*): a conceptual model. *Ecological Modelling*, *303*, 78-86.

Groeneveld, J., Berger, U., Henschke, N., Pakhomov, E.A., Reiss, C.S. & Meyer, B. (2020). Blooms of a key grazer in the Southern Ocean–An individual-based model of *Salpa thompsoni*. *Progress in Oceanography*, *185*, 102339.

Guiet, J., Aumont, O., Poggiale, J.C. & Maury, O. (2016). Effects of lower trophic level biomass and water temperature on fish communities: a modelling study. *Progress in Oceanography*, *146*, 22-37.

Guillaumot, C., Martin, A., Fabri-Ruiz, S., Eléaume, M., & Saucède, T. (2016). Echinoids of the Kerguelen Plateau–occurrence data and environmental setting for past, present, and future species distribution modelling. *ZooKeys*, (630), 1.

Guillaumot, C., Martin, A., Eléaume, M. & Saucède, T. (2018a). Methods for improving species distribution models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau. in press. Mar. Ecol. Prog. Ser. 594:149-164

Guillaumot, C., Fabri-Ruiz, S., Martin, A., Eléaume, M., Danis, B., Féral, J.P., & Saucède, T. (2018b). Benthic species of the Kerguelen Plateau show contrasting distribution shifts in response to environmental changes. *Ecology and Evolution*, 8(12), 6210-6224.

Guillaumot, C., Raymond, B. & Danis, B. (2018c) Marine environmental data layers for Southern Ocean species distribution modelling. Ver. 1, *Australian Antarctic Data Centre* - doi:10.26179/5b8f30e30d4f3

Guillaumot, C., Artois, J., Saucède, T., Demoustier, L., Moreau, C., Eléaume, M., Agüera, A., Danis, B. (2019). Broad-scale species distribution models applied to data-poor areas. *Progress in Oceanography*. 175, 198-207.

Guillaumot, C. (2019a). AmP *Adamussium colbecki*, http://www.bio.vu.nl/thb/deb/deblab/add_my_pet/entries_web/Adamussium_colbecki/Adamussium_colbecki_res.html

Guillaumot, C. (2019b). AmP *Nacella concinna*, https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/entries_web/Nacella_concinna/Nacella_concinna_res.html

Guillaumot, C. (2019c). AmP *Abatus cordatus,* https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/entries_web/Abatus_cordatus/Abatus_cordatus_res.html

Guillaumot, C., Saucède, T., Morley, S.A., Augustine, S., Danis, B. & Kooijman, S. (2020a). Can DEB models infer metabolic differences between intertidal and subtidal morphotypes of the Antarctic limpet *Nacella concinna* (Strebel, 1908)? *Ecological Modelling*, *430*, 109088.

Guillaumot, C., Danis B. & Saucède, T. (2020b). Selecting environmental descriptors is critical to modelling the distribution of Antarctic benthic species. *Polar Biology*. 1-19. 2020a

Guillaumot, C., Moreau, C., Danis, B. & Saucède T. (2020c). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*, 102438.2020b

Guillaumot C, Danis B, Saucède T (in press). Species Distribution Modelling of the Southern Ocean benthos: methods, main limits and some solutions. *Antarctic Science*.

Guillaumot, C., Martin, A., Eléaume, M., Danis, B. & Saucède, T. (2021). SDMPlay. Species Distribution Modelling Playground. https://CRAN.R-project.org/package=SDMPlay

Guille, A. & Lasserre, P. (1979). Consommation d'oxygène de l'oursin *Abatus cordatus* (Verrill) et activité oxydative de son biotope aux îles Kerguelen. Mémoires du Muséum national d'Histoire naturelle, Paris, 43, 211-219.

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E. ... & Wintle, B.A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24(3), 276-292.

Guillera-Arroita, G. (2016). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, *40*(2), 281-295.

Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*(2-3), 147-186.

Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8(9), 993-1009.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. & Zimmermann, N.E. (2006). Using niche-based models to improve the sampling of rare species. *Conservation Biology*, *20*(2), 501-511.

Guisan, A., Graham, C. H., Elith, J., Huettmann, F. & NCEAS Species Distribution Modelling Group. (2007). Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, *13*(3), 332-340.

Guisan, A. & Rahbek, C. (2011). SESAM–a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*. 38(8), 1433-1444.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A. I. ... & Martin, T. G. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16(12), 1424-1435.

Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017). *Habitat suitability and distribution models: with applications in R*. Cambridge University Press.

Gutt, J. (2001). On the direct impact of ice on marine benthic communities, a review. *Polar Biology*, *24*(8), 553-564.

Gutt, J. & Starmans, A. (2001). Quantification of iceberg impact and benthic recolonisation patterns in the Weddell Sea (Antarctica). In *Ecological studies in the Antarctic sea ice zone* (pp. 210-214). Springer, Berlin, Heidelberg.

Gutt, J. & Piepenburg, D. (2003). Scale-dependent impact on diversity of Antarctic benthos caused by grounding of icebergs. *Marine Ecology Progress Series*, *253*, 77-83.

Gutt, J., Fricke, A., Teixido, N., Potthoff, M. & Arntz, W.E. (2006). Mega-epibenthos at Bouvet Island (South Atlantic): a spatially isolated biodiversity hot spot on a tiny geological spot. *Polar Biology*, *29*(2), 97-105.

Gutt, J., Hosie, G. & Stoddart, M. (2010). Marine Life in the Antarctic. In: McIntyre A.D. (Ed.). Life in the World's Oceans: Diversity, Distribution and Abundance. Wiley-Blackwell, Oxford, pp. 203-220.

Gutt, J., Zurell, D., Bracegridle, T., Cheung, W., Clark, M., Convey, P. ... & Griffiths, H. (2012). Correlative and dynamic species distribution modelling for ecological predictions in the Antarctic: a cross-disciplinary concept. *Polar Research*, 31(1), 11091.

Gutt, J., Bertler, N., Bracegirdle, T.J., Buschmann, A., Comiso, J., Hosie, G., ... & Xavier, J. C. (2015). The

Southern Ocean ecosystem under multiple climate change stresses‐an integrated circumpolar assessment. *Global Change Biology*, *21*(4), 1434-1453.

Gutt, J., Isla, E., Bertler, A.N., Bodeker, G.E., Bracegirdle, T.J., Cavanagh, R.D. ... & De Master, D. (2018). Cross-disciplinarity in the advance of Antarctic ecosystem research. *Marine Genomics*, 37, 1-17.

H

Haberle, I., Marn, N., Geček, S. & Klanjšček, T. (2020). Dynamic energy budget of endemic and critically endangered bivalve *Pinna nobilis*: A mechanistic model for informed conservation. *Ecological Modelling*, *434*, 109207.

Habibzadeh, N. & Ludwig, T. (2019). Ensemble of small models for estimating potential abundance of Caucasian grouse (*Lyrurus mlokosiewiczi*) in Iran. *Ornis Fennica*, 96(2), 77-89.

Haelters, J., Dulière, V., Vigin, L. & Degraer S. (2015). Towards a numerical model to simulate the observed displacement of harbour porpoises *Phocoena phocoena* due to pile driving in Belgian waters. *Hydrobiologia, 756*(1), 105-116.

Hair, J., Black, W., Babin, B & Anderson, R. (2014). Multivariate Data Analysis. Seventh Edition. Pearson Education Limited. 729pp.

Halanych, K.M. & Mahon, A.R. (2018). Challenging Dogma Concerning Biogeographic Patterns of Antarctica and the Southern Ocean. *Annual Review of Ecology, Evolution, and Systematics*, 49, 355-378.

Hallgren, W., Beaumont, L., Bowness, A., Chambers, L., Graham, E., Holewa, H., ... & Vanderwal, J. (2016). The biodiversity and climate change virtual laboratory: where ecology meets big data. *Environmental Modelling & Software*, *76*, 182-186.

Halvorsen, R., Mazzoni, S., Dirksen, J.W., Næsset, E., Gobakken, T. & Ohlson, M. (2016). How important are choice of model selection method and spatial autocorrelation of presence data for distribution modelling by MaxEnt? *Ecological Modelling*, *328*, 108-118.

Hamda, N.T., Martin, B., Poletto, J.B., Cocherell, D.E., Fangue, N.A., Van Eenennaam, J., ... & Danner, E. (2019). Applying a simplified energy-budget model to explore the effects of temperature and food availability on the life history of green sturgeon (*Acipenser medirostris*). *Ecological Modelling*, *395*, 1-10.

Hanchet, S.M., Mormède, S. & Dunn, A. (2010). Distribution and relative abundance of Antarctic toothfish (*Dissostichus mawsoni*) on the Ross Sea shelf. *CCAMLR Science*, *17*, 33-51.

Hand, D.J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1),103-123.

Hanegraaf, P.P. (1997) Mass and energy fluxes in micro-organisms according to the Dynamic Energy Budget theory for filaments. PhD Thesis, Vrije Universiteit, Amsterdam.

Hansen, L.K. & Salamon, P. (1990). Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence, 12*(10), 993-1001.

Hansen, J., Sato, M., Russell, G. & Kharecha, P. (2013). Climate sensitivity, sea level and atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(2001), 20120294.

Hao, T., Elith, J., Guillera-Arroita, G. & Lahoz-Monfort, J.J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, *25*(5), 839-852.

Hao, T., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography, 43*(4), 549-558.

Hare, J.A., Wuenschel, M.J. & Kimball, M.E. (2012). Projecting range limits with coupled thermal tolerance-climate change models: an example based on gray snapper (*Lutjanus griseus*) along the US east coast. *PLoS One*, 7(12), e52294.

Harisena, N.V., Groen, T.A., Toxopeus, A.G. & Naimi, B. (2021). When is variable importance estimation in species distribution modelling affected by spatial correlation? *Ecography*.

Harris, D.J. (2015). Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution*, *6*(4), 465-473.

Hassall, C., Hollinshead, J. & Hull, A. (2011). Environmental correlates of plant and invertebrate species richness in ponds. *Biodiversity and Conservation*, *20*(13), 3189-3222.

Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S. & Huth, A. (2012). Connecting dynamic vegetation models to data–an inverse perspective. *Journal of Biogeography*, *39*(12), 2240-2252.

Hartley, S., Harris, R., & Lester, P.J. (2006). Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters*, 9(9), 1068-1079.

Hargens, A.R. & Shabica, S.V. (1973). Protection against lethal freezing temperatures by mucus in an Antarctic limpet. *Cryobiology*, 10(4), 331-337.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning-Data Mining. Inference and Prediction: Springer Series in Statistics.

Havermans, C., Nagy, Z.T., Sonet, G., De Broyer, C. & Martin, P. (2011). DNA barcoding reveals new insights into the diversity of Antarctic species of *Orchomene sensu lato* (Crustacea: Amphipoda: Lysianassoidea). *Deep Sea Research Part II: Topical Studies in Oceanography*, 58(1-2), 230-241.

Hawkins, S.J., Evans, A.J., Dale, A.C., Firth, L.B. & Smith, I.P. (2018). Symbiotic polychaetes revisited: an update of the known species and relationships (1998–2017). *Oceanography and Marine Biology: An Annual Review, Volume 56*, *56*, 371-448.

Hayhoe, K., Edmonds, J., Kopp, R., LeGrande, A., Sanderson, B., Wehner, M. & Wuebbles, D. (2017). Climate models, scenarios, and projections. In *Climate Science Special Report: A Sustained Assessment Activity of the U.S. Global Change Research Program* [Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (eds.)]. U.S. Global Change Research Program, Washington, DC, USA (2017), pp. 186-227.

Hays, G.C., Fossette, S., Katselidis, K.A., Mariani, P. & Schofield, G. (2010). Ontogenetic development of migration: Lagrangian drift trajectories suggest a new paradigm for sea turtles. *Journal of the Royal Society Interface*, *7*(50), 1319-1327.

Headland, R.K. (1994). Historical development of Antarctic tourism. *Annals of Tourism Research*, *21*(2), 269-280.

Hecnar, S. & M'Closkey, R. (1996). Amphibian species richness and distribution in relation to pond water chemistry in south-western Ontario, Canada. *Freshwater Biology*, *36*(1), 7-15.

Hedrick, P.W. (2005) A standardized genetic differentiation measure. *Evolution*, 59, 1633–1638.

Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G. & Körber, J.H. (2007). Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography*, *16*(6), 754-763.

Heikkinen, R.K., Marmion, M. & Luoto, M. (2012). Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, 35(3), 276-288.

Heino, M., Pauli, B.D. & Dieckmann, U. (2015). Fisheries-induced evolution. *Annual Review of Ecology, Evolution, and Systematics*, *46*.

Hellmann, J.J., Byers, J.E., Bierwagen, B.G. & Dukes, J.S. (2008). Five potential consequences of climate change for invasive species. *Conservation Biology*, *22*(3), 534–543.

Hellmann, J.J., Prior, K.M. & Pelini, S.L. (2012). The influence of species interactions on geographic range change under climate change. *Annals of the New York Academy of Sciences*, *1249*(1), 18-28.

Helmuth, B., Veit, R.R. & Holberton, R. (1994). Long-distance dispersal of a subantarctic brooding bivalve (*Gaimardia trapesina*) by kelp-rafting. *Marine Biology*, *120*(3), 421-426.

Hemery, L.G., Galton-Fenzi, B., Ameziane, N., Riddle, M.J., Rintoul, S.R., Beaman, R.J., ... & Eleaume, M. (2011). Predicting habitat preferences for *Anthometrina adriani* (Echinodermata) on the East Antarctic continental shelf. *Marine Ecology Progress Series*, *441*, 105-116.

Hendry, K.R., Meredith, M.P. & Ducklow, H.W. (2018). The marine system of the West Antarctic Peninsula: status and strategy for progress. *Philosophical Transactions of the Royal Society A*. 376:20170179.

Hengl, T., Sierdsema, H., Radović, A. & Dilo, A. (2009). Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. *Ecological Modelling*, *220*(24), 3499-3511.

Henley, S.F., Schofield, O.M., Hendry, K.R., Schloss, I.R., Steinberg, D.K., Moffat, C. ... & Rozema, P.D. (2019). Variability and change in the west Antarctic Peninsula marine system: Research priorities and opportunities. *Progress in Oceanography*, 173, 208-237.

Henschke, N., Pakhomov, E.A., Groeneveld, J. & Meyer, B. (2018). Modelling the life cycle of *Salpa thompsoni*. *Ecological modelling*, 387, 17-26.

Henson, S.A., Yool, A. & Sanders, R. (2015). Variability in efficiency of particulate organic carbon export: A model study. *Global Biogeochemical Cycles*, *29*(1), 33-45.

Hernandez, P.A., Graham, C.H., Master, L.L., & Albert, D.L. (2006). The effect of sample size and

species characteristics on performance of different species distribution modeling methods. *Ecography*, *29*(5), 773-785.

Hibberd, T. & Moore, K. (2009). Field Identification Guide to Heard Island and McDonald Islands Benthic Invertebrates, a guide for scientific observers on board fishing vessels in that area, The Department of Environment, Water, Heritage, and the Arts. Australian Antarctic Division and the Fisheries Research and Development Corporation, Australia, ISBN 9781876934156, pp. 158.

Hibberd, T. (2016). Describing and predicting the spatial distribution of benthic biodiversity in the sub-Antarctic and Antarctic (Doctoral dissertation, University of Tasmania).

Hidalgo, M., Kaplan, D.M., Kerr, L.A., Watson, J.R., Paris, C.B. & Browman, H.I. (2017). Advancing the link between ocean connectivity, ecological function and management challenges. *ICES Journal of Marine Science*, *74*(6), 1702-1707.

Higgs, N.D., Reed, A.J., Hooke, R., Honey, D.J., Heilmayer, O. & Thatje, S. (2009). Growth and reproduction in the Antarctic brooding bivalve *Adacnarca nitens* (Philobryidae) from the Ross Sea. *Marine Biology*, *156*(5), 1073-1081.

Hijmans, R.J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93(3), 679-688.

Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2017). R package "dismo". https://CRAN.R-project.org/package=dismo

Hijmans, R.J. & Elith, J. (2017). Species distribution modeling with R. https://repo.bppt.go.id/cran/web/packages/dismo/vignettes/sdm.pdf

Hijmans, R.J., Van Etten, J., Sumner, M. *et al.* (2019). R package "raster". https://CRAN.R-project.org/package=raster

Hill, S.L., Murphy, E.J., Reid, K., Trathan, P.N. & Constable, A.J. (2006). Modelling Southern Ocean ecosystems: krill, the food-web, and the impacts of harvesting. *Biological Reviews*, *81*(4), 581-608.

Hill, S.L., Keeble, K., Atkinson, A. & Murphy, E.J. (2012). A foodweb model to explore uncertainties in the South Georgia shelf pelagic ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanography*, *59*, 237-252.

Hill, N.A., Foster, S.D., Duhamel, G., Welsford, D., Koubbi, P. & Johnson, C.R. (2017). Model-based mapping of assemblages for ecology and conservation management: A case study of demersal fish on the Kerguelen Plateau. *Diversity and Distributions*, *23*(10), 1216-1230.

Hindell, M.A., Reisinger, R.R., Ropert-Coudert, Y., Hückstädt, L.A., Trathan, P.N., Bornemann, H., ... & Lea, M.A. (2020). Tracking of marine predators to protect Southern Ocean ecosystems. *Nature*, 580(7801), 87-92.

Hinojosa, I.A., Gonzalez, E.R., Macaya, E.C. & Thiel, M. (2010). Macroalgas flotantes en el mar interior de Chiloé, Chile y su fauna asociada con énfasis en peracarida y estados temprano de desarrollo de Decapoda (Crustacea). *Ciencia y Tecnología Del Mar*, 33(2), 71–86.

Hipel K., McLeod A. (1994). Time series modelling of water resources and environmental systems. Elsevier, p1012.

Hobbs, N.T. & Ogle, K. (2011). Introducing data–model assimilation to students of ecology. *Ecological Applications*, *21*(5), 1537-1545.

Hoddell, R.J., Crossley, A.C., Williams, R. & Hosie, G.W. (2000). The distribution of Antarctic pelagic fish and larvae (CCAMLR division 58.4. 1). *Deep Sea Research Part II: Topical Studies in Oceanography*, *47*(12-13), 2519-2541.

Hoey, J.A. & Pinsky, M.L. (2018). Genomic signatures of environmental selection despite near-panmixia in summer flounder. *Evolutionary Applications*, *11*(9), 1732-1747.

Hoffman, J.I., Peck, L.S., Hillyard, G., Zieritz, A. & Clark, M.S. (2010). No evidence for genetic differentiation between Antarctic limpet *Nacella concinna* morphotypes. *Marine Biology*, 157(4), 765-778.

Hofman, R.J. (2019). Stopping overexploitation of living resources on the high seas. *Marine Policy*, 103, 91-100.

Hofmann, E.E., Haskell, A.E., Klinck, J.M. & Lascara, C.M. (2004). Lagrangian modelling studies of Antarctic krill (*Euphausia superba*) swarm formation. *ICES Journal of Marine Science*, *61*(4), 617-631.

Hogeweg, P. & Hesper, B. (1990). Individual-oriented modelling in ecology. *Mathematical and Computer Modelling*, 13(6), 83-90.

Hogg, O.T., Huvenne, V.A., Griffiths, H.J., & Linse, K. (2018). On the ecological relevance of landscape mapping and its application in the spatial planning of

very large marine protected areas. *Science of the Total Environment*, 626, 384-398.

Holdridge, L.R. (1967). Life zone ecology: Tropical science center, San José, Costa Rica.

Holling, C.S. (1966). The strategy of building models of complex ecological systems. *Systems Analysis in Ecology*, 195-214.

Holsman, K.K., Ianelli, J., Aydin, K., Punt, A.E. & Moffitt, E.A. (2016). A comparison of fisheries biological reference points estimated from temperature-specific multi-species and single-species climate-enhanced stock assessment models. *Deep Sea Research Part II: Topical Studies in Oceanography*, 134, 360-378.

Holt, R.D. (2009). Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences*, *106*(Supplement 2), 19659-19665.

Hortal, J., Lobo, J.M. & Jiménez-Valverde, A. (2007). Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conservation Biology*, 21(3), 853-863.

Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M., & Baselga, A. (2008). Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, 117(6), 847-858.

Hortal, J., Lobo, J.M. & Jiménez-Valverde, A. (2012a). Basic questions in biogeography and the (lack of) simplicity of species distributions: putting species distribution models in the right place. *Natureza & Conservação*, 10(2), 108-118.

Hortal, J., De Marco Jr, P., Santos, A.M. & Diniz-Filho, J.A. (2012b). Integrating biogeographical processes and local community assembly. *Journal of Biogeography*, *39*(4), 627-628.

Houlihan, D.F. & Allan, D. (1982). Oxygen consumption of some Antarctic and British gastropods: an evaluation of cold adaptation. *Comparative Biochemistry and Physiology Part A: Physiology*, 73(3), 383-387.

Howard, C., Stephens, P.A., Pearce-Higgins, J.W., Gregory, R.D. & Willis, S.G. (2014). Improving species distribution models: the value of data on abundance. *Methods in Ecology and Evolution*, *5*(6), 506-513.

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., ... & Cuellar, A.A. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, *19*(4), 524-531.

Hughes, K.A. & Convey, P. (2010). The protection of Antarctic terrestrial ecosystems from inter- and intra-continental transfer of non-indigenous species by human activities: A review of current systems and practices. *Global Environmental Change, 20*, 96-112.

Hughes, K.A. & Worland, M.R. (2010). Spatial distribution, habitat preference and colonization status of two alien terrestrial invertebrate species in Antarctica. *Antarctic Science*, *22*(3), 221-231.

Hughes, K.A. & Convey, P. (2014). Alien invasions in Antarctica - is anyone liable? *Polar Research*, *33*(1), 2210

Hughes, K.A., Pertierra, L.R., Molina-Montenegro, M.A. & Convey, P. (2015). Biological invasions in terrestrial Antarctica: what is the current status and can we respond? *Biodiversity and Conservation*, *24*(5), 1031-1055.

Hughes, K.A. & Ashton, G.V. (2017). Breaking the ice: the introduction of biofouling organisms to Antarctica on vessel hulls. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 27(1), 158-164.

Hughes, K.A., Convey, P., Pertierra, L.R., Vega, G.C., Aragón, P. & Olalla-Tárraga, M.Á. (2019). Human-mediated dispersal of terrestrial species between Antarctic biogeographic regions: A preliminary risk assessment. *Journal of Environmental Management*, 232, 73-89.

Hughes, K.A., Pescott, O.L., Peyton, J., Adriaens, T., Cottier-Cook, E. J., Key, G., ... & Belchier, M. (2020). Invasive non-native species likely to threaten biodiversity and ecosystems in the Antarctic Peninsula region. *Global Change Biology*, *26*(4), 2702-2716.

Hunter, R.L. & Halanych, K.M. (2010). Phylogeography of the Antarctic planktotrophic brittle star *Ophionotus victoriae* reveals genetic structure inconsistent with early life history. *Marine Biology*, *157*(8), 1693-1704.

Huntley, M.E. & Niiler, P.P. (1995). Physical control of population dynamics in the Southern Ocean. *ICES Journal of Marine Science*, *52*(3-4), 457-468.

Hutchinson, G.E. (1957). Concluding remarks Cold Spring Harbor Symposia on Quantitative Biology, 22: 415–427. *GS SEARCH*.

Hutchinson, G.E. (1978). An introduction to population ecology. Hew Haven, Connecticut, Yale University Press. 260 p.

Huthnance, J.M. (1991). Physical oceanography of the North Sea. *Ocean and Shoreline Management*, *16*(3-4), 199-231.

I

Iannella, M., Cerasoli, F. & Biondi, M. (2017). Unraveling climate influences on the distribution of the parapatric newts *Lissotriton vulgaris meridionalis* and *L. italicus*. *Frontiers in Zoology*, 14(1), 55.

IAATO, International Association of Antarctica Tour Operators (2019). Tourism statistics. Accessed at http://iaato.org/tourism-statistics.

Ikeda, D.H., Max, T.L., Allan, G.J., Lau, M.K., Shuster, S.M. & Whitham, T.G. (2017). Genetically informed ecological niche models improve climate change predictions. *Global Change Biology*, *23*(1), 164-176.

Ingels, J., Vanreusel, A., Brandt, A., Catarino, A. I., David, B., De Ridder, C., ... & Robert, H. (2012). Possible effects of global environmental changes on Antarctic benthos: a synthesis across five major taxa. *Ecology and Evolution*, 2(2), 453-485.

IPCC (2000). IPCC Special Report. Emissions Scenarios, summary for Policymakers, 27pp. https://www.ipcc.ch/site/assets/uploads/2018/03/sres-en.pdf

IPCC (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp. https://www.ipcc.ch/report/ar5/syr/

IPCC (2015). Climate change 2014: synthesis report. Intergovernmental Panel on Climate Change, Geneva, Switzerland

Iturbide, M., Bedia, J., Herrera, S., del Hierro, O., Pinto, M. & Gutiérrez, J.M. (2015). A framework for species distribution modelling with improved pseudo-absence generation. *Ecological Modelling*, *312*, 166-174.

Iturbide, M., Bedia, J. & Gutiérrez, J.M. (2018). Background sampling and transferability of species distribution model ensembles under climate change. *Global and Planetary Change*, *166*, 19-29.

Ivanescu, A.E., Li, P., George, B., Brown, A.W., Keith, S.W., Raju, D. & Allison, D.B. (2016). The importance of prediction model validation and assessment in obesity and nutrition research. *International Journal of Obesity*, *40*(6), 887-894.

J

Jackson, S.T. & Overpeck, J.T. (2000). Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology*, *26*(S4), 194-220.

Jacobs, S.S., Gordon, A.L. & Amos, A.F. (1979). Effect of glacial ice melting on the Antarctic Surface Water. *Nature*, *277*(5696), 469-471.

Jacobs, S.S. (2004). Bottom water production and its links with the thermohaline circulation. *Antarctic Science*, *16*(4), 427-437.

Jacquet, S.H., Dehairs, F., Savoye, N., Obernosterer, I., Christaki, U., Monnin, C. & Cardinal, D. (2008). Mesopelagic organic carbon remineralization in the Kerguelen Plateau region tracked by biogenic particulate Ba. *Deep Sea Research Part II: Topical Studies in Oceanography*, *55*(5-7), 868-879.

Jacob, U., Terpstra, S. & Brey, T. (2003). High-Antarctic regular sea urchins–the role of depth and feeding in niche separation. *Polar Biology*, *26*(2), 99-104.

Jacob, U., Brey, T., Fetzer, I., Kaehler, S., Mintenbeck, K., Dunton, K., ... & Arntz, W.E. (2006). Towards the trophic structure of the Bouvet Island marine ecosystem. *Polar Biology*, *29*(2), 106-113.

Jager, T. & Kooijman, S. (2005). Modeling receptor kinetics in the analysis of survival data for organophosphorus pesticides. *Environmental Science & Technology*, *39*(21), 8307-8314.

Jager, T. & Zimmer, E.I. (2012). Simplified dynamic energy budget model for analysing ecotoxicity data. *Ecological Modelling*, *225*, 74-81.

Jager, T., Martin, B.T. & Zimmer, E.I. (2013). DEBkiss or the quest for the simplest generic model of animal life history. *Journal of Theoretical Biology*, *328*, 9-18.

Jager, T., Barsi, A., Hamda, N.T., Martin, B.T., Zimmer, E.I. & Ducrot, V. (2014). Dynamic energy budgets in population ecotoxicology: Applications and outlook. *Ecological Modelling*, 280, 140-147.

Jager, T. & Ravagnan, E. (2015). Parameterising a generic model for the dynamic energy budget of Antarctic krill *Euphausia superba*. *Marine Ecology Progress Series*, 519, 115-128.

Jager, T. (2016). DEBkiss. A simple framework for animal energy budgets (Version 1.5). *Leanpub*.

Jager, T., Ravagnan, E. & Dupont, S. (2016). Near-future ocean acidification impacts maintenance costs in sea-urchin larvae: identification of stress factors and tipping points using a DEB modelling

approach. *Journal of Experimental Marine Biology and Ecology*, *474*, 11-17.

Janecki, T., Kidawa, A. & Potocka, M. (2010). The effects of temperature and salinity on vital biological functions of the Antarctic crustacean *Serolis polita*. *Polar Biology*, *33*(8), 1013-1020.

Janosik, A.M., Mahon, A.R., Scheltema, R.S., & Halanych, K.M. (2008). Short Note: Life history of the Antarctic sea star *Labidiaster annulatus* (Asteroidea: Labidiasteridae) revealed by DNA barcoding. *Antarctic Science*, 20(6), 563-564.

Jansen, J., Hill, N.A., Dunstan, P.K., Eléaume, M.P. & Johnson, C.R. (2018). Taxonomic resolution, functional traits, and the influence of species groupings on mapping Antarctic seafloor biodiversity. *Frontiers in Ecology and Evolution*, 6, 81.

Jarnevich, C.S., Stohlgren, T.J., Kumar, S., Morisette, J.T. & Holcombe, T.R. (2015). Caveats for correlative species distribution modeling. *Ecological Informatics*, *29*, 6-15.

Jarnevich, C.S., Talbert, M., Morisette, J., Aldridge, C., Brown, C.S., Kumar, S. ... & Holcombe, T. (2017). Minimizing effects of methodological decisions on interpretation and prediction in species distribution studies: An example with background selection. *Ecological Modelling*, *363*, 48-56.

Jayathilake, P.G., Gupta, P., Li, B., Madsen, C., Oyebamiji, O., González-Cabaleiro, R., Rushton, S., Bridgens, B., Swailes, D., Allen, B. *et al.* (2017). A mechanistic Individual-based Model of microbial communities. *PLoS One*, 12(8), e0181965.

Jerosch, K., Scharf, F.K., Deregibus, D., Campana, G.L., Zacher, K., Pehlke, H., ... & Abele, D. (2019). Ensemble modelling of Antarctic macroalgal habitats exposed to glacial melt in a polar fjord. *Frontiers in Ecology and Evolution*, 7, 207.

Jiménez-Valverde, A. & Lobo, J.M. (2007). Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta oecologica*, *31*(3), 361-369.

Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, 14(6), 885-890.

Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2009). The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Community Ecology*, 10(2), 196-205.

Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P., & Lobo, J.M. (2011). Use of niche models in invasive species risk assessments. *Biological Invasions*, 13(12), 2785-2797.

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21(4), 498-507.

Jiménez-Valverde, A., Rodríguez-Rey, M. & Peña-Aguilera, P. (2020). Climate data source matters in species distribution modelling: the case of the Iberian Peninsula. *Biodiversity and Conservation*, 1-18.

Joblin, S. (2020). *The Legal Status Effossio of the Hydrocarbons of the Southern Ocean Area; Questions of Antarctic Environmental Protection and Possible Minerals Exploitation Under International Law* (Doctoral dissertation, The Australian National University (Australia)).

Johannesson, K. (2003). Evolution in Littorina: ecology matters. *Journal of Sea Research*, 49(2), 107-117.

Johnson, W.S., Stevens, M. & Watling, L. (2001). Reproduction and development of marine peracaridans. *Advances in Marine Biology*. 39: 105-260.

Johnston, E.L., Connell, S.D., Irving, A.D., Pile, A.J. & Gillanders, B.M. (2007). Antarctic patterns of shallow subtidal habitat and inhabitants in Wilke's Land. *Polar Biology*, *30*(6), 781-788.

Jombart, T. (2008). *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.

Jombart ,T. & Ahmed. I. (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071.

Jones, C.D., Anderson, M.E., Balushkin, A.V., Duhamel, G., Eakin, R.R., Eastman, J.T., ... & Detrich, H.W. (2008a). Diversity, relative abundance, new locality records and population structure of Antarctic demersal fishes from the northern Scotia Arc islands and Bouvetøya. *Polar Biology*, *31*(12), 1481-1497.

Jones, W.J., Preston, C.M., Marin, R., Scholin, C.A. & Vrijenhoek, R.C. (2008b). A robotic molecular method for *in situ* detection of marine invertebrate larvae. *Molecular Ecology Resources*, *8*(3), 540-550.

Jost, L. (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, 17, 4015–4026.

Jørgensen, S.E. (1995). Complex ecology in the 21st century. *Complex Ecology: The Part–Whole Relation in Ecosystems. Prentice-Hall, Englewood Cliffs, NJ, pp. xvii–xix*.

Jørgensen, S.E. & Bendoricchio, G. (2001). *Fundamentals of Ecological Modelling* (Vol. 21). Elsevier.

Jovanović, B. (2017). Ingestion of microplastics by fish and its potential consequences from a physical perspective. *Integrated Environmental Assessment and Management*, *13*(3), 510-515.

Jusup, M., Klanjšček, T. & Matsuda, H. (2014). Simple measurements reveal the feeding history, the onset of reproduction, and energy conversion efficiencies in captive bluefin tuna. *Journal of Sea Research*, *94*, 144-155.

Jusup, M., Sousa, T., Domingos, T., Labinac, V., Marn, N., Wang, Z. & Klanjšček, T. (2017). Physics of metabolic organization. *Physics of life reviews*, 20, 1-39.

K

Kaiser, S., Brandão, S.N., Brix, S., Barnes, D.K., Bowden, D.A., Ingels, J. ... & Bax, N. (2013). Patterns, processes and vulnerability of Southern Ocean benthos: a decadal leap in knowledge and understanding. *Marine Biology*, 160(9), 2295-2317.

Kamgar-Parsi, B. & Sander, W.A. (1989). Quantization error in spatial sampling: comparison between square and hexagonal pixels. In *Proceedings CVPR'89: IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 604-611). IEEE.

Kampichler, C. & Sierdsema, H. (2018). On the usefulness of prediction intervals for local species distribution model forecasts. *Ecological Informatics*, *47*, 67-72.

Kamvar, Z.N., Tabima, J.F. & Grünwald, N.J. (2014). *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Kapsenberg, L. & Hofmann, G.E. (2014). Signals of resilience to ocean change: high thermal tolerance of early stage Antarctic sea urchins (*Sterechinus neumayeri*) reared under present-day and future pCO_2 and temperature. *Polar Biology*, *37*(7), 967-980.

Karányi, Z., Holb, I., Hornok, L., Pocsi, I. & Miskei, M. (2013). FSRD: fungal stress response database. *Database*, *2013*.

Kassambara, A. & Mundt, F. (2017). *factoextra*: extract and visualize the results of multivariate data analyses. R package version 1.0.5. https://CRAN.R-project.org/package=factoextra

Kattge, J., Diaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., ... & Wirth, C. (2011). TRY–a global database of plant traits. *Global Change Biology*, *17*(9), 2905-2935.

Kavanaugh, M.T. (2018). Seascape ecology: a review. S.J. Pittman (ed), *Seascape ecology*. John Wiley and Sons Ltd, Hoboken, New Jersey, USA, 526 pp.

Kearney, M. & Porter, W.P. (2004). Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology*, *85*(11), 3119-3131.

Kearney, M., Phillips, B.L., Tracy, C.R., Christian, K.A., Betts, G. & Porter, W.P. (2008). Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography*, *31*(4), 423-434.

Kearney, M., & Porter, W. (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, 12(4), 334-350.

Kearney, M., Porter, W.P., Williams, C., Ritchie, S. & Hoffmann, A.A. (2009). Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. *Functional Ecology*, *23*(3), 528-538.

Kearney, M.R., Wintle, B.A. & Porter, W.P. (2010). Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, *3*(3), 203-213.

Kearney, M.R., Matzelle, A. & Helmuth, B. (2012). Biomechanics meets the ecological niche: the importance of temporal data resolution. *Journal of Experimental Biology*, 215(6), 922-933.

Kearney, M.R. & White, C.R. (2012). Testing metabolic theories. *The American Naturalist*, *180*(5), 546-565.

Kearney, M.R., Domingos, T. & Nisbet, R. (2015). Dynamic energy budget theory: an efficient and general theory for ecology. *Bioscience*, 65(4), 341-341.

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W. & Prodöhl, P.A. (2013). *diveRsity*: a R package for the estimation and exploration of population

genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782-788.

Keil, P., Storch, D. & Jetz, W. (2015). On the decline of biodiversity due to area loss. *Nature Communications*, *6*(1), 1-11.

Keith, D.A., Akçakaya, H.R., Thuiller, W., Midgley, G.F., Pearson, R.G., Phillips, S.J., ... & Rebelo, T.G. (2008). Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, *4*(5), 560-563.

Keller, R., Eckert, C.M. & Clarkson, P.J. (2006). Matrices or node-link diagrams: which visual representation is better for visualising connectivity models? *Information Visualization*, *5*(1), 62-76.

Kennedy, M.C. & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425-464.

Kennicutt, M.C., Chown, S.L., Cassano, J.J., Liggett, D., Massom, R., Peck, L.S. ... & Sutherland, W.J. (2014). Polar research: six priorities for Antarctic science. *Nature News*, 512(7512), 23.

Kennicutt, M.C., Chown, S.L., Cassano, J.J., Liggett, D., Peck, L.S., Massom, R. ... & Allison, I. (2015). A roadmap for Antarctic and Southern Ocean science for the next two decades and beyond. *Antarctic Science*, 27(1), 3-18.

Kennicutt, M.C., Bromwich, D., Liggett, D., Njåstad, B., Peck, L.S, Rintoul, S.R. ... & Cassano, J. (2019). Sustained Antarctic Research: A 21st century imperative. *One Earth*, 1(1), 95-113.

Kidawa, A. & Janecki, T. (2011). Antarctic benthic fauna in the global climate change. *Papers on Global Change IGBP*, *18*(1), 71-86.

Kindt, R. (2017). R Package "BiodiversityR": Package for Community Ecology and Suitability Analysis. Version 2.8-0. *https://cran.r-project.org/web/packages/BiodiversityR/index.html*

King, P.P. & Broderip, W.J. (1832) Description of the Cirrhipeda, Conchifera and Mollusca: In a Collection Formed by the Officers of HMS Adventure and Beagle Employed Between the Years 1826 and 1830 in Surveying the Southern Coasts of South America: Including the Straits of Magalhaens [sic] and the Coast of Tierra Del Fuego. W. Phillips.

King, J.C., Turner, J., Marshall, G.J., Connolley, W.M. & Lachlan-Cope, T.A. (2003). Antarctic Peninsula climate variability and its causes as revealed by analysis of instrumental records. *Antarctic Peninsula climate variability: historical and paleoenvironmental perspectives*, 79, 17-30.

Klanjšček, T., Caswell, H., Neubert, M.G. & Nisbet, R.M. (2006). Integrating dynamic energy budgets into matrix population models. *Ecological Modelling*, 196(3-4), 407-420.

Klanjšček, T., Nisbet, R.M., Priester, J.H. & Holden, P.A. (2013). Dynamic energy budget approach to modeling mechanisms of CdSe quantum dot toxicity. *Ecotoxicology*, *22*(2), 319-330.

Klein, C.J., Steinback, C., Scholz, A.J. & Possingham, H.P. (2008). Effectiveness of marine reserve networks in representing biodiversity and minimizing impact to fishermen: a comparison of two approaches used in California. *Conservation Letters*, *1*(1), 44-51.

Klok, C., Hjorth, M. & Dahllöf, I. (2012a). Qualitative use of Dynamic Energy Budget theory in ecotoxicology: Case study on oil contamination and Arctic copepods. *Journal of Sea Research*, *73*, 24-31.

Klok, C., Ravagnan, E., Grøsvik, B.E., Hjort, M., Hansen, B.H., Farmen, E., ... & Sanni, S. (2012b). Applying dynamic energy budget theory to estimate the impact of oil components in an Arctic food chain. In *SETAC Europe 22nd Annual Meeting, Berlin*.

Klok, C., Wijsman, J.W., Kaag, K. & Foekema, E. (2014a). Effects of CO_2 enrichment on cockle shell growth interpreted with a Dynamic Energy Budget model. *Journal of Sea Research*, *94*, 111-116.

Klok, C., Nordtug, T. & Tamis, J.E. (2014b). Estimating the impact of petroleum substances on survival in early life stages of cod (*Gadus morhua*) using the Dynamic Energy Budget theory. *Marine Environmental Research*, *101*, 60-68.

Knowles, L.L., Carstens, B.C. & Keat, M.L. (2007). Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Current Biology*, *17*(11), 940-946.

Knox, G.A. (2006). Biology of the Southern Ocean. CRC Press

Knutti, R. (2010). The end of model democracy? *Climatic Change*. 102(3-4): 395-404.

Kock, K.H. (1992). *Antarctic Fish and Fisheries*. Cambridge University Press, London.

Kock. K.H. & Kellermann, A. (1991). Reproduction in Antarctic notothenioid fish. *Antarctic Science*, 3, 125–150.

Kock, K.H. & Jones, C.D. (2000). Biological characteristics of Antarctic fish stocks in the southern Scotia Arc region. *CCAMLR Science*, *7*, 1-41.

Kock, K.H. & Belchier, M. (2004). Is the attempt to estimate the biomass of Antarctic fish from a multi-species survey appropriate for all targeted species? *Notothenia rossii* in the Atlantic Ocean sector–revisited. *CCAMLR Science*, *11*, 141-153.

Kock, K.H., Reid, K., Croxall, J. & Nicol, S. (2007). Fisheries in the Southern Ocean: an ecosystem approach. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1488), 2333-2349.

Kooijman, S. (1993). Dynamic Energy Budgets in Biological Systems. Cambridge Univ. Press, Cambridge.

Kooijman, S. (2000). Dynamic Energy and Mass Budgets in Biological Systems. Cambridge Univ. Press, Cambridge.

Kooijman, S., Sousa, T., Pecquerie, L., Van der Meer, J. & Jager, T. (2008). From food-dependent statistics to metabolic parameters, a practical guide to the use of dynamic energy budget theory. *Biological Reviews*, *83*(4), 533-552.

Kooijman, S.A., Baas, J., Bontje, D., Broerse, M., Van Gestel, C.A. & Jager, T. (2009). Ecotoxicological applications of dynamic energy budget theory. In *Ecotoxicology Modeling* (pp. 237-259). Springer, Boston, MA.

Kooijman, S.A. (2010). *Dynamic energy budget theory for metabolic organisation*. Cambridge University press. 510pp.

Kooijman, S.A., Pecquerie, L., Augustine, S. & Jusup, M. (2011). Scenarios for acceleration in fish development and the role of metamorphosis. *Journal of Sea Research*, 66(4), 419-423.

Kooijman, B., Jean, F. & Augustine, S. (2017). AmP *Patella vulgata*, version 2017/01/30, https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/entries_web/Patella_vulgata/Patella_vulgata_res.html

Kool, J.T., Moilanen, A. & Treml, E.A. (2013). Population connectivity: recent advances and new perspectives. *Landscape Ecology*, *28*(2), 165-185.

Korzukhin, M.D., Ter-Mikaelian, M.T. & Wagner, R.G. (1996). Process versus empirical models: which approach for forest ecosystem management? *Canadian Journal of Forest Research*, *26*(5), 879-887.

Koubbi, P., Ozouf-Costaz, C., Goarant, A., Moteki, M., Hulley, P. A., Causse, R., Dettai, A., Duhamel, G., Pruvost, P., Tavernier, E. et al. (2010). Estimating the biodiversity of the East Antarctic shelf and oceanic zone for ecoregionalisation: example of the ichthyofauna of the CEAMARC (Collaborative East Antarctic Marine Census) CAML surveys. *Polar Science*, 4(2), 115-133.

Koubbi, P., Hulley, P.A., Raymond, B., Penot, F., Gasparini, S., Labat, J.P., ... & Mayzaud, P. (2011a). Estimating the biodiversity of the subantarctic Indian part for ecoregionalisation: Part I. Pelagic realm of CCAMLR areas 58.5. 1. and 58.6.

Koubbi, P., Moteki, M., Duhamel, G., Goarant, A., Hulley, P.A., O'Driscoll, R., ... & Hosie, G. (2011b). Ecoregionalization of myctophid fish in the Indian sector of the Southern Ocean: results from generalized dissimilarity models. *Deep Sea Research Part II: Topical Studies in Oceanography*, *58*(1-2), 170-180.

Koubbi, P., Guinet, C., Alloncle, N., Ameziane, N., Azam, C.S., Baudena, A. & Weimerskirch, H. (2016). Ecoregionalisation of the Kerguelen and Crozet islands oceanic zone. Part I: Introduction and Kerguelen oceanic zone. *CCAMLR Document WG-EMM-16/43*.

Kozak, K.H. & Wiens, J. (2006). Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution*, *60*(12), 2604-2621.

Kozak, K.H., Graham, C.H. & Wiens, J.J. (2008). Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution*, *23*(3), 141-148.

Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., ... & Cheyne, S. M. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, 19(11), 1366-1379.

Kruczek, Z., Kruczek, M. & Szromek, A.R. (2018). Possibilities of using the tourism area life cycle model to understand and provide sustainable solution for tourism development in the Antarctic Region. *Sustainability*, *10*(1), 89.

Krüger, L., Ramos, J.A., Xavier, J.C., Gremillet, D., González-Solís, J., Petry, M.V., ... & Paiva, V.H. (2018). Projected distributions of Southern Ocean albatrosses, petrels and fisheries as a consequence of climatic change. *Ecography*, *41*(1), 195-208.

Kühn, I. (2007). Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, 13(1), 66-69.

Kuhn, M. (2012). The *caret* package. R Foundation for Statistical Computing, Vienna, Austria URL https://cran r-project org/package= caret.

Kumar, S., Neven, L.G., Zhu, H. & Zhang, R. (2015). Assessing the global risk of establishment of Cydia

pomonella (Lepidoptera: Tortricidae) using CLIMEX and MaxEnt niche models. *Journal of Economic Entomology*, *108*(4), 1708-1719.

Kuperstein, I., Grieco, L., Cohen, D.P., Thieffry, D., Zinovyev, A. & Barillot, E. (2015). The shortest path is not the one you know: application of biological network resources in precision oncology research. *Mutagenesis*, *30*(2), 191-204.

L

Lacroix, G., Barbut, L. & Volckaert, F.A. (2018). Complex effect of projected sea temperature and wind change on flatfish dispersal. *Global Change Biology*, *24*(1), 85-100.

Lagger, C., Servetto, N., Torre, L. & Sahade, R. (2017). Benthic colonization in newly ice-free soft-bottom areas in an Antarctic fjord. *PloS One*, *12*(11).

Lahoz-Monfort, J.J., Guillera-Arroita, G. & Wintle, B.A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, *23*(4), 504-515.

La Mesa, M., Piñones, A., Catalano, B. & Ashford, J. (2015). Predicting early life connectivity of Antarctic silverfish, an important forage species along the Antarctic Peninsula. *Fisheries Oceanography*, *24*(2), 150-161.

Lamers, M., Haase, D. & Amelung, B. (2008). Facing the elements: analysing trends in Antarctic tourism. *Tourism Review*.

Lang, J. (1967). Contribution à l'étude sédimentologique du golfe du Morbihan: Iles Kerguelen-Terres australes et antarctiques françaises. Ph. D. Dissertation, Université de Paris.

Lang, J. (1971). Contribution à l'étude sédimentologique du Golfe du Morbihan (Iles Kerguelen). Comité National Français des Recherches Antarctiques, 29.

La Peyre, M.K., Bernasconi, S.K., Lavaud, R., Casas, S.M. & La Peyre, J.F. (2020). Eastern oyster clearance and respiration rates in response to acute and chronic exposure to suspended sediment loads. *Journal of Sea Research*, *157*, 101831.

Laurenceau-Cornec, E.C., Trull, T.W., Davies, D.M., Christina, L. & Blain, S. (2015). Phytoplankton morphology controls on marine snow sinking velocity. *Marine Ecology Progress Series*, *520*, 35-56.

Lauzeral, C., Grenouillet, G. & Brosse, S. (2013). Spatial range shape drives the grain size effects in species distribution models. *Ecography*, *36*(7), 778-787.

Lavaud, R., La Peyre, M.K., Casas, S.M., Bacher, C. & La Peyre, J.F. (2017). Integrating the effects of salinity on the physiology of the eastern oyster, *Crassostrea virginica*, in the northern Gulf of Mexico through a Dynamic Energy Budget model. *Ecological Modelling*, *363*, 221-233.

Lavaud, R., Thomas, Y., Pecquerie, L., Benoît, H.P., Guyondet, T., Flye-Sainte-Marie, J. & Chabot, D. (2019). Modeling the impact of hypoxia on the energy budget of Atlantic cod in two populations of the Gulf of Saint-Lawrence, Canada. *Journal of Sea Research*, *143*, 243-253.

Lavaud, R., Filgueira, R., Nadeau, A., Steeves, L. & Guyondet, T. (2020). A Dynamic Energy Budget model for the macroalga *Ulva lactuca*. *Ecological Modelling*, *418*, 108922.

Lavaud, R., La Peyre, M.K., Justic, D. & La Peyre, J.F. (2021). Dynamic Energy Budget modelling to predict eastern oyster growth, reproduction, and mortality under river management and climate change scenarios. *Estuarine, Coastal and Shelf Science*, *251*, 107188.

Lavoie, D.M., Smith, L.D. & Ruiz, G.M. (1999). The potential for intracoastal transfer of non-indigenous species in the ballast water of ships. *Estuarine, Coastal and Shelf Science*, *48*(5), 551-564.

Law Chune, S., Nouel, L., Fernandez, E. & Derval, C. (2019) Product Manual for the Global Ocean Sea Physical Analysis and Forecasting Products GLOBAL_ANALYSIS_FORECAST_PHY_001_024. 30pp.

Lawrence, J.M., Lawrence, A.L. & Giese, A.C. (1966). Role of the gut as a nutrient-storage organ in the purple sea urchin (*Strongylocentrotus purpuratus*). *Physiological Zoology*, 39(4), 281-290.

Lawrence, J.M. & McClintock, J.B. (1994). Energy acquisition and allocation by echinoderms (Echinodermata) in polar seas: adaptations for success. *Echinodermata. Balkema, Rotterdam*, 39-52.

Lawrence, J.M. (2013) Starfish: Biology and ecology of the Asteroidae. JHU Press, Baltimore, 267p.

Lawver, J.A., Cahagan, L. & Coffin, M. (1992) The development of paleo seaways around Antarctica. *Antarctic Research, 56*, 7-30.

Leach, K., Montgomery, W.I. & Reid, N. (2016). Modelling the influence of biotic factors on species distribution patterns. *Ecological Modelling*, 337, 96-106.

Leathwick, J.R. & Austin, M.P. (2001). Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*, *82*(9), 2560-2573.

Leathwick, J.R. (2002). Intra-generic competition among *Nothofagus* in New Zealand's primary indigenous forests. *Biodiversity & Conservation*, *11*(12), 2177-2187.

Leathwick, J.R., Elith, J. & Hastie, T. (2006). Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling*, 199(2), 188-196.

Ledoux, J.B., Tarnowska, K., Gerard, K., Lhuillier, E., Jacquemin, B., Weydmann, A., Féral, J.P. & Chenuil, A. (2012). Fine-scale spatial genetic structure in the brooding sea urchin *Abatus cordatus* suggests vulnerability of the Southern Ocean marine invertebrates facing global change. *Polar Biology*, 35. 611-623.

Legendre, P. & Fortin, M.J. (1989). Spatial pattern and ecological analysis. *Vegetation*, 80(2), 107-138.

Lee, J.E. & Chown, S.L. (2007). Mytilus on the move: transport of an invasive bivalve to the Antarctic. *Marine Ecology Progress Series*, *339*, 307-310.

Lee, J.E. & Chown, S.L. (2009a). Breaching the dispersal barrier to invasion: quantification and management. *Ecological Applications*, *19*(7), 1944-1959.

Lee, J.E. & Chown, S.L. (2009b). Quantifying the propagule load associated with the construction of an Antarctic research station. *Antarctic Science*, *21*(5), 471.

Lee, J.E. & Chown, S.L. (2009c). Temporal development of hull-fouling assemblages associated with an Antarctic supply vessel. *Marine Ecology Progress Series*, *386*, 97-105.

Leese, F., Agrawal, S. & Held, C. (2010). Long-distance island hopping without dispersal stages: transportation across major zoogeographic barriers in a Southern Ocean isopod. *Naturwissenschaften*, 97, 583–594.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74:1659–1673.

Legrand, S. & Dulière, V. (2014). OSERIT: A downstream service dedicated to the Belgian Coastguard Agencies. In: Dahlin H., Flemming N.C., Petersson S.E. (Eds.). Sustainable Operational Oceanography, Proc. 6th Conf. on EuroGOOS, 4-6 October, Sopot (Poland), 181-188.

Leibold, M.A. (1995). The niche concept revisited: mechanistic models and community context. *Ecology*, *76*(5), 1371-1382.

Leis, J.M., Caselle, J.E., Bradbury, I.R., Kristiansen, T., Llopiz, J.K., Miller, M.J., ... & Swearer, S.E. (2013). Does fish larval dispersal differ between high and low latitudes? *Proceedings of the Royal Society B: Biological Sciences*, *280*(1759), 20130327.

Lellouche, J-M., Legalloudec, O., Regnier, C., Levier, B., Greiner, E. & Drevillon, M. (2019). Quality Information Document For Global Sea Physical Analysis and Forecasting Product GLOBAL_ANALYSIS_FORECAST_PHY_001_024. 91pp.

Lennox, R., Choi, K., Harrison, P.M., Paterson, J.E., Peat, T.B., Ward, T.D. & Cooke, S.J. (2015). Improving science-based invasive species management with physiological knowledge, concepts, and tools. *Biological Invasions*, *17*(8), 2213-2227.

Lenz, B., Fogarty, N.D. & Figueiredo, J. (2019). Effects of ocean warming and acidification on fertilization success and early larval development in the green sea urchin *Lytechinus variegatus*. *Marine Pollution Bulletin*, *141*, 70-78.

Leohle, C. (1983). Evaluation of theories and calculation tools in ecology. *Ecological Modelling*, *19*(4), 239-247.

Leroy, B., Delsol, R., Hugueny, B., Meynard, C.N., Barhoumi, C., Barbet-Massin, M., & Bellard, C. (2018). Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography*, 45(9), 1994-2002.

Lett, C., Verley, P., Mullon, C., Parada, C., Brochier, T., Penven, P. & Blanke, B. (2008). A Lagrangian tool for modelling ichthyoplankton dynamics. *Environmental Modelling & Software*, *23*(9), 1210-1214.

Lett, C., Ayata, S.D., Huret, M. & Irisson, J.O. (2010). Biophysical modelling to investigate the effects of climate change on marine population dispersal and connectivity. *Progress in Oceanography*, *87*(1-4), 106-113.

Levin, S.A. (1992). The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, *73*(6), 1943-1967.

Levin, S.A. (1999). Fragile dominion: complexity and the commons. Perseus. *Reading, MA*.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*(4), 421-431.

Lewis, P.N., Hewitt, C.L., Riddle, M. & McMinn, A. (2003). Marine introductions in the Southern Ocean: an unrecognised hazard to biodiversity. *Marine Pollution Bulletin*, *46*(2), 213-223.

Lewis, P.N., Riddle, M.J. & Smith, S.D. (2005). Assisted passage or passive drift: a comparison of alternative transport mechanisms for non-indigenous coastal species into the Southern Ocean. *Antarctic Science, 17*(2), 183-191.

L'Heureux, M.L. & Thompson, W.J. (2006). Observed relationships between the El Niño–Southern Oscillation and the extratropical zonal-mean circulation. *Journal of Climate*, 19, 276–287.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Li, W. & Guo, Q. (2013). How to assess the prediction accuracy of species presence–absence models without absence data? *Ecography*, 36(7), 788-799.

Li, G., Du, S. & Guo, K. (2015). Correction: Evaluation of Limiting Climatic Factors and Simulation of a Climatically Suitable Habitat for Chinese Sea Buckthorn. *PloS One*, 10(8), e0136001

Li, J., Tran, M., & Siwabessy, J. (2016). Selecting optimal random forest predictive models: A case study on predicting the spatial distribution of seabed hardness. *PloS One*, 11(2), e0149089.

Lien, R., Solheim, A., Elverhøi, A. & Rokoengen, K. (1989). Iceberg scouring and sea bed morphology on the eastern Weddell Sea shelf, Antarctica. *Polar Research*, *7*(1), 43-57.

Lieske, D.J., Schmid, M.S. & Mahoney, M. (2018). Ensembles of Ensembles: Combining the Predictions from Multiple Machine Learning Methods. In *Machine Learning for Ecology and Sustainable Natural Resource Management* (pp. 109-121). Springer, Cham.

Lika, K., Kearney, M.R., Freitas, V., van der Veer, H.W., van der Meer, J., Wijsman, J.W. ... & Kooijman, S.A. (2011a). The "covariation method" for estimating the parameters of the standard Dynamic Energy Budget model I: Philosophy and approach. *Journal of Sea Research*, 66(4), 270-277.

Lika, K., Kearney, M.R. & Kooijman, S.A. (2011b) The "co-variation method" for estimating the parameters of the standard Dynamic Energy Budget model II: Properties and preliminary patterns. *Journal of Sea Research*, 66(4), 278-288.

Lika, K. & Kooijman, S.A. (2011). The comparative topology of energy allocation in budget models. *Journal of Sea Research*, *66*(4), 381-391.

Lika, K., Augustine, S. & Kooijman S.A. (2020). The use of augmented loss functions for estimating Dynamic Energy Budget parameters. *Ecological Modelling*, *428*, 109110.

Limpasuvan, V. & Hartmann, D.L. (1999). Eddies and the annular modes of climate variability. *Geophysical Research Letters*, 26, 3133–313. https://climatedataguide.ucar.edu/

Linse, K., Griffiths, H.J., Barnes, D.K., & Clarke, A. (2006). Biodiversity and biogeography of Antarctic and sub-Antarctic mollusca. *Deep Sea Research Part II: Topical Studies in Oceanography*, 53(8-10), 985-1008.

Linse, K. (2014) Bivalvia. In: De Broyer, C., Koubbi, P., Griffiths, H.J., et al. (eds) Biogeographic Atlas of the Southern Ocean. Scientific Committee on Antarctic Research, Cambridge UK, pp 126–128

Lischke, H., Guisan, A., Fischlin, A., Williams, J. & Bugmann, H. (1998). *Vegetation responses to climate change in the Alps: modeling studies* (pp. 309-350). MIT Press. Cambridge, MA, USA.

Lister, K.N., Lamare, M.D. & Burritt, D.J. (2015). Pollutant resilience in embryos of the Antarctic sea urchin *Sterechinus neumayeri* reflects maternal antioxidant status. *Aquatic Toxicology*, *161*, 61-72.

Liu, J. & Curry, J.A. (2010). Accelerated warming of the Southern Ocean and its impacts on the hydrological cycle and sea ice. *Proceedings of the National Academy of Sciences*, *107*(34), 14987-14992.

Liu, C., White, M. & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence‐only data. *Journal of Biogeography*, 40(4), 778-789.

Liu, C., Newell, G. & White, M. (2019). The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. *Ecography*, *42*(3), 535-548.

Liubartseva, S., Coppini, G., Lecci, R. & Clementi, E. (2018). Tracking plastics in the Mediterranean: 2D

Lagrangian model. *Marine Pollution Bulletin*, *129*(1), 151-162.

Lobo, J.M. (2008). More complex distribution models or more representative data? *Biodiversity Informatics*, *5*.

Lobo, J.M., Jiménez-Valverde, A. & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, *33*(1), 103-114.

Loehle, C. & Leblanc, D. (1996) Model-based assessments of climate change effects on forests: a critical review. *Ecological Modelling*, 90,1-31.

Lohrer, A.M., Cummings, V.J. & Thrush, S.F. (2013). Altered sea ice thickness and permanence affects benthic ecosystem functioning in coastal Antarctica. *Ecosystems*, *16*(2), 224-236.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G. & Williams, P.H. (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, *17*(6), 1591-1600.

Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G., & Montiel, O.M. (2008). Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, 35(1), 105-116.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J. & Guisan, A. (2010). Overcoming the rare species modelling paradox: a novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, 143(11), 2647-2657.

Loos, S.A. (2006). *Exploration of MARXAN for utility in Marine Protected Area zoning* (Doctoral dissertation).

Loots, C., Koubbi, P. & Duhamel, G. (2007). Habitat modelling of *Electrona antarctica* (Myctophidae, Pisces) in Kerguelen by generalized additive models and geographic information systems. *Polar Biology*, 30, 951-959.

López-Farrán, Z. / Guillaumot, C. (co-firstauthorship), Vargas-Chacoff, L., Paschke, K., Dulière, V., Danis, B., Poulin, E., Saucède, T., Gerard, K. (in press). Current and predicted invasive capacity of *Halicarcinus planatus* (Fabricius, 1775) in the Antarctic Peninsula. *Global Change Biology*.

López-Farrán, Z., Gerard, K., Saucède, T., Brickle, P., Waters, J., González-Wevar, C., Naretto, J., Poulin, E., Rosenfeld, S., Ojeda, J. & Ceroni C. (2020). Records of *Halicarcinus planatus* at sub-Antarctic regions. Version 1.2. Universidad de Magallanes. Occurrence dataset https://doi.org/10.15468/qnqnn2.

Lorena, A.C., Jacintho, L.F., Siqueira, M.F., De Giovanni, R., Lohmann, L.G., De Carvalho, A.C. & Yamamoto, M. (2011). Comparing machine learning classifiers in potential distribution modelling. *Expert Systems with Applications*, *38*(5), 5268-5275.

Lorenzo-Trueba, J., Voller, V.R. & Paola, C. (2010). Toward a model framework for sedimentary delta growth that accounts for biological processes. *AGUFM*, *2010*, B33D-0427.

Lozier, J.D., Aniello, P. & Hickerson, M.J. (2009). Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *Journal of Biogeography*, *36*(9), 1623-1627.

Lucas, J.S. (1980). Spider crabs of the family Hymenosomatidae (Crustacea; Brachyura) with particular reference to Australian species: systematics and biology. *Records of the Australian Museum*, 33(4), 148–247.

Ludvigsen, M., Eustice, R. & Singh, H. (2006). Photogrammetric models for marine archaeology. In *OCEANS 2006* (pp. 1-6). IEEE.

Luedeling, E., Kindt, R., Huth, N.I. & Koenig, K. (2014). Agroforestry systems in a changing climate—challenges in projecting future performance. *Current Opinion in Environmental Sustainability*, *6*, 1-7.

Luizza, M.W., Wakie, T., Evangelista, P.H. & Jarnevich, C.S. (2016). Integrating local pastoral knowledge, participatory mapping, and species distribution modeling for risk assessment of invasive rubber vine (Cryptostegia grandiflora) in Ethiopia's Afar region. *Ecology and Society*, 21(1), 1-22.

Luoto, M., Pöyry, J., Heikkinen, R.K. & Saarinen, K. (2005). Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography*, 14(6), 575-584.

Luu, K., Bazin, E. & Blum, M.G. (2016). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67-77.

Luyten, P. (2011). COHERENS --- A Coupled Hydrodynamical-Ecological Model for Regional and Shelf Seas: User Documentation. Version 2.0. RBINS-MUMM Report, Royal Belgian Institute of Natural Sciences, 1202pp.

Lynch, H.J., Crosbie, K., Fagan, W.F. & Naveen, R. (2010). Spatial patterns of tour ship traffic in the

Antarctic Peninsula region. *Antarctic Science*, *22*(2), 123-130.

M

Maes, D., Ellis, S., Goffart, P., Cruickshanks, K.L., van Swaay, C.A., Cors, R. ... & De Bruyn, L. (2019). The potential of species distribution modelling for reintroduction projects: the case study of the Chequered Skipper in England. *Journal of Insect Conservation*, *23*(2), 419-431.

Magniez, P. (1919). Modalités de l'incubation chez Abatus cordatus (Verrill), oursin endémique des íles Kerguelen. *Echinoderms: present and past. Balkema, Rotterdam*, 399-403.

Magniez, P. (1980). Le cycle sexuel d'*Abatus cordatus* (Echinoidea: Spatangoida): modalités d'incubation et évolution histologique et biochimique des gonades. Ph. D. Dissertation, Université Pierre et Marie Curie, Paris.

Magniez, P. (1983). Reproductive cycle of the brooding echinoid *Abatus cordatus* (Echinodermata) in Kerguelen (Antarctic Ocean): changes in the organ indices, biochemical composition and caloric content of the gonads. *Marine Biology*, 74(1), 55-64.

Magniez, P. & Féral, J.P. (1988). The effect of somatic and gonadal size on the rate of oxygen consumption in the sub-Antarctic echinoid *Abatus cordatus* (Echinodermata) from Kerguelen. *Comparative Biochemistry and Physiology Part A: Physiology*, 90(3), 429-434.

Mah, C.L., & Blake, D.B. (2012). Global diversity and phylogeny of the Asteroidea (Echinodermata). *PloS One*, 7(4), e35644.

Mainali, K.P., Warren, D.L., Dhileepan, K., McConnachie, A., Strathie, L., Hassan, G. ... & Parmesan, C. (2015). Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. *Global Change Biology*, 21(12), 4464-4480.

Majer, A.P., Petti, M.A., Corbisier, T.N., Ribeiro, A.P., Theophilo, C.Y., de Lima Ferreira, P.A. & Figueira, R.C. (2014). Bioaccumulation of potentially toxic trace elements in benthic organisms of Admiralty Bay (King George Island, Antarctica). *Marine Pollution Bulletin*, *79*(1-2), 321-325.

Malishev, M., Bull, C.M., Kearney, M.R. (2018). An individual-based model of ectotherm movement integrating metabolic and microclimatic constraints. *Methods in Ecology and Evolution*, 9, 472–489.

Malcolm, J.R., Liu, C., Neilson, R.P., Hansen, L. & Hannah, L. (2006). Global warming and extinctions of endemic species from biodiversity hotspots. *Conservation Biology*, 20(2), 538–548.

Maldonado, C., Molina, C.I., Zizka, A., Persson, C., Taylor, C.M., Albán, J., ... & Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography*, 24(8), 973-984.

Mangano, M.C., Giacoletti, A. & Sarà, G. (2019). Dynamic energy budget provides mechanistic derived quantities to implement the ecosystem based management approach. *Journal of Sea Research*, *143*, 272-279.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001). Evaluating presence–absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology, 38*(5), 921-931.

Manel, S., Loiseau, N., Andrello, M., Fietz, K., Goñi, R., Forcada, A., ... & Mouillot, D. (2019). Long-distance benefits of marine reserves: myth or reality? *Trends in Ecology & Evolution*, *34*(4), 342-354.

Marcer, A., Méndez-Vigo, B., Alonso-Blanco, C. & Picó, F.X. (2016). Tackling intraspecific genetic structure in distribution models better reflects species geographical range. *Ecology and Evolution*, *6*(7), 2084-2097.

Marina, T.I., Salinas, V., Cordone, G., Campana, G., Moreira, E., Deregibus, D., ... & Momo, F.R. (2018). The food web of Potter Cove (Antarctica): complexity, structure and function. *Estuarine, Coastal and Shelf Science*, *200*, 141-151.

Marini, M.Â., Barbet-Massin, M., Lopes, L.E. & Jiguet, F. (2010). Predicting the occurrence of rare Brazilian birds with species distribution models. *Journal of Ornithology*, *151*(4), 857-866.

Mariño, J., Augustine, S., Dufour, S. C. & Hurford, A. (2019). Dynamic Energy Budget theory predicts smaller energy reserves in thyasirid bivalves that harbour symbionts. *Journal of Sea Research*, 143, 119-127.

Markowska, M. & Kidawa, A. (2007). Encounters between Antarctic limpets, *Nacella concinna*, and predatory sea stars, *Lysasterias sp*., in laboratory and field experiments. *Marine Biology*, 151(5), 1959-1966.

Marmion, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009). The performance of state-of-the-art modelling techniques depends on geographical

distribution of species. *Ecological Modelling*, *220*(24), 3512-3520.

Marn, N., Jusup, M., Legović, T., Kooijman, S.A. & Klanjšček, T. (2017). Environmental effects on growth, reproduction, and life-history traits of loggerhead turtles. *Ecological Modelling*, *360*, 163-178.

Marn, N., Jusup, M., Catteau, S., Kooijman, S. A. & Klanjšček, T. (2019). Comparative physiological energetics of Mediterranean and North Atlantic loggerhead turtles. *Journal of Sea research*, 143, 100-118.

Marques, G.M., Mateus, M. & Domingos, T. (2014). Can we reach consensus between marine ecological models and DEB theory? A look at primary producers. *Journal of Sea Research*, 94, 92-104

Marques, G.M., Augustine, S., Lika, K., Pecquerie, L., Domingos, T. & Kooijman, S.A. (2018). The AmP project: comparing species on the basis of dynamic energy budget parameters. *PLoS Computational Biology*, 14(5), e1006100.

Marques, G.M., Lika, K., Augustine, S., Pecquerie, L. & Kooijman, S. A. (2019). Fitting Multiple Models to Multiple Data Sets. *Journal of Sea Research*. 143: 48-56

Martin, B.T., Zimmer, E.I., Grimm, V., Jager, T. (2010). DEB-IBM User Manual: Dynamic Energy Budget theory meets individual-based modelling: a generic and accessible implementation. https://www.bio.vu.nl/thb/deb/deblab/debibm/DEB_IBM_manual.pdf.

Martin, B.T., Zimmer, E.I., Grimm, V. & Jager, T. (2012). Dynamic Energy Budget theory meets individual-based modelling: a generic and accessible implementation. *Methods in Ecology and Evolution*, 3, 445-449.

Martin, B.T., Jager, T., Nisbet, R.M., Preuss, T.G., Hammers-Wirtz, M. & Grimm, V. (2013). Extrapolating ecotoxicological effects from individuals to populations: a generic approach based on Dynamic Energy Budget theory and individual-based modeling. *Ecotoxicology*, *22*(3), 574-583.

Martínez, B., Arenas, F., Trilla, A., Viejo, R.M. & Carreño, F. (2015). Combining physiological threshold knowledge to species distribution models is key to improving forecasts of the future niche for macroalgae. *Global Change Biology*, *21*(4), 1422-1433.

Marsh, A.G., Leong, P.K. & Manahan, D.T. (1999). Energy metabolism during embryonic development and larval growth of an Antarctic sea urchin. *Journal of Experimental Biology*, *202*(15), 2041-2050.

Marsh, A.G., Maxson, R.E. & Manahan, D.T. (2001). High macromolecular synthesis with low metabolic cost in Antarctic sea urchin embryos. *Science*, 291(5510), 1950-1952.

Marschoff, E.R., Barrera-Oro, E.R., Alescio, N.S. & Ainley, D.G. (2012). Slow recovery of previously depleted demersal fish at the South Shetland Islands, 1983–2010. *Fisheries Research*, *125*, 206-213.

Marshall, C.E., Glegg, G.A., & Howell, K.L. (2014). Species distribution modelling to support marine conservation planning: The next steps. *Marine Policy*, 45, 330–332.

Martin, B.T., Jager, T., Nisbet, R.M., Preuss, T.G. & Grimm, V. (2013). Predicting population dynamics from the properties of individuals: a cross-level test of dynamic energy budget theory. *The American Naturalist*, *181*(4), 506-519.

Martin, B., Jager, T., Nisbet, R.M., Preuss, T.G. & Grimm, V. (2014). Limitations of extrapolating toxic effects on reproduction to the population level. *Ecological Applications*, *24*(8), 1972-1983.

Martin, R. & Schlüter, M. (2015). Combining system dynamics and agent-based modeling to analyze social-ecological interactions—an example from modeling restoration of a shallow lake. *Frontiers in Environmental Science*, 3, 66.

Martin, A., Trouslard, E., Hautecoeur, M., Blettery, J., Moreau, C., Saucède, T., ... & Eléaume, M. (2019, January). Éco-régionalisation et conservation des communautés benthiques de la zone économique exclusive française des îles Kerguelen, Ecoregionalisation and conservation of benthic communities in the French exclusive economic zone of Kerguelen. In *Second Kerguelen Plateau Symposium on Marine Ecosystems and Fisheries. November 2017, Hobart, Tasmania*.

Martinson, D.G., Stammerjohn, S.E., Iannuzzi, R.A., Smith, R.C. & Vernet, M. (2008). Western Antarctic Peninsula physical oceanography and spatio–temporal variability. *Deep Sea Research Part II: Topical Studies in Oceanography*, *55*(18-19), 1964-1987.

Maruki, T. & Lynch, M. (2017). Genotype calling from population-genomic sequencing data. *G3: Genes, Genomes, Genetics*, *7*(5), 1393-1404.

Maschette, D., Sumner, M. & Raymond, B. (2019) *SOmap*: Southern Ocean maps. https://github.com/AustralianAntarcticDivision/SOmap

Massada, A.B., Syphard, A.D., Stewart, S.I. & Radeloff, V.C. (2013). Wildfire ignition-distribution modelling: a comparative study in the Huron–Manistee National Forest, Michigan, USA. *International Journal of wildland fire*, *22*(2), 174-183.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T.H., Piñero, D. & Emerson, B.C. (2014). Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, *15*(1), 28-41.

Mateo, R.G., Felicísimo, Á.M. & Muñoz, J. (2011). Modelos de distribución de especies: Una revisión sintética. *Revista Chilena de Historia Natural*, 84(2), 217–240.

Mateo, R.G., Felicísimo, Á.M., Pottier, J., Guisan, A. & Muñoz, J. (2012). Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS One*, *7*(3), e32586.

Mathewson, P.D., Moyer-Horner, L., Beever, E.A., Briscoe, N.J., Kearney, M., Yahn, J.M., & Porter, W.P. (2017). Mechanistic variables can enhance predictive models of endotherm distributions: the American pika under current, past, and future climates. *Global Change Biology*, 23(3), 1048-1064.

Matschiner, M., Hanel, R. & Salzburger, W. (2009). Gene flow by larval dispersal in the Antarctic notothenioid fish *Gobionotothen gibberifrons*. *Molecular Ecology*, *18*(12), 2574-2587.

Mauro, A., Arculeo, M. & Parrinello, N. (2003). Morphological and molecular tools in identifying the Mediterranean limpets *Patella caerulea*, *Patella aspera* and *Patella rustica*. *Journal of Experimental Marine Biology and Ecology*, 295(2), 131-143.

Maury, O. & Poggiale, J.C. (2013). From individuals to populations to communities: a dynamic energy budget model of marine ecosystem size-spectrum including life history diversity. *Journal of Theoretical Biology*, *324*, 52-71.

Maxwell, D.L., Stelzenmüller, V., Eastwood, P.D., & Rogers, S.I. (2009). Modelling the spatial distribution of plaice (*Pleuronectes platessa*), sole (*Solea solea*) and thornback ray (*Raja clavata*) in UK waters for marine management and planning. *Journal of Sea Research*, 61(4), 258-267.

Mayewski, P.A., Bracegirdle, T., Goodwin, I., Schneider, D., Bertler, N.A., Birkel, S., ... & Russell, J. (2015). Potential for Southern Hemisphere climate surprises. *Journal of Quaternary Science*, *30*(5), 391-395.

McArthur, R.H. (1972). Geographical ecology: patterns in the distribution of species. *Á Harper and Row*.

McCarthy, A.H., Peck, L.S., Hughes, K.A. & Aldridge, D.C. (2019). Antarctica: The final frontier for marine biological invasions. *Global Change Biology, 25*(7), 2221-2241.

McClanahan, T. & Kurtis, J.D. (1991). Population regulation of the rock-boring sea urchin *Echinometra mathaei* (de Blainville). *Journal of Experimental Marine Biology and Ecology*, 147, 121-146.

McClintock, J.B. (1994). Trophic biology of Antarctic shallow-water echinoderms. *Marine Ecology Progress Series.* 111(1), 191-202.

McClintock, J.B., Angus, R.A., Ho, C., Amsler, C.D. & Baker, B.J. (2008a) A laboratory study of behavioral interactions of the Antarctic keystone sea star *Odontaster validus* with three sympatric predatory sea stars. *Marine Biology*, 154(6), 1077-1084

McClintock, J.B., Ducklow, H. & Fraser, W. (2008b). Ecological Responses to Climate Change on the Antarctic Peninsula: the Peninsula is an icy world that's warming faster than anywhere else on Earth, threatening a rich but delicate biological community. *American Scientist*, *96*(4), 302-310.

McGeoch, M.A., Shaw, J.D., Terauds, A., Lee, J.E. & Chown, S.L. (2015). Monitoring biological invasion across the broader Antarctic: A baseline and indicator framework. Global *Environmental Change*, 32, 108–125.

McInnes, J.C., Jarman, S.N., Lea, M.A., Raymond, B., Deagle, B.E., Phillips, R.A., ... & Alderman, R. (2017). DNA metabarcoding as a marine conservation and management tool: A circumpolar examination of fishery discards in the diet of threatened albatrosses. *Frontiers in Marine Science*, *4*, 277.

McMahon, G., Wiken, E.B. & Gauthier, D.A. (2004). Toward a scientifically rigorous basis for developing mapped ecological regions. *Environmental Management*, *34*(1), S111-S124.

McManus, M.A. & Woodson, C.B. (2012). Plankton distribution and ocean dispersal. *Journal of Experimental Biology*, *215*(6), 1008-1016.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2004). The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, *41*(5), 811-823.

Mearns, L.O., Hulme, M., Carter, T.R., Leemans, R., Lal, M., Whetton, P. ... & Wilby, R. (2001). Climate scenario development. In *Climate change 2001: the science of climate change* (pp. 739-768). Cambrigde University Press. https://www.ipcc.ch/site/assets/uploads/2018/03/TAR-13.pdf

Mee, J.A., Bernatchez, L., Reist, J.D., Rogers, S.M. & Taylor, E.B. (2015). Identifying designatable units for intraspecific conservation prioritization: a hierarchical approach applied to the lake whitefish species complex (*Coregonus spp*.). *Evolutionary Applications*, *8*(5), 423-441.

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.C., Thuiller, W., Araújo, M.B., ... & Zimmermann, N.E. (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, *33*(6), 1038-1048.

Meier, E.S., Edwards Jr, T.C., Kienast, F., Dobbertin, M., & Zimmermann, N.E. (2011). Co-occurrence patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *Fagus sylvatica*. *Journal of Biogeography*, 38(2), 371-382.

Meineri, E., Deville, A.S., Grémillet, D., Gauthier-Clerc, M. & Béchet, A. (2015). Combining correlative and mechanistic habitat suitability models to improve ecological compensation. *Biological Reviews*, *90*(1), 314-329.

Meissner M (1900) Echinoideen. L. Friedrichsen & Company

Melrose, M.J. (1975). The marine fauna of New Zealand: family Hymenosomatidae (Crustacea, Decapoda, Brachyura) (Vol. 34). Wellington, N.Z.

MEPC, Marine Environment Protection Committee (2011) Guidelines for the control and management of ship's biofouling to minimize the transfer of invasive aquatic species. Annex 26. http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Marine-Environment-Protection-Committee-(MEPC)/Documents/MEPC.207(62).pdf

Meredith, M.P. & King, J.C. (2005). Rapid climate change in the ocean west of the Antarctic Peninsula during the second half of the 20th century. *Geophysical Research Letters*, 32(19).

Meredith, M.P. (2019). The global importance of the Southern Ocean and the key role of its freshwater cycle. In Ocean Challenge, vol. 23, No. 2, 49pp, Ed. Angela Colling.

Merow, C., Smith, M.J. & Silander, J.A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36(10), 1058-1069.

Merow, C., Smith, M.J., Edwards Jr, T.C., Guisan, A., McMahon, S.M., Normand, S., ... & Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, *37*(12), 1267-1281.

Merow, C., Wilson, A.M. & Jetz, W. (2017). Integrating occurrence data and expert maps for improved species range predictions. *Global Ecology and Biogeography*, *26*(2), 243-258.

Mesgaran, M.B., Cousens, R.D. & Webber, B.L. (2014). Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity and Distributions*, 20(10), 1147-1159.

Mespoulhé, P. (1992). Morphologie d'un échinide irrégulier subantarctique de l'archipel des Kerguelen: ontogenèse, dimorphisme sexuel et variabilité. Ph. D. Dissertation, Université de Bourgogne, Dijon.

Meynard, C.N. & Quinn, J.F. (2007). Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, *34*(8), 1455-1469.

Michel, L., David, B., Dubois, P., Lepoint, G. & De Ridder, C. (2016). Trophic plasticity of Antarctic echinoids under contrasted environmental conditions. *Polar Biology*. 39, 913-923.

Michel, L.N., Danis, B., Dubois, P., Eleaume, M., Fournier, J., Gallut, C., ... & Lepoint, G. (2019). Increased sea ice cover alters food web structure in East Antarctica. *Scientific Reports*, *9*(1), 1-11.

Milanesi, P., Herrando, S., Pla, M., Villero, D. & Keller, V. (2017). Towards continental bird distribution models: environmental variables for the second European breeding bird atlas and identification of priorities for further surveys. *Vogelwelt*, 137, 53-60.

Miller, J.R. Turner, M.G., Smithwick, E.A., Dent, C.L. & Stanley, E.H. (2004). Spatial extrapolation: the science of predicting ecological patterns and processes. *BioScience*, *54*(4), 310-320.

Miller, S.H. & Morgan, S.G. (2013). Interspecific differences in depth preference: Regulation of larval transport in an upwelling system. *Marine Ecology Progress Series, 476*. 301–306.

Miller, K.J., Baird, H.P., Van Oosterom, J., Mondon, J. & King, C.K. (2018). Complex genetic structure revealed in the circum-Antarctic broadcast spawning

sea urchin *Sterechinus neumayeri*. *Marine Ecology Progress Series*, *601*, 153-166.

Milligan, B.G., Archer, F.I., Ferchaud, A.L., Hand, B.K., Kierepka, E.M. & Waples, R.S. (2018). Disentangling genetic structure for genetic monitoring of complex populations. *Evolutionary Applications*, *11*(7), 1149-1161.

Mitchell, P.J., Monk, J. & Laurenson, L. (2017). Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Methods in Ecology and Evolution*, *8*(1), 12-21.

Miya, T., Gon, O., Mwale, M. & Cheng, C.H. (2013). The effect of habitat temperature on serum antifreeze glycoprotein (AFGP) activity in *Notothenia rossii* (Pisces: Nototheniidae) in the Southern Ocean. *Polar Biology*, *37*(3), 367-373.

Moffat, C. & Meredith, M. (2018). Shelf–ocean exchange and hydrography west of the Antarctic Peninsula: a review. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2122), 20170164.

Moles, J., Figuerola, B., Campanyà-Llovet, N., Monleón-Getino, T., Taboada, S., & Avila, C. (2015). Distribution patterns in Antarctic and Subantarctic echinoderms. *Polar Biology*, 38(6), 799-813.

Molina-Montenegro, M.A., Carrasco-Urra, F., Rodrigo, C., Convey, P., Valladares, F. & Gianoli, E. (2012). Occurrence of the non-native annual bluegrass on the Antarctic mainland and its negative effects on native plants. *Conservation Biology*, *26*(4), 717-723.

Moline, M.A., Claustre, H., Frazer, T.K., Schofield, O. & Vernet, M. (2004). Alteration of the food web along the Antarctic Peninsula in response to a regional warming trend. *Global Change Biology*, *10*(12), 1973-1980.

Moline, M.A., Karnovsky, N.J., Brown, Z., Divoky, G.J., Frazer, T.K., Jacoby, C.A., ... & Fraser, W.R. (2008). High latitude changes in ice dynamics and their impact on polar marine ecosystems. *Annals of the New York Academy of Sciences*, *1134*, 267.

Molloy, S.W., Davis, R.A., Dunlop, J.A. & van Etten, E. (2017). Applying surrogate species presences to correct sample bias in species distribution models: a case study using the Pilbara population of the Northern Quoll. *Nature Conservation,18*, 27-46.

Momigliano, P., Jokinen, H., Calboli, F., Aro, E. & Merilä, J. (2019). Cryptic temporal changes in stock composition explain the decline of a flounder (*Platichthys spp.)* assemblage. *Evolutionary Applications*, *12*(3), 549-559.

Monaco, C.J., Wethey, D.S. & Helmuth, B. (2013). A dynamic energy budget (DEB) model for the keystone predator *Pisaster ochraceus*. *PLoS One*, *9*(8), e104658.

Monaco, C.J. & McQuaid, C.D. (2018). Applicability of Dynamic Energy Budget (DEB) models across steep environmental gradients. *Scientific Reports*, *8*(1), 1-14.

Monaco, C.J., Porporato, E.M., Lathlean, J.A., Tagliarolo, M., Sarà, G. & McQuaid, C.D. (2019). Predicting the performance of cosmopolitan species: dynamic energy budget model skill drops across large spatial scales. *Marine Biology*, *166*(2), 14.

Moon, K.L., Chown, S.L. & Fraser, C.I. (2017). Reconsidering connectivity in the sub-Antarctic. *Biological Reviews*, *92*(4), 2164-2181.

Monk, J. (2014). How long should we ignore imperfect detection of species in the marine environment when modelling their distribution? *Fish and Fisheries*, *15*(2), 352-358.

Monien, D., Monien, P., Brünjes, R., Widmer, T., Kappenberg, A., Busso, A.A., ... & Brumsack, H.J. (2017). Meltwater as a source of potentially bioavailable iron to Antarctica waters. *Antarctic Science*, *29*(3), 277-291.

Montalto, V., Rinaldi, A. & Sara, G. (2015). Life history traits to predict biogeographic species distributions in bivalves. *The Science of Nature*, *102*(9-10), 61.

Montes-Hugo, M., Doney, S.C., Ducklow, H.W., Fraser, W., Martinson, D., Stammerjohn, S.E. & Schofield, O. (2009). Recent changes in phytoplankton communities associated with rapid regional climate change along the western Antarctic Peninsula. *Science*, *323*(5920), 1470-1473.

Moore, J.M., Carvajal, J.I., Rouse, G.W. & Wilson, N.G. (2018). The Antarctic Circumpolar Current isolates and connects: Structured circumpolarity in the sea star *Glabraster antarctica*. *Ecology and Evolution*, *8*(21), 10621-10633.

Morales, N.S., Fernández, I.C. & Baca-González, V. (2017). MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. *PeerJ*, *5*, e3093.

Morales-Castilla, I., Davies, T.J., Pearse, W.D. & Peres-Neto, P. (2017). Combining phylogeny and co-occurrence to improve single species distribution models. *Global Ecology and Biogeography*, *26*(6), 740-752.

Moreau, C., Saucède, T., Jossart, Q., Agüera, A., Brayard, A., & Danis, B. (2017). Reproductive strategy as a piece of the biogeographic puzzle: a case study using Antarctic sea stars (Echinodermata, Asteroidea). *Journal of Biogeography*, 44(4), 848-860.

Moreau, C., Mah, C., Agüera, A., Améziane, N., Barnes, D., Crokaert, G. ... & Jażdżewska, A. (2018). Antarctic and sub-Antarctic Asteroidea database. *ZooKeys*, (747), 141.

Moreau, C., Jossart, Q., Danis, B., Eléaume, M., Christiansen, H., Guillaumot, C., ... & Saucède, T. (2021). The high diversity of Southern Ocean sea stars (Asteroidea) reveals original evolutionary pathways. *Progress in Oceanography*, *190*, 102472.

Moreira, E., Barrera-Oro, E. & La Mesa, M. (2013). Age validation of juvenile *Notothenia rossii* at Potter Cove, South Shetland Islands, using mark-recapture data. *Polar Biology*, *36*(12), 1845-1850.

Moreno-Amat, E., Mateo, R.G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.C. & García-Amorena, I. (2015). Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. *Ecological Modelling*, *312*, 308-317.

Moon, K.L., Chown, S.L. & Fraser, C.I. (2017). Reconsidering connectivity in the sub-Antarctic. *Biological Reviews*, *92*(4), 2164-2181.

Morelle, K. & Lejeune, P. (2015). Seasonal variations of wild boar *Sus scrofa* distribution in agricultural landscapes: a species distribution modelling approach. *European Journal of Wildlife Research*, 61(1), 45-56.

Morin, X. & Thuiller, W. (2009). Comparing niche-and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology*, *90*(5), 1301-1313.

Morley, S.A., Hirse, T., Pörtner, H.O. & Peck, L.S. (2009a) Geographical variation in thermal tolerance within Southern Ocean marine ectotherms. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 153: 154-161.

Morley, S.A., Lurman, G.J., Skepper, J.N., Pörtner, H.O. & Peck, L. S. (2009b). Thermal plasticity of mitochondria: a latitudinal comparison between Southern Ocean molluscs. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 152(3), 423-430.

Morley, S.A., Clark, M.S. & Peck, L. S. (2010). Depth gradients in shell morphology correlate with thermal limits for activity and ice disturbance in Antarctic limpets. *Journal of Experimental Marine Biology and Ecology*, 390(1), 1-5.

Morley, S.A., Lemmon, V., Obermüller, B.E., Spicer, J.I., Clark, M.S. & Peck, L.S. (2011). Duration tenacity: a method for assessing acclimatory capacity of the Antarctic limpet, *Nacella concinna*. *Journal of Experimental Marine Biology and Ecology*, 399(1), 39-42.

Morley, S.A., Martin, S.M., Day, R.W., Ericson, J., Lai, C.H., Lamare, M. ... & Peck, L.S. (2012). Thermal reaction norms and the scale of temperature variation: latitudinal vulnerability of intertidal nacellid limpets to climate change. *PloS One*, 7(12).

Morley, S.A., Lai, C.H., Clarke, A., Tan, K.S., Thorne, M.A. & Peck, L.S. (2014). Limpet feeding rate and the consistency of physiological response to temperature. *Journal of Comparative Physiology Part B*, 184(5), 563-570.

Morley, S.A., Suckling, C.C., Clark, M.S., Cross, E.L. & Peck, L.S. (2016). Long-term effects of altered pH and temperature on the feeding energetics of the Antarctic sea urchin, *Sterechinus neumayeri*. *Biodiversity*, *17*(1-2), 34-45.

Mormède, S., Dunn, A., Hanchet, S. & Parker, S. (2014a). Spatially explicit population dynamics models for Antarctic toothfish in the Ross Sea region. *CCAMLR Science*, *21*, 19-37.

Mormède, S., Dunn, A. & Hanchet, S.M. (2014b). A stock assessment model of Antarctic toothfish (*Dissostichus mawsoni*) in the Ross Sea region incorporating multi-year mark-recapture data. *CCAMLR Science*, *21*, 39-62.

Mormède, S., Irisson, J.O. & Raymond, B. (2014c) Distribution modelling. Biogeographic atlas of the Southern Ocean (ed. by C. De Broyer, P. Koubbi, H.J. Griffiths, B. Raymond, C. d'Udekem d'Acoz, Van dePutte A., B. Danis, B. David, S. Grant,J. Gutt, C. Held, G. Hosie, F. Huettmann, A. Post and Y. Ropert-Coudert), pp. 510. SCAR, Cambridge.

Motreuil, S., Dubois, P., Thellier, T., Marty, G., Marschal, C. & Saucède, T (2018). PROTEKER, rapport de campagne d'été 2018/2019. Impact du changement global sur le benthos et les habitats marins côtiers des Iles Kerguelen. https://doi.org/10.5281/zenodo.4420834

Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., ... & Loreau, M. (2015). Predictive ecology in a changing world. *Journal of Applied Ecology*, *52*(5), 1293-1310.

Moya, F., Saucède, T. & Manjón-Cabeza, M.E. (2012). Environmental control on the structure of

echinoid assemblages in the Bellingshausen Sea (Antarctica). *Polar Biology*, *35*(9), 1343-1357.

Mulcahy, K.A. & Clarke, K.C. (2001). Symbolization of map projection distortion: a review. *Cartography and Geographic Information Science*, *28*(3), 167-182.

Muller, E.B. & Nisbet, R.M. (1997). Modeling the effect of toxicants on the parameters of dynamic energy budget models. In Environmental Toxicology and Risk Assessment: Modeling and Risk Assessment Sixth Volume. ASTM International.

Muller, E.B., Nisbet, R.M. & Berkley, H.A. (2010). Sublethal toxicant effects with dynamic energy budget theory: model formulation. *Ecotoxicology*, *19*(1), 48-60.

Muller, E.B. (2011). Synthesizing units as modeling tool for photosynthesizing organisms with photoinhibition and nutrient limitation. *Ecological Modelling*, *222*(3), 637-644.

Muller, E.B. & Nisbet, R.M. (2014). Dynamic energy budget modeling reveals the potential of future growth and calcification for the coccolithophore *Emiliania huxleyi* in an acidified ocean. *Global Change Biology*, *20*(6), 2031-2038.

Muller, E.B., Klanjšček, T. & Nisbet, R.M. (2019). Inhibition and damage schemes within the synthesizing unit concept of dynamic energy budget theory. *Journal of Sea Research*, *143*, 165-172.

Muñoz, F., Pennino, M.G., Conesa, D., López-Quílez, A. & Bellido, J.M. (2013). Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. *Stochastic Environmental Research and Risk Assessment*, *27*(5), 1171-1180.

MUR MEaSUREs Project JPL. (2015). GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1). Ver. 4.1. PO.DAAC, CA, USA. https://doi.org/10.5067/GHGMR-4FJ04.

Murase, H., Kitakado, T., Hakamada, T., Matsuoka, K., Nishiwaki, S., & Naganobu, M. (2013). Spatial distribution of Antarctic minke whales (*Balaenoptera bonaerensis*) in relation to spatial distributions of krill in the Ross Sea, Antarctica. *Fisheries Oceanography*, 22(3), 154-173.

Murphey, P.C., Guralnick, R P., Glaubitz, R., Neufeld, D., & Ryan, J.A. (2004). Georeferencing of museum collections: A review of problems and automated tools, and the methodology developed by the Mountain and Plains Spatio-Temporal Database-Informatics Initiative (Mapstedi). *PhyloInformatics*, 3, 1-29.

Murphy, E.J., Watkins, J.L., Trathan, P.N., Reid, K., Meredith, M.P., Thorpe, S.E., ... & Fleming, A.H. (2007). Spatial and temporal operation of the Scotia Sea ecosystem: a review of large-scale links in a krill centred food web. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1477), 113-148.

Murphy, E. & Hofmann, E. (2012). End-to-end in Southern Ocean ecosystems. Current Opinion in Environmental Sustainability, 4, 264–271.

Murray, R.J. (1996). Explicit generation of orthogonal grids for ocean models. *Journal of Computational Physics*, 126(2), 251-273.

Muscarella, R., Galante, P.J., Soley‐Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M. & Anderson, R.P. (2014). *ENMeval*: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5(11), 1198-1205.

N

Nachtsheim, D.A., Jerosch, K., Hagen, W., Plötz, J., & Bornemann, H. (2017). Habitat modelling of crabeater seals (*Lobodon carcinophaga*) in the Weddell Sea using the multivariate approach Maxent. *Polar Biology*, 40(5), 961-976.

Naimi, B., Skidmore, A.K., Groen, T.A. & Hamm, N.A. (2011). Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *Journal of Biogeography*, *38*(8), 1497-1509.

Naimi, B., Hamm, N.A., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37(2), 191-203.

Naimi, B. & Araújo, M.B. (2016). *sdm*: a reproducible and extensible R platform for species distribution modelling. *Ecography*, *39*(4), 368-375.

Naimi, B., Araújo, M.B., Naimi, M.B., Naimi, B. & Araujo, M.B. (2019). Package 'sdm'. https://cran.r-project.org/web/packages/sdm/index.html

Navarro, J.M., Détrée, C., Morley, S.A., Cárdenas, L., Ortiz, A., Vargas-Chacoff, L., ... & Gonzalez-Wevar, C. (2020). Evaluating the effects of ocean warming and freshening on the physiological energetics and transcriptomic response of the Antarctic limpet *Nacella concinna*. *Science of the Total Environment*, *748*, 142448.

Near, T.J., Dornburg, A., Kuhn, K.L., Eastman, J.T., Pennington, J.N., Patarnello, T. ... & Jones, C.D. (2012). Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of Sciences*, 109(9), 3434-3439.

Nenzén, H.K., Swab, R.M., Keith, D.A., & Araújo, M.B. (2012). *demoniche*: a R-package for simulating spatially-explicit population dynamics. *Ecography*, *35*(7), 577-580.

Newbold, T. (2010). Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, 34(1), 3-22.

Nicodemus-Johnson, J., Silic, S., Ghigliotti, L., Pisano, E. & Cheng, C.H. (2011). Assembly of the antifreeze glycoprotein/trypsinogen-like protease genomic locus in the Antarctic toothfish *Dissostichus mawsoni* (Norman). *Genomics*, *98*(3), 194-201.

Nicol, S. & Foster, J. (2003). Recent trends in the fishery for Antarctic krill. *Aquatic Living Resources*, *16*(1), 42-45.

Nicolas, J.P., Vogelmann, A.M., Scott, R.C., Wilson, A.B., Cadeddu, M.P., Bromwich, D.H., ... & Powers, H.H. (2017). January 2016 extensive summer melt in West Antarctica favoured by strong El Niño. *Nature Communications*, *8*, 15799.

Niemandt, C., Kovacs, K.M., Lydersen, C., Dyer, B.M., Isaksen, K., Hofmeyr, G.G., ... & de Bruyn, P.N. (2016). Chinstrap and macaroni penguin diet and demography at Nyrøysa, Bouvetøya. *Antarctic Science*, *28*(2), 91.

Nilsson, J.A., Fulton, E.A., Haward, M. & Johnson, C. (2016). Consensus management in Antarctica's high seas–Past success and current challenges. *Marine Policy*, *73*, 172-180.

Nisbet, R.M., Muller, E.B., Lika, K. & Kooijman, S.A. (2000). From molecules to ecosystems through dynamic energy budget models. *Journal of Animal Ecology*, 913-926.

Nisbet, R.M., Jusup, M., Klanjscek, T. & Pecquerie, L. (2012). Integrating dynamic energy budget (DEB) theory with traditional bioenergetic models. *Journal of Experimental Biology*, 215(6), 892-902.

Nixon, S.W. (1981). Remineralization and nutrient cycling in coastal marine ecosystems. In *Estuaries and nutrients* (pp. 111-138). Humana Press.

Nolan, C.P. (1991). Size, shape and shell morphology in the Antarctic limpet *Nacella concinna* at Signy Island, South Orkney Islands. *Journal of Molluscan Studies*, 57(2), 225-238.

Nolan, C.P. & Clarke, A. (1993). Growth in the bivalve *Yoldia eightsi* at Signy Island, Antarctica, determined from internal shell increments and calcium-45 incorporation. *Marine Biology*, *117*(2), 243-250.

Nori, J., Akmentins, M.S., Ghirardi, R., Frutos, N. & Leynaud, G.C. (2011). American bullfrog invasion in Argentina: where should we take urgent measures? *Biodiversity and Conservation*, 20(5), 1125-1132.

Normand, S., Randin, C., Ohlemüller, R. *et al*. (2013) A greener Greenland? Climatic potential and long-term constraints on future expansions of trees and shrubs. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368, 20120479

North, A.W. (2001) Early life history strategies of notothenioids at South Georgia. *Journal of Fish Biology*, 58, 496–505.

North, E.W., Schlag, Z., Hood, R.R., Li, M., Zhong, L., Gross, T. & Kennedy, V.S. (2008). Vertical swimming behaviour influences the dispersal of simulated oyster larvae in a coupled particle-tracking and hydrodynamic model of Chesapeake Bay. *Marine Ecology Progress Series*, 359:99–115.

O

Oberle, B., Ogle, K., Zuluaga, J.C., Sweeney, J. & Zanne, A.E. (2016). A Bayesian model for xylem vessel length accommodates subsampling and reveals skewed distributions in species that dominate seasonal habitats. *Journal of Plant Hydraulics*, 3, e003.

Obermüller, B.E., Morley, S.A., Clark, M.S., Barnes, D.K. & Peck, L.S. (2011). Antarctic intertidal limpet ecophysiology: A winter–summer comparison. *Journal of Experimental Marine Biology and Ecology*, 403(1-2), 39-45.

O'Brien, P.E., Post, A., Romey, R. (2009) Antarctic-wide Geomorphology as an aid to habitat mapping and locating Vulnerable Marine Ecosystems. Science Committee to the Commission of Antartctic Marine Living Resources (SC-CAMLR-XXVIII/10) Workshop on Vulnerable Marine Ecosystems. GeoScience Australia. Conference paper: WS-VME-09/10. La Jolla, CA, USA.

Obryk, M.K., Doran, P.T., Friedlaender, A.S., Gooseff, M.N., Li, W., Morgan-Kiss, R.M., ... & Ducklow, H.W. (2016). Responses of Antarctic marine and freshwater ecosystems to changing ice conditions. *BioScience*, *66*(10), 864-879.

Ocaranza-Barrera, P., González-Wevar, C.A., Guillemin, M.L., Rosenfeld, S. & Mansilla, A. (2019). Molecular divergence between *Iridaea cordata* (Turner) Bory de Saint-Vincent from the Antarctic Peninsula and the Magellan Region. *Journal of Applied Phycology*, *31*(2), 939-949.

Oksanen, J., Blanchet, F.G., Friendly, M. *et al.* (2018) vegan: Community Ecology Package. R package version 2.4-6. https://CRAN.R-project.org/package=vegan

Okuno, J., Miura, H. & Nogi, Y. (2012). Effect of glacial isostasy on the depth of Antarctic continental margin. In *EGU General Assembly Conference Abstracts* (Vol. 14, p. 6877).

Olbers, D., Borowski, D., Völker, C. & Woelff, J.O. (2004). The dynamical balance, transport and circulation of the Antarctic Circumpolar Current. *Antarctic Science*, *16*(4), 439-470.

Olden, J.D., Jackson, D.A. & Peres-Neto, P.R. (2002). Predictive models of fish species distributions: a note on proper validation and chance predictions. *Transactions of the American Fisheries Society*, 131(2), 329-336.

Olden, J.D., Lawler, J.J., & Poff, N.L. (2008). Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology*, 83(2), 171-193.

O'Leary, S.J., Puritz, J.B., Willis, S.C., Hollenbeck, C.M. & Portnoy, D.S. (2018). These aren't the loci you're looking for: principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, 27(16), 3193-3206.

Oliver, J. (2018). A very brief introduction to species distribution models in R. https://jcoliver.github.io/learn-r/011-species-distribution-models.html

Olsen, G. (1954). South Georgian cod, Fischer. *Norsk Hvalfangstidende*, **43**, 373-382.

Orejas, C., Gili, J., López-González, P.J. & Arntz, W. (2001). Feeding strategies and diet composition of four Antarctic cnidarian species. *Polar Biology*, 24, 620–627.

Oreskes, N., Shrader-Frechette, K. & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*(5147), 641-646.

Orsi, A.H., Whitworth., T. & Nowlin Jr, W.D. (1995). On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep Sea Research Part I: Oceanographic Research Papers*, *42*(5), 641-673.

Orsi, A.H., Harris, U. (2019) Fronts of the Antarctic Circumpolar Current - GIS data, Ver. 1, *Australian Antarctic Data Centre* - https://data.aad.gov.au/metadata/records/antarctic_circumpolar_current_fronts.

Ortiz, M., Berrios, F., González, J., Rodríguez-Zaragoza, F. & Gómez, I. (2016). Macroscopic network properties and short-term dynamic simulations in coastal ecological systems at Fildes Bay (King George Island, Antarctica). *Ecological Complexity*, *28*, 145-157.

Ortuño Crespo, G.O., Dunn, D.C., Gianni, M., Gjerde, K., Wright, G. & Halpin, P.N. (2019). High-seas fish biodiversity is slipping through the governance net. *Nature Ecology & Evolution*, *3*(9), 1273-1276.

Osborne, P.E. & Leitão, P.J. (2009). Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Diversity and Distributions*, *15*(4), 671-681.

Owens, H.L., Campbell, L.P., Dornak, L.L., Saupe, E.E., Barve, N., Soberón, J. ... & Peterson, A.T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, 263, 10-18.

P

Pace, D.A. & Manahan, D.T. (2007). Cost of protein synthesis and energy allocation during development of Antarctic sea urchin embryos and larvae. *The Biological Bulletin*, *212*(2), 115-129.

Padilla, A., Zeller, D. & Pauly, D. (2015) The fish and fisheries of Bouvet Island. In: *Marine Fisheries Catches of SubAntarctic Islands, 1950-2010* (eds Palomares MLD, Pauly D), pp. 21–30. Fisheries Centre, University of British Columbia, Vancouver, BC.

Padman, L., Costa, D.P., Dinniman, M.S., Fricker, H.A., Goebel, M.E., Huckstadt, L.A., ... & Scambos, T. (2012). Oceanic controls on the mass balance of Wilkins Ice Shelf, Antarctica. *Journal of Geophysical Research: Oceans*, *117*(C1).

Pagel, J. & Schurr, F.M. (2012). Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, 21(2), 293-304.

Palma, A.T., Poulin, E., Silva, M.G., San Martín, R.B., Muñoz, C.A. & Díaz, A.D. (2007). Antarctic shallow

subtidal echinoderms: is the ecological success of broadcasters related to ice disturbance? *Polar Biology*, *30*(3), 343-350.

Paracuellos, M., Nevado, J.C., Moreno, D., Giménez, A. & Alesina, J.J. (2003). Conservational status and demographic characteristics of *Patella ferruginea* Gmelin, 1791 (Mollusca, gastropoda) on the Alboran Island (Western Mediterranean). *Animal Biodiversity and Conservation*, 26(2), 29-37.

Paradis, E., Strimmer, K., Claude, J., Jobb, G., Opgen-Rhein, R., Dutheil, J., ... & Lemon, J. (2008). The ape package. *Analyses of Phylogenetics and Evolution*. https://ms.mcmaster.ca/~bolker/eeid/2009/Evolution/2009_Day1_phylogenetics/R_phylo_resources/ape.pdf

Paradis, E. (2010). *Pegas*: A R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419–420.

Pardo-Gandarillas, M.C., Ibáñez, C.M., Torres, F.I., Sanhueza, V., Fabres, A., Escobar-Dodero, J., ... & Méndez, M.A. (2018). Phylogeography and species distribution modelling reveal the effects of the Pleistocene ice ages on an intertidal limpet from the south-eastern Pacific. *Journal of Biogeography*, *45*(8), 1751-1767.

Parker, S.J., Mormède, S., Hanchet, S.M., Devries, A., Canese, S. & Ghigliotti, L. (2019). Monitoring Antarctic toothfish in McMurdo Sound to evaluate the Ross Sea region marine protected area. *Antarctic Science*, *31*(4), 195-207.

Parrott, L., Chion, C., Gonzalès, R. & Latombe, G. (2012). Agents, individuals, and networks: modeling methods to inform natural resource management in regional landscapes. *Ecology and Society*, *17*(3).

Pascal P.Y., Reynaud, Y., Poulin, E., De Ridder, C. & Saucède, T. (2021, in press). Feeding in spatangoids: the case of *Abatus cordatus* in the Kerguelen Islands (Southern Ocean). *Polar Biology*.

Paz-Vinas, I., Loot, G., Hermoso, V., Veyssiere, C., Poulet, N., Grenouillet, G. & Blanchet, S. (2018). Systematic conservation planning for intraspecific genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1877), 20172746.

Pauline, C.Y., Sewell, M.A., Matson, P.G., Rivest, E.B., Kapsenberg, L. & Hofmann, G.E. (2013). Growth attenuation with developmental schedule progression in embryos and early larvae of *Sterechinus neumayeri* raised under elevated CO_2. *PloS One*, *8*(1), e52448.

Pearce, J. & Lindenmayer, D. (1998). Bioclimatic analysis to enhance reintroduction biology of the endangered helmeted honeyeater (*Lichenostomus melanops cassidix*) in southeastern Australia. *Restoration Ecology*, *6*(3), 238-243.

Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008). Niche dynamics in space and time. *Trends in Ecology & Evolution*, *23*(3), 149-158.

Pearman, T.R., Robert, K., Callaway, A., Hall, R., Iacono, C.L. & Huvenne, V.A. (2020). Improving the predictive capability of benthic species distribution models by incorporating oceanographic data–towards holistic ecological modelling of a submarine canyon. *Progress in Oceanography*, 102338.

Pearse, J.S. & Giese, A.C. (1966). Food, reproduction and organic constitution of the common Antarctic echinoid *Sterechinus neumayeri* (Meissner, 1900). *The Biological Bulletin*, *130*(3), 387-401.

Pearse, J.S., McClintock, J.B. & Bosch, I. (1991). Reproduction of Antarctic benthic marine invertebrates: tempos, modes, and timing. *American Zoologist*, *31*(1), 65-80.

Pearse, J.S., Mooi, R., Lockhart, S.J. & Brandt, A. (2009). Brooding and species diversity in the Southern Ocean: selection for brooders or speciation within brooding clades? *Smithsonian at the poles: Contributions to international Polar Year Science*.

Pearse, W.D., Barbosa, A.M., Fritz, S.A., Keith, S.A., Harmon, L.J., Harte, J., ... & Davies, T.J. (2018). Building up biogeography: Pattern to process. *Journal of Biogeography*, *45*(6), 1223-1230.

Pearson, R.G. & Dawson, T.P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, *12*(5), 361-371.

Pearson, R.G., Dawson, T.P. & Liu, C. (2004). Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, *27*(3), 285-298.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C. ... & Lees, D.C. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography*, *33*(10), 1704-1711.

Pearson, R.G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, *50*, 54-89.

Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Townsend Peterson, A. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in

Madagascar. *Journal of Biogeography*, *34*(1), 102-117.

Peck, L.S. (1989). Temperature and basal metabolism in two Antarctic marine herbivores. *Journal of Experimental Marine Biology and Ecology*, 127(1), 1-12.

Peck, L.S. & Uglow, R.F. (1990). Two methods for the assessment of the oxygen content of small volumes of seawater. *Journal of Experimental Marine Biology and Ecology*, *141*(1), 53-62.

Peck, L.S. & Bullough, L.W. (1993). Growth and population structure in the infaunal bivalve *Yoldia eightsi* in relation to iceberg activity at Signy Island, Antarctica. *Marine Biology*, *117*(2), 235-241.

Peck, L.S. & Brey, T. (1996) Radiocarbon bomb signals verify biennial growth bands in the shells of 50 year old brachiopods from Antarctica. *Nature*, 380, 206–207.

Peck, L.S. & Conway, L.Z. (2000). The myth of metabolic cold adaptation: oxygen consumption in stenothermal Antarctic bivalves. *Geological Society, London, Special Publications*, *177*(1), 441-450.

Peck, L.S. (2002). Ecophysiology of Antarctic marine ectotherms: limits to life. In Ecological Studies in the Antarctic Sea Ice Zone (pp. 221-230). Springer, Berlin, Heidelberg.

Peck, L.S., Webb, K.E. & Bailey, D.M. (2004). Extreme sensitivity of biological function to temperature in Antarctic marine species. *Functional Ecology*, *18*(5), 625-630.

Peck, L.S. (2005). Prospects for surviving climate change in Antarctic aquatic species. *Frontiers in Zoology*, *2*(1), 1-8.

Peck, L.S., Barnes, D.K. & Willmott, J. (2005). Responses to extreme seasonality in food supply: diet plasticity in Antarctic brachiopods. *Marine Biology*, 147(2), 453–463.

Peck, L.S., Convey, P. & Barnes, D.K. (2006). Environmental constraints on life histories in Antarctic ecosystems: tempos, timings and predictability. *Biological Reviews*, *81*(1), 75-109.

Peck, L.S., Powell, D.K. & Tyler, P.A. (2007). Very slow development in two Antarctic bivalve molluscs, the infaunal clam *Laternula elliptica* and the scallop *Adamussium colbecki*. *Marine Biology*, 150(6), 1191-1197.

Peck, L.S., Massey, A., Thorne, M.A. & Clark, M.S. (2009a). Lack of acclimation in *Ophionotus victoriae*: brittle stars are not fish. *Polar Biology*, *32*(3), 399-402.

Peck, L.S., Clark, M.S., Morley, S.A., Massey, A. & Rossetti, H. (2009b). Animal temperature limits and ecological relevance: effects of size, activity and rates of change. *Functional Ecology*, 23(2), 248-256.

Peck, L.S. (2011). Organisms and responses to environmental change. *Marine Genomics*, *4*(4), 237-243.

Peck, L.S., Morley, S.A., Richard, J. & Clark, M.S. (2014). Acclimation and thermal tolerance in Antarctic marine ectotherms. *Journal of Experimental Biology*, *217*(1), 16-22.

Peck, L.S. (2016). A cold limit to adaptation in the sea. *Trends in Ecology & Evolution*, *31*(1), 13-26.

Peck, L.S., Heiser, S., & Clark, M.S. (2016). Very slow embryonic and larval development in the Antarctic limpet *Nacella polaris*. *Polar Biology*, 39(12), 2273-2280.

Peck, V.L., Oakes, R.L., Harper, E.M., Manno, C. & Tarling, G.A. (2018). Pteropods counter mechanical damage and dissolution through extensive shell repair. *Nature Communications*, *9*(1), 264.

Pecquerie, L., Petitgas, P., & Kooijman, S. A. (2009). Modeling fish growth and reproduction in the context of the Dynamic Energy Budget theory to predict environmental impact on anchovy spawning duration. *Journal of Sea Research*, 62(2-3), 93-105.

Pecquerie, L., Nisbet, R.M., Fablet, R., Lorrain, A. & Kooijman, S.A. (2010). The impact of metabolism on stable isotope dynamics: a theoretical framework. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1557), 3455-3468.

Peel, D., Waples, R.S., Macbeth, G.M., Do, C. & Ovenden, J.R. (2013). Accounting for missing data in the estimation of contemporary genetic effective population size (Ne). *Molecular Ecology Resources*,*13*(2), 243-253.

Peel, S.L., Hill, N.A., Foster, S.D., Wotherspoon, S.J., Ghiglione, C. & Schiaparelli, S. (2019) Reliable species distributions are obtainable with sparse, patchy and biased data by leveraging over species and data types. *Methods in Ecology and Evolution*. *10*(7), 1002-1014.

Peeters, F., Li, J., Straile, D., Rothhaupt, K.O. & Vijverberg, J. (2010). Influence of low and decreasing food levels on Daphnia-algal interactions: Numerical experiments with a new dynamic energy budget model. *Ecological Modelling*, 221(22), 2642-2655.

Pellissier, L., Anne Bråthen, K., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A. ... & Guisan, A. (2010). Species distribution models reveal apparent competitive and facilitative effects of a dominant

species on the distribution of tundra plants. *Ecography, 33*(6), 1004-1014.

Pellissier, L., Rohr, R.P., Ndiribe, C., Pradervand, J.N., Salamin, N., Guisan, A. & Wisz, M. (2013). Combining food web and species distribution models for improved community projections. *Ecology and Evolution, 3*(13), 4572-4583.

Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., ... & Gravel, D. (2018). Comparing species interaction networks along environmental gradients. *Biological Reviews*, *93*(2), 785-800.

Pennino, M.G., Muñoz, F., Conesa, D., López-Quílez, A. & Bellido, J.M. (2014). Bayesian spatio-temporal discard model in a demersal trawl fishery. *Journal of Sea Research*, *90*, 44-53.

Pérez-Portela, R., Bumford, A., Coffman, B., Wedelich, S., Davenport, M., Fogg, A., ... & Oleksiak, M.F. (2018). Genetic homogeneity of the invasive lionfish across the Northwestern Atlantic and the Gulf of Mexico based on single nucleotide polymorphisms. *Scientific Reports*, *8*(1), 1-12.

Perlwitz, J., (2011). Atmospheric science: Tug of war on the jet stream. *Nature Climate Change*, 29–31

Perrault-Hébert M., Girard F. Boucher Y. Fournier R., Mansuy N., Valeria O. (2019). Evaluation of spatiotemporal transferability of wildfire probability across eastern boreal forest of North America. Chapter 1, PhD dissertation. University Montréal.

Pertierra, L.R., Aragón, P., Shaw, J.D., Bergstrom, D.M., Terauds, A. & Olalla-Tárraga, M.Á. (2017). Global thermal niche models of two European grasses show high invasion risks in Antarctica. *Global Change Biology*, 23(7), 2863-2873.

Pertierra, L.R., Bartlett, J.C., Duffy, G.A., Vega, G.C., Hughes, K.A., Hayward, S.A., ... & Aragón, P. (2019). Combining correlative and mechanistic niche models with human activity data to elucidate the invasive potential of a sub‐Antarctic insect. *Journal of Biogeography*, *47*(3), 658-673.

Peters, D.P., Pielke, R.A., Bestelmeyer, B.T., Allen, C.D., Munson-McGee, S. & Havstad, K.M. (2004). Cross-scale interactions, nonlinearities, and forecasting catastrophic events. *Proceedings of the National Academy of Sciences*, *101*(42), 15130-15135.

Peterson, A.T. (2003). Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology*, 78(4), 419–433.

Peterson, A.T. (2006). Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, 3, 59-72.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M. B. (2011). Ecological niches and geographic distributions. Princeton University Press.

Peterson, A.T. & Soberón, J. (2012). Species distribution modeling and ecological niche modeling: getting the concepts right. *Natureza & Conservação*, *10*(2), 102-107.

Petter, G., Weitere, M., Richter, O. & Moenickes, S. (2014). Consequences of altered temperature and food conditions for individuals and populations: a dynamic energy budget analysis for *Corbicula fluminea* in the Rhine. *Freshwater biology*, *59*(4), 832-846.

Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C., & Guisan, A. (2017). Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Global Ecology and Biogeography*, 26(3), 275-287.

Petrou, K., Kranz, S.A., Trimborn, S., Hassler, C.S., Ameijeiras, S.B., Sackett, O., ... & Davidson, A.T. (2016). Southern Ocean phytoplankton physiology in a changing climate. *Journal of Plant Physiology*, *203*, 135-150.

Pfuhl, H.A. & McCave, N.I. (2005). Evidence for late Oligocene establishment of the Antarctic Circumpolar Current. *Earth and Planetary Science Letters, 235*, 715-728.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3-4), 231-259.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009). Sample selection bias and presence‐only distribution models: implications for background and pseudo‐absence data. *Ecological Applications*, 19(1), 181-197.

Phillips, N.D., Reid, N., Thys, T., Harrod, C., Payne, N.L., Morgan, C.A. ... & Houghton, J.D. (2017). Applying species distribution modelling to a data poor, pelagic fish complex: The ocean sunfishes. *Journal of Biogeography*, 44(10), 2176-2187.

Picken, G.B. (1980). The distribution, growth, and reproduction of the Antarctic limpet *Nacella* (*Patinigera*) *concinna* (Strebel, 1908). *Journal of Experimental Marine Biology and Ecology*, 42(1), 71-85.

Picken, G.B. & Allan, D. (1983). Unique spawning behaviour by the Antarctic limpet *Nacella* (*Patinigera*) *concinna* (Strebel, 1908). *Journal of Experimental Marine Biology and Ecology*, 71(3), 283-287.

Piepenburg, D., Buschmann, A., Driemel, A., Grobe, H., Gutt, J., Schumacher, S., ... & Sieger, R. (2017). Seabed images from Southern Ocean shelf regions off the northern Antarctic Peninsula and in the southeastern Weddell Sea. *Earth System Science Data*, *9*(2), 461-469.

Pierce, D. (2019). R package "ncdf4." https://CRAN.R-project.org/package=ncdf4

Pierrat, B. (2011). *Macroécologie des échinides de l'océan Austral: Distribution, Biogéographie et Modélisation* (Doctoral dissertation).

Pierrat, B., Saucède, T., Laffont, R., De Ridder, C., Festeau, A. & David, B. (2012). Large-scale distribution analysis of Antarctic echinoids using ecological niche modelling. *Marine Ecology Progress Series*, 463, 215-230.

Pinkernell, S. & Beszteri, B. (2014). Potential effects of climate change on the distribution range of the main silicate sinker of the Southern Ocean. *Ecology and Evolution*, *4*(16), 3147-3161.

Pinkerton, M.H., Smith, A.N., Raymond, B., Hosie, G.W., Sharp, B., Leathwick, J.R. & Bradford-Grieve, J.M. (2010). Spatial and seasonal distribution of adult *Oithona similis* in the Southern Ocean: predictions using boosted regression trees. *Deep Sea Research Part I: Oceanographic Research Papers*, 57(4), 469-485.

Pinkerton, M.H. & Bradford-Grieve, J.M. (2010). A balanced model of the food web of the Ross Sea, Antarctica. *CCAMLR Science*, *17*, 1-31.

Pinkerton, M.H., Forman, J., Bury, S.J., Brown, J., Horn, P. & O'Driscoll, R.L. (2013). Diet and trophic niche of Antarctic silverfish *Pleuragramma antarcticum* in the Ross Sea, Antarctica. *Journal of Fish Biology*, *82*(1), 141-164.

Pinkerton, M.H. & Bradford-Grieve, J.M. (2014). Characterizing foodweb structure to identify potential ecosystem effects of fishing in the Ross Sea, Antarctica. *ICES Journal of Marine Science*, *71*(7), 1542-1553.

Pinsky, M.L. & Palumbi, S.R. (2014). Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology*, *23*(1), 29-39.

Pinsky, M.L., Saenz-Agudelo, P., Salles, O.C., Almany, G.R., Bode, M., Berumen, M.L., ... & Planes, S. (2017). Marine dispersal scales are congruent over evolutionary and ecological time. *Current Biology*, *27*(1), 149-154.

Piñones, A., Hofmann, E.E., Dinniman, M.S. & Klinck, J.M. (2011). Lagrangian simulation of transport pathways and residence times along the western Antarctic Peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography*, *58*(13-16), 1524-1539.

Piñones, A., Hofmann, E.E., Daly, K.L., Dinniman, M.S. & Klinck, J.M. (2013). Modeling the remote and local connectivity of Antarctic krill populations along the western Antarctic Peninsula. *Marine Ecology Progress Series*, *481*, 69-92.

Pittman, S.J. (Ed.). (2017). *Seascape Ecology*. John Wiley & Sons. 501pp

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., ... & McCarthy, M.A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, *5*(5), 397-406.

Pope, A., Wagner, P., Johnson, R., Shutler, J.D., Baeseman, J. & Newman, L. (2017). Community review of Southern Ocean satellite data needs. *Antarctic Science*, *29*(2), 97-138.

Porfirio, L.L., Harris, R.M., Lefroy, E.C., Hugh, S., Gould, S.F., Lee, G. ... & Mackey, B. (2014). Improving the use of species distribution models in conservation planning and management under climate change. *PLoS One*, *9*(11), e113749.

Pörtner, H.O., Peck, L. & Somero, G. (2007). Thermal limits and adaptation in marine Antarctic ectotherms: an integrative view. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1488), 2233-2258.

Pörtner, H.O. & Knust, R. (2007). Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science*, *315*(5808), 95-97.

Pörtner, H.O. & Farrell, A.P. (2008). Physiology and climate change. *Science*, 690-692.

Post, A., Meijers, A.J., Fraser, A.D., et al. (2014) Environmental Setting. In: Biogeographic Atlas of the Southern Ocean, Scientific Committee on Antarctic Research. Cambridge, SCAR, pp 46–64.

Potocka, M. & Krzemińska, E. (2018). *Trichocera maculipennis* (Diptera) - an invasive species in Maritime Antarctica. *PeerJ*, *6*, e5408.

Poulin, E. & Féral, J.P. (1994). The fiction and the facts of Antarctic brood protecting: population

genetics and evolution of schizasterid echinoids, in: David, B., Guille, A., Féral, J.P., Roux, M. (Eds), Echinoderms through Time, A.A.Balkema Publishers, Rotterdam, pp. 837-843.

Poulin, E. & Féral, J.P. (1995). Pattern of spatial distribution of a brood-protecting schizasterid echinoid, *Abatus cordatus*, endemic to the Kerguelen Islands. *Marine Ecology Progress Series*, 118, 179-186.

Poulin, E. (1996). Signification adaptative et conséquences évolutives de l'incubation chez un invertébré marin benthique subantarctique, *Abatus cordatus* (Verrill, 1876) (Echinodermata: Spatangoida). Ph. D. Dissertation, Université Montpellier II, Montpellier.

Poulin, E. & Féral, J.P. (1998) Genetic structure of the brooding sea urchin *Abatus cordatus*, an endemic of the subantarctic Kerguelen Island. Echinoderms: San Francisco, Mooi & Telford (Eds) © 1998 Balkema, Rotterdam, ISBN 90 5410 929 7.

Poulin, E., Palma, A.T. & Féral, J-P. (2002). Evolutionary versus ecological success in Antarctic benthic invertebrates. *Trends in Ecology & Evolution*, *17*(5), 218-222.

Poulin, E., González-Wevar, C., Díaz, A., Gérard, K. & Hüne, M. (2014). Divergence between Antarctic and South American marine invertebrates: What molecular biology tells us about Scotia Arc geodynamics and the intensification of the Antarctic Circumpolar Current. *Global and Planetary Change*, *123*, 392-399.

Poulsen, A.H., Kawaguchi, S., King, C.K., King, R.A. & Nash, S.M. (2012). Behavioural sensitivity of a key Southern Ocean species (Antarctic krill, *Euphausia superba*) to p, p′-DDE exposure. *Ecotoxicology and Environmental Safety*, *75*, 163-170.

Pousse, E., Flye-Sainte-Marie, J., Alunno-Bruscia, M., Hegaret, H., Rannou, E., Pecquerie, L., ... & Jean, F. (2019). Modelling paralytic shellfish toxins (PST) accumulation in *Crassostrea gigas* by using Dynamic Energy Budgets (DEB). *Journal of Sea Research*, *143*, 152-164.

Pouvreau, S., Bourles, Y., Lefebvre, S., Gangnery, A. & Alunno-Bruscia, M. (2006). Application of a dynamic energy budget model to the Pacific oyster, *Crassostrea gigas*, reared under various environmental conditions. *Journal of Sea Research*, 56(2), 156-167.

Powell, A.W. (1951). Antarctic and subantarctic mollusca: Pelecypoda and Gastropoda. *Discov Rep* (USA) 26:49-196

Prado, L. & Castilla, J.C. (2006). The bioengineer *Perumytilus purpuratus* (Mollusca: Bivalvia) in central Chile: Biodiversity, habitat structural complexity and environmental heterogeneity. *Journal of the Marine Biological Association of the United Kingdom*, 86(2), 417–421.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Provoost. P., Bosch, S. & Appeltans, W. (2019). robis: R Client to access data from the OBIS API. https://CRAN.R-project.org/package=robis

Pulliam, H.R. (2000). On the relationship between niche and distribution. *Ecology Letters*, *3*(4), 349-361.

Q

Qiao, H., Soberón, J. & Peterson, A.T. (2015). No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution*, 6(10), 1126-1136.

R

Radosavljevic, A. & Anderson, R.P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, 41(4), 629-643.

Raes, N. & ter Steege, H. (2007). A null-model for significance testing of presence-only species distribution models. *Ecography*, *30*(5), 727-736.

Raes, N. (2012). Partial versus full species distribution models. *Natureza & Conservação*, *10*(2), 127-138.

Railsback, S.F. & Grimm, V. (2019). Agent-based and Individual-based modelling: A Practical Introduction. Princeton university press.

Ranc, N., Santini, L., Rondinini, C., Boitani, L., Poitevin, F., Angerbjörn, A. & Maiorano, L. (2017). Performance tradeoffs in target-group bias correction for species distribution models. *Ecography*, *40*(9), 1076-1087.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography*, 33(10), 1689-1703.

Raxworthy, C.J., Ingram, C.M. & Pearson, R.G. (2007). Species delimitation applications for

ecological niche modeling: a review and empirical evaluation using Phelsuma day gecko groups from Madagascar. *Systematic Biology*, *56*, 907-923.

Raymond, B. (2012) Polar Environmental Data Layers, Ver. 3, *Australian Antarctic Data Centre* - https://data.aad.gov.au/metadata/records/Polar_Environmental_Data

Raymond, B., Lea, M.A., Patterson, T., Andrews-Goff, V., Sharples, R., Charrassin, J.B., ... & Goldsworthy, S.D. (2015). Important marine habitat off east Antarctica revealed by two decades of multi-species predator tracking. *Ecography*, *38*(2), 121-129.

Razgour, O., Forester, B., Taggart, J.B., Bekaert, M., Juste, J., Ibáñez, C., ... & Manel, S. (2019). Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proceedings of the National Academy of Sciences*, *116*(21), 10418-10423.

Ready, J., Kaschner, K., South, A.B., Eastwood, P.D., Rees, T., Rius, J. ... & Froese, R. (2010). Predicting the distributions of marine organisms at the global scale. *Ecological Modelling*, *221*(3), 467-478.

Real, R., Barbosa, A.M. & Bull, J.W. (2016). Species distributions, quantum theory, and the enhancement of biodiversity measures. *Systematic Biology*, *66*(3), 453-462.

Reed, A.J. & Thatje, S. (2015). Long-term acclimation and potential scope for thermal resilience in Southern Ocean bivalves. *Marine Biology*, *162*(11), 2217-2224.

Reid, G.K., Lefebvre, S., Filgueira, R., Robinson, S.M., Broch, O.J., Dumas, A. & Chopin, T.B. (2020). Performance measures and models for open-water integrated multi-trophic aquaculture. *Reviews in Aquaculture*, *12*(1), 47-75.

Reiss, H., Cunze, S., König, K., Neumann, H. & Kröncke, I. (2011). Species distribution modelling of marine benthos: a North Sea case study. *Marine Ecology Progress Series*, 442, 71-86.

Reiss, H., Birchenough, S., Borja, A., Buhl-Mortensen, L., Craeymeersch, J., Dannheim, J., ... & Populus, J. (2014). Benthos distribution modelling and its relevance for marine ecosystem management. *ICES Journal of Marine Science*, 72(2), 297-315.

Ren, J.S., Ross, A.H., Hadfield, M.G. & Hayden, B.J. (2010). An ecosystem model for estimating potential shellfish culture production in sheltered coastal waters. *Ecological Modelling*, 221(3), 527-539.

Ren, J.S., Stenton-Dozey, J., Plew, D. R., Fang, J. & Gall, M. (2012). An ecosystem model for optimising production in integrated multitrophic aquaculture systems. *Ecological Modelling*, 246, 34-46.

Ren, J.S., Ragg, N.L., Cummings, V.J., & Zhang, J. (2020). Ocean acidification and dynamic energy budget models: Parameterisation and simulations for the green-lipped mussel. *Ecological Modelling*, *426*, 109069.

Renner, A.H., Heywood, K.J. & Thorpe, S.E. (2009). Validation of three global ocean models in the Weddell Sea. *Ocean Modelling*, *30*(1), 1-15.

Renner, A.H., Thorpe, S.E., Heywood, K.J., Murphy, E.J., Watkins, J.L. & Meredith, M.P. (2012). Advective pathways near the tip of the Antarctic Peninsula: Trends, variability and ecosystem implications. *Deep Sea Research Part I: Oceanographic Research Papers*, *63*, 91-101.

Richardson, D.M., Pyšek, P., Rejmánek, M., Barbour, M.G., Panetta, F.D. & West, C.J. (2000). Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions*, *6*(2), 93-107.

Richardson, D.M. & Whittaker, R.J. (2010). Conservation biogeography - foundations, concepts and challenges. *Diversity and Distributions*, 16(3), 313–320.

Richer de Forges, B. (1977). Étude du crabe des îles Kerguelen: *Halicarcinus planatus* (Fabricius). *Comité National Français des Recherches Antarctiques*, *42*, 71-133.

Rico-Villa, B., Bernard, I., Robert, R. & Pouvreau, S. (2010). A Dynamic Energy Budget (DEB) growth model for Pacific oyster larvae, *Crassostrea gigas*. *Aquaculture*, 305(1-4), 84-94.

Ridgeway, G. (2006) Generalized boosted regression models. Documentation on the R Package 'gbm', version 1·5–7. http://www.i-pensieri.com/gregr/gbm.shtml

Ridgeway, G. (2015) gbm: generalized boosted regression models. R package version 2.1.1. https://CRAN.R- project.org/ package=gbm

Rintoul, S., Hughes, C. & Olbers, D. (2001). The Antarctic circumpolar current system. In *In: Ocean Circulation and Climate/G. Siedler, J. Church and J. Gould, eds. New York: Academic Press. p.* (pp. 271-302).

Rintoul, S.R. (2009). Antarctic Circumpolar Current S.R. In *Encyclopedia of Ocean Sciences* (2nd Edition, pp. 178–190).

Ripley, B. (2015). *MASS*: Support Functions and Datasets for Venables and Ripley's MASS. 2015. https://CRAN.R-project.org/package=MASS

Rivière, P., Jaud, T., Siegelman, L., Klein, P., Cotté, C., Le Sommer, J., ... & Guinet, C. (2019). Sub-mesoscale fronts modify elephant seals foraging behavior. *Limnology and Oceanography Letters*, *4*(6), 193-204.

Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G. ... & Warton, D.I. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913-929.

Robertson, R., El-Haj, A., Clarke, A., Peck, L.S. & Taylor, E. (2001). The effects of temperature on metabolic rate and protein synthesis following a meal in the isopod *Glyptonotus antarcticus* Eights (1852). *Polar Biology*, 24(9), 677-686.

Robinson, A.R. & Golnaraghi, M. (1994). The physical and dynamical oceanography of the Mediterranean Sea. In *Ocean processes in climate dynamics: Global and Mediterranean examples* (pp. 255-306). Springer, Dordrecht.

Robinson, T.P., van Klinken, R.D. & Metternicht, G. (2010). Comparison of alternative strategies for invasive species distribution modeling. *Ecological Modelling*, *221*(19), 2261-2269.

Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P. & Richardson, A.J. (2011). Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, *20*(6), 789-802.

Robinson, N.M., Nelson, W.A., Costello, M.J., Sutherland, J.E. & Lundquist, C.J. (2017). A systematic review of marine-based Species Distribution Models (SDMs) with recommendations for best practice. *Frontiers in Marine Science*, *4*, 421.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta, C. ... & Chiarucci, A. (2011). Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, 35(2), 211-226.

Rochette, N.C. & Catchen, J.M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, *12*(12), 2640-2659.

Rochette, N.C., Rivera-Colón, A.G. & Catchen, J.M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737-4754.

Rodda, G.H., Jarnevich, C.S. & Reed, R.N. (2011). Challenges in identifying sites climatically matched to the native ranges of animal invaders. *PloS One*, *6*(2), e14670.

Rodríguez, L., García, J.J., Carreño, F. & Martínez, B. (2019). Integration of physiological knowledge into hybrid species distribution modelling to improve forecast of distributional shifts of tropical corals. *Diversity and Distributions*, *25*(5), 715-728.

Rogers, A.D. (2007). Evolution and biodiversity of Antarctic organisms: a molecular perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1488), 2191-2214.

Rogers, A.D., Johnston, N.M., Murphy, E.J., Clarke, A. (2012) Antarctic ecosystems: an extreme environment in a changing world. John Wiley & Sons, Oxford.

Roos, N.C., Carvalho, A.R., Lopes, P.F. & Pennino, M.G. (2015). Modeling sensitive parrotfish (Labridae: Scarini) habitats along the Brazilian coast. *Marine Environmental Research*, *110*, 92-100.

Ropert-Coudert, Y., Chiaradia, A., Ainley, D., Barbosa, A., Boersma, P.D., Brasso, R., ... & Trathan, P.N. (2019). Happy feet in a hostile world? The future of penguins depends on proactive management of current and expected threats. *Frontiers in Marine Science*, *6*, 248.

Ropert-Coudert, Y., Van de Putte, A. P., Reisinger, R. R., Bornemann, H., Charrassin, J. B., Costa, D. P., ... & Hindell, M. A. (2020). The retrospective analysis of Antarctic tracking data project. *Scientific Data*, *7*(1), 1-11.

Rosenzweig, M.L. (1995). *Species diversity in space and time*. Cambridge University Press.

Ross, R.E., & Howell, K.L. (2013). Use of predictive habitat modelling to assess the distribution and extent of the current protection of 'listed'deep-sea habitats. *Diversity and Distributions*, 19(4), 433-445.

Ryan, K.G., Ralph, P. & McMinn, A. (2004). Acclimation of Antarctic bottom-ice algal communities to lowered salinities during melting. *Polar Biology*, *27*(11), 679-686.

S

Saba, G.K., Fraser, W.R., Saba, V.S., Iannuzzi, R.A., Coleman, K.E., Doney, S.C., ... & Stammerjohn, S.E. (2014). Winter and spring controls on the summer food web of the coastal West Antarctic Peninsula. *Nature Communications*, *5*, 4318.

Sagarin, R. & Pauchard, A. (2012). *Observation and ecology: broadening the scope of science to understand a complex world*. Island Press.

Salama, N.K. & Rabe, B. (2013). Developing models for investigating the environmental transmission of disease-causing agents within open-cage salmon aquaculture. *Aquaculture Environment Interactions*, *4*(2), 91-115.

Salisbury, E.J. (1926). The geographical distribution of plants in relation to climatic factors. *The Geographical Journal*, *67*(4), 312-335.

Sallée, J.B. (2018). Southern ocean warming. *Oceanography*, *31*(2), 52-62.

Sanches, P.F., Pellizzari, F. & Horta, P.A. (2016). Multivariate analyses of Antarctic and sub-Antarctic seaweed distribution patterns: An evaluation of the role of the Antarctic Circumpolar Current. *Journal of Sea Research*, *110*, 29-38.

Sánchez-Fernández, D., Lobo, J.M. & Hernández-Manrique, O.L. (2011). Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. *Diversity and Distributions*, 17(1), 163-171.

Santika, T. (2011). Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. *Global Ecology and Biogeography*, 20(1), 181-192.

Sa Pinto, A., Branco, M., Sayanda, D. & Alexandrino, P. (2008). Patterns of colonization, evolution and gene flow in species of the genus *Patella* in the Macaronesian Islands. *Molecular Ecology*, 17(2), 519-532.

Sarà, G., Kearney, M. & Helmuth, B. (2011). Combining heat-transfer and energy budget models to predict thermal stress in Mediterranean intertidal mussels. *Chemistry and Ecology*, 27(2), 135-145.

Sarà, G., Palmeri, V., Rinaldi, A., Montalto, V. & Helmuth, B. (2013). Predicting biological invasions in marine habitats through eco-physiological mechanistic models: a case study with the bivalve *Brachidontes pharaonis*. *Diversity and Distributions*, 19(10), 1235-1247.

Saraiva, S., Van der Meer, J., Kooijman, S.A., Witbaard, R., Philippart, C.J., Hippler, D. & Parker, R. (2012). Validation of a Dynamic Energy Budget (DEB) model for the blue mussel *Mytilus edulis*. *Marine Ecology Progress Series*, 463, 141-158.

Saraiva, S., Van der Meer, J., Kooijman, S.A. & Ruardij, P. (2014). Bivalves: from individual to population modelling. *Journal of Sea Research*, 94, 71-83.

Saucède, T., Pierrat. B., David. B., In: De Broyer, C., Koubbi, P., Griffiths, H.J., Raymond, B., Udekem d'Acoz, C. d', Van de Putte, A.P., Danis, B., David, B., Grant, S., Gutt, J., Held, C., Hosie, G., Huettmann, F., Post, A., Ropert-Coudert, Y. (eds.): Biogeographic Atlas of the Southern Ocean. Chapter 5.26. Echinoids. 510 pp., (2014). Cambridge, SCAR. ISBN 978-0-948277-28-3

Saucède, T., Guillaumot, C., Michel, L., Fabri-Ruiz, S., Bazin, A., Cabessut, M., ... & Féral, J-P. (2017). Modelling species response to climate change in sub-Antarctic islands: echinoids as a case study for the Kerguelen Plateau. In *The Kerguelen Plateau: Marine Ecosystems and Fisheries* (pp. 95-116).

Saucède, T. (2020). Proteker 8 cruise report. Summer campaign 2019-2020. 7th Nov. 2019 - 3rd Jan. 2020. 39 pp. [in French]. http://www.proteker.net/Campagne-d-ete-2019.html?lang=en

Saupe, E.E., Barve, V., Myers, C.E., Soberón, J., Barve, N., Hensz, C.M., ... & Lira-Noriega, A. (2012). Variation in niche and distribution model performance: the need for a priori assessment of key causal factors. *Ecological Modelling*, *237*, 11-22.

Sax, D.F., Stachowicz, J.J. & Gaines, S.D. (2005). *Species invasions: insights into ecology, evolution and biogeography*. Sinauer Associates Incorporated.

Scales, K.L., Miller, P.I., Ingram, S.N., Hazen, E.L., Bograd, S.J. & Phillips, R.A. (2016). Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models. *Diversity and Distributions*, *22*(2), 212-224.

Scales, K.L., Hazen, E.L., Jacox, M.G., Castruccio, F., Maxwell, S.M., Lewison, R.L. & Bograd, S.J. (2018). Fisheries bycatch risk to marine megafauna is intensified in Lagrangian coherent structures. *Proceedings of the National Academy of Sciences*, *115*(28), 7362-7367.

Schapire, R.E. (1990). The strength of weak learnability. *Machine Learning, 5*(2): 197–227.

Schatt, P. (1985). Développement et croissance embryonnaire de l'oursin incubant *Abatus cordatus* (Echinoidea: Spatangoida). Ph. D. Dissertation, Université Pierre et Marie Curie, Paris.

Schatt, P. & Féral, J-P. (1991). The brooding cycle of *Abatus cordatus* (Echinodermata: Spatangoida) at Kerguelen Islands. *Polar Biology*, 11(5), 283-292.

Schatt, P. & Féral, J-P. (1996). Completely direct development of *Abatus cordatus*, a brooding schizasterid (Echinodermata: Echinoidea) from Kerguelen, with description of perigastrulation, a hypothetical new mode of gastrulation. *The Biological Bulletin*, 190, 24-44.

Schiaparelli, S., Danis, B., Wadley, V. & Stoddart, D.M. (2013). The Census of Antarctic Marine Life: the first available baseline for Antarctic marine biodiversity. In *Adaptation and Evolution in Marine Environments, Volume 2* (pp. 3-19). Springer, Berlin, Heidelberg.

Schiaparelli, S., Linse, K. (2014) Gastropoda. In: De Broyer C, Koubbi P, Griffiths HJ, et al. (eds) Biogeographic Atlas of the Southern Ocean. Scientific Committee on Antarctic Research, Cambridge UK, pp 122–125.

Schloss, I.R., Abele, D., Moreau, S., Demers, S., Bers, A.V., González, O. & Ferreyra, G.A. (2012). Response of phytoplankton dynamics to 19-year (1991–2009) climate trends in Potter Cove (Antarctica). *Journal of Marine Systems*, 92(1), 53-66.

Schlosser, P., Bullister, J.L. & Bayer, R. (1991). Studies of deep water formation and circulation in the Weddell Sea using natural and anthropogenic tracers. *Marine Chemistry*, *35*(1-4), 97-122.

Schofield, O., Ducklow, H.W., Martinson, D.G., Meredith, M.P., Moline, M.A. & Fraser, W.R. (2010). How do polar marine ecosystems respond to rapid climate change? *Science*, *328*(5985), 1520-1523.

Schofield, O., Saba, G., Coleman, K., Carvalho, F., Couto, N., Ducklow, H. ... & Montes-Hugo, M. (2017). Decadal variability in coastal phytoplankton community composition in a changing West Antarctic Peninsula. *Deep Sea Research Part I: Oceanographic Research Papers*, 124, 42-54.

Schofield, O., Brown, M., Kohut, J., Nardelli, S., Saba, G., Waite, N. & Ducklow, H. (2018). Changes in the upper ocean mixed layer and phytoplankton productivity along the West Antarctic Peninsula. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2122), 20170173.

Schouten, R., Vesk, P.A. & Kearney, M.R. (2020). Integrating dynamic plant growth models and microclimates for species distribution modelling. *Ecological Modelling*, *435*, 109262.

Schram, J.B., Schoenrock, K.M., McClintock, J.B., Amsler, C.D. & Angus, R.A. (2015). Multi-frequency observations of seawater carbonate chemistry on the central coast of the western Antarctic Peninsula. *Polar Research*, *34*(1), 25582.

Schwarzkopf, L., Caley, M.J. & Kearney, M.R. (2016). One lump or two? Explaining a major latitudinal transition in reproductive allocation in a viviparous lizard. *Functional Ecology*, *30*(8), 1373-1383.

Schweiger, O., Settele, J., Kudrna, O., Klotz, S. & Kühn, I. (2008). Climate change can cause spatial mismatch of trophically interacting species. *Ecology*, *89*(12), 3472-3479.

Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., ... & Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, *21*(1), 88-99.

Scott, G.K., Fletcher, G.L. & Davies, P.L. (1986). Fish antifreeze proteins: recent gene evolution. *Canadian Journal of Fisheries and Aquatic Sciences*, *43*(5), 1028-1034.

Segurado, P., & Araujo, M. B. (2004). An evaluation of methods for modelling species distributions. *Journal of Biogeography*, *31*(10), 1555-1568.

Segurado, P.A., Araújo, M.B. & Kunin, W.E. (2006). Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, 43(3), 433-444.

Selkoe, K.A., Henzler, C.M. & Gaines, S.D. (2008). Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, *9*(4), 363-377.

Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., ... & Lexer, M.J. (2017). Forest disturbances under climate change. *Nature Climate Change*, *7*(6), 395.

Seo, C., Thorne, J.H., Hannah, L. & Thuiller, W. (2009). Scale effects in species distribution models: implications for conservation planning under climate change. *Biology Letters*, 5(1), 39-43.

Serpa, D., Ferreira, P.P., Ferreira, H., da Fonseca, L.C., Dinis, M.T. & Duarte, P. (2013). Modelling the growth of white seabream (*Diplodus sargus*) and gilthead seabream (*Sparus aurata*) in semi-intensive earth production ponds using the Dynamic Energy Budget approach. *Journal of Sea Research*, 76, 135-145.

Sexton, M., 2005. The Construction of a Bathymetric Grid for the Heard Island - Kerguelen Plateau

Region. Record 2005/xx, Geoscience Australia, Canberra, Australia, pp. 26.

Shabica, S.V. (1971). General ecology of Antarctic limpet *Patinigera polaris*. *Antarctic Journal of the United States*, 6(5), 160.

Shabica, S.V. (1976) The natural history of the Antarctic limpet *Patinigera polaris* (Hombron and Jacquinot). Ph. D. Thesis, Oregon State University (Corvallis, Oregon, USA) p. 294

Shafer, A.B., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W. & Wolf, J.B. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907-917.

Sharp, B.R.; Parker, S.J.; Pinkerton, M.H. (lead authors); also B.B. Breen, V. Cummings, A. Dunn, S.M. Grant, S.M. Hanchet, H.J.R. Keys, S.J. Lockhart, P.O'B. Lyver, R.L. O'Driscoll, M.J.M. Williams, P.R. Wilson. (2010). Bioregionalisation and Spatial Ecosystem Processes in the Ross Sea Region. CCAMLR Working Group paper. WG-EMM-10/30. For access: http://www.ccamlr.org/en/wg-emm-10/30.

Shcherbich, L.B. (1975) Method of determining age and onset of sexual maturity in the marbled Antarctic cod *Notothenia rossii marmorata*. *Journal of Ichthyology*, 15, 82-87.

Sheen, K.L., Brearley, J.A., Naveira Garabato, A.C., Smeed, D.A., Waterman, S., Ledwell, J.R. ... & Watson, A.J. (2013). Rates and mechanisms of turbulent dissipation and mixing in the Southern Ocean: Results from the Diapycnal and Isopycnal Mixing Experiment in the Southern Ocean (DIMES). *Journal of Geophysical Research: Oceans*, *118*(6), 2774-2792.

Sherborne, N. & Galic, N. (2020). Modeling Sublethal Effects of Chemicals: Application of a Simplified Dynamic Energy Budget Model to Standard Ecotoxicity Data. *Environmental Science & Technology*, *54*(12), 7420-7429.

Shilling, F.M. & Manahan, D.T. (1994). Energy metabolism and amino acid transport during early development of Antarctic and temperate echinoderms. *The Biological Bulletin*, *187*(3), 398-407.

Shin, S.C., Ahn, D.H. Kim, S.J., Pyo, C.W., Lee, H., Kim, M.K., ... & Park, H. (2014). The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome Biology*, *15*(9), 1-14.

Shulman, M.J. (1990). Aggression among sea urchins on Carribean coral reefs. *Journal of Experimental Marine Biology and Ecology*, 140(3), 197-207.

Sillero, N. (2011). What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling*, *222*(8), 1343-1346.

Sillero, N., & Barbosa, A. M. (2020). Common mistakes in ecological niche models. *International Journal of Geographical Information Science*, *35*(2), 213-226.

Silva, B.P., Silva, M.L., Avalos, F.A., de Menezes, M.D. & Curi, N. (2019). Digital soil mapping including additional point sampling in Posses ecosystem services pilot watershed, southeastern Brazil. *Scientific Reports*, 9(1), 1-12.

Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. (2005). ROCR: visualizing classifier performance in R. *Bioinformatics*, *21*(20), 3940-3941.

Singer, A., Johst, K., Banitz, T., Fowler, M.S., Groeneveld, J., Gutiérrez, A.G., ... & Meyer, K.M. (2016). Community dynamics under environmental change: How can next generation mechanistic models improve projections of species distributions? *Ecological Modelling*, *326*, 63-74.

Smale, D.A. & Barnes, D.K. (2008). Likely responses of the Antarctic benthos to climate-related changes in physical disturbance during the 21st century, based primarily on evidence from the West Antarctic Peninsula region. *Ecography*, 31(3), 289-305.

Smeraldo, S., Di Febbraro, M., Bosso, L., Flaquer, C., Guixé, D., Lisón, F., ... & Russo, D. (2018). Ignoring seasonal changes in the ecological niche of non-migratory species may lead to biases in potential distribution models: lessons from bats. *Biodiversity and Conservation*, 27(9), 2425-2441.

Smith, C.R., Mincks, S. & DeMaster, D.J. (2006). A synthesis of bentho-pelagic coupling on the Antarctic shelf: food banks, ecosystem inertia and global climate change. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(8-10), 875-894.

Smith, R.J., Eastwood, P.D., Ota, Y. & Rogers, S.I. (2009). Developing best practice for using Marxan to locate Marine Protected Areas in European waters. *ICES Journal of Marine Science*, *66*(1), 188-194.

Smith, R.I. & Richardson, M. (2011). Fuegian plants in Antarctica: natural or anthropogenically assisted immigrants? *Biological Invasions*, *13*(1), 1-5.

Smith, A.B. (2013). On evaluating species distribution models with random background sites in place of absences when test presences disproportionately

sample suitable habitat. *Diversity and Distributions*, *19*(7), 867-872.

Smith Jr, K.L., Sherman, A.D., Shaw, T.J. & Sprintall, J. (2013). Icebergs as unique Lagrangian ecosystems in polar seas. *Annual Review of Marine Science*, *5*, 269-287.

SNAP (2020). ESA Science Toolbox Exploitation Platform (SNAP). http://step.esa.int/main/download/

Snelgrove, P.V. & Butman, C.A. (1994). Animal-sediment relationships revisited: cause versus effect. *Oceanography and Marine Biology: An Annual Review*, *32*, 111-177

Snickars, M., Gullström, M., Sundblad, G., Bergström, U., Downie, A.L., Lindegarth, M. & Mattila, J. (2014). Species–environment relationships and potential for distribution modelling in coastal waters. *Journal of Sea Research*, *85*, 116-125.

Soberón, J. & Peterson, A.T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*. 2, 1-10.

Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, *10*(12), 1115-1123.

Soberón, J. & Nakamura, M. (2009). Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences*, *106* (Supplement 2), 19644-19650.

Soberón, J.M. (2010). Niche and area of distribution modeling: a population ecology perspective. *Ecography*, *33*(1), 159-167.

Soberón, J. & Arroyo-Peña, B. (2017). Are fundamental niches larger than the realized? Testing a 50-year-old prediction by Hutchinson. *PloS One*, *12*(4), e0175138.

Sousa, T., Domingos, T., Kooijman, S. (2008). From empirical patterns to theory: a formal metabolic theory of life. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1502), 2453-2464.

Sousa-Silva, R., Alves, P., Honrado, J. & Lomba, A. (2014). Improving the assessment and reporting on rare and endangered species through species distribution models. *Global Ecology and Conservation*, *2*, 226-237.

Souster, T.A., Morley, S.A. & Peck, L.S. (2018). Seasonality of oxygen consumption in five common Antarctic benthic marine invertebrates. *Polar Biology*, 41(5), 897-908.

Southwell, C.J., Kerry, K.R. & Ensor, P.H. (2005). Predicting the distribution of crabeater seals *Lobodon carcinophaga* off east Antarctica during the breeding season. *Marine Ecology Progress Series*, *299*, 297-309.

Srivastava, V., Lafond, V. & Griess, V.C. (2019). Species distribution models (SDM): applications, benefits and challenges in invasive species management. *CAB Rev*, *14*(020), 1-13.

Stainthorp, R. & Kooijman, S.A. (2017). AmP *Sterechinus neumayeri*, version 2017/07/06. http://www.bio.vu.nl/thb/deb/deblab/add_my_pet/entries_web/Sterechinus_neumayeri/Sterechinus_neumayeri_res.html

Stammerjohn, S.E., Martinson, D.G., Smith, R.C. & Iannuzzi, R.A. (2008). Sea ice in the western Antarctic Peninsula region: Spatio-temporal variability from ecological and climate change perspectives. *Deep Sea Research Part II: Topical Studies in Oceanography*, *55*(18-19), 2041-2058.

Stammerjohn, S., Massom, R., Rind, D. & Martinson, D. (2012). Regions of rapid sea ice change: An inter-hemispheric seasonal comparison. *Geophysical Research Letters*, 39(6), L06501.

Stanwell-Smith, D. & Clarke, A. (1998). The Timing of reproduction in the Antarctic limpet *Nacella concinna* (Strebel, 1908) (Patellidae) at Signy Island, in relation to environmental variables. *Journal of Molluscan Studies*, 64(1), 123-127.

Stanwell-Smith, D. & Peck, L.S. (1998). Temperature and embryonic development in relation to spawning and field occurrence of larvae of three Antarctic echinoderms. *The Biological Bulletin*, *194*(1), 44-52.

Stanwell-Smith, D., Peck, L.S., Clarke, A., Murray, A.W. & Todd, C.D. (1999). The distribution, abundance and seasonality of pelagic marine invertebrate larvae in the maritime Antarctic. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *354*(1382), 471-484.

Staveley, T.A., Perry, D., Lindborg, R. & Gullström, M. (2017). Seascape structure and complexity influence temperate seagrass fish assemblage composition. *Ecography*, *40*(8), 936-946.

Starfield A.M., Smith, K.A. & Andrew L. (1990). *How to model it: Problem solving for the computer age*. Mc Graw-Hill.

Stavrakidis-Zachou, O., Papandroulakis, N. & Lika, K. (2019). A DEB model for European sea bass (*Dicentrarchus labrax*): Parameterisation and application in aquaculture. *Journal of Sea Research*, *143*, 262-271.

Stebbing, T.R. (1914). IX.—Stalk-eyed Crustacea Malacostraca of the Scottish National Antarctic Expedition. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 50(2), 253-307.

Stenni, B., Curran, M.A., Abram, N., Orsi, A., Goursaud, S., Masson-Delmotte, V. ... & Steig, E.J. (2017). Antarctic climate variability on regional and continental scales over the last 2000 years. *Climate Past*, 13, 1609-1634.

Stevenson, A., Mitchell, F. & Davies, J.S. (2015). Predation has no competition: Factors influencing space and resource use by echinoids in deep-sea coral habitats, as evidenced by continuous video transects. *Marine Ecology*, 36(4), 1454-1467.

Steyerberg, E.W., Bleeker, S.E., Moll, H.A., Grobbee, D.E. & Moons, K.G. (2003). Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*, *56*(5), 441-447.

Støa, B., Halvorsen, R., Mazzoni, S. & Gusarov, V.I. (2018). Sampling bias in presence-only data used for species distribution modelling: theory and methods for detecting sample bias and its effects on models. *Sommerfeltia*, *38*(1), 1-53.

Stock, A., Subramaniam, A., Van Dijken, G.L., Wedding, L.M., Arrigo, K.R., Mills, M.M. ... & Micheli, F. (2020). Comparison of Cloud-Filling Algorithms for Marine Satellite Data. *Remote Sensing*, *12*(20), 3313.

Stockwell, D.R., & Peterson, A.T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148(1), 1-13.

Storey, J.D., Bass, A.J., Dabney, A. & Robinson, D. (2019) qvalue: Q-value estimation for false discovery rate control. https://rdrr.io/bioc/qvalue/

Strauss, T., Gabsi, F., Hammers-Wirtz, M., Thorbek, P. & Preuss, T.G. (2017). The power of hybrid modelling: An example from aquatic ecosystems. *Ecological Modelling*, *364*, 77-88.

Strebel, H. (1908). Dei Gastropoden. Wissenschaftliche Ergebnisse der Schwedischen Südpolar-Expedition, 1901-1903. 6:1-112.

Strobel, A., Bennecke, S., Leo, E., Mintenbeck, K., Pörtner, H.O. & Mark, F.C. (2012). Metabolic shifts in the Antarctic fish *Notothenia rossii* in response to rising temperature and PCO_2. *Frontiers in Zoology*, *9*(1), 1-15.

Strugnell, J.M., Watts, P.C., Smith, P.J. & Allcock, A.L. (2012). Persistent genetic signatures of historic climatic events in an Antarctic octopus. *Molecular Ecology*, *21*(11), 2775-2787.

Suckling, C.C., Clark, M.S., Richard, J., Morley, S.A., Thorne, M.A., Harper, E.M. & Peck, L.S. (2015). Adult acclimation to combined temperature and pH stressors significantly enhances reproductive outcomes compared to short-term exposures. *Journal of Animal Ecology*, *84*(3), 773-784.

Sussarellu, R., Suquet, M., Thomas, Y., Lambert, C., Fabioux, C., Pernet, M.E., ... & Corporeau, C. (2016). Oyster reproduction is affected by exposure to polystyrene microplastics. *Proceedings of the National Academy of Sciences*, *113*(9), 2430-2435.

Sutter, R.D., Wainscott, S.B., Boetsch, J.R., Palmer, C.J. & Rugg, D.J. (2015). Practical guidance for integrating data management into long-term ecological monitoring projects. *Wildlife Society Bulletin*, *39*(3), 451-463.

Suda, C.N., Vani, G.S., de Oliveira, M.F., Rodrigues, E. & Lavrado, H.P. (2015). The biology and ecology of the Antarctic limpet *Nacella concinna*. *Polar Biology*, 38(12), 1949-1969.

Sundqvist, L., Keenan, K., Zackrisson, M., Prodöhl, P. & Kleinhans, D. (2016). Directional genetic differentiation and relative migration. *Ecology and Evolution*, *6*(11), 3461-3475.

Swanson, A.K., Dobrowski, S.Z., Finley, A.O., Thorne, J.H. & Schwartz, M.K. (2013). Spatial regression methods capture prediction uncertainty in species distribution model projections through time. *Global Ecology and Biogeography*, *22*(2), 242-251.

Synes, N.W. & Osborne, P.E. (2011). Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Global Ecology and Biogeography*, 20(6), 904-914.

Syfert, M.M., Smith, M.J. & Coomes, D.A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PloS One*, *8*(2).

Syfert, M.M., Joppa, L., Smith, M.J., Coomes, D.A., Bachman, S.P. & Brummitt, N.A. (2014). Using species distribution models to inform IUCN Red List assessments. *Biological Conservation*, *177*, 174-184.

T

Tagliarolo, M., Montalto, V., Sarà, G., Lathlean, J.A. & McQuaid, C.D. (2016). Low temperature trumps high food availability to determine the distribution of intertidal mussels *Perna perna* in South Africa. *Marine Ecology Progress Series*, *558*, 51-63.

Talluto, M.V., Boulangeat, I., Ameztegui, A., Aubin, I., Berteaux, D., Butler, A. ... & Liénard, J. (2016). Cross-scale integration of knowledge for predicting species ranges: a metamodelling framework. *Global Ecology and Biogeography*, *25*(2), 238-249.

Tavares, M. & De Melo, G.A. (2004). Discovery of the first known benthic invasive species in the Southern Ocean: The North Atlantic spider crab *Hyas araneus* found in the Antarctic Peninsula. *Antarctic Science*, *16*(2), 129–131.

Teal, L.R., van Hal, R., van Kooten, T., Ruardij, P. & Rijnsdorp, A.D. (2012). Bio-energetics underpins the spatial response of North Sea plaice (*Pleuronectes platessa L.*) and sole (*Solea solea L.*) to climate change. *Global Change Biology*, 18(11), 3291-3305.

Telford, R.J. & Birks, H.J. (2009). Evaluation of transfer functions in spatially structured environments. *Quaternary Science Reviews, 28*(13-14), 1309-1316.

Teschke, K., Bester, M.N., Bornemann, H., Brandt, A., Brtnik, P., De Broyer, C., ... & Griffiths, H.J. (2015). Scientific background document in support of the development of a CCAMLR MPA in the Weddell Sea (Antarctica)–Version 2014.

Teschke, K., Pehlke, H., Deininger, M., Jerosch, K. & Brey, T. (2016). Scientific background document in support of the development of a CCAMLR MPA in the Weddell Sea (Antarctica) -Version 2016- Part C: Data analysis and MPA scenario development. *CCAMLR WG-EMM-16*, (WG-EMM).

Teschke, K., Pehlke, H. & Brey, T. (2017). Scientific background document in support of the development of a CCAMLR MPA in the Weddell Sea (Antarctica)–Version 2017–Reflection of the recommendations by WG-EMM-16 and SC-CAMLRXXXV.

Teschke, K., Pehlke, H., Siegel, V., Bornemann, H., Knust, R. & Brey, T. (2019). An integrated compilation of data sources for the development of a marine protected area in the Weddell Sea. *Earth System Science Data*, *12*(2), 1003-1023.

Tessarolo, G., Rangel, T.F., Araújo, M.B. & Hortal, J. (2014). Uncertainty associated with survey design in Species Distribution Models. *Diversity and Distributions*, 20(11), 1258-1269.

Tessarolo, G., Ladle, R., Rangel, T. & Hortal, J. (2017). Temporal degradation of data limits biodiversity research. *Ecology and Evolution*, 7(17), 6863-6870.

Thatje, S. & Fuentes, V. (2003). First record of anomuran and brachyuran larvae (Crustacea: Decapoda) from Antarctic waters. *Polar Biology*, *26*, 279–282.

Thatje, S., Anger, K., Calcagno, J.A., Lovrich, G.A., Pörtner, H.O. & Arntz, W.E. (2005a). Challenging the cold: crabs reconquer the Antarctic. *Ecology*, *86*(3), 619-625.

Thatje, S., Hillenbrand, C. D., & Larter, R. (2005b). On the origin of Antarctic marine benthic community structure. *Trends in Ecology & Evolution*, 20(10), 534–540.

Thatje, S. (2012). Effects of capability for dispersal on the evolution of diversity in Antarctic benthos. *Integrative and Comparative Ecology*, 52(4): 470-482.

Thiers, L., Delord, K., Bost, C.A., Guinet, C. & Weimerskirch, H. (2017). Important marine sectors for the top predator community around Kerguelen Archipelago. *Polar Biology*, *40*(2), 365-378.

Thomalla, S.J., Fauchereau, N., Swart, S. & Monteiro, P.M. (2011). Regional scale characteristics of the seasonal cycle of chlorophyll in the Southern Ocean. *Biogeosciences*, *8*(10), 2849-2866.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., … Williams, S.E. (2004). Extinction risk from climate change. *Nature*, 427(6970), 145–148.

Thomas, Y., Mazurié, J., Alunno-Bruscia, M., Bacher, C., Bouget, J.F., Gohin, F. ... & Struski, C. (2011). Modelling spatio-temporal variability of *Mytilus edulis* (L.) growth by forcing a dynamic energy budget model with satellite-derived environmental data. *Journal of Sea Research*, 66(4), 308-317.

Thomas, C.J., Lambrechts, J., Wolanski, E., Traag, V.A., Blondel, V.D., Deleersnijder, E. & Hanert, E. (2014). Numerical modelling and graph theory tools to study ecological connectivity in the Great Barrier Reef. *Ecological Modelling*, *272*, 160-174.

Thomas, Y., Pouvreau, S., Alunno-Bruscia, M., Barillé, L., Gohin, F., Bryère, P. & Gernez, P. (2015). Global change and climate-driven invasion of the Pacific oyster (*Crassostrea gigas*) along European coasts: a bioenergetics modelling approach. *Journal of Biogeography*, *43*(3), 568-579.

Thomas, Y., Dumas, F. & Andréfouët, S. (2016). Larval connectivity of pearl oyster through biophysical modelling; evidence of food limitation and broodstock effect. *Estuarine, Coastal and Shelf Science*, 182, 283-293.

Thomas, Y. & Bacher, C. (2018). Assessing the sensitivity of bivalve populations to global warming using an individual-based modelling approach. *Global Change Biology*, 24(10), 4581-4597.

Thomas, Y., Razafimahefa, N.R., Ménesguen, A. & Bacher, C. (2020). Multi-scale interaction processes modulate the population response of a benthic species to global warming. *Ecological Modelling*, *436*, 109295.

Thompson, A., Sanders, J., Tandstad, M., Carocci, F. & Fuller, J. (2016). Vulnerable marine ecosystems: Processes and practices in the high seas. *FAO Fisheries and Aquaculture Technical Paper*, (595), I.

Thorpe, S.E., Heywood, K.J., Stevens, D.P. & Brandon, M.A. (2004). Tracking passive drifters in a high resolution ocean model: implications for interannual variability of larval krill transport to South Georgia. *Deep Sea Research Part I: Oceanographic Research Papers*, *51*(7), 909-920.

Thorrold, S.R., Jones, G.P., Hellberg, M.E., Burton, R.S., Swearer, S.E., Neigel, J.E., ... & Warner, R.R. (2002). Quantifying larval retention and connectivity in marine populations with artificial and natural markers. *Bulletin of Marine Science*, *70*(1), 291-308.

Thuiller, W., Vayreda, J., Pino, J., Sabate, S., Lavorel, S. & Gracia, C. (2003). Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology and Biogeography*, *12*(4), 313-325.

Thuiller, W., Brotons, L., Araújo, M.B. & Lavorel, S. (2004). Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, 27(2), 165–172.

Thuiller, W., Richardson, D.M., Pysek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. (2005). Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, *11*(12), 2234-2250.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009). BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369-373.

Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffers, K. & Gravel, D. (2013). A road map for integrating eco-evolutionary processes into biodiversity models. *Ecology Letters*, *16*, 94-105.

Thuiller, W., Georges, D., Engler, R. & Breiner, F. (2016) biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7. https://CRAN.R-project.org/package=biomod2

Thuiller, W., Georges, D. & Engler, R. (2018). Ensemble platform for species distribution modeling package biomod2 version3.3 https://CRAN. Rproject. org/package= biomod2.

Thyrring, J., Bundgaard, A. & Sejr, M.K. (2017). Seasonal acclimation and latitudinal adaptation are of the same magnitude in *Mytilus edulis* and *Mytilus trossulus* mitochondrial respiration. *Polar Biology*, *40*(9), 1885-1891.

Tigano, A. & Friesen, V.L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, *25*(10), 2144-2164.

Tikhonov, G., Opedal, Ø.H., Abrego, N., Lehikoinen, A., de Jonge, M.M., Oksanen, J. & Ovaskainen, O. (2020). Joint species distribution modelling with the r-package *Hmsc*. *Methods in Ecology and Evolution*, *11*(3), 442-447.

Tingley, R., Vallinoto, M., Sequeira, F., & Kearney, M.R. (2014). Realized niche shift during a global biological invasion. *Proceedings of the National Academy of Sciences*, 201405766.

Tirunelveli, G., Gordon, R. & Pistorius, S. (2002). Comparison of square-pixel and hexagonal-pixel resolution in image processing. In *IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Conference Proceedings (Cat. No. 02CH37373)* (Vol. 2, pp. 867-872). IEEE.

Titeux, N., Maes, D., Van Daele, T., Onkelinx, T., Heikkinen, R.K., Romo, H. ... & Schweiger, O. (2017). The need for large-scale distribution data to estimate regional changes in species richness under future climate change. *Diversity and Distributions*, 23(12), 1393-1407.

Torres, L.G., Sutton, P.J., Thompson, D.R., Delord, K., Weimerskirch, H., Sagar, P.M. ... & Phillips, R.A. (2015). Poor transferability of species distribution models for a pelagic predator, the grey petrel, indicates contrasting habitat preferences across ocean basins. *PLoS One*, 10(3), e0120014.

Trathan, P.N., Collins, M.A., Grant, S.M., Belchier, M., Barnes, D.K., Brown, J. & Staniland, I.J. (2014). The South Georgia and the South Sandwich Islands MPA: protecting a biodiverse oceanic island chain situated in the flow of the Antarctic Circumpolar Current. *Advances in Marine Biology*, *69*, 15-78.

Trathan, P.N. & Grant, S.M. (2020). The South Orkney Islands Southern Shelf Marine Protected Area: Towards the establishment of marine spatial protection within international waters in the Southern

Ocean. In *Marine Protected Areas* (pp. 67-98). Elsevier.

Treml, E.A., Halpin, P.N., Urban, D.L. & Pratson, L.F. (2008). Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecology*, *23*(1), 19-36.

Trolle, D., Elliott, J.A., Mooij, W.M., Janse, J.H., Bolding, K., Hamilton, D.P. & Jeppesen, E. (2014). Advancing projections of phytoplankton responses to climate change through ensemble modelling. *Environmental Modelling & Software*, *61*, 371-379.

Troost, T.A., Wijsman, J.W., Saraiva, S. & Freitas, V. (2010). Modelling shellfish growth with dynamic energy budget models: an application for cockles and mussels in the Oosterschelde (southwest Netherlands). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1557), 3567-3577.

Tropsha, A., Gramatica, P. & Gombar, V.K. (2003). The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, *22*(1), 69-77.

Trull, T.W., Passmore, A., Davies, D.M., Smit, T., Berry, K. & Tilbrook, B. (2018). Distribution of planktonic biogenic carbonate organisms in the Southern Ocean south of Australia: a baseline for ocean acidification impact assessment. *Biogeosciences*, *15*(1), 31.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. (2007). A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, *13*(4), 397-405.

Turner, J.T. (2002). Zooplankton fecal pellets, marine snow and sinking phytoplankton blooms. *Aquatic Microbial Ecology*, *27*(1), 57-102.

Turner, J., Bindschadler, R., Convey, P., et al. (eds) (2009) Antarctic climate change and the environment: [a contribution to the International Polar Year 2007-2008]. Scientific Committee on Antarctic Research, Cambridge.

Turner, J., Barrand, N.E., Bracegirdle, T.J., Convey, P., Hodgson, D.A., Jarvis, M., ... & Shanklin, J. (2014). Antarctic climate change and the environment: an update. *Polar Record*, 50(3), 237-259.

Turner, J., Lu, H., White, I., King, J.C., Phillips, T., Hosking, J.S. ... & Deb, P. (2016). Absence of 21st century warming on Antarctic Peninsula consistent with natural variability. *Nature*, 535(7612), 411.

Tyler, P.A., Young, C.M. & Clarke, A. (2000). Temperature and pressure tolerances of embryos and larvae of the Antarctic sea urchin *Sterechinus neumayeri* (Echinodermata: Echinoidea): potential for deep-sea invasion from high latitudes. *Marine Ecology Progress Series*, *192*, 173-180.

U

USGS. Department of the Interior. Landsat 8 Data User Handbook. LSDS-1574 version 5.0. 2019.

Uri, S., Rogers, W.P. & Kirkham, R.M. (1992). Modelling the bathymetry of the Antarctic continental shelf. In *Sixth International Symposium on Antarctic Earth Sciences* (pp. 763-771).

Uthicke, S., Liddy, M., Nguyen, H.D. & Byrne, M. (2014). Coral Reefs, 33(3), 831-845.

V

Václavík, T., & Meentemeyer, R.K. (2009). Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling*, 220(23), 3248-3258.

Václavík, T. & Meentemeyer, R.K. (2012). Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions*, *18*(1), 73-83.

Valavanis, V.D., Pierce, G.J., Zuur, A.F., Palialexis, A., Saveliev, A., Katara, I., & Wang, J. (2008). Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. In *Essential Fish Habitat Mapping in the Mediterranean* (pp. 5-20). Springer, Dordrecht.

Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2018). blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *bioRxiv*, 357798.

Vale, C.G., Tarroso, P. & Brito, J.C. (2014). Predicting species distribution at range margins: testing the effects of study area extent, resolution and threshold selection in the Sahara–Sahel transition zone. *Diversity and Distributions*, *20*(1), 20-33.

Vanden Berghe, E. (2013). Report of the project 'Turning OBIS data into information'; project funded by the Census of Marine Life International Cosmos Prize Fund. Rutgers University. 62 pp.

Van de Putte, A.P., Jackson, G.D., Pakhomov, E., Flores, H. & Volckaert, F.A. (2010). Distribution of

squid and fish in the pelagic zone of the Cosmonaut Sea and Prydz Bay region during the BROKE-West campaign. *Deep Sea Research Part II: Topical Studies in Oceanography*, *57*(9-10), 956-967.

Van de Putte, A.P., Janko, K., Kasparova, E., Maes, G.E., Rock, J., Koubbi, P., ... & Marshall, C. (2012). Comparative phylogeography of three trematomid fishes reveals contrasting genetic structure patterns in benthic and pelagic species. *Marine Genomics*, *8*, 23-34.

van der Meer, J. (2006). An introduction to Dynamic Energy Budget (DEB) models with special emphasis on parameter estimation. *Journal of Sea Research*, 56(2), 85-102.

van der Meer, J. & Kooijman, S.A. (2014). Inference on energetics of deep-sea fish that cannot be aged: The case of the hagfish. *Journal of Sea Research*, *94*, 138-143.

van der Meer, J., Klok, C., Kearney, M.R., Wijsman, J.W. & Kooijman, S.A. (2014). 35 years of DEB research. *Journal of Sea Research*. 94(1-4).

Van der Putten, W.H., Macel, M., & Visser, M.E. (2010). Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1549), 2025-2034.

Vandersteen, W. (2011). Detecting gene expression profiles associated with environmental stressors within an ecological context. *Molecular Ecology*, *20*(7), 1322-1323.

van der Veer, H.W., Cardoso, J.F. & van der Meer, J. (2006). The estimation of DEB parameters for various Northeast Atlantic bivalve species. *Journal of Sea Research*, 56(2), 107-124.

van de Wolfshaar, K.E., Barbut, L. & Lacroix, G. (2021). From spawning to first-year recruitment: the fate of juvenile sole growth and survival under future climate conditions in the North Sea. *ICES Journal of Marine Science*.

Van Dongen, S. (2006). Prior specification in Bayesian statistics: three cautionary tales. *Journal of Theoretical Biology*, *242*(1), 90-100.

van Proosdij, A.S., Sosef, M.S., Wieringa, J.J. & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, 39(6), 542-552.

Van Sebille, E., Griffies, S.M., Abernathey, R., Adams, T.P., Berloff, P., Biastoch, A., ... & Deleersnijder, E. (2018). Lagrangian ocean analysis: Fundamentals and practices. *Ocean Modelling*, *121*, 49-75.

Vapnik, V. (1998). Statistical learning theory Wiley. New York, 1, 624.

Vargas-Chacoff, L., Astola, A., Arjona, F.J., Martín del Río, M.P., García-Cózar, F., Mancera, J.M. & Martínez-Rodríguez, G. (2009). Pituitary gene and protein expression under experimental variation on salinity and temperature in gilthead sea bream *Sparus aurata*. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 154(3), 303–308.

Varisco, M., Colombo, J., Isola, T. & Vinuesa, J. (2016). Growth and maturity of the spider crab *Halicarcinus planatus* (Brachyura: Hymenosomatidae) females in the southwestern Atlantic Ocean. Can these parameters be influenced by the population sex ratio? *Marine Biology Research*, *12*(6), 647-655.

Vaughan, I.P. & Ormerod, S.J. (2003). Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conservation Biology*, *17*(6), 1601-1611.

Vaughan, D.G., Marshall, G.J., Connolley, W.M., Parkinson, C., Mulvaney, R., Hodgson, D.A. ... & Turner, J. (2003). Recent rapid regional climate warming on the Antarctic Peninsula. *Climatic change*, 60(3), 243-274.

Veloz, S.D. (2009). Spatially auto-correlated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36(12), 2290-2299.

Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S*. Springer, New York.

Venables, H., Meredith, M.P., Atkinson, A. & Ward, P. (2012). Fronts and habitat zones in the Scotia Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, *59*, 14-24.

Venables, H.J., Clarke, A. & Meredith, M.P. (2013). Wintertime controls on summer stratification and productivity at the western Antarctic Peninsula. *Limnology and Oceanography*, *58*(3), 1035-1047.

Venette, R.C., Kriticos, D.J., Magarey, R.D., Koch, F.H., Baker, R.H., Worner, S.P., ... & Pedlar, J. (2010). Pest risk maps for invasive alien species: a roadmap for improvement. *BioScience*, 60(5), 349-362.

Vermeij, G.J. (1973). Morphological patterns in high-intertidal gastropods: adaptive strategies and their limitations. *Marine Biology*, 20(4), 319-346.

Vernet, M., Martinson, D., Iannuzzi, R., Stammerjohn, S., Kozlowski, W., Sines, K., ... & Garibotti, I. (2008). Primary production within the sea-ice zone west of the Antarctic Peninsula: I—Sea ice, summer mixed layer, and irradiance. *Deep Sea Research Part II: Topical Studies in Oceanography*, *55*(18-19), 2068-2085.

Vernet, M., Geibert, W., Hoppema, M., Brown, P.J., Haas, C., Hellmer, H.H., ... & Brearley, J.A. (2019). The Weddell Gyre, Southern Ocean: present knowledge and future challenges. *Reviews of Geophysics*, *57*(3), 623-708.

Villari, C., Herms, D.A., Whitehill, J.G., Cipollini, D. & Bonello, P. (2016). Progress and gaps in understanding mechanisms of ash tree resistance to emerald ash borer, a model for wood-boring insects that kill angiosperms. *New Phytologist*, *209*(1), 63-79.

Vincenot, C.E., Giannino, F., Rietkerk, M., Moriya, K. & Mazzoleni, S. (2011). Theoretical considerations on the combined use of System Dynamics and individual-based modeling in ecology. *Ecological Modelling*, 222(1), 210-218.

Vincenot, C.E. & Moriya, K. (2011). Impact of the topology of metapopulations on the resurgence of epidemics rendered by a new multiscale hybrid modeling approach. *Ecological Informatics*, 6(3-4), 177-186.

Vincenot, C.E., Mazzoleni, S. & Parrott, L. (Eds.). (2017). Hybrid solutions for the modelling of complex environmental systems. Frontiers Media SA.

Vinuesa, J.H. & Ferrari, L. (2008). Reproduction of *Halicarcinus planatus* (crustacea, decapoda, hymenosomatidae) in the Deseado River estuary, southwestern Atlantic Ocean. *Marine Biology*, *154*(2), 345-351.

Vlaeminck, K., Viaene, K.P.J., Van Sprang, P., Baken, S., De Schamphelaere, K.A.C. (2019). The Use of Mechanistic Population Models in Metal Risk Assessment: Combined Effects of Copper and Food Source on *Lymnaea stagnalis* Populations. *Environmental Toxicology and Chemistry*, 38, 1104-1119.

Volonterio, O., de León, R.P., Convey, P. & Krzemińska, E. (2013). First record of Trichoceridae (Diptera) in the maritime Antarctic. *Polar Biology*, *36*(8), 1125-1131.

von Humboldt, A. (1807). Voyage de Humboldt et Bonpland (Première partie. Physique Générale, et relation historique du voyage. Premier Volume, Contenant Essai sur la Géographie des plantes, accompagné d'un Tableau physique des régions équinoxiales, et servant d'introduction à l'Ouvrage). *Paris: Chez Fr. Schœll*.

W

Wace, N. (1990). Antarctica: a new tourist destination. *Applied Geography*, *10*(4), 327-341.

Walker, A.J. (1972). Introduction to the ecology of the antarctic limpet *Patinigera polaris* (Hombron and Jacquinot) at Signy Island, South Orkney Islands. British Antarctic Survey Bulletin. 28:49-71

Waller, C.L., Worland, M.R., Convey, P. & Barnes, D.K. (2006). Ecophysiological strategies of Antarctic intertidal invertebrates faced with freezing stress. *Polar Biology*, 29(12), 1077-1083.

Waller, C.L., Overall, A., Fitzcharles, E.M. & Griffiths, H. (2017). First report of *Laternula elliptica* in the Antarctic intertidal zone. *Polar Biology*, 40(1), 227-230.

Walsh, J.E. (2009). A comparison of Arctic and Antarctic climate change, present and future. *Antarctic Science*, *21*(3), 179-188.

Walsh, J.R., Carpenter, S.R. & Vander Zanden, M.J. (2016). Invasive species triggers a massive loss of ecosystem services through a trophic cascade. *Proceedings of the National Academy of Sciences, 113*(15), 4081-4085.

Walsh, E. & Hudiburg, T.W. (2018). A Framework for Forest Landscape and Habitat Suitability Model Integration to Evaluate Forest Ecosystem Response to Climate Change. *AGUFM*, *2018*, GC11G-0989.

Walther, G.R., Roques, A., Hulme, P.E., Sykes, M.T., Pyšek, P., Kühn, I., ... & Settele, J. (2009). Alien species in a warmer world: risks and opportunities. *Trends in Ecology & Evolution*, *24*(12), 686-693.

Waples, R.S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, *7*(2), 167.

Waples, R.S. & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*(6), 1419-1439.

Waples, R.S. & Do, C. (2008). LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, *8*(4), 753-756.

Ward, G., Hastie, T., Barry, S., Elith, J., & Leathwick, J.R. (2009). Presence‐only data and the EM algorithm. *Biometrics*, 65(2), 554-563.

Warren, D.L. & Seifert, S.N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*(2), 335-342.

Warren, D.L., Wright, A.N., Seifert, S.N. & Shaffer, H.B. (2014). Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Diversity and Distributions*, *20*(3), 334-343.

Waters, J.M., King, T.M., Fraser, C.I. & Craw, D. (2018). Crossing the front: contrasting storm-forced dispersal dynamics revealed by biological, geological and genetic analysis of beach-cast kelp. *Journal of the Royal Society Interface*, *15*(140), 20180046.

Watling, J.I., Brandt, L.A., Bucklin, D.N., Fujisaki, I., Mazzotti, F.J., Romañach, S.S. & Speroterra, C. (2015). Performance metrics and variance partitioning reveal sources of uncertainty in species distribution models. *Ecological Modelling*, *309*, 48-59.

Watson, A.J., Ledwell, J.R., Messias, M.J., King, B.A., Mackay, N., Meredith, M.P., ... & Garabato, A.C. N. (2013). Rapid cross-density ocean mixing at mid-depths in the Drake Passage measured by tracer release. *Nature*, *501*(7467), 408-411.

Watts, M.E., Ball, I.R., Stewart, R.S., Klein, C.J., Wilson, K., Steinback, C., ... & Possingham, H. P. (2009). Marxan with Zones: Software for optimal conservation based land-and sea-use zoning. *Environmental Modelling & Software*, *24*(12), 1513-1521.

Wei, J.H. (1988). Morphological and genetic variation in natural populations of Antarctic *limpet Nacella concinna* (Doctoral dissertation, University of Wales (UCNW, Bangor: Ocean Sciences)).

Weihe, E. & Abele, D. (2008). Differences in the physiological response of inter-and subtidal Antarctic limpets *Nacella concinna* to aerial exposure. *Aquatic Biology*, 4(2), 155-166.

Weinert, M., Mathis, M., Kröncke, I., Neumann, H., Pohlmann, T. & Reiss, H. (2016). Modelling climate change effects on benthos: Distributional shifts in the North Sea from 2001 to 2099. *Estuarine, Coastal and Shelf Science*, *175*, 157-168.

Weir, B.S. & Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370.

Wenger, S.J. & Olden, J.D. (2012). Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3(2), 260-267.

Wernberg, T., Smale, D.A., Tuya, F., Thomsen, M.S., Langlois, T.J., De Bettignies, T., ... & Rousseaux, C.S. (2013). An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. *Nature Climate Change*, 3(1), 78.

West, G.B., Woodruff, W.H. & Brown, J.H. (2002). Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proceedings of the National Academy of Sciences*, *99*(suppl 1), 2473-2478.

Westermeyer, W.E. (2020). *The politics of mineral resource development in Antarctica: alternative regimes for the future*. Routledge.

White, M.G. (1998). Development, dispersal and recruitment: a paradox for survival among Antarctic fish. In *Fishes of Antarctica* (pp. 53-62). Springer, Milano.

Whitehouse, M.J., Meredith, M.P., Rothery, P., Atkinson, A., Ward, P. & Korb, R.E. (2008). Rapid warming of the ocean around South Georgia, Southern Ocean, during the 20th century: forcings, characteristics and implications for lower trophic levels. *Deep Sea Research Part I: Oceanographic Research Papers*, *55*(10), 1218-1228.

White-Newsome, J.L., Brines, S.J., Brown, D.G., Dvonch, J.T., Gronlund, C.J., Zhang, K., ... & O'Neill, M.S. (2013). Validating satellite-derived land surface temperature with in situ measurements: A public health perspective. *Environmental Health Perspectives, 121*(8), 925-931.

Whittaker, R.J., Willis, K.J. & Field, R. (2001). Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, *28*(4), 453-470.

Whittingham, M.J., Stephens, P.A., Bradbury, R.B., & Freckleton, R.P. (2006). Why do we still use stepwise modelling in ecology and behaviour?. *Journal of Animal Ecology*, 75(5), 1182-1189.

Wiencke, C., Amsler, C.D. & Clayton, M.N. (2014) Chapter 5.1 Macroalgae. In: *Biogeographic Atlas of the Southern Ocean* (eds De Broyer C, Koubbi P, Griffiths H, et al.), pp. 66–73. Scientific Committee on Antarctic Research, Cambridge.

Wiens, J.A. (1989). Spatial scaling in ecology. *Functional Ecology*, *3*(4), 385-397.

Wijsman, J.W. & Smaal, A.C. (2011). Growth of cockles (*Cerastoderma edule*) in the Oosterschelde described by a Dynamic Energy Budget model. *Journal of Sea Research*, *66*(4), 372-380.

Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. http://ccl.northwestern.edu/netlogo/

Wiley, E.O., McNyset, K.M., Peterson, A.T., Robins, C.R. & Stewart, A.M. (2003). Niche modeling perspective on geographic range predictions in the marine environment using a machine-learning algorithm.

Williams, J.W. & Jackson, S.T. (2007). Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, 5(9), 475-482.

Williams, J.W., Jackson, S.T. & Kutzbach, J.E. (2007). Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences*, 104(14), 5738-5742.

Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., O'Brien, J.M. & Schwartz, M.W. (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, *15*(4), 565-576.

Wilson, K.A., Westphal, M.I., Possingham, H.P., & Elith, J. (2005). Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation*, *122*(1), 99-112.

Winter, D.J. (2012). *MMOD*: a R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, *12*(6), 1158-1160.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & NCEAS Predicting Species Distributions Working Group. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*(5), 763-773.

Wisz, M.S. & Guisan, A. (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecology*, 9(1), 8.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., ... & Heikkinen, R.K. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews*, 88(1), 15-30.

Wittmann, M.E., Barnes, M.A., Jerde, C.L., Jones, L.A., & Lodge, D.M. (2016). Confronting species distribution model predictions with species functional traits. *Ecology and Evolution*, 6(4), 873-879.

WOCE (2013). https://www.nodc.noaa.gov/OC5/woa13/woa13data.html

Wolcott, T.G. (1973). Physiological ecology and intertidal zonation in limpets (Acmaea): a critical look at" limiting factors". *The Biological Bulletin*, 145(2), 389-422.

Wood, S., Paris, C.B., Ridgwell, A. & Hendy, E.J. (2013). Modelling dispersal and connectivity of broadcast spawning corals at the global scale. *Global Ecology and Biogeography*, *23*(1), 1-11.

Woodin, S.A., Hilbish, T.J., Helmuth, B., Jones, S.J. & Wethey, D.S. (2013). Climate change, species distribution models, and physiological performance metrics: predicting when biogeographic models are likely to fail. *Ecology and Evolution*, *3*(10), 3334-3346.

WoRMS Editorial Board (2016) World Register of Marine Species. http://www.marinespecies.org

Wu, J. (1999). Hierarchy and scaling: extrapolating information along a scaling ladder. *Canadian Journal of Remote Sensing*, *25*(4), 367-380.

Wu, J. & David, J.L. (2002). A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. *Ecological Modelling*, *153*(1-2), 7-26.

Wüest, R.O., Zimmermann, N.E., Zurell, D., Alexander, J.M., Fritz, S.A., Hof, C., ... & Karger, D.N. (2020). Macroecology in the age of Big Data–Where to go from here? *Journal of Biogeography*, *47*(1), 1-12.

Wunsch, C. (2002). What is the thermohaline circulation? *Science*, *298*(5596), 1179-1181.

X

Xavier, J.C., Raymond, B., Jones, D.C. & Griffiths, H. (2015). Biogeography of Cephalopods in the Southern Ocean using habitat suitability prediction models. *Ecosystems*, 19, 220–247.

Xavier, J.C., Brandt, A., Ropert-Coudert, Y., Badhe, R., Gutt, J., Havermans, C. ... & Kennicutt, M.C. (2016). Future challenges in Southern Ocean ecology research. *Frontiers in Marine Science*, 3, 94.

Xiao, N. (2017). ggsci: scientific journal and sci-fi themed color palettes for "ggplot2". R package

version 2.8. https://CRAN.R-project.org/package=ggsci

Xuereb, A., Benestan, L., Normandeau, É., Daigle, R.M., Curtis, J.M., Bernatchez, L. & Fortin, M.J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Molecular Ecology*, *27*(10), 2347-2364.

Y

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H. & Veran, S. (2013). Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, *4*(3), 236-243.

Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., ... & Dormann, C.F. (2018). Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33(10), 790-802.

Young, J.S., Peck, L. S., & Matheson, T. (2006). The effects of temperature on walking and righting in temperate and Antarctic crustaceans. *Polar Biology*, *29*(11), 978-987.

Young, E.F., Rock, J., Meredith, M.P., Belchier, M., Murphy, E.J. & Carvalho, G.R. (2012). Physical and behavioural influences on larval fish retention: contrasting patterns in two Antarctic fishes. *Marine Ecology Progress Series*, *465*, 201-215.

Young, E.F.,Thorpe, S.E., Banglawala, N. & Murphy E.J. (2014), Variability in transport pathways on and around the South Georgia shelf, Southern Ocean: Implications for recruitment and retention, *Journal of Geophysical Research Oceans, 119*, 241–252,

Young, E.F., Belchier, M., Hauser, L., Horsburgh, G.J., Meredith, M.P., Murphy, E.J., ... & Carvalho, G.R. (2015). Oceanography and life history predict contrasting genetic population structure in two Antarctic fish species. *Evolutionary Applications*, *8*(5), 486-509.

Young, E.F., Tysklind, N., Meredith, M.P., de Bruyn, M., Belchier, M., Murphy, E.J. & Carvalho, G.R. (2018). Stepping stones to isolation: Impacts of a changing climate on the connectivity of fragmented fish populations. *Evolutionary Applications*, *11*(6), 978-994.

Z

Zachos, J.C., Dickens, G.R. & Zeebe, R.E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, *451*(7176), 279-283.

Zacharias, M.A., Gerber, L.R. & Hyrenbach, K.D. (2006). Review of the Southern Ocean Sanctuary: marine protected areas in the context of the International Whaling Commission Sanctuary Programme. *Journal of Cetacean Research and Management*, *8*(1), 1-12.

Zambianchi, E., Trani, M. & Falco, P. (2017). Lagrangian transport of marine litter in the Mediterranean Sea. *Frontiers in Environmental Science*, *5*, 5.

Zaniewski, A.E., Lehmann, A. & Overton, J.M. (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, *157*(2-3), 261-280.

Zenteno, L., Cárdenas, L., Valdivia, N., Gómez, I., Höfer, J., Garrido, I. & Pardo, L.M. (2019). Unraveling the multiple bottom-up supplies of an Antarctic nearshore benthic community. *Progress in Oceanography*, *174*, 55-63.

Zhang, X. & Mahadevan, S. (2019). Ensemble machine learning models for aviation incident risk prediction. *Decision Support Systems*, *116*, 48-63.

Zhao, K., Wulder, M.A., Hu, T., Bright, R., Wu, Q., Qin, H., ... & Brown, M. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sensing of Environment*, 232, 111181.

Zhou, Z.H. (2012). *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC.

Zhu, G.P. & Peterson, A.T. (2017). Do consensus models outperform individual models? Transferability evaluations of diverse modeling approaches for an invasive moth. *Biological Invasions, 19*(9), 2519-2532.

Zuckerberg, B., Fink, D., La Sorte, F.A., Hochachka, W.M. & Kelling, S. (2016). Novel seasonal land cover associations for eastern North American forest birds identified through dynamic species distribution modelling. *Diversity and Distributions*, 22(6), 717-730.

Zurell, D., Jeltsch, F., Dormann, C.F. & Schröder, B. (2009). Static species distribution models in dynamically changing systems: how good can predictions really be? *Ecography*, *32*(5), 733-744.

Zurell, D., Thuiller, W., Pagel, J., Cabral, J.S., Münkemüller, T., Gravel, D., ... & Zimmermann, N.E.

(2016). Benchmarking novel approaches for modelling species range dynamics. *Global Change Biology*, *22*(8), 2651-2664.

Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T., & Wüest, R.O. (2020). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, 47(1), 101-113.

LIST OF FIGURES

INTRODUCTION

CHAPTER 1 — Guillaumot et al. (2020a) – *Nacella concinna* DEB model –

Arnould-Pétré et al. (2020) – *Abatus cordatus* DEB-IBM model –

Guillaumot et al. (2020b) – SDM Choice of descriptors –

Guillaumot et al. (2020c) – SDM and extrapolation –

INTRODUCTION

CHAPTER 1 Guillaumot et al. (2020a) – *Nacella concinna* DEB model –

Arnould-Pétré et al. (2020) – *Abatus cordatus* DEB-IBM model –

CHAPTER 2 Guillaumot et al. (2019) – SDM cross-validation procedures –

Guillaumot et al. (2020b) – SDM Choice of descriptors –


THESIS MATERIAL
© Jenny Wynyslowski

THESIS MATERIAL

In the matter of replicating results of this PhD, scripts used to generate the models relating to the different studies were annotated and shared:

For Chapter 1 (DEB modelling), the Matlab codes for *Nacella concinna* DEB model (Guillaumot et al. 2020a) are available on the Add-my-Pet platform (https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/entries_web/Nacella_concinna/Nacella_concinna_res.html).
For the NetLogo codes of the DEB-IBM of *Abatus cordatus* (Arnould-Pétré et al. 2020), a specific page was created on the Netlogo platform (http://modelingcommons.org/browse/one_model/6201). Input data and some guidelines are provided to help you implement the model.

For Chapter 2 (SDM modelling), R codes corresponding to the four studies are available on my Github page (https://github.com/charleneguillaumot/THESIS). In addition, these codes were compiled into the *SDMPlay* R package (https://CRAN.R-project.org/package=SDMPlay) as simplified functions. Have a specific look at the package vignettes: they were created purposely to apply these different functions and help beginners to generate their first SDM for Southern Ocean case studies.

For Chapter 3 (Integrated approaches), R codes to generate SDM models for *Halicarcinus planatus* case study (López-Farrán / Guillaumot et al. in press) come from the SDMPlay codes. For Fabri-Ruiz et al. (2021) analysis, you need to directly contact Salomé Fabri-Ruiz (salome.fabriruiz@gmail.com). Finally for the last study (Guillaumot et al. submitted), codes related to the simple GLM models, integrated DEB-SDM and integrated Bayesian approaches are available on my GitHub page (https://github.com/charleneguillaumot/THESIS). Don't hesitate to contact me for any issue (charleneguillaumot21@gmail.com).

Finally, **Chapter 4 (dispersal models)** mainly relies on Valérie Dulière's codes (vduliere@naturalsciences.be).

Tutorial for SDMPlay: 1/ Compute Species Distribution Models

2021-02-01

Species distribution modelling (SDM) has been developed for several years to address conservation issues, to assess the direct impact of human activities on ecosystems and to predict the potential distribution shifts of invasive species (see Elith et al. 2006, Pearson 2007, Elith and Leathwick 2009). SDM relates species occurrences with environmental information and can predict species distribution on their entire occupied space. **This approach has been increasingly applied to Southern Ocean case studies, but requires corrections in such a context, due to the broad scale area, the limited number of presence records available and the spatial and temporal aggregations of these datasets.**

SDMPlay is a pedagogic package that will allow you to compute SDMs, to understand the overall method, and to produce model outputs. The package, along with its associated vignettes, highlights the different steps of model calibration and describes how to choose the best method to generate accurate and relevant outputs. SDMPlay proposes codes to apply a popular machine learning approach, BRT (Boosted Regression Trees) and introduces MaxEnt (Maximum Entropy). It contains occurrences of marine species and environmental descriptor datasets as examples associated with several vignette tutorials.

- **Tutorial #1/ Compute Species Distribution Models**
  Focusses on data structure, data preparation and general model computing.
- **Tutorial #2/ SDM outputs**
  Presents the main outputs you can generate with your SDM.
- **Tutorial #3/ Importance of model calibration**
  Highlights the procedure to accurately calibrate your model and proposes some methods to limit the influence of several biases.
- **Tutorial #4/ Spatial cross-validation**
  Cross-validation is a method to validate your model. When working with presence data spatially aggregated, the cross-validation procedure should be adapted. This tutorial provides some elements to apply this method (refering to Guillaumot et al. 2019).
- **Tutorial #5/ Spatial extrapolation**
  Models can extrapolate when projected on broad scale areas. This tutorial provides codes to calculate extrapolation scores and generate extrapolation maps that could be associated to SDM maps (refering to Guillaumot et al. 2020).

Data overview

Occurrence records

In the package, you can download the presence-only records of two echinoid species of the Kerguelen Plateau, *Brisaster antarcticus* and *Ctenocidaris nutrix* and the presence-only records of the sea stars *Odontaster validus* and *Glabraster antarctica*, distributed at the scale of the Southern Ocean. These species present contrasting ecological niches, with different feeding preferences and reproductive behaviours (David et al. 2005, Mah and Blake 2012). The complete dataset of Kerguelen echinoid species is available in Guillaumot et al. (2016), the complete dataset of Southern Ocean sea stars is available in the updated database of Moreau et al. (2018).

```
library(SDMPlay)
data("ctenocidaris.nutrix") # Species distributed on the Kerguelen Plateau, table with
                            # longitude, latitude and several other columns
head(ctenocidaris.nutrix)
```

```
##    id     scientific.name scientific.name.authorship
## 56  1 Ctenocidaris_nutrix             (Thomson 1876)
## 57  2 Ctenocidaris_nutrix             (Thomson 1876)
## 58  3 Ctenocidaris_nutrix             (Thomson 1876)
## 59  4 Ctenocidaris_nutrix             (Thomson 1876)
## 60  5 Ctenocidaris_nutrix             (Thomson 1876)
## 61  6 Ctenocidaris_nutrix             (Thomson 1876)
##                          genus                        family
## 56 Ctenocidaris Mortensen 1910 Ctenocidarinae Mortensen 1928
## 57 Ctenocidaris Mortensen 1910 Ctenocidarinae Mortensen 1928
## 58 Ctenocidaris Mortensen 1910 Ctenocidarinae Mortensen 1928
## 59 Ctenocidaris Mortensen 1910 Ctenocidarinae Mortensen 1928
## 60 Ctenocidaris Mortensen 1910 Ctenocidarinae Mortensen 1928
## 61 Ctenocidaris Mortensen 1910 Ctenocidarinae Mortensen 1928
##    order.and.higher.taxonomic.rank decimal.Longitude decimal.Latitude depth
## 56            Cidaroida Claus 1880          67.13167        -48.98500   315
## 57            Cidaroida Claus 1880          67.33167        -49.44167   301
## 58            Cidaroida Claus 1880          67.51167        -49.00500   206
## 59            Cidaroida Claus 1880          67.54167        -48.11667   365
## 60            Cidaroida Claus 1880          67.88500        -49.46667   191
## 61            Cidaroida Claus 1880          68.05833        -49.06667   178
##    year       campaign             reference          vessel
## 56 1975 MD04 (BENTHOS) De Ridder et al. 1992 Marion Dufresne
## 57 1975 MD04 (BENTHOS) De Ridder et al. 1992 Marion Dufresne
## 58 1975 MD04 (BENTHOS) De Ridder et al. 1992 Marion Dufresne
## 59 1975 MD04 (BENTHOS) De Ridder et al. 1992 Marion Dufresne
## 60 1975 MD04 (BENTHOS) De Ridder et al. 1992 Marion Dufresne
## 61 1975 MD04 (BENTHOS) De Ridder et al. 1992 Marion Dufresne
```

You can similarly load data for Southern Ocean distributed species

```
data("Odontaster.validus") # Species distributed around the entire Southern Ocean, table
                           # with longitude and latitude only
head(Odontaster.validus)
```

```
##   longitude latitude
## 1  166.6492 -77.8504
## 2  166.6492 -77.8504
## 3  166.6492 -77.8504
## 4  166.6492 -77.8504
## 5  166.6492 -77.8504
## 6  166.4818 -77.4319
```

In which concerns environmental descriptors, two regions are presented in this package: the Kerguelen Plateau and the Southern Ocean. For the Kerguelen Plateau, the environmental dataset compiles 15 environmental descriptors, displayed in a raster format, for three time periods [1965-1974], [2005-2012], and for the future climatic scenario AIB (IPCC, 4th report 2007) for 2200. Grid-cell pixels are set at a 0.1° resolution

and data were not interpolated (presence of N/A values in the area). Extra metadata and environmental layers are available in Guillaumot et al. (2016).

Environmental layers

Load the raster stacks

```
library(raster)
data(predictors1965_1974)
data(predictors2005_2012)
data(predictors2200AIB)
```

Observe their content

```
predictors2005_2012
```

```
## class      : RasterStack
## dimensions : 100, 179, 17900, 15  (nrow, ncol, ncell, nlayers)
## resolution : 0.1, 0.1  (x, y)
## extent     : 63, 80.9, -56, -46  (xmin, xmax, ymin, ymax)
## crs        : +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0
## names      :         depth, seasurfac//_2005_2012, seasurfac//_2005_2012, seafloor_//_2005_2012, sea
## min values : -4.977000e+03,          3.263100e-01,         -4.002820e+00,         -2.985100e-01,
## max values :    -1.0000000,             9.0129099,            -1.2873000,             4.8823700,
```

```
names(predictors2005_2012)
```

```
##  [1] "depth"
##  [2] "seasurface_temperature_mean_2005_2012"
##  [3] "seasurface_temperature_amplitude_2005_2012"
##  [4] "seafloor_temperature_mean_2005_2012"
##  [5] "seafloor_temperature_amplitude_2005_2012"
##  [6] "seasurface_salinity_mean_2005_2012"
##  [7] "seasurface_salinity_amplitude_2005_2012"
##  [8] "seafloor_salinity_mean_2005_2012"
##  [9] "seafloor_salinity_amplitude_2005_2012"
## [10] "chlorophyla_summer_mean_2005_2012"
## [11] "geomorphology"
## [12] "sediments"
## [13] "slope"
## [14] "seafloor_oxygen_mean_2005_2012"
## [15] "roughness"
```

You can select only a part of the layers with the 'subset' function of the raster package.

```
layer_ex <- raster::subset(predictors2005_2012, c(1:4))
plot(layer_ex, cex.axis=0.7, cex.main=0.8,legend.width=1, legend.shrink=0.5)
```

As you can notice (Fig. 1), particularly for seafloor layers, maps are incomplete and contain an important number of missing values (N/A), because data were not interpolated in space. You can interpolate your data if you want using functions provided in the raster package, but you need to be aware that it will complexify interpretation afterwards.

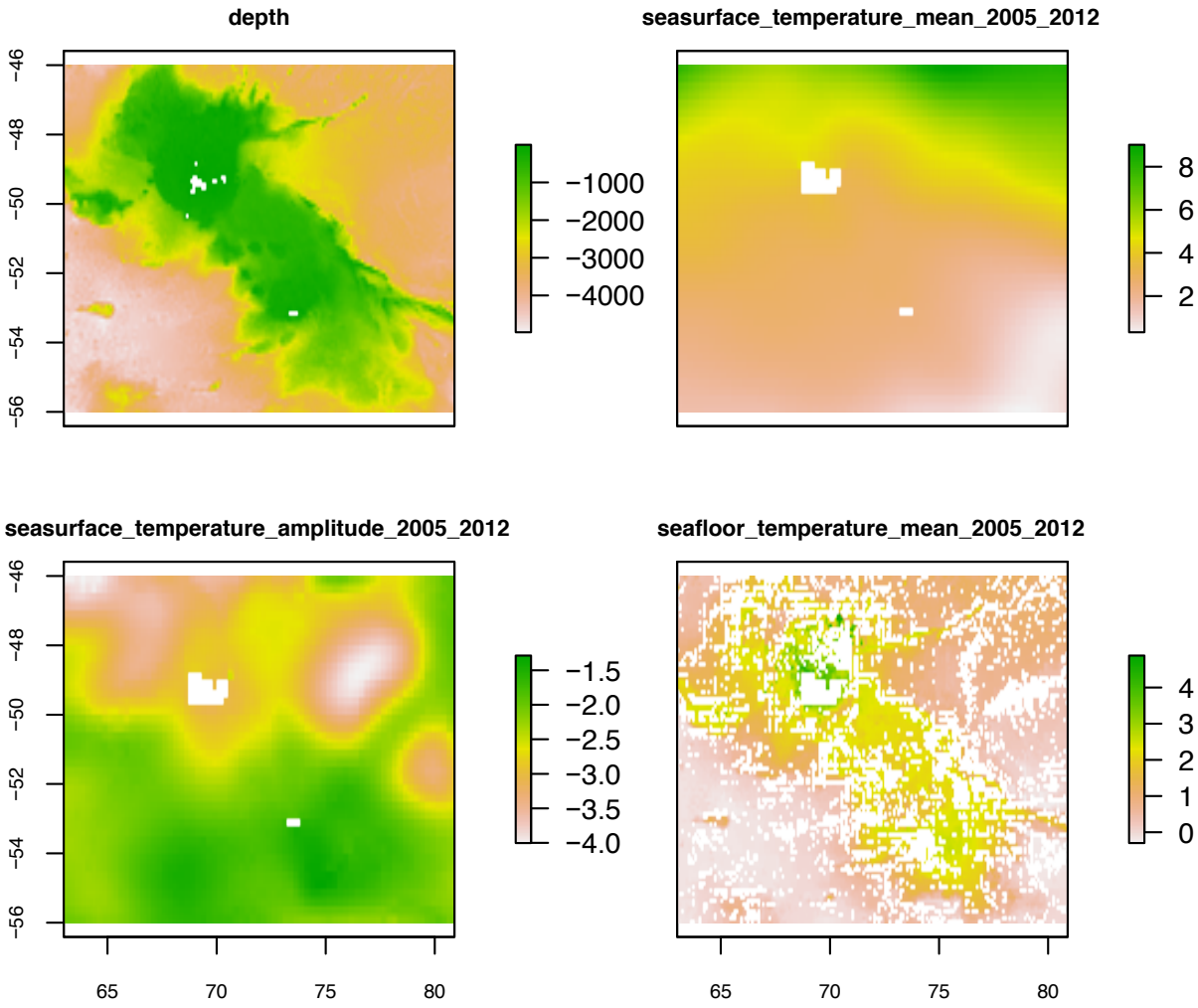


Figure 1: Some environmental descriptors on the Kerguelen Plateau area

At the scale of the Southern Ocean, some layers are available on the AAD website (click to access to the link).

Download the following dataset, select, open and stack some layers according to your study species:

Three of these layers are provided in the internal data of the SDMPlay package, to help you immediately generate simple examples.

```
data("depth_SO")
data("ice_cover_mean_SO")
data("seafloor_temp_2005_2012_mean_SO")

predictors_stack_SO <- stack(depth_SO,ice_cover_mean_SO,seafloor_temp_2005_2012_mean_SO)
names(predictors_stack_SO)<-c("depth","ice_cover_mean","seafloor_temp_mean")
```

If you have downloaded the environmental descriptors on your computer and stored them in a folder (called for example "environmental_layers"), you can do:

```
library(ncdf4)
depth <- raster("environmental_layers/depth.nc")
seafloor_temp_mean <- raster("environmental_layers/seafloor_temp_2005_2012_mean.nc")

predictors_stack <- stack(depth,seafloor_temp_mean)
```

```
# Plot the raster layers, create nice color palettes (Fig. 2)
library(RColorBrewer)
my.palette.oranges <- brewer.pal(n = 9, name = "Oranges")
my.palette.blue <- rev(brewer.pal(n = 9, name = "Blues"))

plot(raster::subset(predictors_stack_SO,1), col=my.palette.blue)
points(worldmap, type="l")
```

We also advice you to have a look at the SOmap (click to link) R package for nice plotting of Southern Ocean maps

```
remotes::install_github("AustralianAntarcticDivision/SOmap")
library(SOmap)
SOmap()
# see the tutorial at the above website to customize your plot
```

Prepare model inputs

The first step after loading and checking your data is to adapt your dataset for modelling. Model algorithms require a table containing the environmental values associated with occurrence data.

Usually, SDMs are calibrated with presence and absence records. If you don't have access to absence records (or you don't rely in them!), it is necessary to sample in the projected area a set of points to define the background environmental conditions (termed background or pseudo-absence records)(Pearce and Boyce 2006). In this case, the SDM will be calibrated with presence AND background data instead of presences/absences.

Several background sampling methods exist (Phillips et al. 2009), and its choice depends on the sampling pattern of the presence-only records and on the scientific questions. In this first vignette, we will give

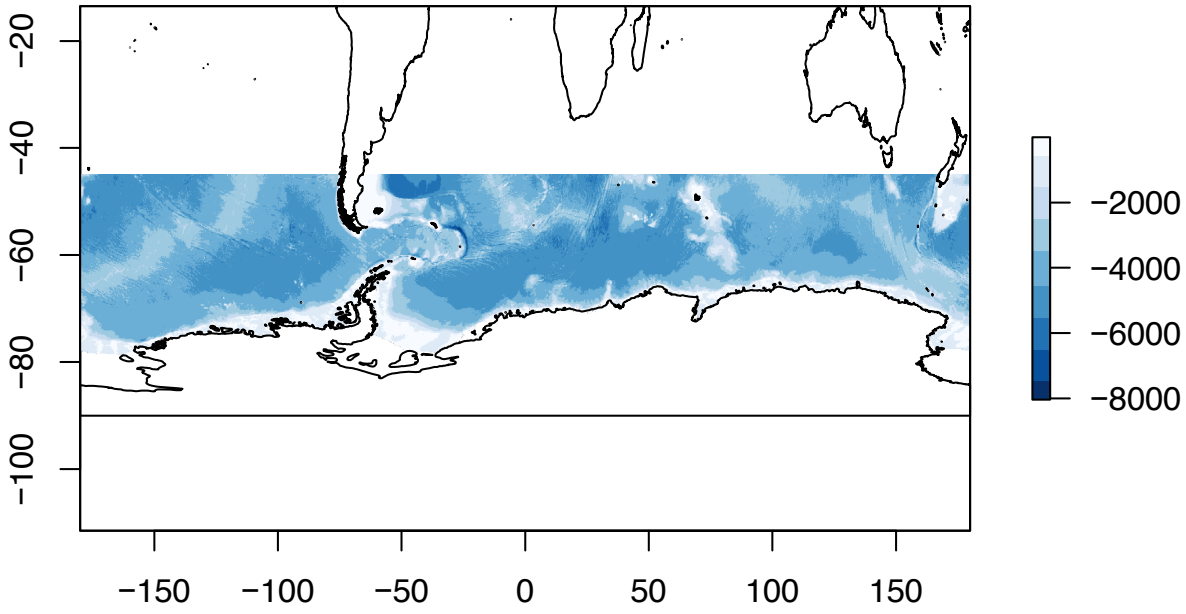


Figure 2: Bathymetry layer at the Southern Ocean extent

the example of a random sampling of the background records but see Tutorial #3/ "Importance of model calibration" for further implementations.

The first step before running a SDM is to create a table such as presented below. The `SDMtab` function can automatize its creation.

ID*	Longitude	Latitude	Depth	...	Temperature
1	63.33	-48.26	-480	...	1.4
1	64.13	-48.57	-104	...	1.2
...	...	...	...	...	...
0	67.32	-47.23	-1013	...	2.5
0	67.90	-55.45	-98	...	4.3

*ID corresponds to presence data (ID=1), or background data (ID=0) (or absences if you have absences !).

To create this table, apply the following steps:

Extract longitude and latitude values

You need to work with longitude and latitude data only -> clean your initial dataframe.

```
ctenocidaris.nutrix.occ <- ctenocidaris.nutrix[,c(7,8)] # longitude (first column),
                                                        # latitude (second column)
head(ctenocidaris.nutrix.occ)
```

```
##    decimal.Longitude decimal.Latitude
## 56          67.13167        -48.98500
## 57          67.33167        -49.44167
## 58          67.51167        -49.00500
## 59          67.54167        -48.11667
## 60          67.88500        -49.46667
## 61          68.05833        -49.06667
```

Create your SDMtab dataframe

```
SDMtable_ctenocidaris <- SDMPlay:::SDMtab(xydata=ctenocidaris.nutrix.occ,
       predictors=predictors2005_2012,
       unique.data=FALSE,
       same=TRUE)
```

When `unique.data= TRUE`, presence-only duplicates located on a same grid-cell pixel will be removed from the `xydata` variable.

`same` and `background.nb` functions refer to the sampling of background data: `background.nb`, indicates the specific number of background data to be sampled, while `same` is a shortcut that makes the number of background data similar to the number of presence-only data available. You can refer to Barbet-Massin et al. (2012) to choose the most appropriate number of background data to sample for your case study.

We can display the beginning and the end of the first columns of this new `SDMtab` object:

```
head(SDMtable_ctenocidaris[,c(1:5)])
```

```
##   id longitude latitude depth seasurface_temperature_mean_2005_2012
## 1  1     67.15   -48.95  -653                                4.24109
## 2  1     67.35   -49.45  -204                                3.89770
## 3  1     67.55   -49.05  -168                                4.06841
## 4  1     67.55   -48.15  -355                                4.65109
## 5  1     67.85   -49.45  -136                                3.86259
## 6  1     68.05   -49.05  -155                                4.03309
```

```
tail(SDMtable_ctenocidaris[,c(1:5)])
```

```
##     id longitude latitude depth seasurface_temperature_mean_2005_2012
## 245  0     71.25   -49.35  -380                                3.207595
## 246  0     66.65   -48.15  -599                                4.565590
## 247  0     70.15   -53.85 -3487                                2.440090
## 248  0     69.65   -47.95  -135                                4.675900
## 249  0     80.25   -52.65 -3428                                1.077700
## 250  0     65.25   -50.95 -3385                                3.428560
```

The dataframe combines environmental values of the 125 presence-only data available (ID=1) and environmental values associated with 125 background data randomly sampled in the area (ID=0).

You can display the sampled data on a map (Fig.3):

```
# nice colors
bluepalette<-colorRampPalette(c("blue4","blue","dodgerblue", "deepskyblue",
                                "lightskyblue"))(800)

# map
data("worldmap")
# Isolate depth layer from the environmental stack (Kerguelen data)
# Use it as a nice background for your figure
depth <- subset(predictors2005_2012,1)

# Extract background coordinates from SDMtable
background.occ <- subset(SDMtable_ctenocidaris,SDMtable_ctenocidaris$id==0)[,c(2,3)]

# plot the result (Fig.3)
plot(depth, col=bluepalette, cex=0.8,legend.width=0.5, legend.shrink=0.4,
     legend.args=list(text='Depth (m)', side=3, font=2, cex=0.8))
points(worldmap, type="l")
points(ctenocidaris.nutrix.occ, pch= 20, col="black")
points(background.occ, pch= 20, col="red")
legend("bottomleft", pch=20, col=c("black", "red"), legend=c("presence-only data","background data"),
       cex=0.6, bg="white")
```

You can assess the quality of your dataset with the `SDMdata.quality` function. This function estimates the percentage of presence-only records that fall on grid-cell pixels containing non-informative values (N/A). It estimates the quality of your dataset.

```
head(SDMdata.quality(SDMtable_ctenocidaris))
```

```
##                                        NA.percent (%)
```

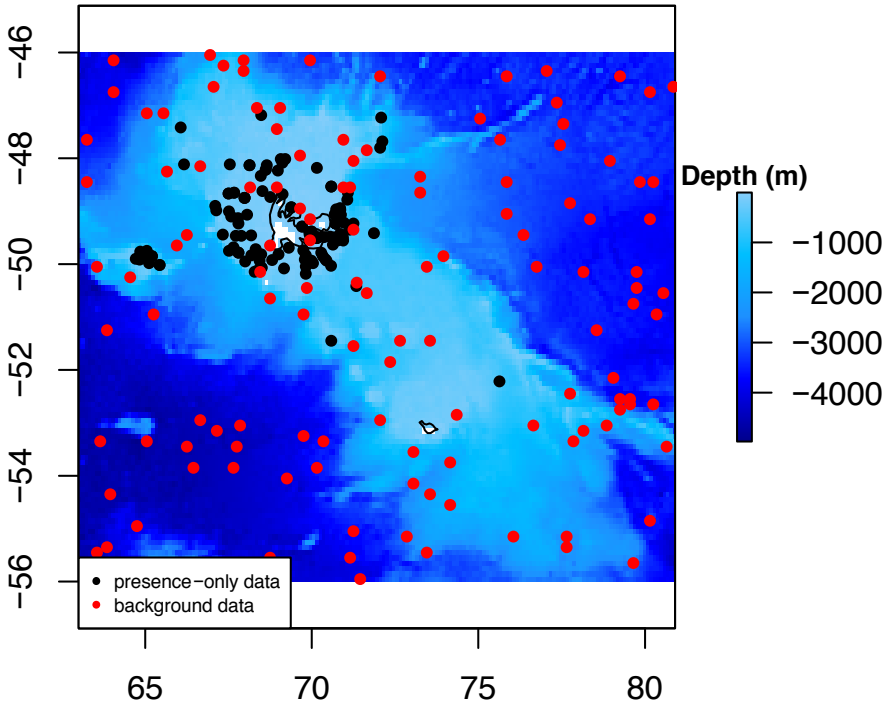


Figure 3: Presence data of Ctenocidaris nutrix and sampled background records on the Kerguelen Plateau

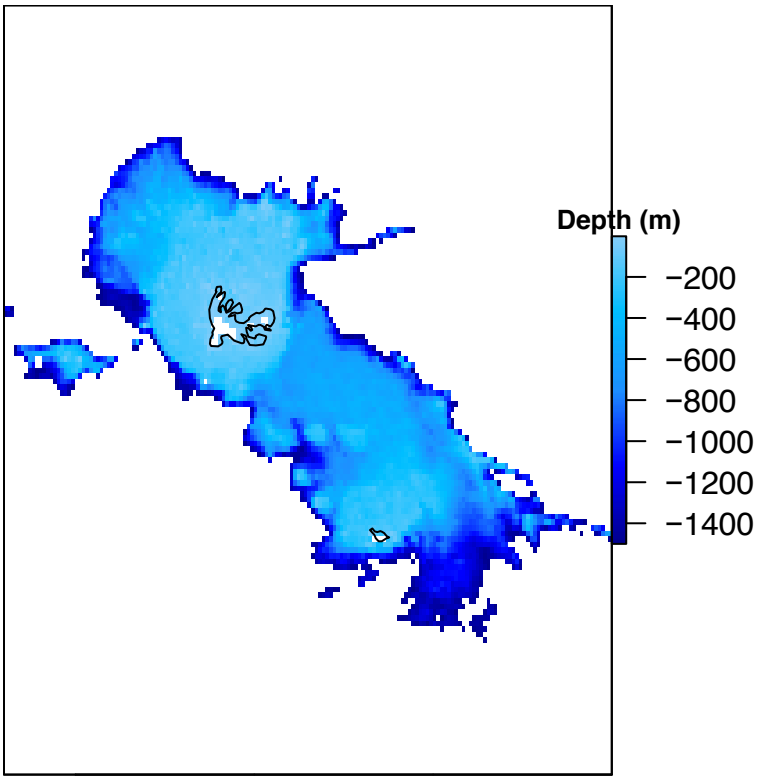


Figure 4: Restraining the area to 1,500m depth

```
## depth                                          1.2
## seasurface_temperature_mean_2005_2012          5.6
## seasurface_temperature_amplitude_2005_2012     5.6
## seafloor_temperature_mean_2005_2012           32.0
## seafloor_temperature_amplitude_2005_2012      32.0
## seasurface_salinity_mean_2005_2012             5.6
```

A last calibration step that you can perform before modelling is delineating the modelled area (Fig.4). The `delim.area` function can be used to restrict in geography and/or depth the environmental descriptor layers. This step can play an important role to enhance modelling performances by limiting the extent of extrapolation.

```
# restrict to 1500m depth
predictors2005_2012_1500m <- SDMPlay:::delim.area(predictors2005_2012, longmin=62, longmax=80,
                                                  latmin=-55, latmax=-45, interval=c(0,-1500))
# plot the new layer (Fig.4)
plot(subset(predictors2005_2012_1500m,1), col=bluepalette,legend.width=0.5, legend.shrink=0.4,
     legend.args=list(text='Depth (m)', side=3, font=2, cex=0.8))
points(worldmap, type="l")
```

You can focus your background sampling on this restrained environment (Fig.5). Run again the `SDMtab` code with these changes. The function will omit the N/A pixels when selecting the random background data.

```
SDMtable_ctenocidaris_1500 <- SDMtab(xydata=ctenocidaris.nutrix.occ,
       predictors=predictors2005_2012_1500m,
```

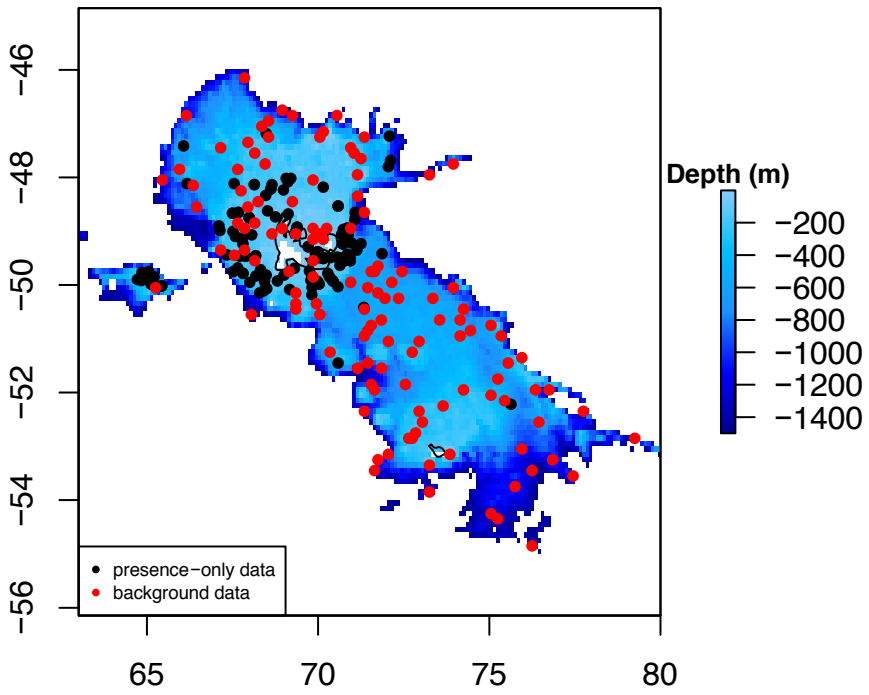


Figure 5: Restraining the area and the background sampling to 1,500m depth

```
        unique.data=FALSE,
        same=TRUE)
```

```
# Observe the changes (Fig.5)
background.occ_1500 <- subset(SDMtable_ctenocidaris_1500,SDMtable_ctenocidaris_1500$id== 0)[,c(2,3)]
plot(subset(predictors2005_2012_1500m,1), col=bluepalette, cex=0.8, legend.width=0.5,
     legend.shrink=0.4,
     legend.args=list(text='Depth (m)', side=3, font=2, cex=0.8))
points(worldmap, type="l")
points(ctenocidaris.nutrix.occ, pch= 20, col="black")
points(background.occ_1500, pch= 20, col="red")
legend("bottomleft", pch=20, col=c("black", "red"), legend=c("presence-only data",
                                                             "background data"), cex=0.6)
```

Perform species distribution models

Once you have built your `SDMtab` dataframe, you can easily perform models using the `compute.brt` or `compute.maxent` functions.

```
compute.brt(x, proj.predictors, tc = 2, lr = 0.001, bf = 0.75,
           n.trees = 50, step.size = n.trees)
compute.maxent(x, proj.predictors)
```

The fonctions require two main parameters, `x` which correspond to the `SDMtab` object previously created and `proj.predictors` being the `RasterStack` containing the environmental descriptors on which you want to project your model. The other arguments aim at calibrating the model. You can refere to Elith et al. (2008) and Elith et al. (2011) to choose the parameters according to your dataset. BRT arguments are explained in the gbm package.

Example for BRT

Predict species distribution on the Kerguelen Plateau, for [2005-2012]

```
Cteno_model_2005_2012 <- SDMPlay:::compute.brt(x=SDMtable_ctenocidaris_1500,
                                               proj.predictors=predictors2005_2012_1500m,
                                               tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

While the function is uploading, you can observe that the `gbm` function, called by `SDMPlay`, calculates the regression trees until reaching the best estimation of the predicted deviance. See Tutorial #3/"Importance of model calibration" for more information about the choice and the influence of these parameters on model predictions. See also Elith et al. (2008) for details on these parameters.

Afterwards, different outputs can be produced (see Tutorial #2/ SDM outputs for detailed applications). Here we will just provide the example of the distribution map:

```
# display nice colors
palettecolor <- colorRampPalette(c("deepskyblue", "darkseagreen","lightgreen",
                                   "green","yellow","gold","orange", "red","firebrick"))(100)
# plot the results (Fig.6)
plot(Cteno_model_2005_2012$raster.prediction,col=palettecolor, main="Projection for [2005-2012]",
     cex.axis= 0.7,
     legend.width=0.5, legend.shrink=0.25,
     legend.args=list(text='Distribution probability', side=3, font=2, cex=0.8))
points(worldmap, type="l")
```

The output of your model cannot extrapolate on the grid-cell pixels from which it does not know environmental values (N/A pixels). Choose the option of interpolating your `RasterStack` layers before modelling or when projecting if you want to obtain smoother prediction maps. The map gives you the species distribution probabilities contained between 0 and 1.

Project on other time periods

If you want to project your model on another time period and infere your species distribution for other environmental conditions, you just need to change the `proj.predictors` in `compute.brt`, and replace it by a stack of future layers. The fonction will do the relationship between the environmental descriptors used for creating the model (the stack of predictors of present conditions that you have used to create the SDMtab matrtix) and the one for projecting (your future stack in this case). Be careful ! You must ensure that the extent, number, order and names of your future raster layers (stacked) are similar than the ones you have for the present time period.

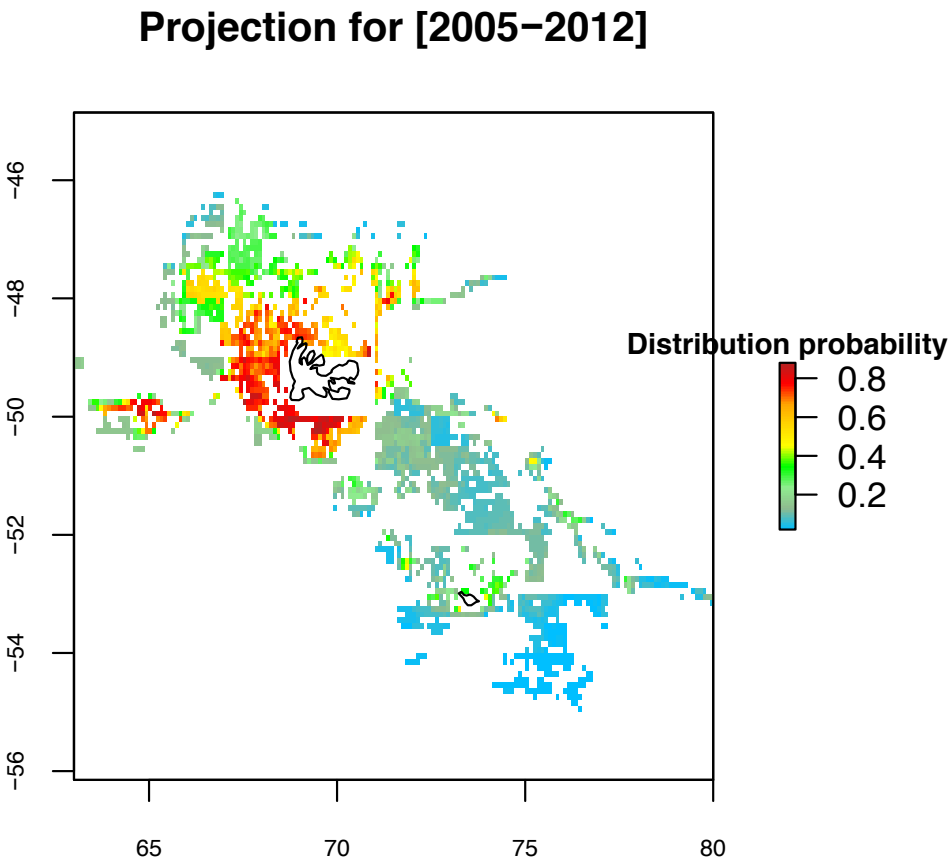


Figure 6: Projection for [2005-2012], Ctenocidaris nutrix predicted distribution, BRT

Example for MaxEnt

The procedure for MaxEnt algorithm is similar to BRT. `compute.maxent` uses the functionalities of the *dismo* `maxent` function. This function calls MaxEnt species distribution software, which is a java program that can be downloaded here (click to access to the link). In order to run `compute.maxent`, put the `maxent.jar` file downloaded at this address in the *java* folder of the *dismo* package (path obtained with the `system.file('java', package='dismo')` command). For issues with Java installation, consult dismo and rJava packages.

MaxEnt model outputs are similar to BRT, you can compute maps, response plots, environmental descriptor contributions. Refere to the example section of the function for more details.

Go further

SDMPlay provides extra fonctions to go further in your modelling work. You can perform null models with `null.model`, evaluate modelling performance and define probability threshold with `SDMeval`, calculate extrapolation and test different cross-validation procedures. See the following tutorials and the examples provided within the different functions for further details.

References

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2), 327-338.

David, B., Choné, T., Mooi, R. & de Ridder C. (2005). *Antarctic echinoidea* (Vol. 10). ARG Gantner.

Elith, J., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J Hijmans, R., Huettmann, F., ... & A Loiselle, B. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151.

Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.

Elith, J. & Leathwick, J.R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*, 17(1), 43-57.

Guillaumot, C., Martin, A., Fabri-Ruiz, S., Eléaume, M. & Saucède, T. (2016). Echinoids of the Kerguelen Plateau–occurrence data and environmental setting for past, present, and future species distribution modelling. *ZooKeys*, (630), 1.

Guillaumot, C., Artois, J., Saucède, T., Demoustier, L., Moreau, C., Eléaume, M. ... & Danis, B. (2019). Broad-scale species distribution models applied to data-poor areas. *Progress in Oceanography*, 175, 198-207.

Guillaumot, C., Moreau, C., Danis, B. & Saucède, T. (2020). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*, 188, 102438.

Mah, C.L. & Blake, D.B. (2012). Global diversity and phylogeny of the Asteroidea (Echinodermata). *PloS One*, 7(4), e35644.

Moreau, C., Mah, C., Agüera, A., Améziane, N., Barnes, D., Crokaert, G., ... & Jażdżewska, A. (2018). Antarctic and sub-Antarctic Asteroidea database. *ZooKeys*, (747), 141.

Pearce, J.L. & Boyce, M.S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, 43(3), 405-412.

Pearson, R.G. (2007). Species' distribution modeling for conservation educators and practitioners. Synthesis. *American Museum of Natural History*, 50.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181-197.

Tutorial for SDMPlay: 2/ Generate SDM outputs

2021-02-01

Species distribution modelling (SDM) has been developed for several years to address conservation issues, to assess the direct impact of human activities on ecosystems and to predict the potential distribution shifts of invasive species (see Elith et al. 2006, Pearson 2007, Elith and Leathwick 2009). SDM relates species occurrences with environmental information and can predict species distribution on their entire occupied space. **This approach has been increasingly applied to Southern Ocean case studies, but requires corrections in such a context, due to the broad scale area, the limited number of presence records available and the spatial and temporal aggregations of these datasets.**

SDMPlay is a pedagogic package that will allow you to compute SDMs, to understand the overall method, and to produce model outputs. The package, along with its associated vignettes, highlights the different steps of model calibration and describes how to choose the best method to generate accurate and relevant outputs. SDMPlay proposes codes to apply a popular machine learning approach, BRT (Boosted Regression Trees) and introduces MaxEnt (Maximum Entropy). It contains occurrences of marine species and environmental descriptor datasets as examples associated with several vignette tutorials.

Objectives of tutorial #2/ Generate SDM outputs

First, basic approaches to generate SDM outputs are provided (prediction maps, contribution percentages of the environmental descriptors, response plots, interactions between variables). Second, classic tools to evaluate model performance are supplied (Area Under the Curve, True Skill Statistics, Biserial Pearson Correlation) and are completed with tools to perform null models (Raes and ter Steege 2007, van Proosdij et al. 2016).

See also. . .

- **Tutorial #1/ Compute Species Distribution Models**
  Focusses on data structure, data preparation and general model computing.
- **Tutorial #3/ Importance of model calibration**
  Highlights the procedure to accurately calibrate your model and proposes some methods to limit the influence of several biases.
- **Tutorial #4/ Spatial cross-validation**
  Cross-validation is a method to validate your model. When working with presence data spatially aggregated, the cross-validation procedure should be adapted. This tutorial provides some elements to apply this method (refering to Guillaumot et al. 2019).
- **Tutorial #5/ Spatial extrapolation**
  Models can extrapolate when projected on broad scale areas. This tutorial provides codes to calculate extrapolation scores and generate extrapolation maps that could be associated to SDM maps (refering to Guillaumot et al. 2020).

Let's start with Tutorial #2 !

After running your model (follow Tutorial #1), several model outputs can be presented and used for model interpretation. First, you can plot the map of your results (Fig. 1):

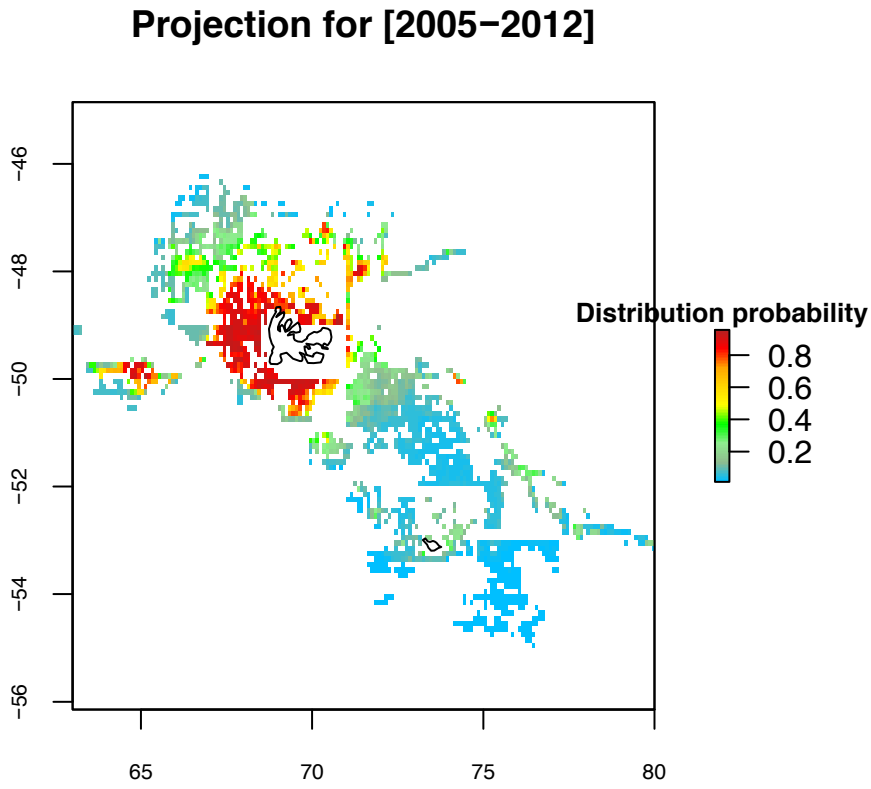


Figure 1: Model predictions for Ctenocidaris nutrix, for [2005-2012], on the Kerguelen Plateau area, with BRT

```
# Plot the map of your results (Fig. 1)
palettecolor <- colorRampPalette(c("deepskyblue", "darkseagreen","lightgreen","green",
                                   "yellow","gold","orange", "red","firebrick"))(100)

plot(Cteno_model_2005_2012$raster.prediction, col=palettecolor, main="Projection for [2005-2012]",
     cex.axis= 0.7,
     legend.width=0.5, legend.shrink=0.25,
     legend.args=list(text='Distribution probability', side=3, font=2, cex=0.8))
points(worldmap, type="l")
```

Contribution of the different environmental descriptors

The '$response' part of the produced model variable also provides several information that you can use to study your model, among which the contribution of each environmental descriptor to the model (Fig. 2).

```
contributions <- Cteno_model_2005_2012$response$contributions
b <- barplot(contributions[,2], ylab="Contribution (%)")
text(b-0.1, par("usr")[3] - 0.025, srt = 45, adj = 1, labels=contributions[,1],cex=0.5,xpd=T)
```

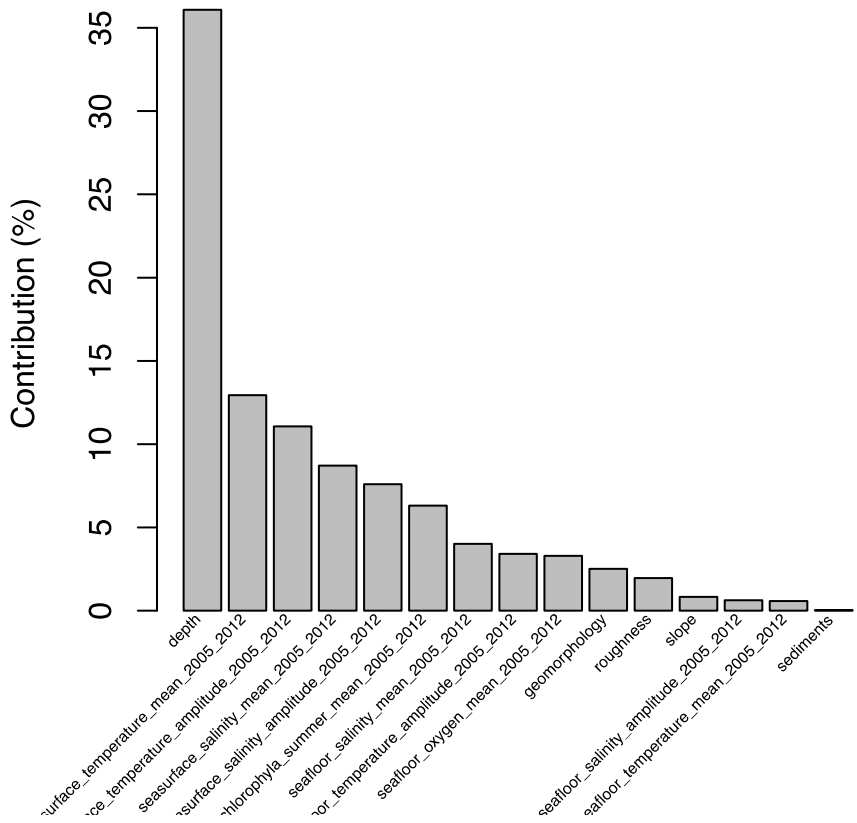


Figure 2: Percentage of contribution of each environmental descriptors to the model

Response plots

Response plots are useful indicators of environmental preferential values for the species. The y axis contains distribution probabilities predicted by the model and response plots associate these values with environmental data ('x' axis) (Fig. 3). You can use the quick and simple function below to generate them, or you can manually create them, with the extract function of the *raster* package: extract the values of your model predictions for each latitude-longitude pixels (these probability values will be your y axis) and also extract the values of the environment at the corresponding pixels (e.g. take the temperature layer, temperature values will be your x axis). Plot y~x and fit a polynom to observe the trend.

```
# Figure 3:
library(dismo)
gbm.plot(Cteno_model_2005_2012$response,n.plots=12,cex.axis=0.6,cex.lab=0.7, smooth=TRUE)
```

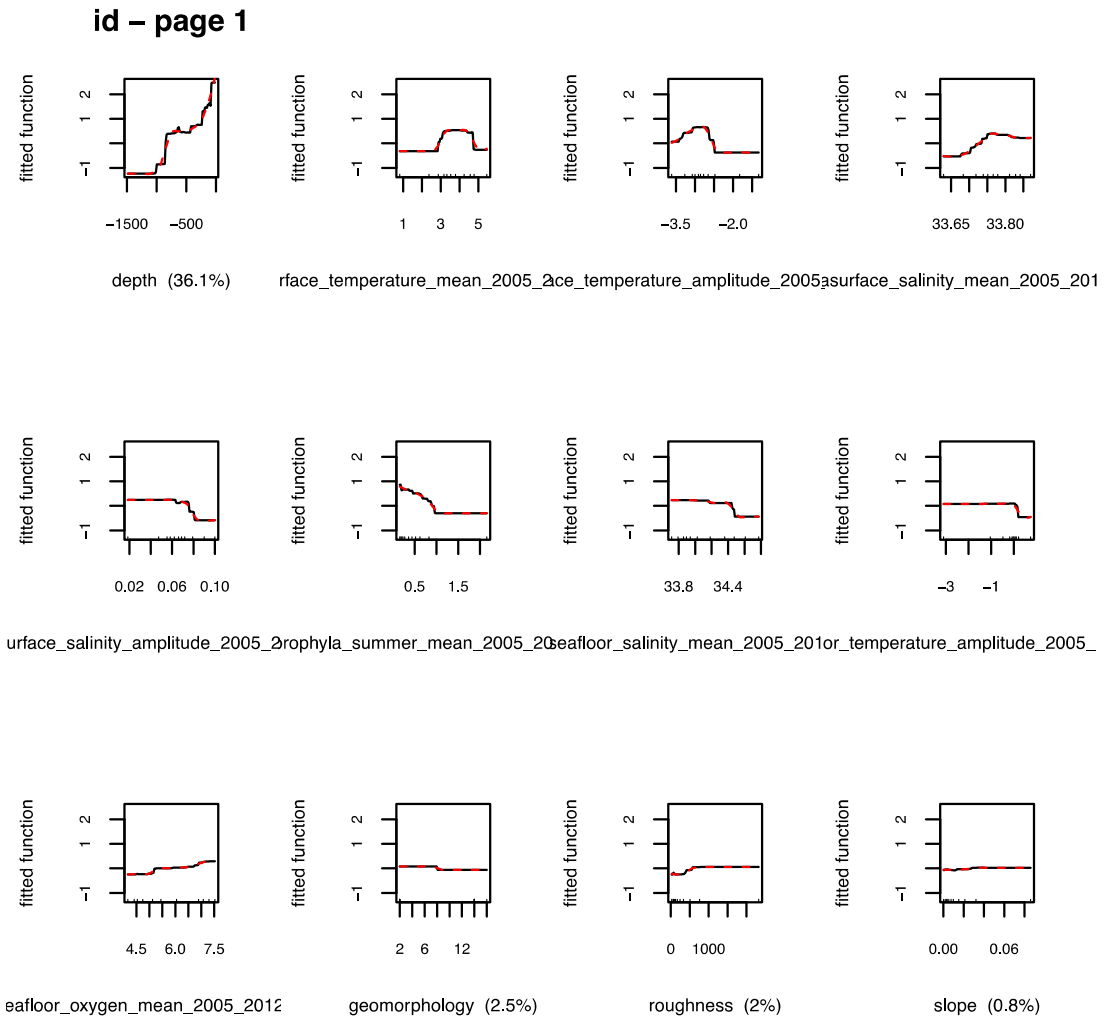


Figure 3: Response plots= Partial dependence plot, that represent how the environmental conditions (x axis) influence the predicted probabilities (y axis)

Interactions between variables

Environmental variables can be related between each other. You can represent these interactions by 3D plots with gbm (Fig. 4).

```
interactions <- gbm.interactions(Cteno_model_2005_2012$response)
head(interactions$rank.list[,c(5,2,4)])
```

```
##   int.size                                 var1.names
## 1     4.41   seafloor_temperature_amplitude_2005_2012
## 2     4.07          chlorophyla_summer_mean_2005_2012
## 3     2.95    seasurface_salinity_amplitude_2005_2012
## 4     2.66 seasurface_temperature_amplitude_2005_2012
## 5     2.10 seasurface_temperature_amplitude_2005_2012
## 6     1.53                                  roughness
##                              var2.names
## 1                                 depth
## 2                                 depth
## 3                                 depth
## 4                                 depth
## 5 seasurface_temperature_mean_2005_2012
## 6                                 depth
```

```
# Plot interactions in 3D (Fig. 4)
gbm.perspec(Cteno_model_2005_2012$response,interactions$rank.list[1,1],
            interactions$rank.list[1,3], cex.lab=0.6, cex.axis=0.6,par(mar=c(0,0,0,0)))
```

Binarize model predictions

Instead of representing model predictions by a distribution probability contained between 0 and 1, you can threshold these probabilities into suitable/unsuitable areas for species distribution. Several thresholds exist, up to you to choose the best one (Liu et al. 2013). We will give here the example of the MaxSSS (Maximum Sensitivity plus Specificity) threshold that is adapted to work with presence-only datasets (Liu et al. 2013).

```
Cteno_model_2005_2012$eval.stats$maxSSS
```

```
## [1] 0.5154886
```

```
maxSSS <- Cteno_model_2005_2012$eval.stats$maxSSS

# Plot binary map predictions (Fig. 5)
plot(Cteno_model_2005_2012$raster.prediction, col=c("lightblue","red"), breaks=c(0, maxSSS ,1),
     main="Projection for [2005-2012]",
     cex.axis= 0.7,
     legend.width=0.5, legend.shrink=0.25,
     legend.args=list(text='Distribution probability', side=3, font=2, cex=0.7))
points(worldmap, type="l")
```

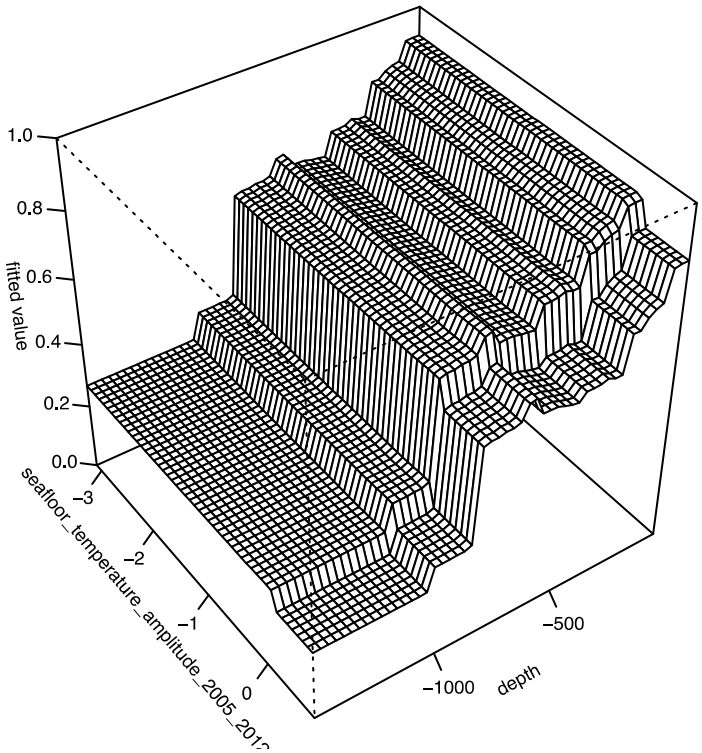


Figure 4: Detail of the interaction between two environmental descriptors: seasurface temperature amplitude and seasurface temperature mean

Model statistics: assess model performance and accuracy

Several statistics enable to evaluate a model. Please refer to literature for further details on each of these statistics (Fielding and Bell 1997, Allouche et al. 2006, Elith et al. 2006).

- **Area under the Curve (AUC)**
- **True Skill Statistics (TSS)**
- **Biserial Pearson Correlation (COR)**

You can also have a look at https://ipa-tys.github.io/ROCR/articles/ROCR.html

```
# To obtain these statistics, go into the statistic part of your model
Cteno_model_2005_2012$eval.stats$AUC #AUC
```

```
## [1] 0.959616
```

```
Cteno_model_2005_2012$eval.stats$TSS # TSS
```

```
## [1] 0.6515557
```

```
Cteno_model_2005_2012$eval.stats$COR # COR
```

```
## [1] 0.8299933
```

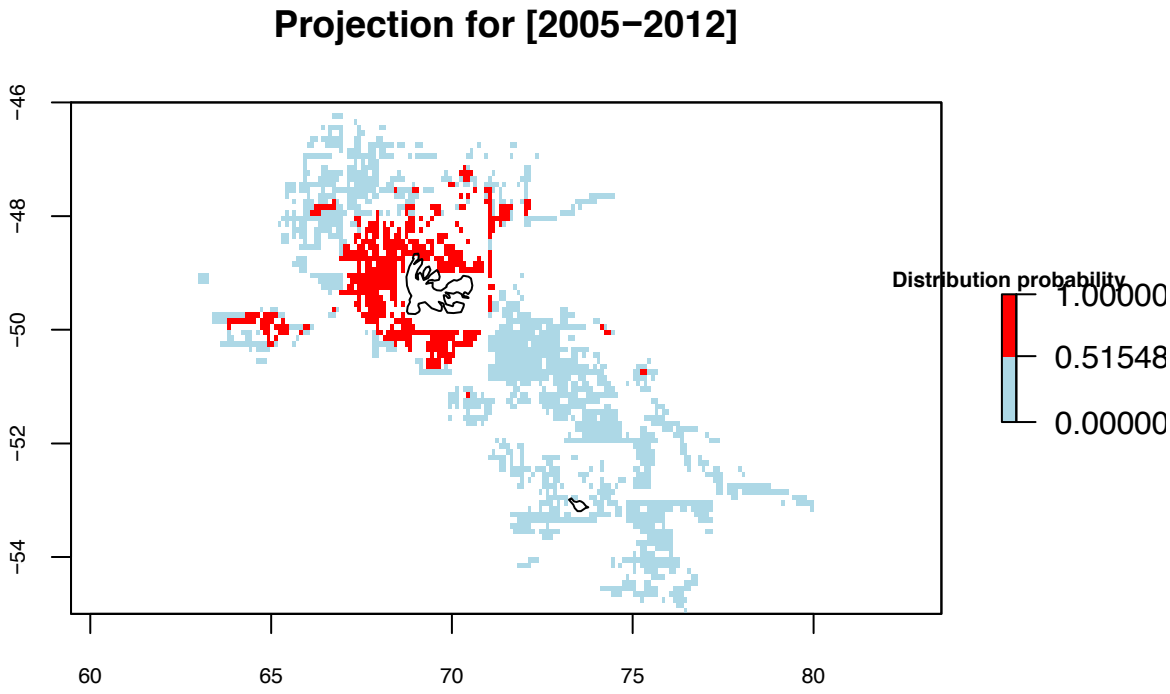


Figure 5: Predicted distribution probabilities for Ctenocidaris nutrix, for [2005-2012], using BRT, represented by suitable (red) or unsuitable (blue) areas.

- **Correctly classified test data (%)**
  Before running your model you can split your dataset into two subsets: one for training your model and the remaining data, that you independently keep, to test it. Once your model has run, you can extract the distribution probabilities predicted by the model at the location of your test data. You can then evaluate the proportion of test data for which the distribution probability is contained into the suitability threshold you have chosen (i.e. MaxSSS here). You can compile this with the following lines:

```
# Run your model with a subset of your presence-only records
# There are several methods to split your dataset into several subsets, the example of
# a simple random splitting (70% training/30% test) will be given here.
# But see Tutorial #4 for more examples.

library(raster)
library(dismo)
library(SDMPlay)
data("ctenocidaris.nutrix")
ctenocidaris.nutrix.occ <- ctenocidaris.nutrix[,c(7,8)] # longitude (first column),
                                                        # latitude (second column)
# Split your dataset into test and training subsets
idx <- sample.int(nrow(ctenocidaris.nutrix.occ),
                  size= round(nrow(ctenocidaris.nutrix.occ)*70/100), replace=F )
presence_data <- ctenocidaris.nutrix.occ[idx,]
test_data <- ctenocidaris.nutrix.occ[-idx,]

# plot training and test data on top of the bathymetry layer (Fig. 6)
data(predictors2005_2012)
data("worldmap")

library(RColorBrewer)
my.palette.blue <- rev(brewer.pal(n = 9, name = "Blues"))

plot(subset(predictors2005_2012,1), col=my.palette.blue)
points(worldmap, type="l")
points(presence_data, pch=20, col="pink")
points(test_data, pch=20, col="green")
legend("bottomright", pch=20, col=c("pink", "green"), legend=c("training","test"), bg="white")
```

```
# Build your model as precedently, except that you will use only your subset of presence data
# (i.e. your training subset)
predictors2005_2012_1500m <- SDMPlay:::delim.area(predictors2005_2012, longmin=62,
                                                  longmax=80, latmin=-55, latmax=-45,
                                                  interval=c(0,-1500))
SDMtable_ctenocidaris_1500 <- SDMPlay:::SDMtab(xydata=presence_data,
       predictors=predictors2005_2012_1500m,
       unique.data=FALSE,
       same=TRUE)
background.occ_1500 <- subset(SDMtable_ctenocidaris_1500,SDMtable_ctenocidaris_1500$id==0)[,c(2,3)]

Cteno_model_2005_2012 <- SDMPlay:::compute.brt(x=SDMtable_ctenocidaris_1500,
                                               proj.predictors=predictors2005_2012_1500m,
                                               tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

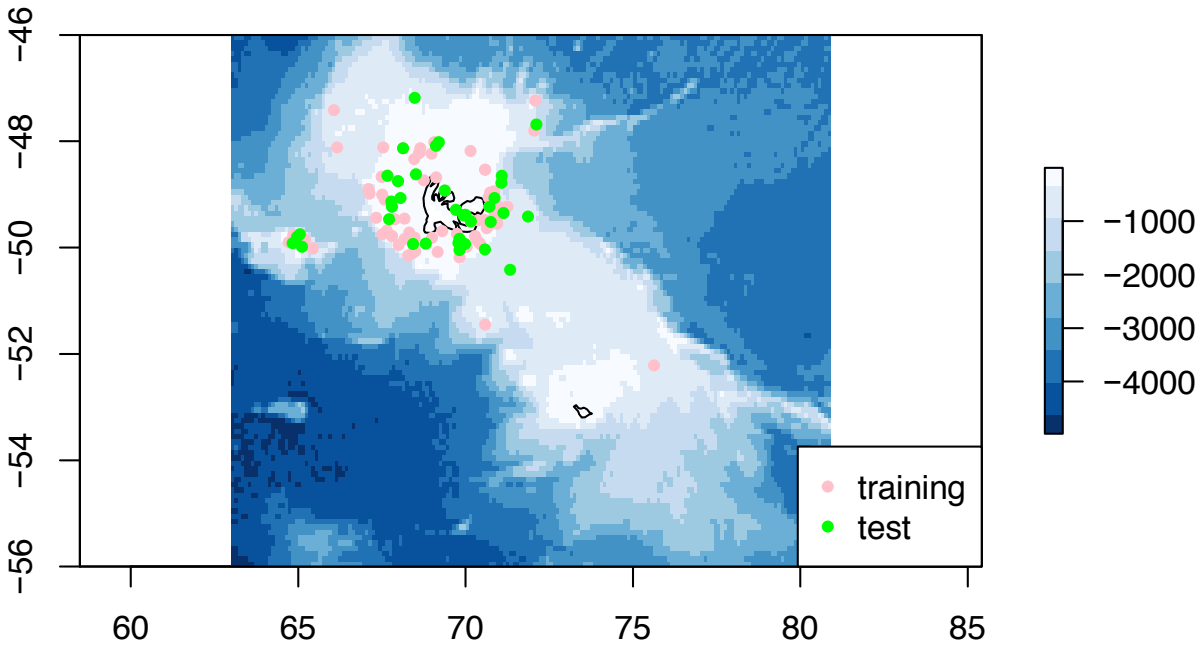


Figure 6: Plot created subsets of training and test data on top of the bathymetry layer

```
# you can then plot your predictions and the test data subset on top of that (Fig.7)

palettecolor <- colorRampPalette(c("deepskyblue", "darkseagreen","lightgreen","green",
                                   "yellow","gold","orange", "red","firebrick"))(100)
plot(Cteno_model_2005_2012$raster.prediction,col=palettecolor, main="Projection for [2005-2012]",
     cex.axis= 0.7,
     legend.width=0.5, legend.shrink=0.25,
     legend.args=list(text='Distribution probability', side=3, font=2, cex=0.8))
points(worldmap, type="l")
points(test_data, pch=20, col="black")
```

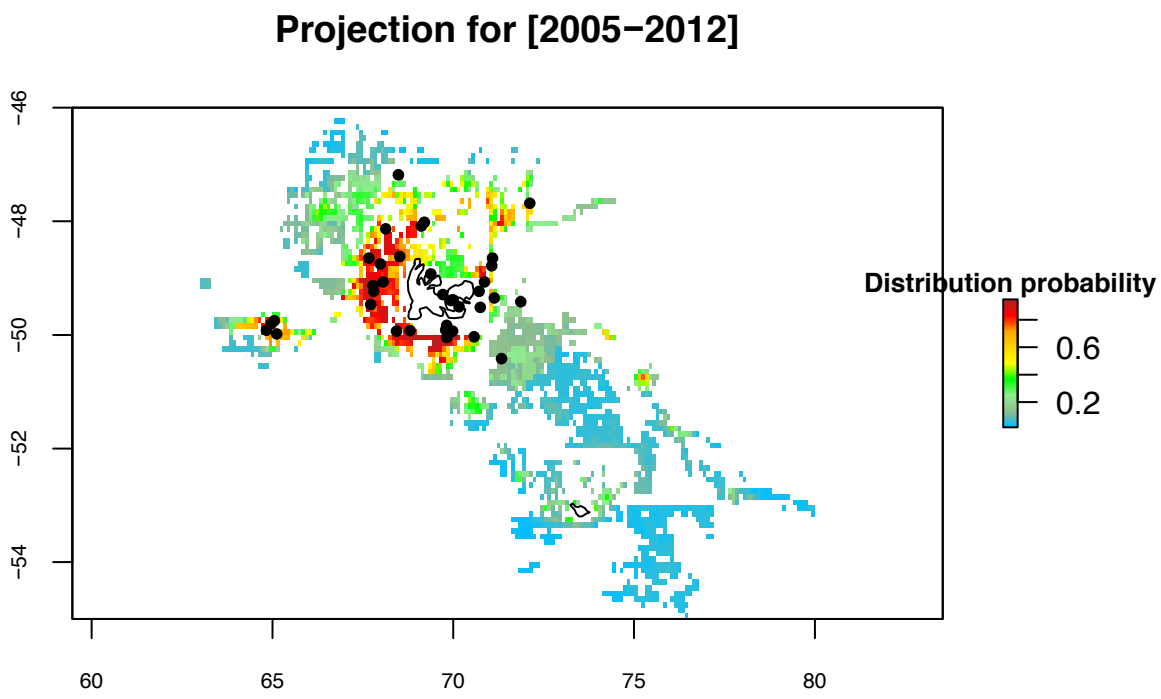


Figure 7: Similar model predictions as for Figure 1, except that only your subset of presence data is used to generate the model and test data (black dots) are presented on top of predictions

```
# Measure the proportion of test data that correctly fall into "suitable areas"
# (suitability defined by the MaxSSS threshold)

values_test_pres <- extract(Cteno_model_2005_2012$raster.prediction,test_data)
maxSSS <- Cteno_model_2005_2012$eval.stats$maxSSS
100*length(which(values_test_pres>maxSSS))/(length(values_test_pres)
                                            -length(which(is.na(values_test_pres))))
```

```
## [1] 66.66667
```

```
# -> you actually assess the proportion of test data that fall on a predicted "suitable"
#    probability, over the total number of test data that do not fall on a NA pixel
```

Null model comparisons

Another way of interpreting your model results are to compare them with results from a null model. Null models are defined in (Raes and ter Steege 2007, van Proosdij et al. 2016) and also applied in Guillaumot et al. (2018). The principle is to generate (case #1) a model that simulates probability distribution when following the presence record sampling bias or (case #2) a model that simulates presence data sampled at random over the entire study area.

A null model #2 is expected to produce distribution maps of equal suitability over the entire study area. If sampling is spatially biased, it is expected that the null model would deviate from the predictions of the standard model (Raes and ter Steege 2007) and of the null model #1. Have a look at the 'null.model' function contained in this package for more details and examples.

References

Allouche, O., Tsoar, A. & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223-1232.

Elith, J., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., ... & A Loiselle, B. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151.

Elith, J. & Leathwick, J.R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.

Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 38-49.

Guillaumot, C., Martin, A., Eléaume, M. & Saucède, T. (2018). Methods for improving species distribution models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau. *Marine Ecology Progress Series*, 594, 149-164.

Guillaumot, C., Artois, J., Saucède, T., Demoustier, L., Moreau, C., Eléaume, M., ... & Danis, B. (2019). Broad-scale species distribution models applied to data-poor areas. *Progress in Oceanography*, 175, 198-207.

Guillaumot, C., Moreau, C., Danis, B. & Saucède, T. (2020). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*, 188, 102438.

Liu, C., White, M. & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, 40(4), 778-789.

Pearson, R.G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, 50.

Raes, N. & ter Steege, H. (2007). A null-model for significance testing of presence-only species distribution models. *Ecography*, 30(5), 727-736.

van Proosdij, A.S., Sosef, M.S., Wieringa, J.J. & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, 39(6), 542-552.

Tutorial for SDMPlay: 3/ Importance of model calibration

2021-02-26

Species distribution modelling (SDM) has been developed for several years to address conservation issues, to assess the direct impact of human activities on ecosystems and to predict the potential distribution shifts of invasive species (see Elith et al. 2006, Pearson 2007, Elith and Leathwick 2009). SDM relates species occurrences with environmental information and can predict species distribution on their entire occupied space. **This approach has been increasingly applied to Southern Ocean case studies, but requires corrections in such a context, due to the broad scale area, the limited number of presence records available and the spatial and temporal aggregations of these datasets.**

SDMPlay is a pedagogic package that will allow you to compute SDMs, to understand the overall method, and to produce model outputs. The package, along with its associated vignettes, highlights the different steps of model calibration and describes how to choose the best method to generate accurate and relevant outputs. SDMPlay proposes codes to apply a popular machine learning approach, BRT (Boosted Regression Trees) and introduces MaxEnt (Maximum Entropy). It contains occurrences of marine species and environmental descriptor datasets as examples associated with several vignette tutorials.

Objectives of tutorial #3/ Importance of model calibration When generating a model, an ensemble of elements should be checked to properly implement the model and generate the most accurate predictions. This implies to perform several tests to ensure that the method is appropriate to each modelled case study. This tutorial aims at illustrating the influence of the number of presence records on predictions and the influence of georeferencing errors that may be contained within presence datasets. It also highlights the importance of selecting the number and location of background data accordingly to each case study. An example of the influence of spatial resolution of environmental descriptors on model outputs is also given. Finally, an analysis to illustrate the consequences of the choice of BRT parameters is presented.

See also...

- **Tutorial #1/ Compute Species Distribution Models**
  Focusses on data structure, data preparation and general model computing.
- **Tutorial #2/ SDM outputs**
  Presents the main outputs you can generate with your SDM.
- **Tutorial #4/ Spatial cross-validation**
  Cross-validation is a method to validate your model. When working with presence data spatially aggregated, the cross-validation procedure should be adapted. This tutorial provides some elements to apply this method (refering to Guillaumot et al. 2019).
- **Tutorial #5/ Spatial extrapolation**
  Models can extrapolate when projected on broad scale areas. This tutorial provides codes to calculate extrapolation scores and generate extrapolation maps that could be associated to SDM maps (refering to Guillaumot et al. 2020).

Let's start with Tutorial #3 !...

Influence of the number of presence data and georeferencing errors on model outputs

The number of presence records used to generate a SDM is crucial since it characterises the potential of the dataset to describe the species occupied space (i.e. the ability of accurately representing the species ecology).

The following example highlights the influence of the number of presence-only records on model outputs. The distribution of the sea urchin *Ctenocidaris nutrix* is modelled using either the entire dataset available (125 records) or a subset of this dataset (60% of these data = 75 records).

```
library(SDMPlay)
#open occurrence records
data("ctenocidaris.nutrix") # Species distributed on the Kerguelen Plateau,
                            # table with longitude and latitude only
all_data <- ctenocidaris.nutrix[,c("decimal.Longitude","decimal.Latitude")]

#subset a part of the dataset
data_60percent <- all_data[sample(nrow(all_data), nrow(all_data)*60/100, replace=F),]
```

```
# open environmental layers
library(raster)
data("worldmap")
data(predictors2005_2012)
names(predictors2005_2012)
```

```
##  [1] "depth"
##  [2] "seasurface_temperature_mean_2005_2012"
##  [3] "seasurface_temperature_amplitude_2005_2012"
##  [4] "seafloor_temperature_mean_2005_2012"
##  [5] "seafloor_temperature_amplitude_2005_2012"
##  [6] "seasurface_salinity_mean_2005_2012"
##  [7] "seasurface_salinity_amplitude_2005_2012"
##  [8] "seafloor_salinity_mean_2005_2012"
##  [9] "seafloor_salinity_amplitude_2005_2012"
## [10] "chlorophyla_summer_mean_2005_2012"
## [11] "geomorphology"
## [12] "sediments"
## [13] "slope"
## [14] "seafloor_oxygen_mean_2005_2012"
## [15] "roughness"
```

```
# create your input matrices
SDMtable_all_data<- SDMPlay:::SDMtab(xydata=all_data,predictors=predictors2005_2012,
                                     unique.data=FALSE,same=TRUE)

SDMtable_data_60percent<- SDMPlay:::SDMtab(xydata=data_60percent,
                                           predictors=predictors2005_2012,
                                           unique.data=FALSE,same=TRUE)
```

```
# Launch models
Model_all_data <- SDMPlay:::compute.brt(x=SDMtable_all_data,
                                        proj.predictors=predictors2005_2012,
                                        tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

```
Model_data_60percent <- SDMPlay:::compute.brt(x=SDMtable_data_60percent,
                                              proj.predictors=predictors2005_2012,
                                              tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

```
# Compare maps (Fig. 1)
palettecolor <- colorRampPalette(c("deepskyblue", "darkseagreen","lightgreen","green",
                                   "yellow","gold","orange", "red","firebrick"))(100)

par(mfrow=c(1,2))
plot(Model_all_data$raster.prediction, col=palettecolor, zlim=c(0,1),
     main="Projection for all data",
     cex.axis= 0.7,
     legend.width=0.5, legend.shrink=0.5,
     legend.args=list(text='Distribution probability', side=3, font=2))
points(worldmap, type="l")

plot(Model_data_60percent$raster.prediction, col=palettecolor, zlim=c(0,1),
     main="Projection for 60% of data",
     cex.axis= 0.7, legend=FALSE)
points(worldmap, type="l")
```

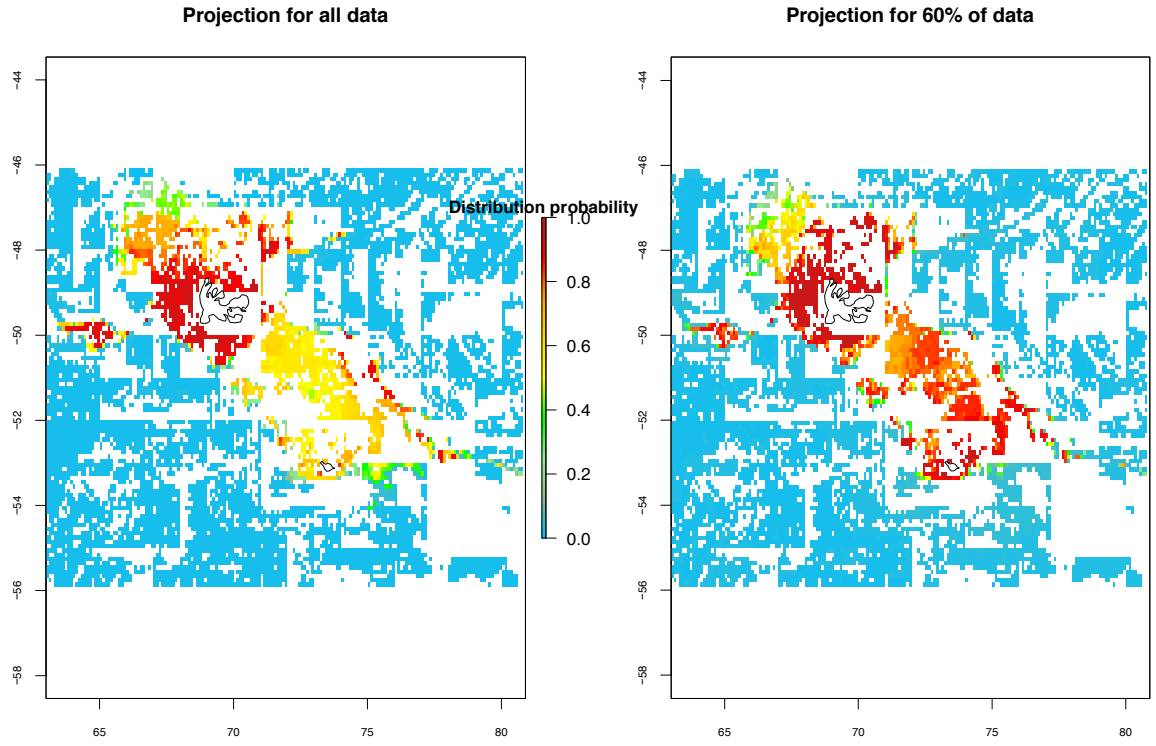


Figure 1: Comparison of model predictions for Ctenocidaris nutrix (2005-2012 environmental conditions), using all presence data available (left) or a subset (60% of data, right).

We can see that model predictions are different when using a different number of presence records to generate the model (Fig. 1). These shifts are expected to be even more pronounced if the presence-only records would not be as aggregated in a same area as they are in this example.

Now, let's generate a second analysis where georeferencing errors are introduced in the dataset (15% of the total presence records are modified, i.e. 19 presence records). Model outputs are compared (Fig. 2). The differents steps are similar to the previous analysis (and skipped from screen).

```r
# Generate random errors in the presence records dataset
all_data <- ctenocidaris.nutrix[,c("decimal.Longitude","decimal.Latitude")]
idx <- sample.int(nrow(all_data), size= round(nrow(all_data)*15/100), replace=F )
data_15percent <- all_data[idx,]
remains <- all_data[-idx,]

perc15_data_errors <- matrix(data = NA, ncol=2, nrow=nrow(data_15percent))
colnames(perc15_data_errors) <- colnames(all_data)

for (i in 1:nrow(data_15percent)){
  perc15_data_errors[i,1] <- data_15percent[i,1]+runif(1,1,4) # random changes
                                # between 1 and 4° in latitude or longitude
  perc15_data_errors[i,2] <- data_15percent[i,2]+runif(1,1,4)
}
all_data_errors <- rbind(perc15_data_errors,remains)
```

```r
# Create your input matrices
SDMtable_data_errors<- SDMPlay:::SDMtab(xydata=all_data_errors,predictors=predictors2005_2012,
                                        unique.data=FALSE,same=TRUE)

# Launch models
Model_data_errors <- SDMPlay:::compute.brt(x=SDMtable_data_errors,
                                           proj.predictors=predictors2005_2012,
                                           tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

```r
# Compare maps (Fig. 2)
par(mfrow=c(1,2))
plot(Model_all_data$raster.prediction,col=palettecolor, zlim=c(0,1),
     main="Projection for all data",
     cex.axis= 0.7,
     legend.width=0.5, legend.shrink=0.25,
     legend.args=list(text='Distribution probability', side=3, font=2))
points(worldmap, type="l")

plot(Model_data_errors$raster.prediction,col=palettecolor, zlim=c(0,1),
     main="With 15% of georeferencing errors",
     cex.axis= 0.7, legend=FALSE)
points(worldmap, type="l")
```

These examples highlight the importance of gathering an exhaustive presence dataset and to check for georeferecing errors within these datasets (because they are often a combinaison between several sources). Model outputs can be more or less influenced by these biases.

Influence of the number and distribution of background data on model outputs

In the context of calibrating SDM based on presence-only records, background records are used to represent the environmental boundaries on which the model will be projected. The choice of the number of presence records has an important influence on model outputs and contrasts between the chosen algorithm (Barbet-Massin et al. 2012). For BRT, it is adviced to use a number of background records as close as possible as the number of available presence records (Barbet-Massin et al. 2012). We will illustrate contrasts between models outputs for models calibrated with 125 or 500 background records.

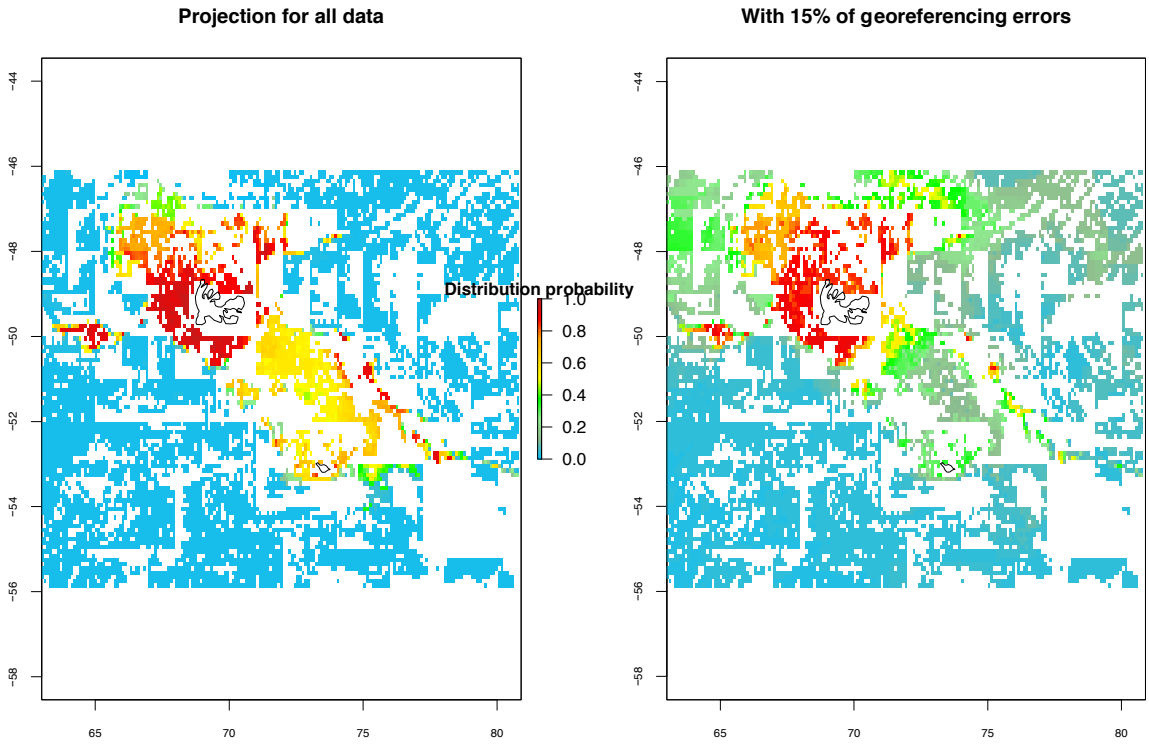


Figure 2: Comparison of model predictions for Ctenocidaris nutrix (2005-2012 environmental conditions), using the original dataset (left) or a dataset with 15% georeferencing errors (right)

```
# Create your input matrices
SDMtable_125bgr<- SDMPlay:::SDMtab(xydata=all_data,
                                   predictors=predictors2005_2012,
                                   unique.data=FALSE, same=TRUE)

SDMtable_500bgr<- SDMPlay:::SDMtab(xydata=all_data,
                                   predictors=predictors2005_2012,
                                   unique.data=FALSE, same=FALSE, background.nb=500)
```

```
# Launch models
Model_125bgr <- SDMPlay:::compute.brt(x=SDMtable_125bgr,
                                      proj.predictors=predictors2005_2012,
                                      tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

```
Model_500bgr <- SDMPlay:::compute.brt(x=SDMtable_500bgr,
                                      proj.predictors=predictors2005_2012,
                                      tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

```
# Compare maps (Fig. 3)
par(mfrow=c(1,2))
plot(Model_125bgr$raster.prediction,col=palettecolor, zlim=c(0,1),
     main="Model 125 background records",
     cex.axis= 0.7,
     legend.width=0.5, legend.shrink=0.25,
     legend.args=list(text='Distribution probability', side=3, font=2))
points(worldmap, type="l")

plot(Model_500bgr$raster.prediction,col=palettecolor, zlim=c(0,1),
     main="Model 500 background records",
     cex.axis= 0.7, legend=FALSE)
points(worldmap, type="l")
```

Another point: background data sampling can be methodologically used to reduce the influence of spatial aggregation of presence records on model outputs (Phillips et al. 2009). Several methods have been developed so far (Phillips et al. 2009, Guillaumot et al. 2018). One of them consists in building a kernel density estimation (KDE) layer to represent the spatial sampling bias, based on visited areas for example (Venables and Ripley 2002). Background data are then sampled according to the weighting scheme of the KDE layer, in order to reduce discrepancies between presence-only records and background data (Phillips et al. 2009, Barbet-Massin et al. 2012, Guillaumot et al. 2018). The comparison between models generated with and without 'targeted' background data samples is illustrated below.

```
# Create a KDE layer
# As explained above, you can use a set of data that describes the pixels
# that have been visited so far or use the distribution of another species
# to generate your KDE.
# We will provide here a simple construction based on the species presence records

KDE_layer <- raster(MASS::kde2d(all_data$decimal.Longitude,
                                all_data$decimal.Latitude,
                                n=c(ncol(predictors2005_2012),
                                    nrow(predictors2005_2012)),
                                lims=c( 63, 81, -56,-46))) # number of pixels and
```

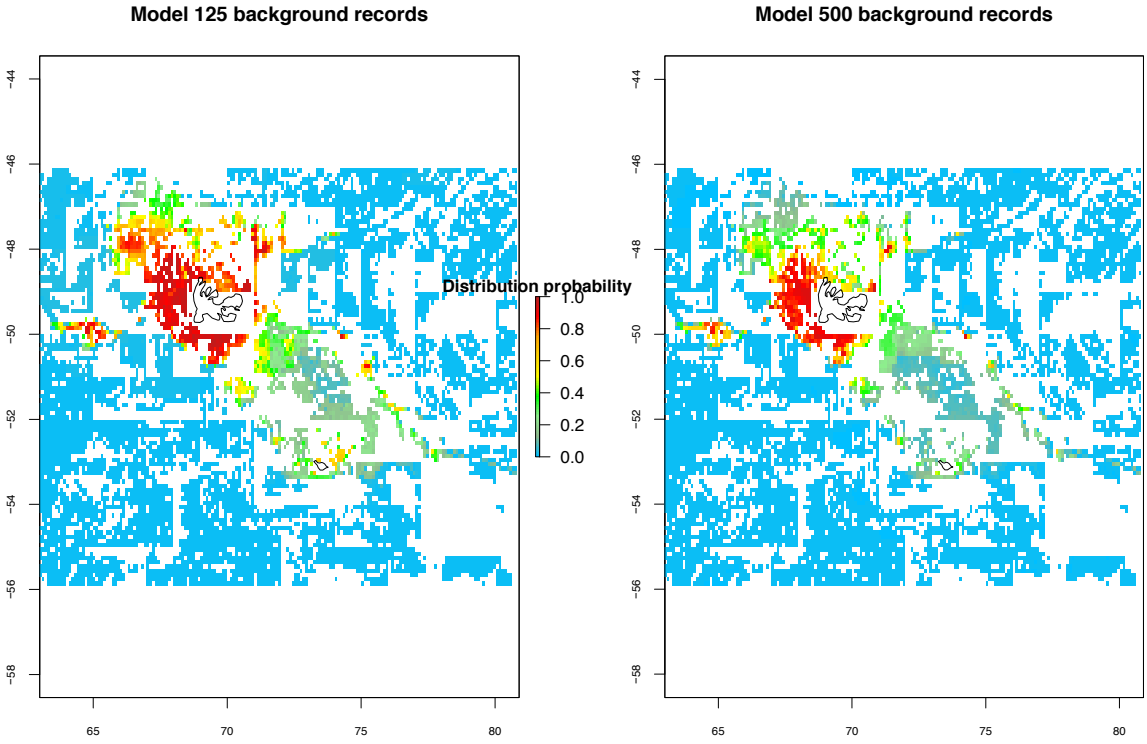


Figure 3: Comparison of model predictions for Ctenocidaris nutrix (2005-2012 environmental conditions), using 125 or 500 background records to calibrate the model. Contrasts between model predictions are clearly important.

```
                                        #extent (latitude/longitude) of the projection area

extent(KDE_layer) <- extent(predictors2005_2012)
KDE_layer <- mask(KDE_layer, subset(predictors2005_2012,1)) # mask the KDE layer by
                                                            # NA values (continental areas)
# Plot the KDE layer (Fig.4)
plot(KDE_layer)
points(all_data, pch=20)
points(worldmap, type="l")
```

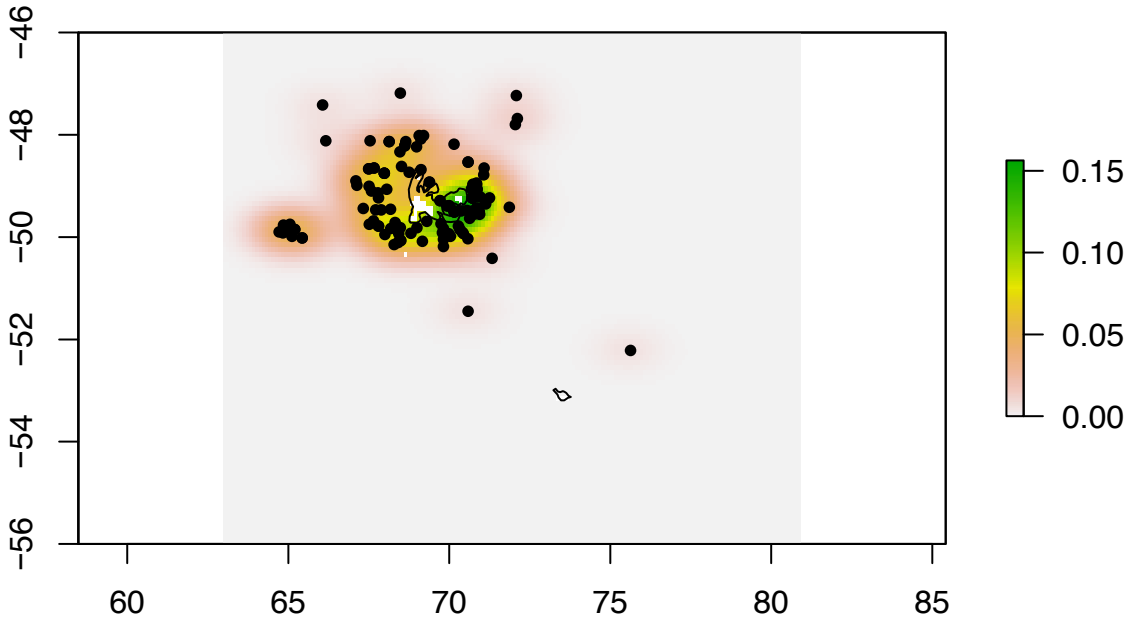


Figure 4: KDE layer

```
# Create your matrice and generate your model
library(SDMPlay)
library(raster)
library(dismo)

SDMtable_KDE<- SDMPlay:::SDMtab(xydata=all_data, predictors=predictors2005_2012,
                                unique.data=FALSE,same=TRUE, KDE=KDE_layer)
background_detail_KDE <- subset(SDMtable_KDE, SDMtable_KDE$id==0)[,c(2,3)]
background_detail <- subset(SDMtable_all_data,SDMtable_all_data$id==0)[,c(2,3)]

Model_KDE <- SDMPlay:::compute.brt(x=SDMtable_KDE,
                                   proj.predictors=predictors2005_2012,
```

```
                                    tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

```
# Compare background records samples (Fig. 5)
bluepalette<-colorRampPalette(c("blue4","blue","dodgerblue", "deepskyblue","lightskyblue"))(800)

par(mfrow=c(1,2))
plot(subset(predictors2005_2012,1), col=bluepalette, main="Background sampling without KDE",
     cex.axis= 0.7)
points(worldmap, type="l")
points(background_detail, pch=20)

plot(subset(predictors2005_2012,1), col=bluepalette, main="Background sampling with KDE",
     cex.axis= 0.7)
points(worldmap, type="l")
points(background_detail_KDE, pch=20)
```

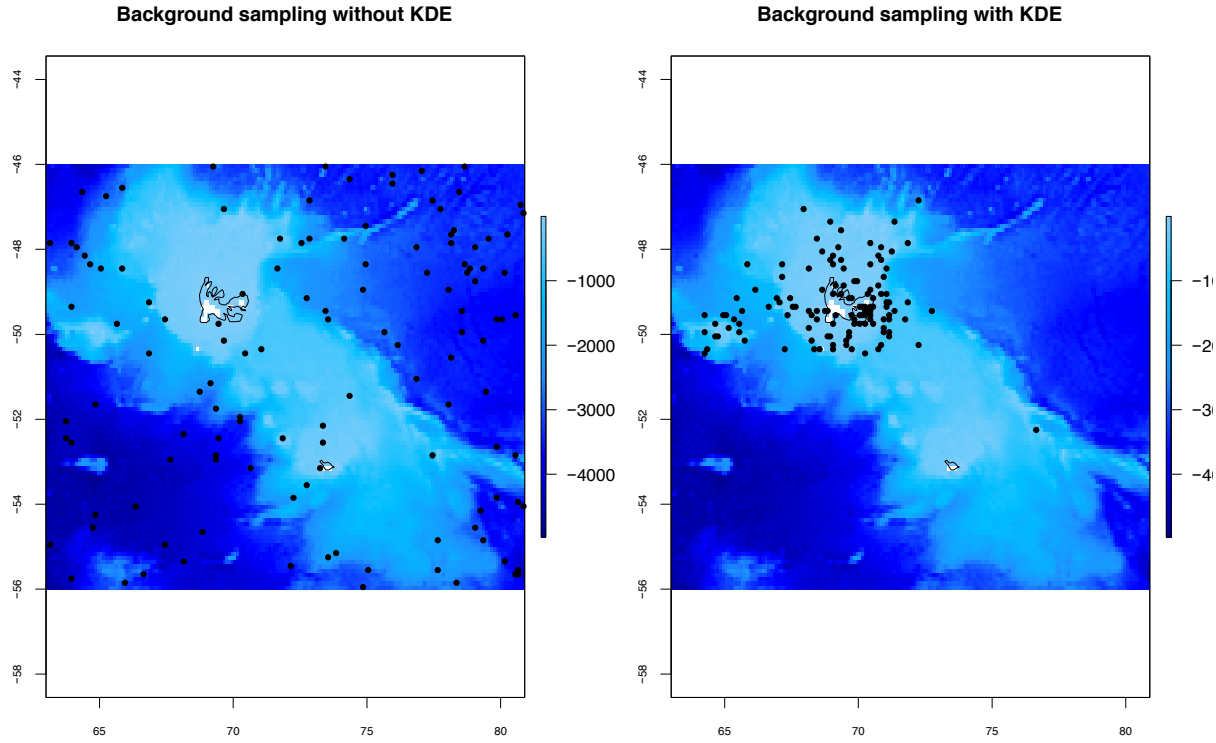


Figure 5: Comparison of background data samplings without (left) or with (right) KDE sampling.

```
# Compare SDM maps (Fig. 6)
par(mfrow=c(1,2))
plot(Model_all_data$raster.prediction, col=palettecolor, zlim=c(0,1), main="Model without KDE",
     cex.axis= 0.7,
     legend.width=0.5, legend.shrink=0.25,
     legend.args=list(text='Distribution probability', side=3, font=2))
points(worldmap, type="l")
```

```
plot(Model_KDE$raster.prediction,col=palettecolor, zlim=c(0,1), main="Model with KDE ",
     cex.axis= 0.7, legend=FALSE)
points(worldmap, type="l")
```

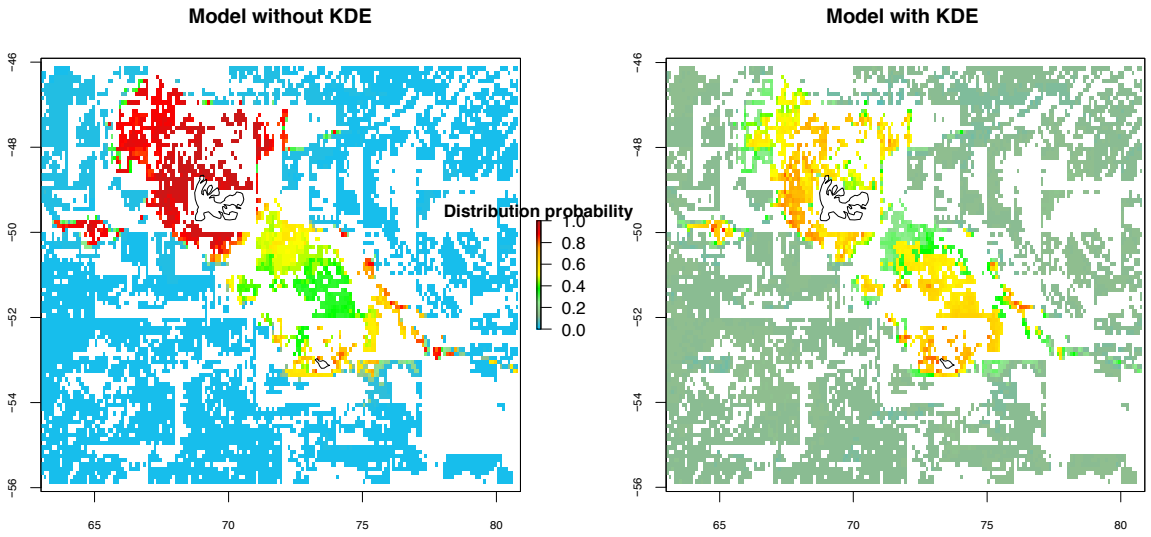


Figure 6: Comparison of model predictions, with background data sampled without (left) or with (right) a KDE scheme

Influence of the spatial resolution of environmental descriptors on model outputs

When generating a SDM, one of the first step is to collect and select the list of environmental descriptors that will be used for the modelling analysis (see Tutorial # 1). The choice of the grid-cell pixel resolution has its importance for prediction accuracy, relevance and interpretation. Let's compare two model outputs if calibrated with environmental descriptors at 0.1° resolution or 10 times coarser.

```
# Create environmental predictors with a spatial resolution 10 times coarser (Fig. 7)
predictors2005_2012_10X <- raster::aggregate(predictors2005_2012, 10, na.rm=T)
plot(subset(predictors2005_2012_10X, c(1:3)))
```

```
# Create the SDMtab matrix
SDMtable_10X<- SDMPlay:::SDMtab(xydata=all_data,
                                predictors=predictors2005_2012_10X,
                                unique.data=FALSE,same=TRUE)
SDMtable_10X[is.nan(as.matrix(SDMtable_10X))]<- NA
```

```
# Launch the model
Model_10X <- SDMPlay:::compute.brt(x=SDMtable_10X,
                                   proj.predictors=predictors2005_2012_10X,
                                   tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

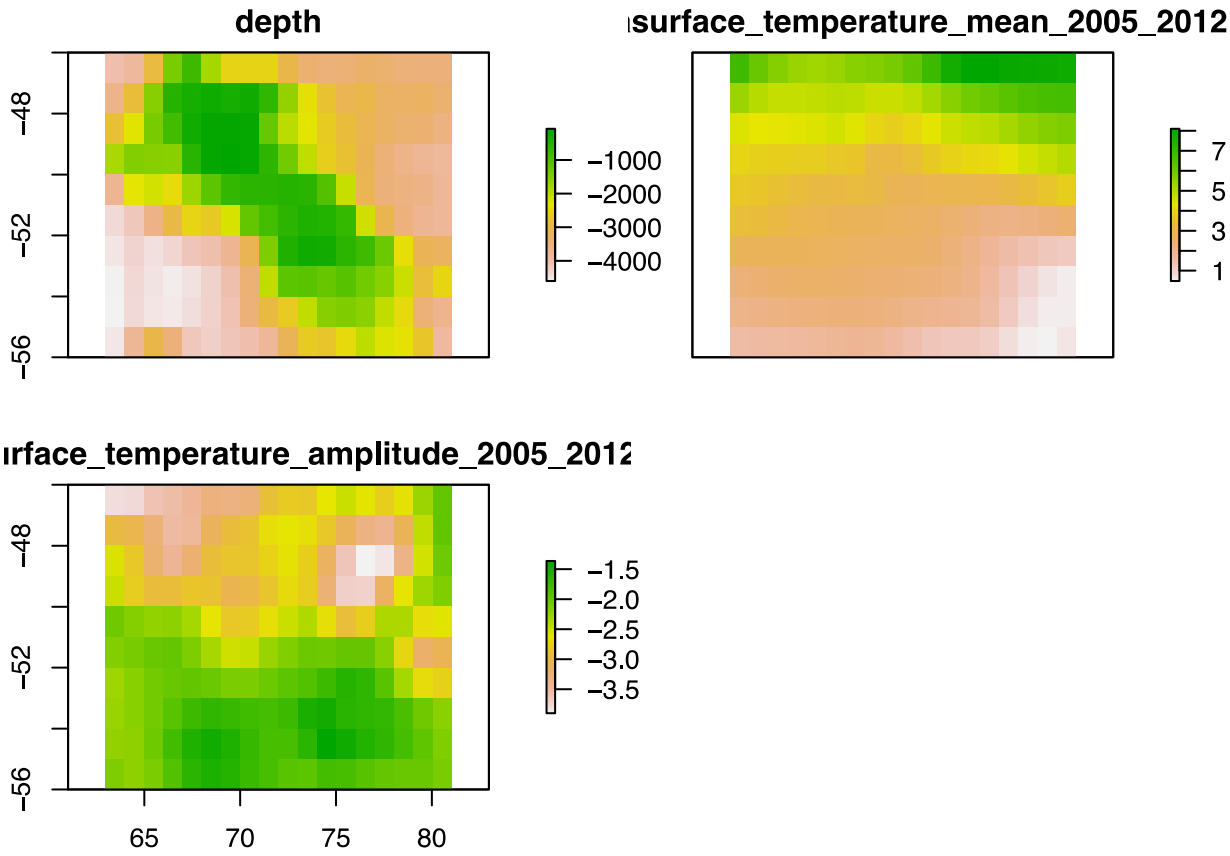


Figure 7: Environmental descriptors, resolution 10 times coarser (1°)

```
# Compare SDM maps (Fig. 8)
par(mfrow=c(1,2))
plot(Model_all_data$raster.prediction, col=palettecolor, zlim=c(0,1),
     main="Model at 0.1° resolution", cex.axis= 0.7,
legend.width=0.5, legend.shrink=0.25,
legend.args=list(text='Distribution probability', side=3, font=2))
points(worldmap, type="l")

plot(Model_10X$raster.prediction, col=palettecolor, zlim=c(0,1),
     main="Model at 1° resolution", cex.axis= 0.7, legend=FALSE)
points(worldmap, type="l")
```

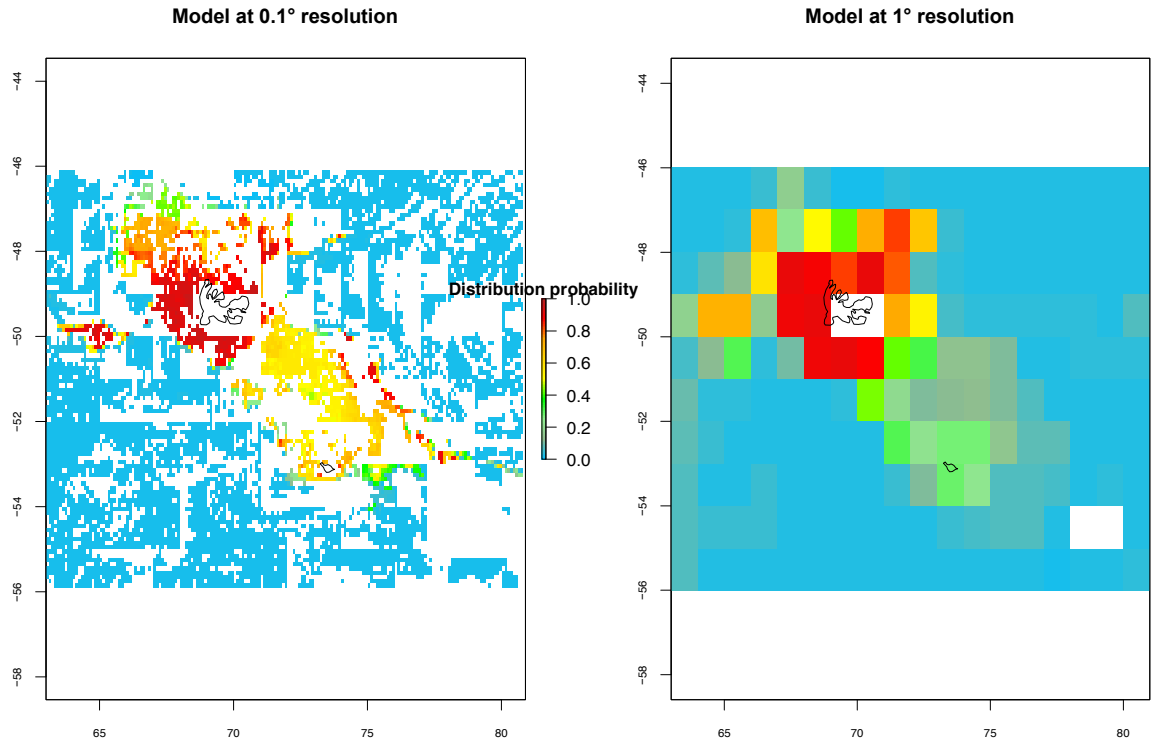


Figure 8: Compare predictions of Ctenocidaris nutrix with environmental descriptors at 0.1° resolution (left) and 1° resolution (right)

Influence of the choice of BRT parameters on model outputs

In Elith et al.(2008), a tutorial is provided to explain the importance of each BRT parameter in model construction (bag fraction bg, learning rate lr, tree complexity tc). A simple analysis can be run during model calibration in order to select the best combination of BRT parameters to generate the most accurate model.

```
library(dismo)
library(gbm)
```

```r
# In order to perform this analysis, we need to go into details of the BRT calculations
# (that are made simpler when you use the compute.brt function). To do so, we will
# rather use the gbm.step function of the dismo package.

# Do several models with several BRT parameters
model1 <- dismo::gbm.step(data = SDMtable_all_data, gbm.x = 4:ncol(SDMtable_all_data),
                          gbm.y = 1, tree.complexity = 2, learning.rate = 0.001,
                          bag.fraction = 0.75, n.trees = 500, step.size = 500)

model2 <- dismo::gbm.step(data = SDMtable_all_data, gbm.x = 4:ncol(SDMtable_all_data),
                          gbm.y = 1, tree.complexity = 3, learning.rate = 0.0001,
                          bag.fraction = 0.75, n.trees = 500, step.size = 500)

model3 <- dismo::gbm.step(data = SDMtable_all_data, gbm.x = 4:ncol(SDMtable_all_data),
                          gbm.y = 1, tree.complexity = 3, learning.rate = 0.0005,
                          bag.fraction = 0.75, n.trees = 500, step.size = 500)

model4 <- dismo::gbm.step(data = SDMtable_all_data, gbm.x = 4:ncol(SDMtable_all_data),
                          gbm.y = 1, tree.complexity = 4, learning.rate = 0.0005,
                          bag.fraction = 0.85, n.trees = 500, step.size = 500)

tree.list1 <- seq(100, model1$gbm.call$best.trees, by = 100)

tree.list2 <- seq(100, model2$gbm.call$best.trees, by = 100)

tree.list3 <- seq(100, model3$gbm.call$best.trees, by = 100)

tree.list4 <- seq(100, model4$gbm.call$best.trees, by = 100)

pred1 <- predict.gbm(model1,SDMtable_all_data, n.trees = tree.list1, "response")
pred2 <- predict.gbm(model2,SDMtable_all_data, n.trees = tree.list2, "response")
pred3 <- predict.gbm(model3,SDMtable_all_data, n.trees = tree.list3, "response")
pred4 <- predict.gbm(model4,SDMtable_all_data, n.trees = tree.list4, "response")

# comparison of predictions with observed values : measure of predicted deviance
graphe.deviance1 <- rep(0,max(tree.list1)/100)
for (i in 1:length(graphe.deviance1)) {
  graphe.deviance1 [i] <- calc.deviance(SDMtable_all_data$id, pred1[,i],calc.mean=T)
}
graphe.deviance2 <- rep(0,max(tree.list2)/100)
for (i in 1:length(graphe.deviance2)) {
  graphe.deviance2 [i] <- calc.deviance(SDMtable_all_data$id, pred2[,i],calc.mean=T)
}
graphe.deviance3 <- rep(0,max(tree.list3)/100)
for (i in 1:length(graphe.deviance3)) {
  graphe.deviance3 [i] <- calc.deviance(SDMtable_all_data$id, pred3[,i],calc.mean=T)
}
graphe.deviance4 <- rep(0,max(tree.list4)/100)
for (i in 1:length(graphe.deviance4)) {
  graphe.deviance4 [i] <- calc.deviance(SDMtable_all_data$id, pred4[,i],calc.mean=T)
}
```

Finally, plot the predicted deviance of each of the case study according to the number of trees necessary to reach the lowest deviance (Fig. 9). This plot will serve as the decision criteria for the best set of parameter to choose: the set for which the predicted deviance is the lowest for the lowest number of trees will be the best (Elith et al. 2008).

```
# Figure 9
plot(tree.list1,graphe.deviance1,xlim = c(-100,10000),
     ylim=c(0,1),type='l', xlab = "Number of trees",
     ylab = "Predictive deviance", cex.lab = 1.5, col="black")
lines(tree.list2,graphe.deviance2,col="blue")
lines(tree.list3,graphe.deviance3,col="red")
lines(tree.list4,graphe.deviance4,col="green",
      xlab = "Number of trees",
      ylab = "Predictive deviance")
legend("topright",legend=c("tc2 lr0.001 bf=0.75","tc3 lr0.0001 bf=0.75",
                           "tc3 lr0.0005 bf=0.75","tc4 lr0.0005 bf=0.85"),
       col=c("black","blue","red", "green"),pch=16,cex=1.5)
```

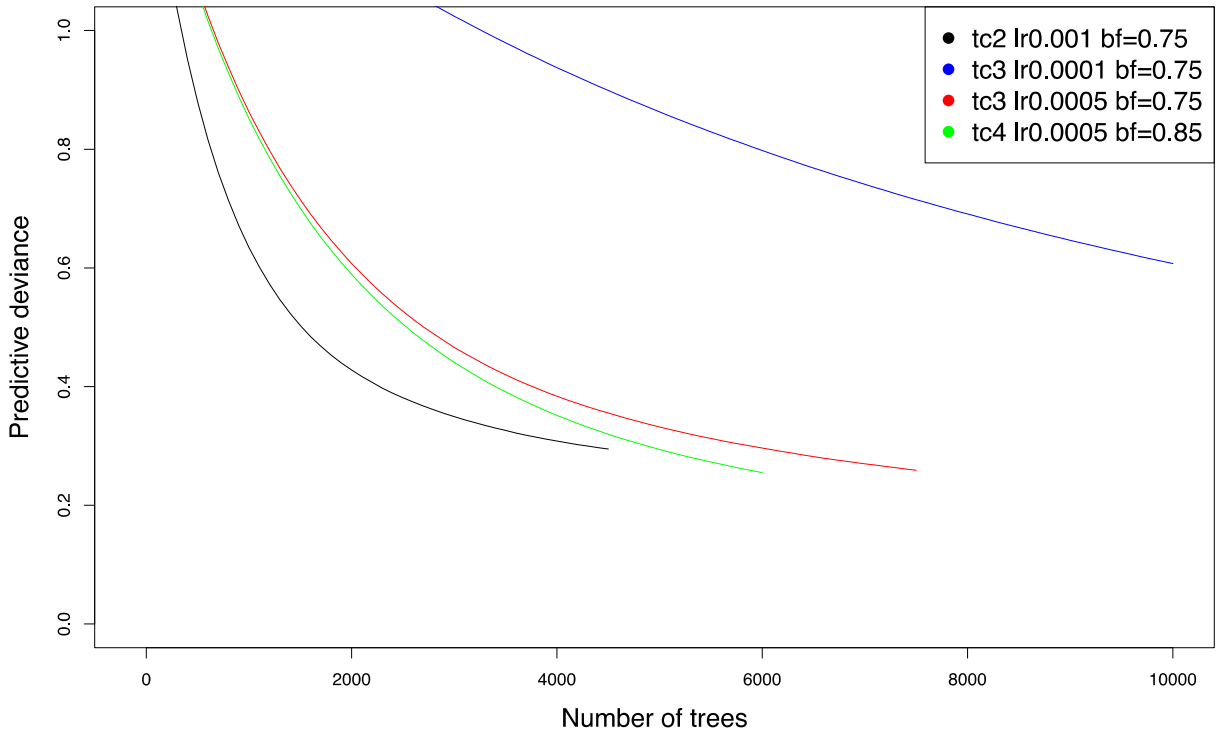


Figure 9: Procedure to choose the best set of BRT parameters (Elith et al. 2008). The set of parameters for which the predictive deviance is the lowest for the minimal number of trees is the optimal

Of course, you should generate more possibilities than these 4 models to figure out the best set of parameters. In this example, the blue line (Fig. 9) has the highest predictive deviance and consequently is not interesting, it also needs a (too) high number of trees to explain model deviance. That is because of the learning rate that has a very low value. Among the three other lines, they all reach more or less the same predicted deviance threshold (around 0.3) but the black line requires less trees to reach it. The best set of BRT parameters

(bg, tc, lr) is therefore the one corresponding to this black line (choosing the green one can also be justified as the predictive deviance is a bit lower). Refer to Elith et al. (2008) for more technical details about the influence of each BRT parameter.

REFERENCES

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3(2), 327-338.

Elith, J., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J Hijmans, R., Huettmann, F., ... & A Loiselle, B. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151.

Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.

Elith, J. & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.

Guillaumot, C., Martin, A., Eléaume, M. & Saucède, T. (2018). Methods for improving species distribution models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau. *Marine Ecology Progress Series*, 594, 149-164.

Guillaumot, C., Artois, J., Saucède, T., Demoustier, L., Moreau, C., Eléaume, M., ... & Danis, B. (2019). Broad-scale species distribution models applied to data-poor areas. *Progress in Oceanography*, 175, 198-207.

Guillaumot, C., Moreau, C., Danis, B. & Saucède, T. (2020). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*, 188, 102438.

Pearson, R.G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, 50.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181-197.

Venables, W.N. & Ripley, B.D. (2002). Random and mixed effects. *In Modern applied statistics with S (pp. 271-300). Springer*, New York, NY.

Tutorial for SDMPlay: 4/ Spatial cross-validation

2021-02-26

Species distribution modelling (SDM) has been developed for several years to address conservation issues, to assess the direct impact of human activities on ecosystems and to predict the potential distribution shifts of invasive species (see Elith et al. 2006, Pearson 2007, Elith and Leathwick 2009). SDM relates species occurrences with environmental information and can predict species distribution on their entire occupied space. **This approach has been increasingly applied to Southern Ocean case studies, but requires corrections in such a context, due to the broad scale area, the limited number of presence records available and the spatial and temporal aggregations of these datasets.**

SDMPlay is a pedagogic package that will allow you to compute SDMs, to understand the overall method, and to produce model outputs. The package, along with its associated vignettes, highlights the different steps of model calibration and describes how to choose the best method to generate accurate and relevant outputs. SDMPlay proposes codes to apply a popular machine learning approach, BRT (Boosted Regression Trees) and introduces MaxEnt (Maximum Entropy). It contains occurrences of marine species and environmental descriptor datasets as examples associated with several vignette tutorials.

Objectives of tutorial #4/ Spatial cross-validation

Cross-validation is a method to evaluate your model. It consists in splitting your initial occurrence dataset into a subset to train the model, and another independent subset to test it. Generally, the random cross-validation procedure randomly partitions the data (e.g. 70% to train, 30% to test). However, when working with presence data spatially aggregated, you may violate the "independency" assumption you made with your test data, as they are spatially very close to the data you have used for training. The evaluation of your model will be consequently biased. In such cases, the spatial cross-validation procedure should be used: the splitting into training and test is not random, but spatially separated. This tutorial provides some elements to apply this method (refering to Guillaumot et al. 2019).

See also...

- **Tutorial #1/ Compute Species Distribution Models**
  Focusses on data structure, data preparation and general model computing.
- **Tutorial #2/ SDM outputs**
  Presents the main outputs you can generate with your SDM.
- **Tutorial #3/ Importance of model calibration**
  Highlights the procedure to accurately calibrate your model and proposes some methods to limit the influence of several biases.
- **Tutorial #5/ Spatial extrapolation**
  Models can extrapolate when projected on broad scale areas. This tutorial provides codes to calculate extrapolation scores and generate extrapolation maps that could be associated to SDM maps (refering to Guillaumot et al. 2020).

Let's begin !

We will work with the *Odontaster validus* case study. Load the occurrence records and environmental layers.

```
library(SDMPlay)
data("Odontaster.validus") # Species distributed around the entire Southern Ocean, table
                           # with longitude and latitude only
#head(Odontaster.validus)
```

```
library(SDMPlay)
library(raster)
data("depth_SO")
data("ice_cover_mean_SO")
data("seafloor_temp_2005_2012_mean_SO")

predictors_stack_SO <- raster::stack(depth_SO,ice_cover_mean_SO,
                                     seafloor_temp_2005_2012_mean_SO)
names(predictors_stack_SO)<-c("depth","ice_cover_mean","seafloor_temp_mean")
predictors_stack_SO
```

```
## class      : RasterStack
## dimensions : 350, 3600, 1260000, 3  (nrow, ncol, ncell, nlayers)
## resolution : 0.1, 0.1  (x, y)
## extent     : -180, 180, -80, -45  (xmin, xmax, ymin, ymax)
## crs        : NA
## names      : depth, ice_cover_mean, seafloor_temp_mean
```

```
 library(RColorBrewer)
 data("worldmap")
 my.palette.oranges <- brewer.pal(n = 9, name = "Oranges")
 my.palette.blue <- rev(brewer.pal(n = 9, name = "Blues"))

#You can load the SOmap package to generate nice figures
#remotes::install_github("AustralianAntarcticDivision/SOmap")
library(SOmap)
```

```
# create your input matrix
SDMtable <- SDMPlay:::SDMtab(xydata=Odontaster.validus,
                             predictors=predictors_stack_SO,
                             unique.data=FALSE,
                             same=TRUE)
head(SDMtable)
```

```
##   id longitude latitude     depth ice_cover_mean seafloor_temp_mean
## 1  1    166.65   -77.85 -112.5210       0.386371                 NA
## 2  1    166.65   -77.85 -112.5210       0.386371                 NA
## 3  1    166.65   -77.85 -112.5210       0.386371                 NA
## 4  1    166.65   -77.85 -112.5210       0.386371                 NA
## 5  1    166.65   -77.85 -112.5210       0.386371                 NA
## 6  1    166.45   -77.45 -384.7778       0.384223                 NA
```

```
# sort presence and background data
presences <- subset(SDMtable,SDMtable$id==1)
head(presences)
```

```
##   id longitude latitude     depth ice_cover_mean seafloor_temp_mean
## 1  1    166.65   -77.85 -112.5210       0.386371                NA
## 2  1    166.65   -77.85 -112.5210       0.386371                NA
## 3  1    166.65   -77.85 -112.5210       0.386371                NA
## 4  1    166.65   -77.85 -112.5210       0.386371                NA
## 5  1    166.65   -77.85 -112.5210       0.386371                NA
## 6  1    166.45   -77.45 -384.7778       0.384223                NA
```

```
background <- subset(SDMtable,SDMtable$id==0)
head(background)
```

```
##     id longitude latitude      depth ice_cover_mean seafloor_temp_mean
## 327  0   -106.55   -72.85  -594.4722     0.74236727                NA
## 328  0     41.95   -60.55 -5343.5557     0.19950417         0.7883283
## 329  0    -51.65   -74.65  -459.8889     0.93878442                NA
## 330  0   -139.85   -46.05 -5115.8613     0.00000000         2.9465933
## 331  0   -144.05   -58.45 -3422.0833     0.00271775         0.8161358
## 332  0    161.65   -49.75 -3771.8333     0.00000000         2.8946247
```

Generate a model with a standard random cross-validation

Randomly select 70% of your presence and background data to train your model and the 30% remaining is kept for testing

```
random_choice <- sample (seq(1:nrow(presences)), size=round((70/100)*nrow(presences)))
remaining <- seq(1:nrow(presences))[-random_choice]
presence_training <- presences[random_choice,]
presence_test <- presences[remaining,]

random_choice2 <- sample (seq(1:nrow(background)), size=round((70/100)*nrow(background)))
remaining_back <- seq(1:nrow(background))[-random_choice2]
background_training <- background[random_choice2,]
background_test <- background[remaining_back,]

SDMtable_training <- rbind(presence_training,background_training)
head(SDMtable_training)
```

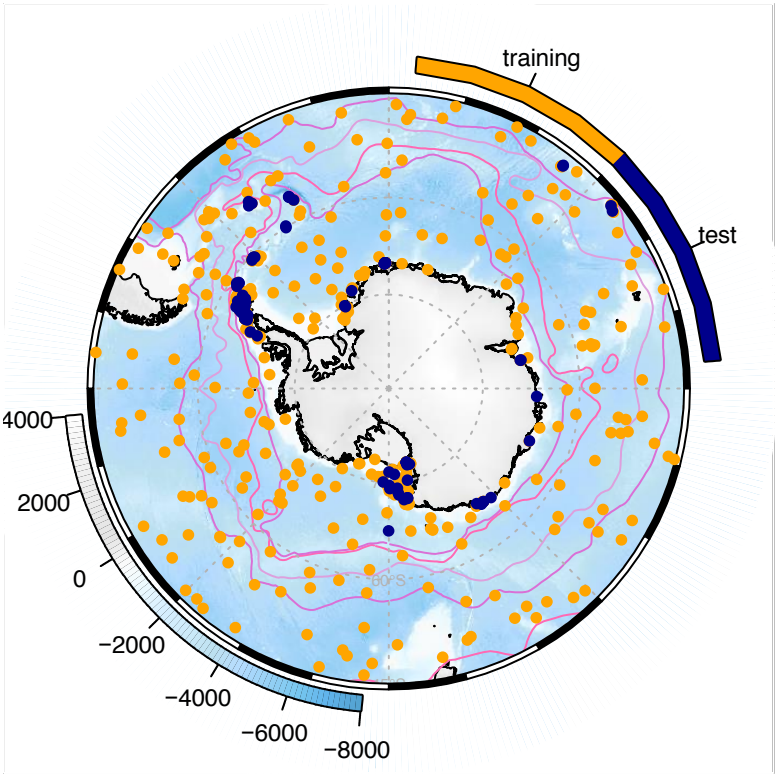
```
##     id longitude latitude      depth ice_cover_mean seafloor_temp_mean
## 221  1    -62.55   -64.25  -57.57156      0.1387166         -0.5271336
## 227  1    -64.15   -64.75 -143.76871      0.2105842          0.4973567
## 2    1    166.65   -77.85 -112.52102      0.3863710                NA
## 119  1    -58.25   -62.95 -390.63889      0.1800755         -1.3562200
## 103  1    -56.15   -62.65 -375.50000      0.2268452         -1.5231029
## 206  1    -67.35   -65.55 -168.61111      0.2568383          0.8559521
```

```
location_presence_test <- presence_test[,c(2,3)]
head(location_presence_test)
```

```
##    longitude latitude
## 1     166.65   -77.85
## 6     166.45   -77.45
```

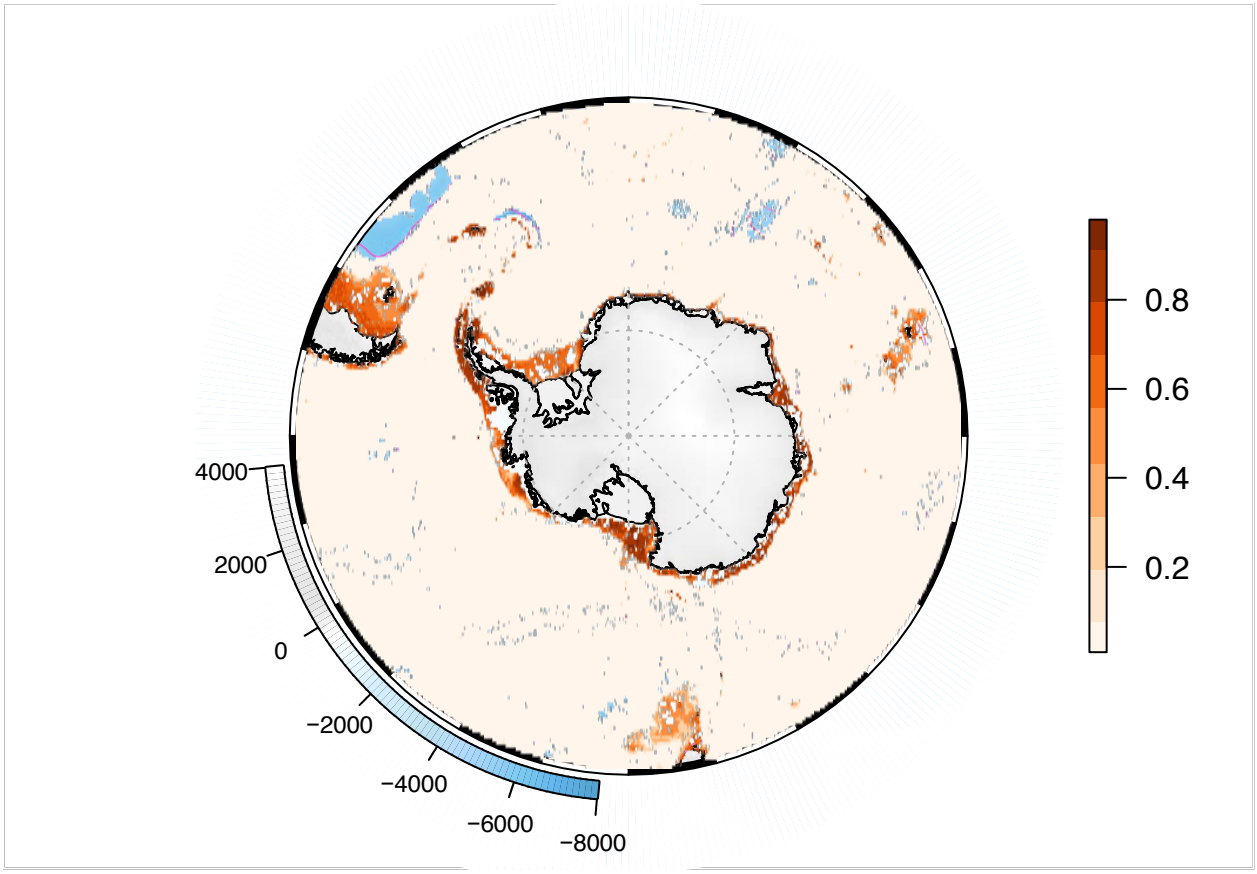
```
## 10    -36.45   -54.25
## 11    -37.45   -54.25
## 13    -45.55   -60.55
## 17    -57.95   -63.35
```

```
# Plot training and test data for the random cross-validation example
basemap <- SOmap(bathy_legend= T, graticules= T, fronts= T, border_width= 0.8)
plot(basemap)
SOplot(SDMtable_training[,2],SDMtable_training[,3], col="orange", pch=20)
SOplot(location_presence_test[,1],location_presence_test[,2], col="darkblue", pch=20)
SOleg(col=c("orange","darkblue"),position = "topright",
      tlabs = c("training","test"), type = "discrete")
```




```
#Run your model
Model_output <- SDMPlay:::compute.brt(x=SDMtable_training,
                                      proj.predictors=predictors_stack_SO,
                                      tc = 2, lr = 0.001, bf = 0.75, n.trees = 500)
```

```
# Plot the result
Model_output_map <- Model_output$raster.prediction
crs(Model_output_map)<-"+proj=longlat + ellps=WGS84"

yy <- SOproj(Model_output_map)
plot(basemap)
plot(yy, col=my.palette.oranges, add=T)
```



Finally, you can evaluate your predictions with your test data. Binarize your predictions with the maxSSS threshold (see Tutorial # 2/ Model outputs) and evaluate if your test data correctly fall into suitable areas.

```
maxSSS <- Model_output$eval.stats$maxSSS

# extract predictions at test data location
extracted_values <- raster::extract(Model_output_map,location_presence_test)
head(extracted_values)
```

```
## [1]        NA        NA 0.9197329 0.9239488        NA        NA
```

```
# compare values with the maxSSS value and evaluate the percentage of
# correctly classified presence test data
100* length (which(na.omit(extracted_values) >= maxSSS)) / length(na.omit(extracted_values))
```

```
## [1] 100
```

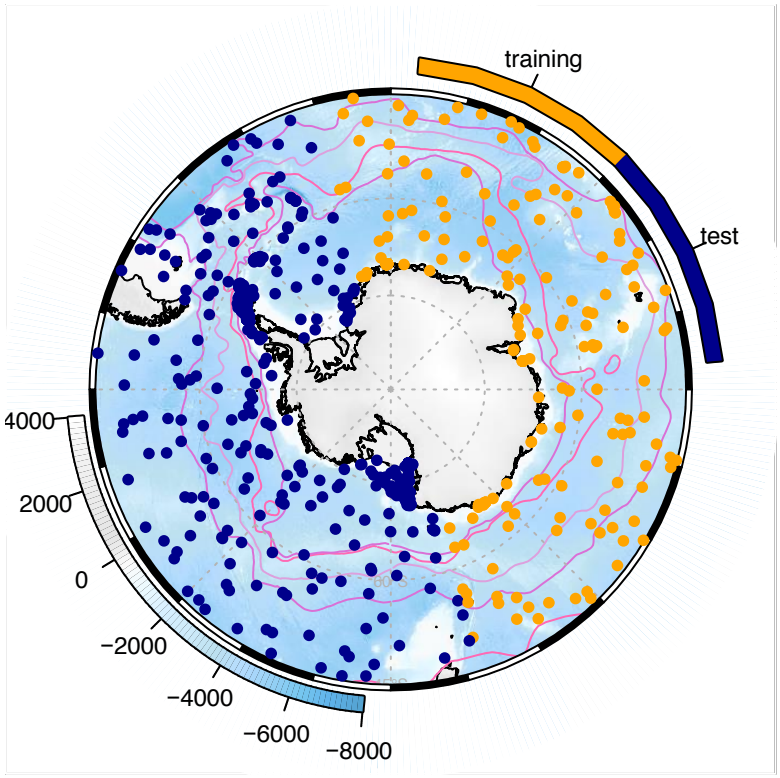
Generate a model with a spatial cross-validation

This time, you will make the training/ test partition with a spatial condition. You can split your environment into 2, 3, 4, 5... areas that will contain either training or test data. See Guillaumot et al. (2019) for further details.

Example of partition into 2 areas

```
  # Tag presence and background datasets with a new variable
  idP <- which(SDMtable$id == 1) #  id of presence data

  partition_function <- SDMPlay:::clock2(SDMtable[idP, c("longitude", "latitude")],
                                         SDMtable [-idP, c("longitude", "latitude")])

  # Generate a variable that will contain the spatial splitting information (factor format)
  MyFold <- rep(NA, nrow(SDMtable)) # training/test data will be labelled 1 or 2

  MyFold[idP] <- partition_function$occ.grp # splitting within the presence data
  MyFold[-idP] <- partition_function$bg.coords.grp # splitting within the background data
```

```
# Plot training and test data
basemap <- SOmap(bathy_legend= T, graticules= T, fronts= T, border_width= 0.8)
plot(basemap)
SOplot(SDMtable[,c(2,3)], col=c("orange","darkblue")[as.factor(MyFold)], pch=20)
SOleg(col=c("orange","darkblue"), position = "topright",
      tlabs = c("training","test"), type = "discrete")
```

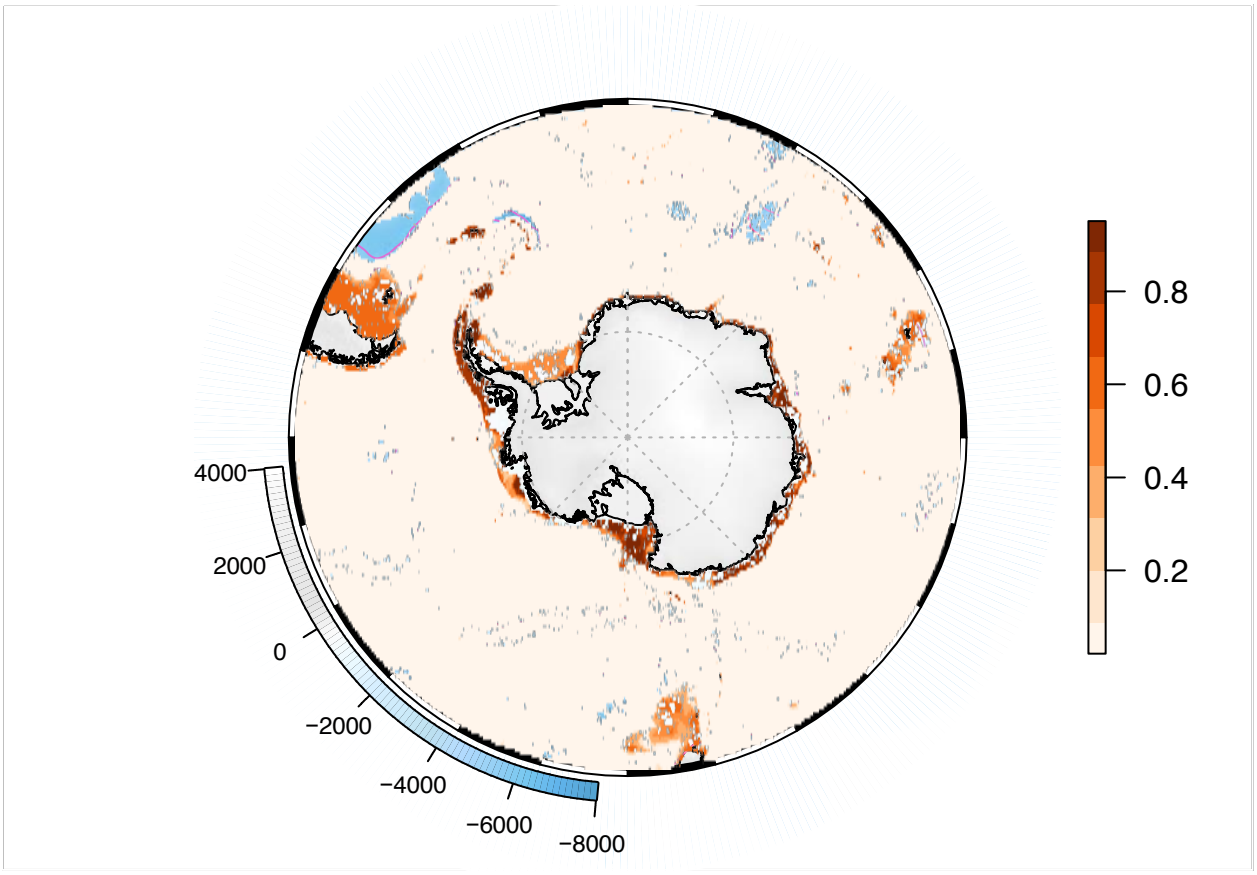


Notice that if you run the code several time, the partition changes: it depends on random selections. You can therefore run the code several time in a loop to make several splitting replicates (see one example below).

```
#Run your model with the splitting, fill the 2 new arguments (n.folds and fold.vector)
# that indicate to the function that you are using a spatial cross-validation procedure
```

```
Model_output <- SDMPlay:::compute.brt(x=SDMtable,
                                      proj.predictors=predictors_stack_SO,
                                      tc = 2, lr = 0.001, bf = 0.75, n.trees = 500,
                                      n.folds = 2,
                                      fold.vector = MyFold)
```

```
# Plot the result
Model_output_map <- Model_output$raster.prediction
crs(Model_output_map)<-"+proj=longlat + ellps=WGS84"

yy <- SOproj(Model_output_map)
plot(basemap)
plot(yy, col=my.palette.oranges, add=T)
```



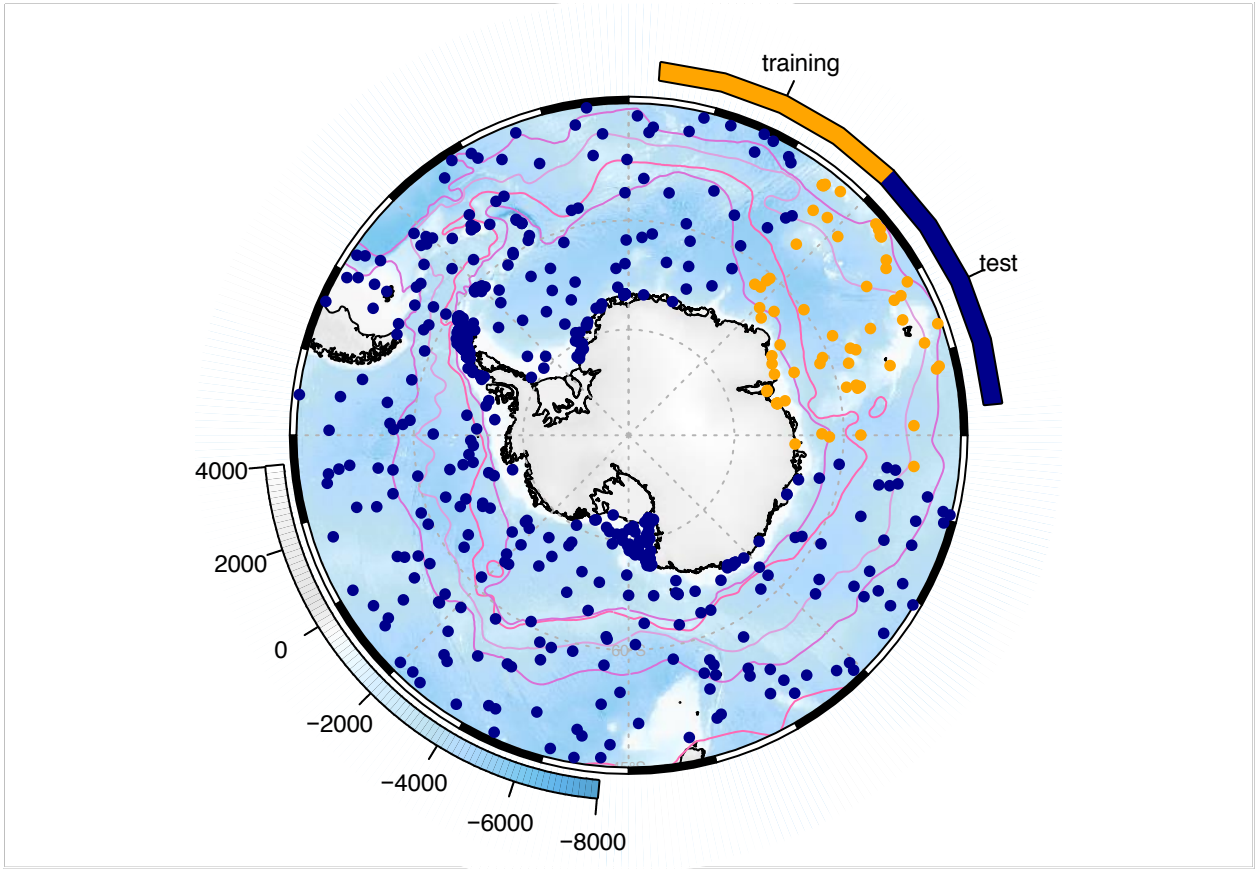

```
# Calculate evaluation scores
maxSSS <- Model_output$eval.stats$maxSSS

# extract predictions at test data location
location_presence_test <- SDMtable[as.factor(MyFold)==1,c(2,3)]
extracted_values <- raster::extract(Model_output_map,location_presence_test)
#extracted_values

# compare the values with the maxSSS value and evaluate the percentage
# of correctly classified presence test data
100* length (which(na.omit(extracted_values) >= maxSSS)) / length(na.omit(extracted_values))
```

```
## [1] 14
```

You can choose and play with 3, 4 or 6 areas !

Several functions to do so are provided in this package (clock3, clock4, clock6). Make some trials. Your results will be more robust with several replicates each time. Don't forget to also apply all the advices given in the previous tutorials (these were not applied here because of computering matters).

```
  idP <- which(SDMtable$id == 1)
  partition_function <- SDMPlay:::clock6(SDMtable[idP, c("longitude", "latitude")],
                                         SDMtable [-idP, c("longitude", "latitude")])

  MyFold <- rep(NA, nrow(SDMtable))
  MyFold[idP] <- partition_function$occ.grp
  MyFold[-idP] <- partition_function$bg.coords.grp


  basemap <- SOmap( bathy_legend = T, graticules = T, fronts = T, border_width = 0.8)
plot(basemap)
SOplot(SDMtable[,c(2,3)],
       col=c("orange","darkblue","darkblue","darkblue","darkblue","darkblue")[as.factor(MyFold)],
       pch=20)
SOleg(col=c("orange","darkblue"), position = "topright",
      tlabs = c("training","test"), type = "discrete")
```



The model will consider at each time that among the 6 areas, one will be used for training and the other

ones for testing. You need to manually transform the labelling of the MyFold if you want 2 or 3 out of the 6 areas to be used for training.

Example of a loop

```
# create an empty stack that you will fill with produced maps during the loop
stack.pred <- subset(predictors_stack_SO, 1);values(stack.pred) <- NA

nb_replicates <- 3
for (i in nb_replicates){
  idP <- which(SDMtable$id == 1)
  partition_function <- SDMPlay:::clock6(SDMtable[idP, c("longitude", "latitude")],
                                         SDMtable [-idP, c("longitude", "latitude")])

  MyFold <- rep(NA, nrow(SDMtable))
  MyFold[idP] <- partition_function$occ.grp
  MyFold[-idP] <- partition_function$bg.coords.grp

  Model_output <- SDMPlay:::compute.brt(x=SDMtable,
                                        proj.predictors=predictors_stack_SO,
                                        tc = 2, lr = 0.001, bf = 0.75, n.trees = 500,
                                        n.folds = 6,
                                        fold.vector = MyFold)

  stack.pred <- stack(stack.pred, Model_output$raster.prediction)
}
```

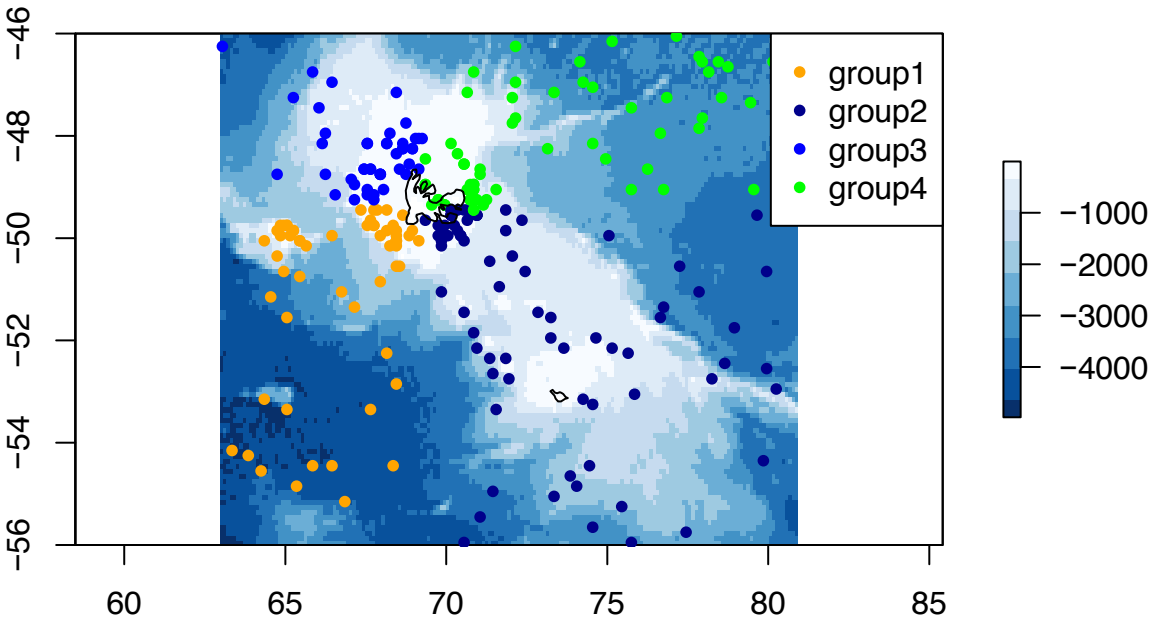
```
# at the end of the loop, calculate the average predictions of your 3 replicates

average_pred <- calc(stack.pred, mean)

# This is a raster layer, you can then plot it as previously !
```

Spatial cross-validation without a Southern Ocean circle shape

If you are working in another area than the entire Southern Ocean and you want to split your occurrence records without this "clock" shape, your can have a look at the ENMeval R package (Muscarella et al. 2014) and their proposed "get.block" function

```
library(SDMPlay)
library(ENMeval)
library(dismo)
library(raster)

data("ctenocidaris.nutrix")
data(predictors2005_2012)
ctenocidaris.nutrix.occ <- ctenocidaris.nutrix[,c(7,8)]

SDMtable_ctenocidaris <- SDMPlay:::SDMtab(xydata=ctenocidaris.nutrix.occ,
                                          predictors=predictors2005_2012,
                                          unique.data=FALSE,
                                          same=TRUE)
```

```
idP <- which(SDMtable_ctenocidaris$id == 1)
partition_function <- ENMeval::get.block(SDMtable_ctenocidaris[idP, c("longitude", "latitude")],
                                         SDMtable_ctenocidaris [-idP, c("longitude", "latitude")])

MyFold <- rep(NA, nrow(SDMtable_ctenocidaris))
MyFold[idP] <- partition_function$occ.grp
MyFold[-idP] <- partition_function$bg.grp

plot(subset(predictors2005_2012,1), col=my.palette.blue)
points(SDMtable_ctenocidaris[,c("longitude", "latitude")],
     col=c("orange","darkblue","blue","green")[as.factor(MyFold)],
     pch=20)
points(worldmap, type="l")
legend("topright", legend=c("group1","group2","group3","group4"),
       col=c("orange","darkblue","blue","green"),
       pch=20, bg = "white")
```



REFERENCES

Elith, J., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J Hijmans, R., Huettmann, F., ... & A Loiselle, B. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151.

Elith, J. & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.

Guillaumot, C., Artois, J., Saucède, T., Demoustier, L., Moreau, C., Eléaume, M., ... & Danis, B. (2019). Broad-scale species distribution models applied to data-poor areas. *Progress in Oceanography*, 175, 198-207.

Guillaumot, C., Moreau, C., Danis, B. & Saucède, T. (2020). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*, 188, 102438.

Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M. & Anderson, R.P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5(11), 1198-1205.

Pearson, R.G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, 50.

Tutorial for SDMPlay: 5/ Model extrapolation

2021-02-26

Species distribution modelling (SDM) has been developed for several years to address conservation issues, to assess the direct impact of human activities on ecosystems and to predict the potential distribution shifts of invasive species (see Elith et al. 2006, Pearson 2007, Elith and Leathwick 2009). SDM relates species occurrences with environmental information and can predict species distribution on their entire occupied space. **This approach has been increasingly applied to Southern Ocean case studies, but requires corrections in such a context, due to the broad scale area, the limited number of presence records available and the spatial and temporal aggregations of these datasets.**

SDMPlay is a pedagogic package that will allow you to compute SDMs, to understand the overall method, and to produce model outputs. The package, along with its associated vignettes, highlights the different steps of model calibration and describes how to choose the best method to generate accurate and relevant outputs. SDMPlay proposes codes to apply a popular machine learning approach, BRT (Boosted Regression Trees) and introduces MaxEnt (Maximum Entropy). It contains occurrences of marine species and environmental descriptor datasets as examples associated with several vignette tutorials.

Objectives of tutorial #5/ Spatial extrapolation

Considering the reference dataset of environmental conditions for which species presence-only records are modelled, extrapolation corresponds to the part of the projection area for which one environmental value at least falls outside of the reference dataset. Because occurrence sampling is limited in the Southern Ocean and that environmental conditions are contrasting in this broad scale area, it was observed that model predictions are actually extrapolation in proportions up to 75% in some cases ! (Guillaumot et al. 2020). It is therefore important to measure extrapolation and to provide extrapolation uncertainty maps along with model predictions. Here, we used the Multivariate Environmental Similarity Surface (MESS) index to quantify model uncertainty associated to extrapolation (Elith et al. 2010), but other approaches exist (Owens et al. 2013).

See also...

- **Tutorial #1/ Compute Species Distribution Models**
  Focusses on data structure, data preparation and general model computing.
- **Tutorial #2/ SDM outputs**
  Presents the main outputs you can generate with your SDM.
- **Tutorial #3/ Importance of model calibration**
  Highlights the procedure to accurately calibrate your model and proposes some methods to limit the influence of several biases.
- **Tutorial #4/ Spatial cross-validation**
  Cross-validation is a method to validate your model. When working with presence data spatially aggregated, the cross-validation procedure should be adapted. This tutorial provides some elements to apply this method (refering to Guillaumot et al. 2019).

Let's begin ! Evaluate extrapolation of your case study!

We will work with the *Ctenocidaris nutrix* case study. Load the occurrence records and environmental layers.

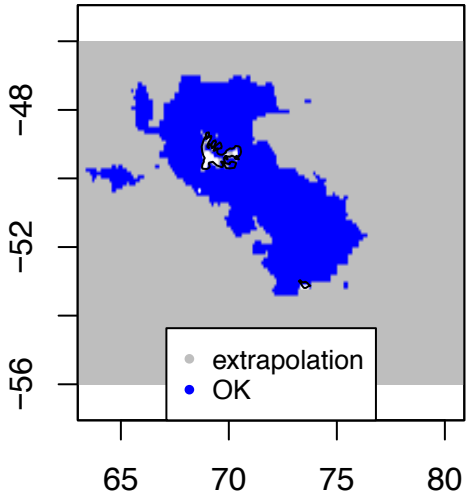
```
library(SDMPlay)
library(raster)
data("ctenocidaris.nutrix")
data(predictors2005_2012)
ctenocidaris.nutrix.occ <- ctenocidaris.nutrix[,c(7,8)]

# we will just keep 3 descriptors for the example
predictors2005_2012 <- raster::subset(predictors2005_2012,c(1:3))
```

```
 library(RColorBrewer)
 data("worldmap")
 my.palette.oranges <- brewer.pal(n = 9, name = "Oranges")
 my.palette.blue <- rev(brewer.pal(n = 9, name = "Blues"))
```

```
library(dismo)
# Measure the extrapolation score (MESS: Multivariate Environmental Similarity Surface)

# Extract environmental values at presence-only records location
envi_scores <- raster::extract(predictors2005_2012, ctenocidaris.nutrix.occ)
#head(envi_scores)

# Calculate MESS score
x <- dismo::mess(predictors2005_2012,envi_scores)

y <- x; values(y)<- values(x)>0
y <- reclassify(y,cbind(FALSE,0)) # MODEL EXTRAPOLATES (MESS <0)
y <- reclassify(y,cbind(TRUE,1)) # MODEL DOES NOT EXTRAPOLATE (MESS >0)

MESS_layer <- mask(y,subset(predictors2005_2012,1)) # remove land pixel (layer #1 = depth)
```

MESS_layer is a RasterLayer with values contained between 0 and 1. The model extrapolates for 0 and does not extrapolate for 1 value.

```
# Plot the result
plot (MESS_layer, col=c("grey","blue"), legend=F)
points(worldmap, type="l")
legend("bottom", legend=c("extrapolation","OK"),
         col=c("grey","blue"),
         pch=20, bg = "white", cex=0.8)
```



You can then overlap this layer to SDM predictions when interpreting your results or preparing your maps for your publications! See examples in Guillaumot et al. (2020).

```
# Calculate the proportion of the area where extrapolation occurs
MESS<- reclassify(MESS_layer,cbind(1,NA))

# compare the number of pixels = 0 to the number of total pixels of the area
length(which(!is.na(values(MESS))))*100 / length(which(!is.na(values(subset(predictors2005_2012,1)))))
```

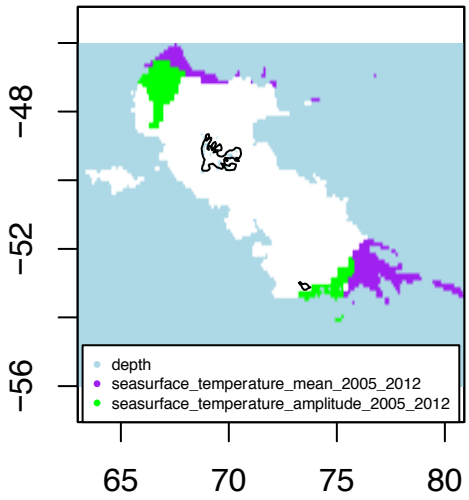
```
## [1] 82.89967
```

Assess which environmental descriptors are responsible for extrapolation at each pixel

```
# create an empty raster to initiate a Rasterstack
stack_amelio_MESS <- subset(predictors2005_2012,1); values(stack_amelio_MESS) <- NA

# Loop to calculate the value of dissimilarity of each environmental descriptor
# For each pixel, it will be determined if extrapolation occurs for each environmental descriptor

for (k in 1:nlayers(predictors2005_2012)){
  presvals <- raster::extract(subset(predictors2005_2012, k),
                              ctenocidaris.nutrix.occ)
  x_amelio <- dismo::mess(subset(predictors2005_2012, k),presvals)
  stack_amelio_MESS <- stack(stack_amelio_MESS,x_amelio)
}
```

```
# Delete the first layer of the stack that was empty (initialization)
stack_amelio_MESS <- dropLayer(stack_amelio_MESS,1)
names(stack_amelio_MESS) <- names(predictors2005_2012)

# Search for the environmental layer that is responsible for the lower MESS score
# (i.e. responsible for the extrapolation)

MESS_amelio <- which.min(stack_amelio_MESS)
MESS_amelio <- mask(MESS_amelio, MESS) # keep only areas where extrapolation occurs
```

```
# Plot the result

plot (MESS_amelio, col=c("lightblue","purple", "green"), legend=F)
points(worldmap, type="l")
legend("bottom", legend=names(predictors2005_2012),
         col=c("lightblue","purple", "green"),
         pch=20, bg = "white", cex=0.5)
```




```
# Calculate the contribution of each environmental descriptor in extrapolation
table_mess_amelio <- matrix(data=NA, nrow = 1, ncol= nlayers(predictors2005_2012))
colnames(table_mess_amelio) <-c("depth","mean_temp","amplitude_temp")

for (k2 in 1:nlayers(predictors2005_2012)){
  table_mess_amelio[1,k2] <- length(which(values(MESS_amelio)==k2))*100/
    length(which(!is.na(values(subset(predictors2005_2012,1)))))
}

table_mess_amelio
```

```
##         depth mean_temp amplitude_temp
## [1,] 78.68054  2.344581       1.874545
```

REFERENCES

Elith, J., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J Hijmans, R., Huettmann, F., ... & A Loiselle, B. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151.

Elith, J. & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330-342.

Guillaumot, C., Artois, J., Saucède, T., Demoustier, L., Moreau, C., Eléaume, M., ... & Danis, B. (2019). Broad-scale species distribution models applied to data-poor areas. *Progress in Oceanography*, 175, 198-207.

Guillaumot, C., Moreau, C., Danis, B. & Saucède, T. (2020). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*, 188, 102438.

Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., ... & Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, 263, 10-18.

Pearson, R.G. (2007). Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, 50.

APPENDIX

PLOS ONE

RESEARCH ARTICLE

A Dynamic Energy Budget (DEB) model to describe *Laternula elliptica* (King, 1832) seasonal feeding and metabolism

Antonio Agüera[1]*, In-Young Ahn[2], Charlène Guillaumot[1], Bruno Danis[1]

1 Laboratoire de Biologie Marine CP160/15. Université Libre de Bruxelles, Brussels, Belgium, **2** Korea Polar Research Institute (KOPRI), Yeonsu-gu, Incheon, Republic of Korea

* antonio.aguera@gmail.com

Citation: Agüera A, Ahn I-Y, Guillaumot C, Danis B (2017) A Dynamic Energy Budget (DEB) model to describe *Laternula elliptica* (King, 1832) seasonal feeding and metabolism. PLoS ONE 12(8): e0183848. https://doi.org/10.1371/journal.pone.0183848

Editor: De-Hua Wang, Institute of Zoology, CHINA

Received: April 16, 2017

Accepted: August 11, 2017

Published: August 29, 2017

Data Availability Statement: All relevant data is contained the publications referred and/or contained within this manuscript and its supplements.

Funding: This work is funded by the project vERSO (from the Belgian Office of Science, BELSPO contract no BR/132/A1/vERSO). This fund covered the time invested by the researchers at the University Libre de Bruxelles on this work. Marian Cove field data used was provided and previously published by author I-YA. Marian Cove data collection was funded by KOPRI (PE17070).

Abstract

Antarctic marine organisms are adapted to an extreme environment, characterized by a very low but stable temperature and a strong seasonality in food availability arousing from variations in day length. Ocean organisms are particularly vulnerable to global climate change with some regions being impacted by temperature increase and changes in primary production. Climate change also affects the biotic components of marine ecosystems and has an impact on the distribution and seasonal physiology of Antarctic marine organisms. Knowledge on the impact of climate change in key species is highly important because their performance affects ecosystem functioning. To predict the effects of climate change on marine ecosystems, a holistic understanding of the life history and physiology of Antarctic key species is urgently needed. DEB (Dynamic Energy Budget) theory captures the metabolic processes of an organism through its entire life cycle as a function of temperature and food availability. The DEB model is a tool that can be used to model lifetime feeding, growth, reproduction, and their responses to changes in biotic and abiotic conditions. In this study, we estimate the DEB model parameters for the bivalve *Laternula elliptica* using literature-extracted and field data. The DEB model we present here aims at better understanding the biology of *L. elliptica* and its levels of adaptation to its habitat with a special focus on food seasonality. The model parameters describe a metabolism specifically adapted to low temperatures, with a low maintenance cost and a high capacity to uptake and mobilise energy, providing this organism with a level of energetic performance matching that of related species from temperate regions. It was also found that *L. elliptica* has a large energy reserve that allows enduring long periods of starvation. Additionally, we applied DEB parameters to time-series data on biological traits (organism condition, gonad growth) to describe the effect of a varying environment in food and temperature on the organism condition and energy use. The DEB model developed here for *L. elliptica* allowed us to improve benchmark knowledge on the ecophysiology of this key species, providing new insights in the role of food availability and temperature on its life cycle and reproduction strategy.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Antarctica includes some of the most challenging habitats on Earth [1]. They are characterised by low temperatures and a very marked seasonality in day length, leading to large variations in ice cover and phytoplankton biomass [2]. Adaptation to such conditions has resulted in organisms generally displaying a poor capacity to cope with temperature elevations [3], yet capable of surviving low-food availability over long periods [4]. These characteristics raise concern about their capacity to face ongoing global climate change. It is now largely accepted that Southern Ocean ecosystems are particularly vulnerable to global warming as some regions are challenged by rapid temperature rise [5,6]. Recent research shows that global warming induces cascading effects, causing a wide variety of changes in the structure and functioning of Antarctic marine ecosystems. The variation in the duration of seasonal sea ice cover, marine-terminating glacier retreat [7], the increase in seasonal ice scouring on sea bottom or highly fluctuating salinity due to glacial melt water introduction [6,8] have for instance been shown to induce changes in key processes for Antarctic ecosystems such as primary production [9] and causing ecosystem structural shifts [10]. Climate change is influencing both physical and biotic components of marine ecosystems, and will have an impact on the distribution and population dynamics of Antarctic marine organisms. Ultimately, life history, distribution and abundance of species reflect the action of metabolic processes in the context of varying environments [11]. To assess the potential effects of climate change on Antarctic benthic marine ecosystems, an in-depth knowledge of metabolic processes is needed. This knowledge will provide a valuable benchmark to quantify species population dynamics, performance, and functional role within a given ecosystem, as well as a ground-truthing ongoing modeling efforts [12].

Dynamic Energy Budget (DEB) theory provides first-principle models describing the processes of energy and matter-uptake and their use for maintenance, development, growth and reproduction of a broad range of organisms [13,14]. DEB theory allows establishing links between the physiology of a model organism and its environment by capturing all the metabolic processes of the organisms through their life cycle as function of matter-uptake and temperature [14]. Derived from DEB theory, DEB models [14] can describe the underlying physiological processes based on first principles (e.g. mass-energy conservation laws, linkage of processes to volume or surface, homeostasis of compounds) [15] common to all life forms. Therefore the DEB model becomes a tool that can be used to model lifetime feeding, growth, reproduction, and their responses to changes in combination of biotic and abiotic conditions [16,17]. The DEB model is a useful tool to fully integrate all organism processes, offering a complete overview of a species physiology and life cycle [17], and its parameters can be used to increase our knowledge on particular processes and adaptations integrating the mechanistic framework underlying the DEB theory [17,18]. This approach addresses the necessity of incorporating species physiology (and actual limitations) in predictive models, which is promising for the description of complex impacts of environmental variations on life history and biological traits.

In this study, we estimated the DEB parameters for *Laternula elliptica*, a large-sized infaunal suspension-feeding bivalve with a circumpolar distribution [19]. *L. elliptica* is particularly common in shallow waters (less than 30m) where it is often found in high densities (up to 170 ind.m^{-2} in Marian Cove, King George Is.) in soft sediments, representing a high biomass (289.9 g ash free dry weight m^{-2}) [19–21]. *L. elliptica* is a key species in Antarctic coastal benthic ecosystems, strongly influencing efficiencies of bentho-pelagic coupling processes [21]. Due to its role in transferring organic carbon from the water column to the benthic realm, *L. elliptica* enriches the surrounding sediments, sustaining associated biota [21,22]. Due to its

abundance and key-role, the energetic performance of *L. elliptica* populations has an important impact on associated ecosystems. *L. elliptica* has been broadly used as an experimental model and abundant research literature is available on its growth, gametogenesis, metabolism, feeding, thermal and acidification tolerances [21,23–29]. These studies focus on specific physiological processes, but do not provide an overarching view of the biology/ecology of *L. elliptica*. Altogether, these studies describe *L. elliptica* as a "typical" Antarctic organism with low metabolic rate, extended lifespan, long larval development, and relatively extended gametogenesis. A few analyses have lead to the development of population models, in an attempt to describe the effect of *L. elliptica* population dynamics on its ecosystem [26,30,31]. However, these models do not rely on mechanistic principles, and as such they are limited to describing the observed variability. There are still several gaps in our knowledge of the life history and population dynamics of *L. elliptica* such as the role of food and temperature on growth and reproduction, although this bivalve inhabits areas where food availability is highly variable and heavily influenced by environmental conditions such as ice-cover, ice-scouring or land sediment run-off [24,32]. To the best of our knowledge, no comprehensive study has tried to provide a tool to assess how such a variable environment affects *L. elliptica* performance and *in extenso* its role within the ecosystem.

The aim of the present study was to use DEB theory to mechanistically describe the adult life cycle of *L. elliptica*. Our work provides a quantitative model that can be used to better understand the physiological condition of *L. elliptica* and its specific adaptations to its environment. Moreover, the DEB parameters obtained here were explored to delineate the seasonal variability of animal condition exposed to varying environmental conditions of food and temperature. This work responds to a growing need to quantify and predict the effects of environmental changes on the population dynamics of key species from the Southern Ocean. Our DEB model should help to fill some of knowledge gaps in the life history of *L. elliptica*. The generality of DEB models will also allow comparisons with closely related temperate and tropical species to confirm or infirm the existence of possibly unique adaptations of the energy budget in this Antarctic species. Moreover, the links established by DEB theory with environmental resources will help to describe the seasonal food availability and its dependencies.

Material and methods

Model description

Dynamic Energy Budget (DEB) theory describes the processes of energy and matter-uptake throughout life [14]. The DEB model divides the mass and energy of an organism into four state variables: reserves (E), structural volume (V), maturity (E_H) and reproduction buffer (E_R). Energy enters the organism as food (X) and is assimilated at a rate of $\dot{p}_A$ into reserves. The mobilisation rate ($\dot{p}_C$) regulates the energy mobilised from the reserves to cover somatic maintenance ($\dot{p}_M$), structural growth ($\dot{p}_G$), maturity maintenance ($\dot{p}_J$), maturation ($\dot{p}_R$) (immature individuals) and reproduction ($\dot{p}_R$) (mature individuals). κ is the proportion of the mobilised energy diverted to $\dot{p}_M$ and $\dot{p}_G$, while the rest is used for $\dot{p}_J$ and $\dot{p}_R$. In DEB, assimilation is a function of food availability, following a functional response of Holling type II. Mobilisation however depends on the amount of energy stored into the reserves (see S1 File for detailed description of DEB assumptions, schematic representation and notation).

DEB theory assumes isomorphism (animal shape does not change with growth) [14]. The ratio between physical size, structural volume and surface remains constant as the animal grows. Marine benthic animals with pelagic larval stages undergo metamorphosis, which leads to a change in morphometry (or body shape) and therefore our DEB model incorporates one

shape coefficient for the adult (δ_M) and a different one for the D-larvae ($\delta_{M.larv}$). DEB model intends to describe the entire life cycle with the same set of parameters. It has been reported that *Laternula elliptica* has an encapsulated larval stage [33,34], during which the larva develops by consuming egg reserves within the capsule, without external feeding. After hatching out, the D-shaped larva settles on the bottom, starts feeding, and matures into an adult. In this work, we do not intend to describe the larval development of *L. elliptica*, however we consider this characteristic to help the model parametrisation with a realistic birth event of a D-larva with a different shape coefficient than that of the adult clam. Although we do not know if the larval development is accelerated, we considered a model parametrisation including an accelerated stage as this is a general characteristic of bivalves [14,35].

Estimation of DEB model parameters

DEB model parameters are derived from data determined for natural populations and experimental studies where the effects of controlled variables (e.g. temperature or food level) on growth, metabolic rate, reproductive output of individuals were measured [15,36]. There is a vast literature-based knowledge about several aspects of *L. elliptica* population dynamics in several localities around Antarctica [23,24,26,37,38]. However, little work has been carried out experimentally, under controlled conditions. In this study, we extracted literature-based data for parameter estimation (see Table 1 for a parameter list and corresponding definition and units). This data was used in combination with data gathered from Marian Cove between 1998 and 1999 on gonad development and animal condition for a single population. Detailed description of this data gathering and processing can be found in Ahn et al. [24].

Starting values for some DEB parameters were obtained directly from experimental studies and field observations, as follows:

Temperature sensitivity: Arrhenius parameters. DEB theory integrates the Arrhenius concept of enzyme activation to account for the sensitivity of metabolic rates to temperature [14] (S1 File). Arrhenius temperature (T_A) can be calculated from observed values of rates, such as metabolic or growth rates. The DEB model uses a curve for temperature sensitivity given by 5 parameters. However, it is possible to use a three-parameter function considering only the lower limit of the temperature range (S1 File). There were no conclusive data to determine the upper temperature limit parameters for *L. elliptica*. Therefore, this study only attempted the parametrisation of the three-parameter Arrhenius function.

Parameters were obtained by adjusting the Arrhenius function to the scaled values on oxygen consumption measured by Peck et al. [39] by means of a non-linear least squares regression using the package minpack.lm [40] and R v.3.15 [41].

Post-metamorphic shape coefficient. A post-metamorphic shape coefficient (δ_M) was calculated based on the relationship between the body ash-free dry weight (W_d) (after subtraction of the weight of the gonad) and the animal's shell length (L_w) by fitting the equation $W_d = (\delta_M \cdot L_w)^3$ by means of a weighted least squares regression [15]. The post-metamorphic shape coefficient was calculated from observations of individuals collected by Ahn et al. [24] in Marian Cove and the data provided by Ahn and Shim [23].

The covariation method for DEB parameter estimation. DEB models are very rich in parameters, however the proportion which can be calculated directly from empirical observations is very limited [15,42]. The covariation method was used for further estimation of the DEB model parameters [42,43] applied with MATLAB® (2015a) using the toolbox package DEBtool (available at http://www.bio.vu.nl/thb/deb/deblab/debtool/). DEB models combine different mechanisms and principles to describe all the processes during the life cycle of an organism [42]. The covariation method uses experimental and field observations of different

Table 1. DEB parameters values for *Laternula elliptica*. These parameters are given for a temperature of 273.15 K.

Parameter	Symbol	Value	Units
Basic DEB parameters			
Maximum structural length[1]	L_m	8.426	cm
Maximum surface area-specific assimilation rate [1]	$\{\dot{p}_{Am}\}$	87.752	J d^{-1} cm^{-2}
Volume-specific cost of maintenance [2]	$[\dot{p}_M]$	6.861	J d^{-1} cm^{-3}
Volume-specific cost of structure [1]	$[E_G]$	2371	J cm^{-3}
Fraction of energy allocated to somatic maintenance and growth [1]	κ	0.659	-
Maturity at birth [1]	E_H^b	3.371	J
Maturity at puberty (onset first gametogenesis) [1]	E_H^p	2116	J
Scaled functional response at Marian Cove[1]	f_{MC}	0.332	-
Scaled functional response at Potter Cove[1]	f_{PC}	0.384	-
Scaled functional response at Rothera[3]	f_R	0.8	-
DEB compound parameters			
Energy conductance [1]	$\dot{v}$	0.023	cm d^{-1}
Maturity maintenance rate coefficient [1]	$\dot{k}_J$	0.001	d^{-1}
Shape coefficients			
Post-metamorphic [2]	δ_M	0.341	-
Pre-metamorphic [1]	$\delta_{M.Irv}$	7.227	-
Temperature sensitivity			
Arrhenius temperature [2]	T_A	4832±1306	K
Arrhenius temperature at lower limit [2]	T_{AL}	19966±1.5x10^5	K
Lower temperature limit[2]	T_L	271±1.74	K
Conversion parameters			
Density of structure [3]	d_V	0.09	g cm^{-3}
Weight-energy coupler for reserves [3]	ρ_E	4.35x10^{-5}	g J^{-1}
Molecular weight of reserves [3]	w_E	23.9	g mol^{-1}
Chemical potential of reserves [3]	$\bar{\mu}_E$	550	kJ mol^{-1}

[1] Estimated using the covariation method

[2] Estimated from data.

[3] Fixed

https://doi.org/10.1371/journal.pone.0183848.t001

life stages and approximates the parameters using a Nelder-Mead numerical optimization to minimize the difference between observed and predicted values based on a weighted least-squares criterion [42]. Input data is all connected through the different parameters and the combination of data from different developmental stages and processes at different food availabilities result in a robust prediction of DEB parameters [42]. The parameters previously approximated were used as starting values in the covariation method. The parameters without experimental estimation included pseudo-data, and their starting values were yielded from DEB theory and closely related species [17]. The covariation method is completed with direct observations and data yielded from experiments for which it will approximate the parameters. Two different types of observations can be used in the covariation method [42,43]: zero-variate data represent single data points for a range of different physiological observations; uni-variate data comprise paired data of an independent variable and a dependent variable. The level of fitness of the covariation method is given by the mean absolute relative error (MRE) [44] among all data points and sets used in the parameter estimation. The procedure outputs an MRE value for each zero- and uni-variate variables to help assess the fitness of the predictions

Table 2. Zero-variate data used for the estimation of the DEB model parameters. MRE: mean absolute relative error.

Variable		Obs.	T (K)	Pred.	Units	MRE	Reference
Age at birth[1]	a_b	23	274.15	21.6	d	0.06	[34]
Age at puberty[2]	a_p	730	274.15	557	d	0.23	[49]
D-larva shell length at birth[1]	L_b	0.02	n/a	0.02	cm	0.03	[34]
Shell length at puberty[2]	L_p	2.87	n/a	3.19	cm	0.11	[53]
Maximum shell length[3]	L_i	8.7	n/a	8.19	cm	0.06	[23]
Dry weight at puberty[2]*	dW_p	0.18	n/a	0.19	g	0.05	[23]
Maximum dry weight[3]*	dW_i	3.21	n/a	3.19	g	0.01	[23]
Gonadosomatic Index[4]	GSI	0.22	272.7	0.21	-	0.03	[24]

[1] birth is set at the moment the animal starts or is able to feed. D-larva

[2] start of first gametogenesis.

[3] maximum size reached by the species when there is no food limitation. Taken as the upper 95% quantile of population size

[4] maximum gonad index for an animal of the maximum size, gonad index being defined as gonad weight/total wet weight.

* dry weights correspond to ash free dry weights.

https://doi.org/10.1371/journal.pone.0183848.t002

to each data set. A list of the zero-variate data used in the estimation of the DEB model parameters can be found in Table 2, zero-variate data corresponding to the population from Marian Cove [23] except for the data on development (age and length at birth) which belong to the population from Rothera station (Marguerite Bay) [34]. Uni-variate data are represented in the results and their origin referenced in Fig 1, and include data from different populations: length-weight from Marian Cove, oxygen comsumption from Rothera and size at age from Potter Cove (King George Island) (S2 File for locations). Using data from different locations allowed to include populations with different condition, feeding at different food levels [17].

DEB model parameters application: Describing seasonal metabolism and food availability

Model parameters were applied to describe the seasonal metabolism of *L. elliptica* from Marian Cove using the data on animal condition and gonad development during the years 1998 and 1999 [24]. This application aimed at exploring the parameters' performance on field data that was not used during parameter estimation and additionally to get an insight on the variability of available food for *L. elliptica*.

To describe the seasonal changes in energy reserves (E) we used observations on shell length and weight after removing the gonad as a proxy to the amount of reserves using DEB parameters (Eq 1) from samples taken during the years 1998 and 1999 (see [24] for details on sampling)

$$E = W - L^3 \quad (1)$$

where W is weight (g), L is structural length (cm, related to physical length through δ_M) (Table 1, S1 File). L^3 is the structural volume (V).

For simplification and easy handling, we used the scaled version of the DEB state variables and we used the scaled energy reserve (Eq 2).

$$e = \frac{W - L^3}{E_m \cdot L^3} \quad (2)$$

where e is the scaled energy reserve (unitless), E_m is the maximum energy density (J cm^{-3}) and

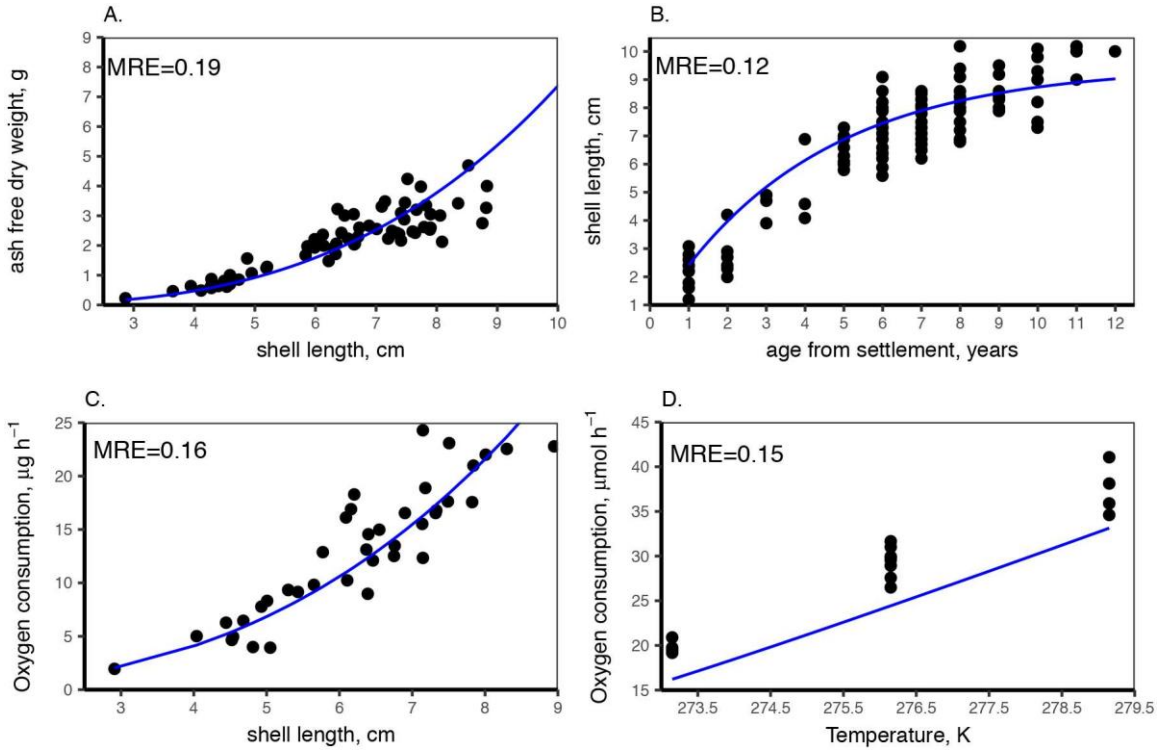


Fig 1. DEB model outputs and uni-variate data used for DEB parameter estimation. A. Total ash free dry weight (no gonads) as a function of shell length [23]. **B.** Size at age from shell growth rings at Potter Cove [26]. **C.** Respiration at size at Marian Cove [23] at a temperature of 274.15 K. **D.** Respiration at temperature for a standard individual of 7.5cm shell length [39] excluding assimilation (starved individuals). Dots are data from field observations or laboratory experiments. Blue line represents DEB model output. MRE is the mean absolute relative error.

https://doi.org/10.1371/journal.pone.0183848.g001

L is again structural length (cm). The denominator is the maximum energy reserve for an organism structural length L without food limitation.

Scaled energy reserves were approximated by applying a smoother to the observed values of e at each sampling event using the function *gam* from the package mgcv v.1.8–15 [45] and R v.3.15 [41]. The smoother was optimised for a limited number of knots to avoid overfitting, while still describing possible seasonal patterns [46] and was used to predict values of e for each day of the sampling period.

Predicted daily values of e from the *gam* smoother were then used to simulate the gonad growth during the same timeframe and to explore how the changes in energy reserves and seasonal temperature affected the investment towards reproduction by *L. elliptica* during the years 1998 and 1999. Although DEB theory specifies that state variables (V, E, E_H and E_R) cannot be measured empirically, it is possible to relate the reproduction buffer (E_R) with the gonad tissue imposing some handling rules [47,48]. Here, some assumptions were taken to describe the gonad growth and the spawning event in summer 1999: (1) E_R was assumed to be always in the gonad, therefore at constant conditions of temperature, reserves level and animal size, the gonad will grow linearly as reproduction flux ($\dot{p}_R$) is constantly transformed into gonad and accumulated [14]; (2) to account for the spawning event during 1999 a fixed date was assumed for the start of spawning (mid-December), which will last until March [24,25]; (3) only half of the gonad mass at the onset of spawning will be released during the spawning season, the rest of the gonad is resorbed and reused for further reproduction events [25,49];

(4) the released amount of gonad at a given time follows a logistic curve, which parameters were calculated based on the proportion of specimens with gonads in spawning stage [25,49] during the spawning season, allowing to take into account the variability of the spawning intensity during the spawning season; (5) the reproduction flux ($\dot{p}_R$) is directly released (i.e. not accumulated in the gonad) during the spawning season.

To further link the observed biological traits measured in the field (gonad development, loss of weight) to available resources, we used the rates of change observed in the approximated scaled reserves (e) to assess food availability. In DEB, energy reserves are determined by the availability of food resources in the environment and therefore reserve changes are directly related to the products available for consumption in the environment. To account for the variability of food the DEB model uses a scaled version of the Holling's type II functional response [50], f, to account for the effects of food availability on feeding (S1 File). The scaled functional response f directly relates to the assimilation flux $\dot{p}_A$ and therefore provides information on the energy acquired by the organism (Eq 3, S1 File).

$$\dot{p}_A = \{\dot{p}_{Am}\} \cdot L^2 \cdot f \tag{3}$$

where $\{\dot{p}_{Am}\}$ is the maximum surface-area specific assimilation (Table 1)

It was possible to use e to approximate the scaled functional response f. The scaled reserves (e) tends to be in equilibrium with available food, when this happens $e = f$ and therefore e is often used as a proxy to account for food availability [14]. This is probably not the case most of the time in *L. elliptica*, due to the hypothesized seasonal variability of food [24] and the slow metabolism of the organism [37,39]. At the natural temperature of their habitat, it may take months or years for polar organisms to reach equilibrium with available resources when these are kept constant [17]. It is still possible to relate f and e using the DEB theory dynamic of the reserves (S1 File). Hence, the rate of change of the reserves is directly related to food availability through Eq 4.

$$\frac{\Delta e}{\Delta t} = (f - e) \cdot \dot{v} \cdot L^{-1}, \tag{4}$$

where $\dot{v}$ is energy conductance (Table 1).

Finite differences of the previously calculated *gam* smoother were then used to describe the change in e (Δe) by day (Δt). From thereon, Eq 4 was used to yield the scaled functional response (f). The scaled functional response cannot be quantitatively related to food density due to the lack of information on feeding and clearance rates of *L. elliptica* at different food densities. However, it offered a quantitative assessment of the energy assimilated by the organism during 1998–1999 and a scale of available resources that relates to animal condition and metabolism during the years 1998 and 1999 and how resources influenced the potential growth and reproductive outputs.

To gain insight in the factors that may have determined f during 1998 and 1999 we analysed the observed variability of f against potential food sources. *L. elliptica* is a suspension feeder, and therefore the availability of food resources depends on the nutrients available in the water column [51]. Chlorophyll concentration is often used as a proxy for quantifying food available to suspension feeders, however for a population located at 30m depth, sea surface chlorophyll concentration may not be such a good proxy in this respect [24]. In this study, we analyse the yielded f against surface measured chlorophyll concentration and the particulate organic carbon flux at 30m depth measured at Marian Cove [52], while still highly related to chlorophyll concentration, this flux could be a better proxy to the amount of food reaching *L. elliptica*. Moreover we also explored the effect of lithogenic sediment particles [52] in f. Lithogenic

particles have no nutrient value, however they are known to compete with food particles, reducing *L. elliptica* feeding efficiency [21,32]. Linear models were used with the aim to test the variation of observed *f* that could be explained by chlorophyll, POC and lithogenic fluxes. Measurements of POC and lithogenic fluxes were obtained from [52] and consisted of monthly averaged rates. Chlorophyll measurements were obtained from [24] and consisted of weekly measurements on sea surface. For analysis mean *f* and chlorophyll values were calculated for the same time frame covered for each POC and lithogenic flux value given by [52]. Chlorophyll and POC were highly correlated and were never used together. Initial models considered chlorophyll or POC with lithogenic flux and their interactions. Model selection was performed by Akaike Information Criterion (AIC) and likelihood ratio test using R v.3.15 [41].

Results

DEB parameters

Adjusting the Arrhenius function to the respiration data from Peck et al. [39] yielded a T_A of 4832 K, a lower limit temperature (T_L) of 271 K and an Arrhenius temperature at the lower limit (T_{AL}) of 19660 K (see S1 File). These values were used as fixed values in the covariation method to determine the other DEB parameters. Further parametrisation was not necessary as these parameters accurately described the temperature sensitivity detected in other observations of *Laternula elliptica*.

The calculated value of the post-metamorphic shape coefficient yielded from observations of shell length and wet weight relationship was of 0.33±0.02 (mean ± sd). This value was set as free within the covariation method, which gave back a definitive value of 0.341 (Table 1).

Parameter estimations are detailed in Table 1. The total fit of the covariation method resulted in a MRE (mean absolute relative error) of 0.100. In general, the estimated model parameters accurately describe the data used for their estimation (Table 2, Fig 1). During model parametrisation, it was assumed that the observed organism condition was the result of an average food availability for the population. The scaled functional responses (*f*) in Table 1 are considered in equilibrium with the organism reserve and therefore $e = f$. Comparison of data from different populations allowed us to understand the food limitation for the population at Marian Cove and Potter Cove with an average scaled functional response of 0.33 and 0.38 respectively, when compared with the population at Rothera. Animals from Rothera [39] showed the best condition (length—weight and higher metabolic rate) so the food level at Rothera was taken as maximum reference, we fixed the scaled functional response for the animals in Rothera to $f = 0.8$ (instead of the possible maximum $f = 1$) because it was rather unprovable that the animals from Rothera point were fed '*ad libitum*' considering the seasonal variability in the area [37]. Although the effect of temperature is well described by the calculated Arrhenius curve, the DEB model consistently underestimated respiration at temperature for the organisms at Rothera (Fig 1D). *L. elliptica* used in these experiments were starved [39], as such we assumed that there were no contribution of assimilation to the oxygen consumption [14]. It is possible that the starvation period was not long enough and there was still some contribution from assimilation explaining the consistent underestimation.

Seasonal variability of the scaled functional response in Marian Cove was assessed in more details during the model exploration.

Model exploration

DEB parameters relate to a model organism with a large storage capacity, where energy reserves of animals with no limitation in food supply compose more than 65% of the total dry

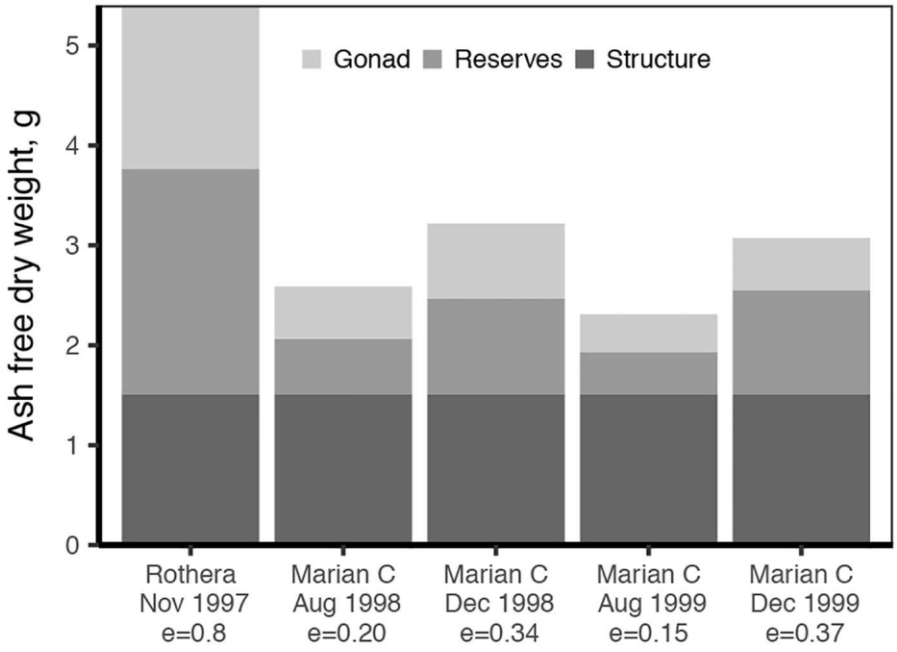


Fig 2. DEB state variables as biomass for different levels of scaled reserves (*e*), for a standard individual of 7.5cm shell length. Gonad weight for Rothera was probably overestimated as e = 0.8 was used for the whole gonad growth season.

https://doi.org/10.1371/journal.pone.0183848.g002

mass (excluding gonads). However natural populations never fed '*ad libitum*' due to large variability of food resources, and their reserves are greatly reduced: that is specially the case for the animals at Marian Cove where reserves compose between 20% and 41% of total mass (excluding gonads, Fig 2). Reserves for the animals at Rothera were much higher suggesting better feeding conditions.

DEB model parameters were explored further by comparing model outputs with time-series of biological traits of *L. elliptica* from Marian Cove, comprising the raw data collected by Ahn et al. [24] during the years 1998 and 1999 on animal condition, gonad development and seawater temperature (Fig 3). The approximation to the scaled energy (e) from length—weight data (using Eqs 1 and 2 and *gam* smoother S3 File) showed a seasonal variation of energy reserves during the year 1998–1999 (Fig 3B). Reserves are lower during spring and winter months, and reach their maximum during autumn-summer. Reserves varied seasonally by approximately a 50%. The minimum reserve level occurred during the winter 1999 (Figs 2 and 3B).

A description of the metabolism pattern was possible by using the approximated e and the estimated DEB parameters and theory to describe the main energy fluxes during 1998–1999. These fluxes relate to organism reserves, size and environmental temperature following DEB theory (S1 File). The reduction of energy reserves reduced significantly the energy mobilisation ($\dot{p}_c$), which is directly related to the size of reserves (S1 File), during the winter months (Fig 3C). Although maintenance costs ($\dot{p}_M + \dot{p}_j$) decreased during winter due to temperature (Fig 3C), the decrease of $\dot{p}_c$ due to the lowered reserves was steeper and in the middle of the winter $\dot{p}_c$ was barely enough to cover maintenance. The situation was worse during winter 1999 with reserves being lower than in 1998. In this case, the organisms were not able to cover maintenance (Fig 3C).

Further exploration of the DEB parameters and estimated e was done by simulating the gonad growth from March 1998 to December 1999. Predicted gonad dynamics described the

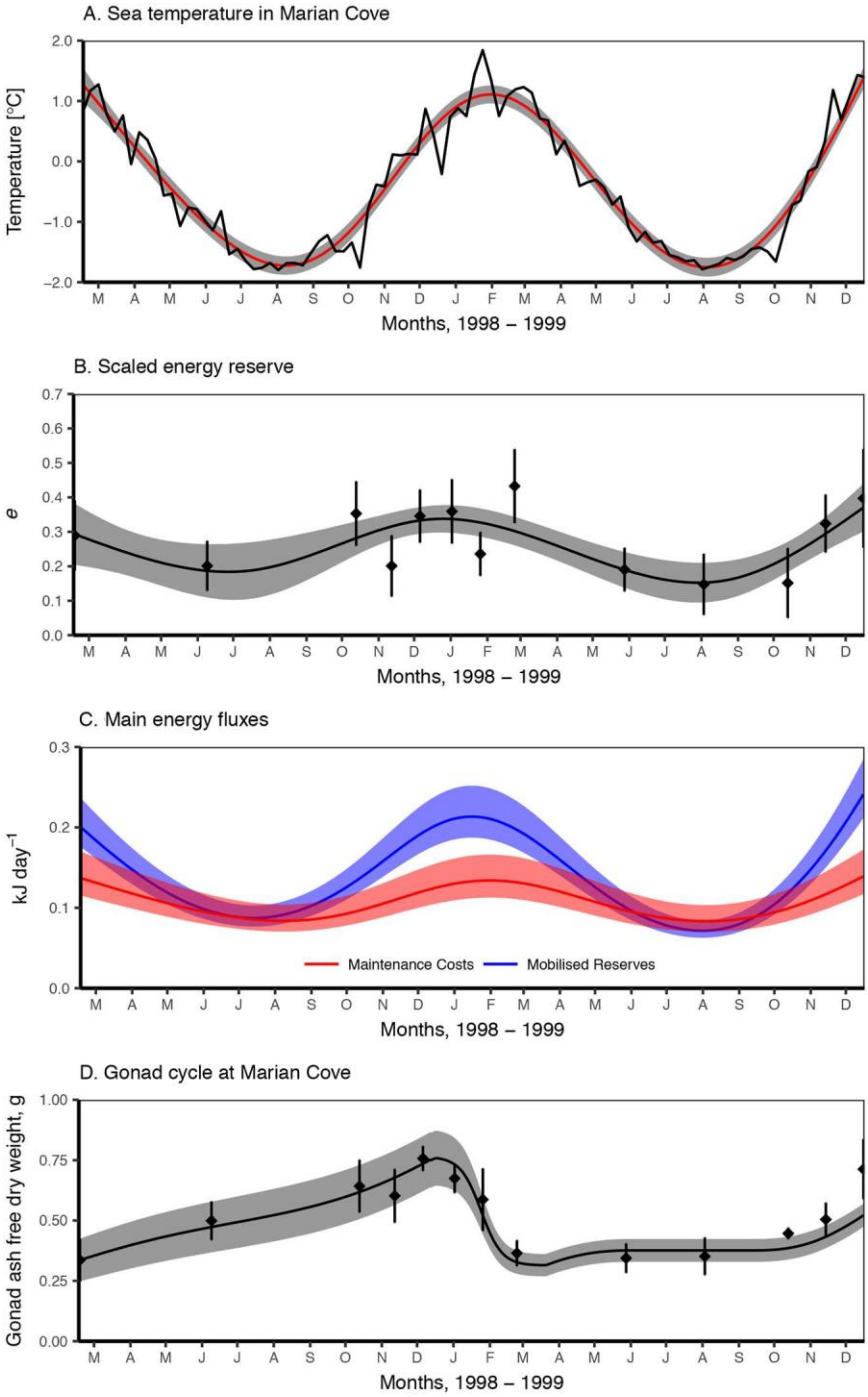


Fig 3. Estimated seasonal variation of energy reserves, metabolism and reconstructed gonad growth of *L. elliptica* at Marian Cove during 1998–1999. A. Temperature measured in the field. Red line is a fitted smoother (See S3 File), shaded area is the 95% confident interval (ci) of the smoother from [24]. **B.** Reconstructed scaled energy reserve. Dots and bars are mean values of *e* and the 95% ci yielded directly from field observations of length-weight (Eq 2). Line and shaded area respectively correspond to the mean and 95% ci of *gam* smoother (see Methods and S3 File). **C.** In blue calculated mobilisation flux ($\dot{p}_C$) for current energy reserves (*e*) (see S1 File for formulation, figure 3.B). In red calculated maintenance costs, as somatic maintenance ($\dot{p}_M$) plus maturity maintenance ($\dot{p}_J$) for a population with shell length distribution observed by Ahn et al. [24]. Both corrected for current temperature at day (Fig 3A). **D.** Gonad ash free dry weights. Dots and bars are the mean and the 95% ci of the field observations [24]. Line and shaded areas are DEB model predictions mean and 95% ci considering a population with the same shell length distribution as used by Ahn

et al. [24] and using the approximated *e* from Fig 3A. Gonad production was corrected by temperature (see S1 File).

https://doi.org/10.1371/journal.pone.0183848.g003

observed patterns in the field (Fig 3D). Higher reserves during 1998 resulted in the gonad growing all year long (Fig 3D), with a perceptible slow down during June and July 1998, due to the decrease in energy mobilisation (Fig 3C). When $\dot{p}_c$ is larger than maintenance the organism, can grow, and invest on reproduction, that does not happen in 1999 when there was an energy deficit, this is reflected by no increase on gonad size during the winter of 1999 (Fig 3C and 3D) as DEB prioritises maintenance over gonad growth [14]. Gonad growth accelerated during the last months of 1999 in response to the increase of *e* and temperature (Fig 3B, 3D and 3A).

The observed variation of reserves was analysed using DEB theory reserve dynamics (S1 File). The scaled functional response during the years 1998 and 1999 was reconstructed applying DEB reserve dynamics (Eq 4) to the finite differences in *e* given by the *gam* smoother (Fig 3B, S3 File) (Fig 4A). Reconstructed *f* (Fig 4A) showed how food resources varied seasonally between 1998–1999 with minimum values during winter and increasing fast during spring and early summer. As expected from observed *e*, *f* and therefore assimilation was lowest during winter 1999, when *f* decreased rapidly during summer and autumn and increased very fast during spring. The linear model (S3 File) considering chlorophyll concentration (Fig 4B), the amount of lithogenic particles in the sediment flux (Fig 4D) and their interaction provided the best fit for the observed *f*, explaining 45% of variability (see S3 File for model parameters and residual plots). The variability of *f* was directly related to chlorophyll concentration (S3 File, p-value < 0.01, Fig 4B) and indirectly to the lithogenic particles sedimentation (p-value < 0.01, Fig 4D). An alternative model considering POC flux (Fig 4C) instead of chlorophyll concentration and the amount of lithogenic particles in the sediment flux (Fig 4D), was the next best fit and explained 37% of the variability. The same negative effect was observed with the lithogenic flux, while POC flux was directly correlated with *f* (p-value < 0.01, S3 File).

Discussion

The present study provides the parametrisation of a DEB model for *Laternula elliptica*, a common Antarctic suspension feeding bivalve which plays a key role both structurally and functionally in shallow Antarctic marine ecosystems [21]. DEB theory and the parameters estimated here provide a mechanistic model that can be effectively used to understand the physiological condition of *L. elliptica* and the inherent adaptations to its habitat. In general DEB parameters describe an organism specially adapted to the characteristics of Antarctic environment, such as year-round low temperatures and extreme seasonality in food availability [2,54]. The exploration of DEB parameters against field observations recorded for two consecutive years allowed to gain insight on the variability of resources and the seasonal metabolism of *L. elliptica* and promoted the characterization of the model performance against data independent of that used for parameter estimation. Moreover, this data was used to link the observed variability of resources to sedimentation fluxes.

Dynamic Energy Budget model parameters for *L. elliptica* assigned a large reserve (≈65% of total weight, Fig 2) enough to survive long starvation, which benefitted from the low maintenance costs ($\dot{p}_M$) at the low-temperature conditions. Considering the temperature effect, $\dot{p}_M$ is larger but not very different from the values for other marine bivalves (at 20˚C *L. elliptica* [$\dot{p}_M$] is 31.2 $J \cdot cm^{-3} \cdot d^{-1}$ compared with the mean for marine bivalves of ≈24.54 $J \cdot cm^{-3} \cdot d^{-1}$, from the "Add my pet" webportal: http://www.bio.vu.nl/thb/deb/deblab/add_my_pet/), suggesting no adaptation to reduce the maintenance costs. *L. elliptica* has a large capacity to mobilise energy

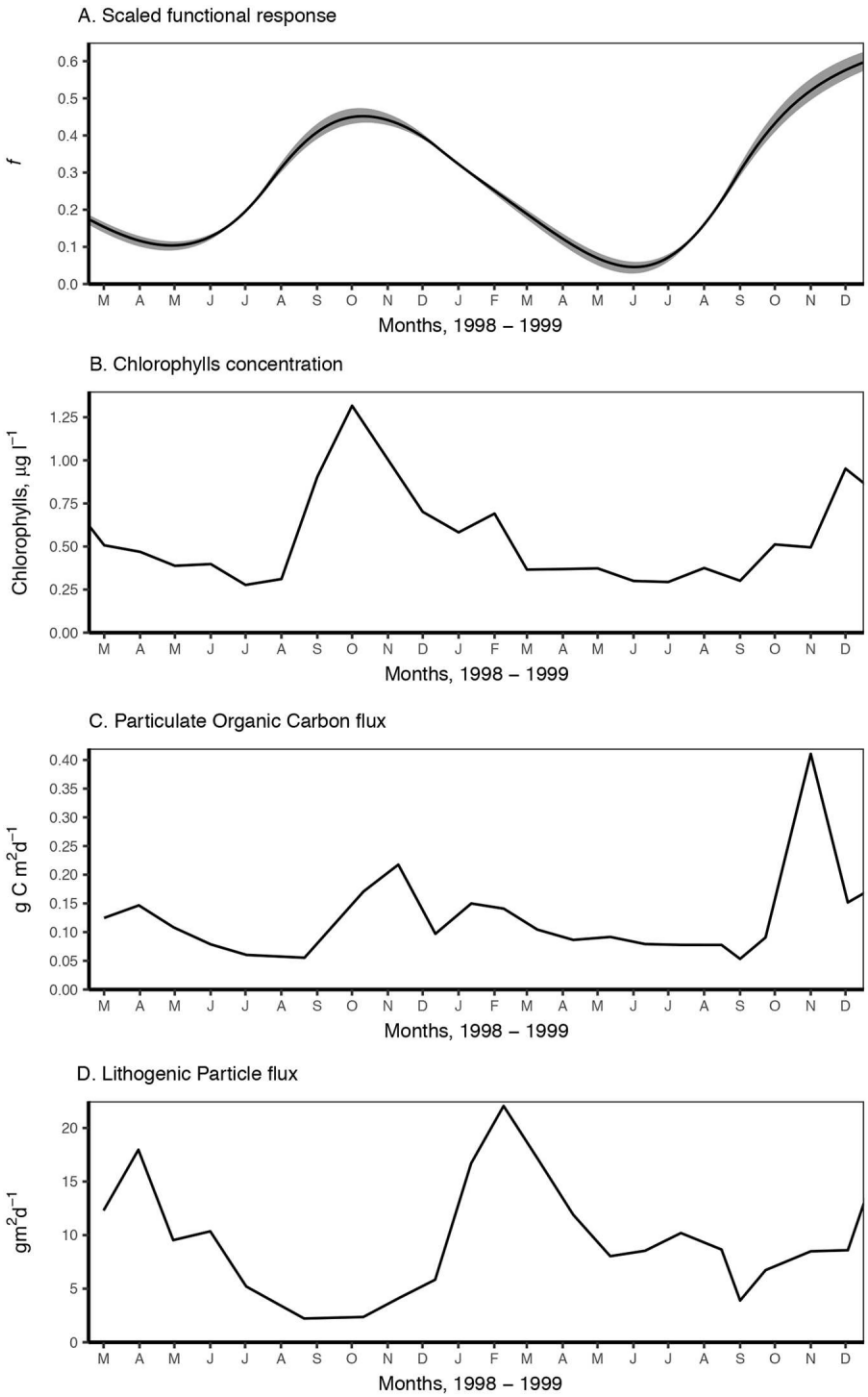


Fig 4. Estimated food levels for *L. elliptica* in Marian Cove during 1998–1999 and measured sediment fluxes (Khim et al. [52]). A. Reconstructed scaled functional response (*f*) from energy reserves dynamics considering temperature. For both, lines are mean and shaded area is the 95% ci. B. Chlorophylls concentration [24] C. Particulate organic carbon flux. D. Lithogenic particle flux. C & D. measured at Marian Cove at 30m depth. For details see Khim et al. [52].

https://doi.org/10.1371/journal.pone.0183848.g004

($\dot{v}$) that allows the organisms to keep a high supply of energy from the reserves even at low temperatures and when reserves are low. The energy conductance of *L. elliptica* at 0˚ is directly comparable (without correction for the low temperature) to that of organisms from habitats with a temperature close to 20˚C ($\dot{v} = 0.02$ cm d^{-1}, "Add my pet" webportal). This allows *L. elliptica* to keep the same level of energy mobilisation at 0˚C, as that of temperate organisms at 20˚C, keeping activity to the comparable levels [21]. The large capacity to build reserves associated with low maintenance costs allows *L. elliptica* to grow fast at low temperatures and even when food is not abundant, reaching its maximum size after 9 to 10 years [26]. Another important adaptation described by a DEB parameter is the high assimilation power ($\{\dot{p}_{Am}\}$) of *L. elliptica* (at 20˚C *L. elliptica* $\{\dot{p}_{Am}\}$ is 399 J·cm^{-2}·d^{-1} compared with the mean for marine bivalves of ≈20.5 J·cm^{-2}·d^{-1}, from the "Add my pet" webportal). Even if it was not possible to relate it to other characteristics such as the clearance rate or the functional response, this parameter indicates a high capacity to make use of the available food resources, allowing *L. elliptica* to build up reserves efficiently when food is available. It also allows for a large body as size is determined by the ratio between assimilation power and maintenance costs of structure (S1 File), with *L. elliptica* being one of the only two large bivalves species in Antarctica [24]. Altogether, the DEB parameters obtained in this study fit to an organism especially adapted to cold environments where food is limited being associated with extreme seasonality in day length [1].

The use of a time-series on *L. elliptica* condition and gonad growth alongside DEB parameters took advantage of one of the key characteristics of DEB theory and models: linking observed biological/physiological traits to environmental temperature and food resources. At the same time, we explored how the DEB parameters estimated here can assess biological traits in varying conditions. Data on weight-length and gonad development shows that energy reserves levels are seasonal (Figs 3 and 4). A longer period of food shortage happened in 1999 and animals lost a 25% of mass during the period of March through August 1999 [24]. With the use of DEB parameters on the same data the present study provided an assessment of the food available and how it relates to observed energy reserves variability. The considerable weight loss in winter 1999 related to a prolonged period of low energy uptake, during which *L. elliptica* uses up to 66% of its reserves (Figs 2 and 3B). Along with the severe weight loss gonad development was also affected. Gonad development seemed retarded in 1999, while during the same months in 1998 gonad was maturing continuously (Fig 3C and 3D). The high capacity to adapt to available resources by *L. elliptica* allowed to increase its energy reserves fast at the end of the winter resulting in a fast increase on the investment on reproduction and the consequent gonad growth at the end of 1999.

The exploration of biological traits showed that the scaled functional response (f) during the years 1998 and 1999 followed a seasonal pattern. The determination of f presented here allowed us to assess quantitatively the seasonal assimilation flux (S1 File). The assimilation flux is directly related to food quantity and quality through the scaled version of a Holling's type II functional response [14], therefore f is a scaled quantification of the amount of food available and the capacity of the organism to uptake it. The results showed than f is never high in Marian Cove (max of 0.4 over 1) and considerably lower than that of Rothera in November 1997 (Fig 2); however, it is considerably larger in the autumn-summer periods of 1998 and 1999 (f = 0.35–0.4) compared to spring-winter (f = 0.15–0.2). This study intended to link observed f to field measurements of environmental variables that may be able to describe the food availability for *L. elliptica* and its capacity to make use of this food. *L. elliptica* is a suspension-feeder and animal condition and reproduction cycle have been previously linked to the spring phytoplankton bloom [24]. The results here agreed with the observations by Ahn et al. [24] as

surface seasonal chlorophyll explained the seasonal food variability (Fig 4, S3 File). However, food availability is a complex variable, difficult to quantify, and chlorophyll concentration in surface water may not suppose always a good link as the amount reaching the bottom will depend on oceanographic conditions which varied for different locations, seasons, etc. Benthic suspension feeders like *L. elliptica* may rely primarily on benthic food materials [55]. Although these benthic food materials need to be resuspended into the water column to be accessible [26], amount of which may not be accounted by measurement of surface chlorophyll concentration. The sedimentation flux of particulate organic carbon (POC) was also a good proxy to *L. elliptica* scaled functional response, POC flux in Marian Cove being related to the phytoplankton bloom [52] however, it was measured at the depth where the animals were sampled and it also considers sedimentation of resuspended particles. However, chlorophyll and POC flux were not the only factors describing *L. elliptica* seasonal feeding, the analysis of the sediment fluxes found a strong negative effect of the lithogenic particle flux on *L. elliptica* scaled functional response. A negative effect of sediment load on condition and growth of *L. elliptica* have been reported before in the field [32] and a reduction in assimilation has been assessed in experiments [22]. *L. elliptica* feeds while filtering the water, so, when the concentration of particles without nutrition value increase relatively to the particulate carbon the effective concentration of food decreases, in the extreme cases a high concentration of particles may clog the filtering system [22,56,57]. Although two years may not be a long enough time-series to establish robust correlations, the results clearly highlight the importance of nourishing mass as well as its concentration within the total suspended particles, showing that the effects of sediment run-off are already affecting the seasonal feeding of *L. elliptica* population in Marian Cove. Establishing a link of nutritional state and assimilation of *L. elliptica* with variables such as POC is important to understand how environmental variations affect *L. elliptica* performances and to describe differences among populations. The difference between Marian Cove and Rothera *e* was due to different levels of food that could be related to a larger different primary production, and therefore of POC flux between the locations and also for a smaller lithogenic flux in Rothera as it is not influenced by a land-terminating glacier [2,9,24,58].

Laternula elliptica is a key organism in Antarctic shallow soft bottoms. Its capacity to enhance the carbon flux to the sediment by the biodeposition of faeces and pseudofaeces [21] creates an enriched area and reduces suspended matter that sustains associated benthic biota [22]. As such, the performance and population dynamics of *L. elliptica* has consequences for its ecosystem. Exploring the DEB model allowed to show how the life of *L. elliptica* was affected by the balance between the efficient use of food available during spring-summer season and the starvation during a food depleted winter, due to the bivalve's low maintenance metabolism. Antarctic coastal ecosystems are changing fast, especially in the Western Antarctic Peninsula [6]. Because of global warming, ice dynamics are already changing. The land glacier retreat and the decrease of sea-ice season duration [7] induce a cascade effect on coastal environments, decreasing salinity, increasing ice scouring and land sediment run-off [58,59]. All these changes are affecting primary production, POC fluxes, disturbance, resuspension and sediment load in coastal waters [9,58–60] and therefore they are changing the dynamics of the food abundance for *L. elliptica*. If food in summer is available in a quantity or long enough to build up a reserve large enough it has the potential to jeopardize the capacity of *L. elliptica* to survive winter starvation. The decrease observed in 1999 already resulted in a significant mass loss and slowed down gonad production (Fig 3B and 3D), if that situation was to be prolonged, reproduction in 1999 may have been affected too. Moreover, Antarctic coastal bottom water is also warming up as a consequence of global change, although the rate of change is small compared with that of air temperature [5,6], a raise in temperature during winter with very low food will increase the maintenance costs while decreasing the time *L. elliptica* can withstand

starvation (Fig 3C). On the contrary, an increase of food availability will improve condition, favouring growth and reproduction and buffering any non-lethal effect of warming.

Conclusion

Key species have a large impact on their ecosystems which rely on them for provision of food, space or protection among others [47,61,62]. In this context, knowledge on the physiological performance of key species under varying environmental conditions is fundamental to understand how such variations might affect the ecosystem. Changes in physiological performance will result in changes on key species population dynamics, affecting recruitment and/or survival [63,64] and particularly species key functions that will impact ecosystem structure and functioning [65]. In the case of *Laternula elliptica*, changes of its population density or feeding activity will impact bentho-pelagic coupling function and therefore will have a cascade effect on the biota that depends on the sediment enrichment. The DEB parameters estimated in this study provide detailed information on the metabolic strategy of *L. elliptica* and provided a mechanistic link between the organism's physiology and its environment. DEB parameters successfully describe the observed field variation of the population in condition and allocation to gonad, becoming a powerful and robust tool to understand the effects of varying environments on *L. elliptica* performance. Further knowledge on filtration feeding and deposition rates at different densities of food will allow to use this DEB model to assess the impact of *L. elliptica* through quantification of its function as a bentho-pelagic coupler. Moreover, this study provides a link between organism condition and energy uptake and important environmental variables: chlorophyll, POC and lithogenic fluxes which are important oceanographic variables for which large scale measurements and estimations are available. This link allows to derive *L. elliptica* performance directly from knowledge on these fluxes and temperature making it possible to express organism biological traits in a spatially-explicit context, to develop mechanistic species distribution models [66], which can be applied to study a range of present and future scenarios to predict future species performance and distribution.

Supporting information

S1 File. General description of standard DEB model assumptions and notation.
(PDF)

S2 File. Map with the localities mentioned.
(PDF)

S3 File. Statistical models summaries and residual plots.
(PDF)

Acknowledgments

This is contribution no.19 to the vERSO project (www.versoproject.be), funded by the Belgian Science Policy Office (BELSPO, contract no BR/132/A1/vERSO). Original data collection funded by the Korea Polar Research Institute (PE17070).

Author Contributions

Conceptualization: Antonio Agüera.

Data curation: In-Young Ahn, Charlène Guillaumot.

Formal analysis: Antonio Agüera.

Funding acquisition: Bruno Danis.

Investigation: In-Young Ahn.

Methodology: Antonio Agüera.

Project administration: Bruno Danis.

Resources: In-Young Ahn, Charlène Guillaumot, Bruno Danis.

Writing – original draft: Antonio Agüera.

Writing – review & editing: Antonio Agüera, In-Young Ahn, Charlène Guillaumot, Bruno Danis.

References

1. Barnes DKA, Clarke A. Antarctic marine biology. Curr Biol. Elsevier; 2011; 21: R451–R457. https://doi.org/10.1016/j.cub.2011.04.012 PMID: 21683895
2. Clarke A, Meredith MP, Wallace MI, Brandon MA, Thomas DN. Seasonal and interannual variability in temperature, chlorophyll and macronutrients in northern Marguerite Bay, Antarctica. Deep Res Part II Top Stud Oceanogr. 2008; 55: 1988–2006. https://doi.org/10.1016/j.dsr2.2008.04.035
3. Peck LS. Ecophysiology of Antarctic marine ectotherms: limits to life. Polar Biol. 2002; 25: 31–40.
4. McClintock J. Trophic biology of antarctic shallow-water echinoderms. Mar Ecol Prog Ser. 1994; 111: 191–202. https://doi.org/10.3354/meps111191
5. IPCC. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Barros VR, Field CB, Dokken DJ, Mastrandrea MD, Mach KJ, Bilir TE, et al., editors. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA; 2014.
6. Turner J, Barrand NE, Bracegridle TJ, Dle TJ, Convey P, Hodgson DA, et al. Antarctic climate change and the environment: an update. Polar Rec (Gr Brit). 2014; 50: 237–259. https://doi.org/10.1017/S0032247413000296
7. Stammerjohn SE, Martinson DG, Smith RC, Iannuzzi RA. Sea ice in the western Antarctic Peninsula region: Spatio-temporal variability from ecological and climate change perspectives. Deep Sea Res Part II Top Stud Oceanogr. 2008; 55: 2041–2058. https://doi.org/10.1016/j.dsr2.2008.04.026
8. Barnes DKA, Souster T. Reduced survival of Antarctic benthos linked to climate-induced iceberg scouring. Nat Clim Chang. Nature Publishing Group; 2011; 1: 365–368. Available: http://dx.doi.org/10.1038/nclimate1232
9. Montes-Hugo M, Doney SC, Ducklow HW, Fraser W, Martinson D, Stammerjohn SE, et al. Recent Changes in Phytoplankton Communities Associated with Rapid Regional Climate Change Along the Western Antarctic Peninsula. Science (80-). 2009; 323: 1470–1473. https://doi.org/10.1126/science.1164533 PMID: 19286554
10. Sahade R, Lagger C, Torre L, Momo F, Monien P, Schloss I, et al. Climate change and glacier retreat drive shifts in an Antarctic benthic ecosystem. Sci Adv. 2015; 1: e1500050. https://doi.org/10.1126/sciadv.1500050 PMID: 26702429
11. Kearney M. Metabolic theory, life history and the distribution of a terrestrial ectotherm. Funct Ecol. 2012; 26: 167–179. https://doi.org/10.1111/j.1365-2435.2011.01917.x
12. Kearney MR, Wintle BA, Porter WP. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. Conserv Lett. 2010; 3: 203–213. https://doi.org/10.1111/j.1755-263X.2010.00097.x
13. Kooijman SALM. Energy budgets can explain body size relations. J Theor Biol. 1986; 121: 269–282. https://doi.org/10.1016/S0022-5193(86)80107-2
14. Kooijman SALM. Dynamic Energy Budget theory for metabolic organisation. Third. Cambridge: Cambridge University Press; 2010.
15. van der Meer J. An introduction to Dynamic Energy Budget (DEB) models with special emphasis on parameter estimation. J Sea Res. 2006; 56: 85–102. https://doi.org/10.1016/j.seares.2006.03.001
16. Jusup M, Klanjscek T, Matsuda H, Kooijman SALM. A Full Lifecycle Bioenergetic Model for Bluefin Tuna. Planque R, editor. PLoS One. 2011; 6: e21903. https://doi.org/10.1371/journal.pone.0021903 PMID: 21779352

17. Agüera A, Collard M, Jossart Q, Moreau C, Danis B. Parameter Estimations of Dynamic Energy Budget (DEB) Model over the Life History of a Key Antarctic Species: The Antarctic Sea Star *Odontaster validus* Koehler, 1906. Thuesen E V., editor. PLoS One. Public Library of Science; 2015; 10: e0140078. https://doi.org/10.1371/journal.pone.0140078 PMID: 26451918
18. Pecquerie L, Fablet R, De Pontual H, Bonhommeau S, Alunno-Bruscia M, Petitgas P, et al. Reconstructing individual food and growth histories from biogenic carbonates. Mar Ecol Prog Ser. 2012; 447: 151–164. https://doi.org/10.3354/meps09492
19. Ahn I-Y. Ecology of the Antarctic Bivalve *Laternula elliptica* (King and Broderip) in Collins Harbor, King George Island: Benthic Environment and an adaptative strategy. Mem Natl Inst Polar Res. 1994; 50: 1–10.
20. Ahn I-Y, Chung H, Choi K-S. Some Ecological and Physiological Features of the Antarctic Clam, *Laternula elliptica* (King and Broderip) in a Nearshore Habitat on King George Island. Ocean Polar Res. Korea Institute of Ocean Science & Technology; 2001; 23: 419–424.
21. Ahn I-Y. Enhanced particle flux through the biodeposition by the Antarctic suspension-feeding bivalve *Laternula elliptica* in Marian Cove, King George Island. J Exp Mar Bio Ecol. 1993; 171: 75–90. https://doi.org/10.1016/0022-0981(93)90141-A
22. Tatián M, Sahade R, Mercuri G, Fuentes VL, Antacli JC, Stellfeldt A, et al. Feeding ecology of benthic filter-feeders at Potter Cove, an Antarctic coastal ecosystem. Polar Biol. 2008; 31: 509–517. https://doi.org/10.1007/s00300-007-0379-7
23. Ahn I-Y, Shim JH. Summer metabolism of the Antarctic clam, *Laternula elliptica* (King and Broderip) in Maxwell Bay, King George Island and its implications. J Exp Mar Bio Ecol. 1998; 224: 253–264. https://doi.org/10.1016/S0022-0981(97)00201-3
24. Ahn I-Y, Surh J, Park Y-G, Kwon H, Choi K-S, Kang S-H, et al. Growth and seasonal energetics of the Antarctic bivalve *Laternula elliptica* from King George Island, Antarctica. Mar Ecol Prog Ser. 2003; 257: 99–110. https://doi.org/10.3354/meps257099
25. Kang D-H, Ahn I-Y, Choi K-S. The annual reproductive pattern of the Antarctic clam, *Laternula elliptica* from Marian Cove, King George Island. Polar Biol. 2009; 32: 517–528. https://doi.org/10.1007/s00300-008-0544-7
26. Urban H-J, Mercuri G. Population dynamics of the bivalve *Laternula elliptica* from Potter Cove, King George Island, South Shetland Islands. Antarct Sci. 1998; 10: 153–160. https://doi.org/10.1017/S0954102098000200
27. Peck LS, Morley SA, Pörtner HO, Clark MS. Thermal limits of burrowing capacity are linked to oxygen availability and size in the Antarctic clam *Laternula elliptica*. Oecologia. 2007; 154: 479–484. https://doi.org/10.1007/s00442-007-0858-0 PMID: 17899201
28. Morley SA, Martin SM, Bates AE, Clark MS, Ericson J, Lamare M, et al. Spatial and temporal variation in the heat tolerance limits of two abundant Southern Ocean invertebrates. Mar Ecol Prog Ser. 2012; 450: 81–92. https://doi.org/10.3354/meps09577
29. Bylenga CH, Cummings VJ, Ryan KG. Fertilisation and larval development in an Antarctic bivalve, *Laternula elliptica*, under reduced pH and elevated temperatures. Mar Ecol Prog Ser. 2015; 536: 187–201. https://doi.org/10.3354/meps11436
30. Momo F, Kowalke J, Schloss I, Mercuri G, Ferreyra G. The role of *Laternula elliptica* in the energy budget of Potter Cove (King George Island, Antarctica). Ecol Modell. 2002; 155: 43–51. https://doi.org/10.1016/S0304-3800(02)00081-9
31. Brey T, Voigt M, Jenkins K, Ahn I-Y. The bivalve *Laternula elliptica* at King George Island—a biological recorder of climate forcing in the West Antarctic peninsula region. J Mar Syst. Elsevier B.V.; 2011; 88: 542–552. https://doi.org/10.1016/j.jmarsys.2011.07.004
32. Philipp EER, Husmann G, Abele D. The impact of sediment deposition and iceberg scour on the Antarctic soft shell clam *Laternula elliptica* at King George Island, Antarctica. Antarct Sci. 2011; 23: 127–138. https://doi.org/10.1017/S0954102010000970
33. Ansell AD, Harvey R. Proteceted Larval Development in the Antarctic bivalve *Laternula elliptica* (King and Broderip) (Anomalodesmata: Laternulidae). J Molluscan Stud. 1997; 63: 285–286.
34. Peck LS, Powell DK, Tyler PA. Very slow development in two Antarctic bivalve molluscs, the infaunal clam *Laternula elliptica* and the scallop *Adamussium colbecki*. Mar Biol. 2007; 150: 1191–1197. https://doi.org/10.1007/s00227-006-0428-8
35. Kooijman SALM. Metabolic acceleration in animal ontogeny: An evolutionary perspective. J Sea Res. Elsevier B.V.; 2014; 94: 128–137. https://doi.org/10.1016/j.seares.2014.06.005
36. van der Veer HW, Cardoso JFMF, van der Meer J. The estimation of DEB parameters for various Northeast Atlantic bivalve species. J Sea Res. 2006; 56: 107–124. https://doi.org/10.1016/j.seares.2006.03.005

37. Brockington S. The seasonal energetics of the Antarctic bivalve *Laternula elliptica* (King and Broderip) at Rothera Point, Adelaide Island. Polar Biol. Springer-Verlag; 2001; 24: 523–530. https://doi.org/10.1007/s003000100251

38. Guy CI, Cummings VJ, Lohrer AM, Gamito S, Thrush SF. Population trajectories for the Antarctic bivalve Laternula elliptica: Identifying demographic bottlenecks in differing environmental futures. Polar Biol. 2014; 37: 541–553. https://doi.org/10.1007/s00300-014-1456-3

39. Peck LS, Pörtner HO, Hardewig I. Metabolic Demand, Oxygen Supply, and Critical Temperatures in the Antarctic Bivalve *Laternula elliptica*. Physiol Biochem Zool. 2002; 75: 123–133. https://doi.org/10.1086/340990 PMID: 12024288

40. Elzhov T V, Mullen KM, Spiess A-N, Bolker B. minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds [Internet]. 2013. http://cran.r-project.org/package=minpack.lm

41. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2014. http://www.r-project.org/

42. Lika K, Kearney MR, Freitas V, van der Veer HW, van der Meer J, Wijsman JWM, et al. The "covariation method" for estimating the parameters of the standard Dynamic Energy Budget model I: Philosophy and approach. J Sea Res. Elsevier B.V.; 2011; 66: 270–277. https://doi.org/10.1016/j.seares.2011.07.010

43. Lika K, Kearney MR, Kooijman SALM. The "covariation method" for estimating the parameters of the standard Dynamic Energy Budget model II: Properties and preliminary patterns. J Sea Res. Elsevier B.V.; 2011; 66: 278–288. https://doi.org/10.1016/j.seares.2011.09.004

44. Jusup M, Klanjšček T, Matsuda H. Simple measurements reveal the feeding history, the onset of reproduction, and energy conversion efficiencies in captive bluefin tuna. J Sea Res. 2014; 94: 144–155. https://doi.org/10.1016/j.seares.2014.09.002

45. Wood S. Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation. 2016.

46. Zuur AF. A beginner's guide to Generalized Additive Models with R. 2nd ed. Zuur AF, editor. Newburgh, Scotland: Highland Statistics; 2012.

47. Agüera A, van de Koppel J, Jansen JM, Smaal AC, Bouma TJ. Beyond food: a foundation species facilitates its own predator. Oikos. 2015; 124: 1367–1373. https://doi.org/10.1111/oik.01949

48. Pecquerie L, Petitgas P, Kooijman SALM. Modeling fish growth and reproduction in the context of the Dynamic Energy Budget theory to predict environmental impact on anchovy spawning duration. J Sea Res. Elsevier B.V.; 2009; 62: 93–105. https://doi.org/10.1016/j.seares.2009.06.002

49. Kang D-H, Ahn I-Y, Choi K-S. Quantitative assessment of reproductive condition of the Antarctic clam, *Laternula elliptica* (King & Broderip), using image analysis. Invertebr Reprod Dev. 2003; 44: 71–78. https://doi.org/10.1080/07924259.2003.9652555

50. Holling CS. Some Characteristics of Simple Types of Predation and Parasitism. Can Entomol. 1959; 91: 385–398. https://doi.org/10.4039/Ent91385-7

51. Ahn I-Y. Feeding ecology of the Antarctic lamellibranch *Laternula elliptica* (Laternulidae) in Marian Cove and vicinity, King George Island, during one austral summer. In: Battaglia B, Valencia J, Walton D, editors. Antarctic Communities: Species, Structure and Survival. Cambridge: Cambridge University Press; 1997. pp. 142–151.

52. Khim BK, Shim J, Yoon HI, Kang YC, Jang YH. Lithogenic and biogenic particle deposition in an Antarctic coastal environment (Marian Cove, King George Island): Seasonal patterns from a sediment trap study. Estuar Coast Shelf Sci. 2007; 73: 111–122. https://doi.org/10.1016/j.ecss.2006.12.015

53. Bigatti G, Penchaszadeh PE, Mercuri G. Aspects of the gonadal cycle in the Antartic bivalve *Laternula elliptica*. J Shellfish Res. 2001; 20: 283–287.

54. Hunt BM, Hoefling K, Cheng C-HC. Annual warming episodes in seawater temperatures in McMurdo Sound in relationship to endogenous ice in notothenioid fish. Antarct Sci. 2003; 15: 333–338. https://doi.org/10.1017/S0954102003001342

55. Ahn I-Y, Moon H-W, Jeon M, Kang S. First Record of Massive Blooming of Benthic Diatoms and Their Association with Megabenthic Filter Feeders on the Shallow Seafloor of an Antarctic Fjord: Does Glacier Melting Fuel the Bloom? 2016; 51: 273–279.

56. Österling ME, Arvidsson BL, Greenberg LA. Habitat degradation and the decline of the threatened mussel *Margaritifera margaritifera*: influence of turbidity and sedimentation on the mussel and its host. J Appl Ecol. 2010; 47: 759–768. https://doi.org/10.1111/j.1365-2664.2010.01827.x

57. Kowalke J. Energieumsatze bethischer Fitrierer der Potter Cove (King George Island, Antatktis). Bremerhaven; 1998.

58. Meredith MP, Stammerjohn SE, Venables HJ, Ducklow HW, Martinson DG, Iannuzzi RA, et al. Changing distributions of sea ice melt and meteoric water west of the Antarctic Peninsula. Deep Sea Res Part II Top Stud Oceanogr. 2017; 139: 40–57. https://doi.org/10.1016/j.dsr2.2016.04.019

59. Massom RA, Stammerjohn SE. Antarctic sea ice change and variability—Physical and ecological implications. Polar Sci. 2010; 4: 149–186. https://doi.org/10.1016/j.polar.2010.05.001

60. Bers AV, Momo F, Schloss IR, Abele D. Analysis of trends and sudden changes in long-term environmental data from King George Island (Antarctica): Relationships between global climatic oscillations and local system response. Clim Change. 2013; 116: 789–803. https://doi.org/10.1007/s10584-012-0523-4

61. Bertness MD, Callaway R. Positive Interactions in communities. Trends Ecol Evol. 1994; 9: 191–193. https://doi.org/10.1016/0169-5347(94)90088-4 PMID: 21236818

62. Paine RT. Food Web Complexity and Species Diversity. Am Nat. 1966; 100: 65–75.

63. Agüera A, Schellekens T, Jansen JM, Smaal AC. Effects of osmotic stress on predation behaviour of *Asterias rubens* L. J Sea Res. Elsevier B.V.; 2015; 99: 9–16. https://doi.org/10.1016/j.seares.2015.01.003

64. Gedan KB, Bertness MD. How will warming affect the salt marsh foundation species *Spartina patens* and its ecological role? Oecologia. 2010; 164: 479–487. https://doi.org/10.1007/s00442-010-1661-x PMID: 20490551

65. Bruno JF, Stachowicz JJ, Bertness MD. Inclusion of facilitation into ecological theory. Trends Ecol Evol. 2003; 18: 119–125. https://doi.org/10.1016/S0169-5347(02)00045-9

66. Kearney M, Simpson SJ, Raubenheimer D, Helmuth B. Modelling the ecological niche from functional traits. Philos Trans R Soc B Biol Sci. 2010; 365: 3469–3483. https://doi.org/10.1098/rstb.2010.0034 PMID: 20921046

Vol. 594: 149–164, 2018
https://doi.org/10.3354/meps12538

MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser

Published April 26

Methods for improving species distribution models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau

Charlène Guillaumot[1,*], Alexis Martin[2], Marc Eléaume[3], Thomas Saucède[4]

[1]Laboratoire de Biologie Marine, Université Libre de Bruxelles, Avenue FD Roosevelt 50, CP 160/15, 1150 Bruxelles, Belgique
[2]Département adaptation du vivant, Muséum national d'Histoire naturelle, UMR BOREA 7208, 57 rue Cuvier, 75231 Paris Cedex 05, France
[3]Département Origine et Évolution, Muséum national d'Histoire naturelle, UMR ISYEB 7205, 57 rue Cuvier, 75231 Paris Cedex 05, France
[4]Biogéosciences, UMR 6282, Université Bourgogne Franche-Comté, CNRS, 6 bd Gabriel 21000 Dijon, France

ABSTRACT: Species distribution models (SDMs) are essential tools to aid conservation biologists in evaluating the combined effects of environmental change and human activities on natural habitats and for the development of relevant conservation plans. However, modeling species distributions over vast and remote regions is often challenging due to poor and heterogeneous data sets, and this raises questions regarding the relevance of the modeling procedures. In recent years, there have been many methodological developments in SDM procedures using virtual species and broad data sets, but few solutions have been proposed to deal with poor or heterogeneous data. In the present work, we address this methodological challenge by studying the performance of different modeling procedures based on 4 real species, using presence-only data compiled from various oceanographic surveys on the Kerguelen Plateau (Southern Ocean). We followed a practical protocol to test for the reliability and performance of the models and to correct for limited and aggregated data, as well as accounting for spatial and temporal sampling biases. Our results show that producing reliable SDMs is feasible as long as the amount and quality of available data allow testing and correcting for these biases. However, we found that SDMs could be corrected for spatial and temporal heterogeneities in only 1 of the 4 species we examined, highlighting the need to consider all potential biases when modeling species distributions. Finally, we show that model reliability and performance also depend on the interaction between the incompleteness of the data and species niches, with the distribution of narrow-niche species being less sensitive to data gaps than species occupying wider niches.

KEY WORDS: Species distribution modeling · Model performance · Historical datasets · Kerguelen Plateau · Presence-only data

INTRODUCTION

Species distribution models (SDMs) are essential tools used by conservation biologists for understanding species distribution patterns and their drivers (see Guillera-Arroita et al. 2015 for a review), assessing the combined effects of environmental change and direct human pressure (i.e. economic activities including tourism) on natural habitats (Gutt et al. 2012), defining conservation priorities (Vierod et al. 2014, Greathead et al. 2015) and developing relevant management plans (Reiss et al. 2015, Koubbi et al. 2016). SDMs allow scientists to interpolate the known distribution of single species, assemblages or communities

*Corresponding author: charleneguillaumot21@gmail.com

(Ferrier & Guisan 2006) to little-accessed or undersampled areas (Reiss et al. 2011, Robinson et al. 2011) and help improve our knowledge of the distribution of rare species (McCune 2016).

In regions subject to rapid environmental change and significant anthropogenic activities, SDMs can be useful tools in planning conservation measures (Guisan et al. 2013, Reiss et al. 2015). However, modeling species distributions over vast and remote areas is challenging and raises questions regarding the relevance of this method compared to more traditional and qualitative approaches (Koubbi et al. 2016). In such regions, our knowledge of species distributions is usually based on historical and heterogeneous presence-only data sets, which may include many gaps, and may induce methodological biases that affect the level of SDM performance (Loiselle et al. 2008, Costa et al. 2010, Newbold 2010). The use of historical data in SDMs has been widely discussed (Reutter et al. 2003, Hortal et al. 2007, 2008); for instance, regarding the spatial and temporal heterogeneities induced by the use of different sampling strategies. Limitations to SDM performance are mainly due to uncertainties in data location and detection (Costa et al. 2010, Naimi et al. 2014, Tessarolo et al. 2014), overestimations of habitat suitability in intensively sampled areas (Guillera-Arroita et al. 2015) and artefacts in niche descriptions (Hortal et al. 2008). The lack of available data from remote areas also constitutes a limitation to SDMs, which are restricted to presence-only data and are regarded as being less reliable and less efficient than presence–absence and abundance-based models (Brotons et al. 2004). Over the past few years, many methodological developments in SDM procedures have been produced to correct for such biases (Dormann 2007, Phillips et al. 2009, Barbet-Massin et al. 2012), but no single corrective procedure has emerged (Qiao et al. 2015) and few practical solutions have been proposed to deal with poor and heterogeneous data sets.

Our knowledge of species distribution in the Southern Ocean is still patchy (Koubbi et al. 2016). Therefore, the growing interest of marine biologists and biogeographers in the region has led to the conception of collaborative projects compiling past and present marine biodiversity data in information networks such as the SCAR-Marine Biodiversity Information Network (SCAR-MarBIN) (Griffiths et al. 2011), the Biogeographic Atlas of the Southern Ocean (De Broyer et al. 2014) and other open access databases (Danis et al. 2013, Gutt et al. 2013, Van de Putte et al. 2014). However, running SDMs in the region still requires a significant data compilation effort (Guillaumot et al. 2016) to complement the existing open access data sources and to check for data quality. In addition, modeling Southern Ocean species distributions poses auxiliary problems due to the paucity of data and model performances that can vary with ecological niche width (Qiao et al. 2015). Recent works have developed methodologies to adapt SDMs to rare species and poorly sampled areas, but none have been tested for the Southern Ocean (Pokharel et al. 2016, Phillips et al. 2017).

In this work, we analysed the reliability of modeling procedures with regards to the heterogeneous nature of data available and the gaps in our knowledge of species distributions. We compiled echinoid presence-only data collected from several ancient and recent oceanographic campaigns that have been carried out on the Kerguelen Plateau (sub-Antarctic region) over the past 145 yr. The distributions of 4 echinoid species with contrasting ecological niches were modeled and the reliability and performance of the modeling procedures were tested. We propose methodological procedures to correct for spatial and temporal biases and assess the sensitivity of modeling procedures to a species' ecological niche width. This is the first methodological approach to correct for potential biases in SDMs in the Southern Ocean. Our objective is to offer useful perspectives for future modeling, along with a practical and transferable protocol to test for the reliability and performance of modeling procedures.

MATERIALS AND METHODS

Biological data

Species occurrence data were taken from Guillaumot et al. (2016) and Pierrat et al. (2012). The data set includes presence-only data of echinoid species collected during 19 scientific cruises carried out on the Kerguelen Plateau (46 to 56° S, 63 to 81° E) since 1872 (Fig. 1). Fig. 1B illustrates the expeditions that mainly contributed to the dataset. The full list is available in Guillaumot et al. (2016). Scientific objectives, dates, sampling effort, gears and surveyed areas differed between cruises, leading to spatial and temporal heterogeneities (Guillaumot et al. 2016). From this data set, 4 echinoid species with contrasting ecological preferences and a high number of presence-only records were selected. Species included 2 sediment feeders of the family Schizasteridae (1 shallow water species, *Abatus cordatus*, and a deeper one, *Brisaster antarcticus*), 1 carnivorous/detritivorous and eury-

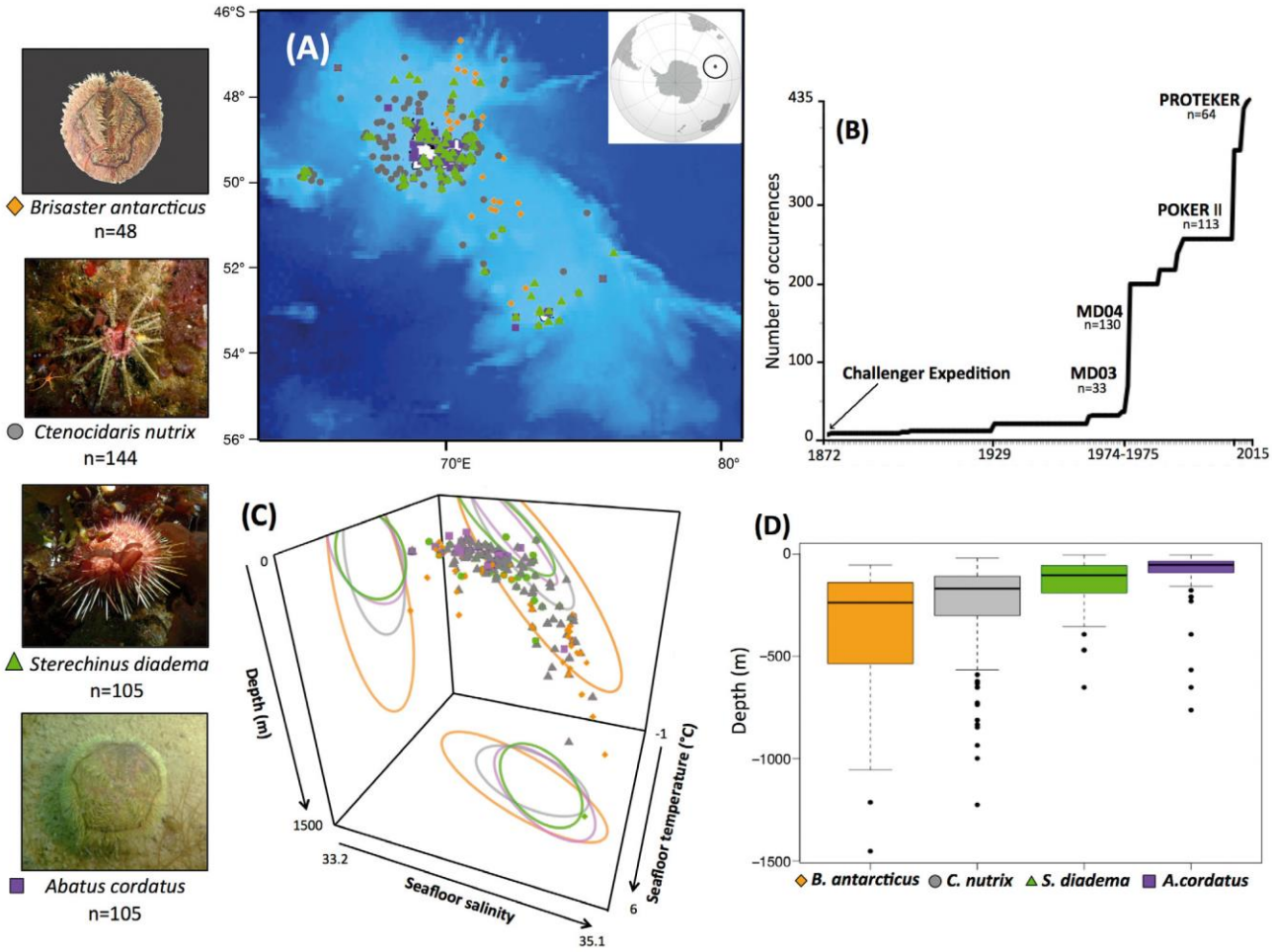


Fig. 1. (A) Occurrence data of the 4 studied echinoid species over the Kerguelen Plateau: *Brisaster antarcticus*, *Ctenocidaris nutrix*, *Sterechinus diadema*, *Abatus cordatus*. (B) Sampling effort (in presence-only records) through time by main scientific cruises during which the 4 studied species were collected on the Kerguelen Plateau. (C) Species presence data plotted according to depth , seafloor salinity and seafloor temperature on the Kerguelen Plateau with projection of standardized distribution ellipsoids (see Jackson et al. 2011 for details) on bivariate plots. (D) Species depth range over the Kerguelen Plateau based on occurrence data (solid line: median; box: upper and lower quartiles; whiskers: 75 ± 1.5% interquartile range; dots: outliers)

bathic species of the family Cidaridae, *Ctenocidaris nutrix* and 1 omnivorous and eurybathic species of Echinidae, *Sterechinus diadema* (David et al. 2005) (Fig. 1). *A. cordatus* is a coastal species endemic to the Kerguelen Plateau, *B. antarcticus* is known to occur in the Kerguelen and Crozet archipelagoes and has broader environmental preferences than *A. cordatus*, and *C. nutrix* and *S. diadema* are widespread in the Southern Ocean and have contrasting environmental preferences (Fig. 1).

Environmental descriptors

Environmental descriptors were taken from Guillaumot et al. (2016). The data set covers the geographic extent of the Kerguelen Plateau and comprises environmental data encompassing 6 decades (1955–2012). Environmental data are available at a grid cell resolution of 10 km. Environmental layers include no-data pixels, particularly in seafloor-related descriptors. Data were not interpolated to avoid potential biases due to interpolation procedures.

Collinearity between descriptors can alter modeling performances (Phillips et al. 2006) because collinear data may (1) inflate standard errors, (2) induce the violation of residual independency during model validation and (3) generate noise that can be interpreted as a link between descriptors (Dormann et al. 2013). To reduce the collinearity effect, we computed the variance inflation factor (VIF) and Spearman correlation coefficient (r_S) between all available descriptors from Guillaumot et al. (2016). VIF analysis was performed in a stepwise procedure using the 'vifstep'

function in the R package 'usdm' (Naimi et al. 2014). Descriptor pairs with high VIF and r_S values were omitted based on the commonly used thresholds of VIF < 5 and r_S < 0.85 (Pierrat et al. 2012, Dormann et al. 2013, Duque-Lazo et al. 2016). Environmental descriptors finally selected to model species distribution are given in Table 1.

Environmental changes were tested between 1955 and 2012. The comparison of pixel values between periods was generated using a Wilcoxon signed-rank test with the Bonferroni correction.

Analytical procedure

The flow chart of Fig. 2 details the analytical procedure used in the present work.

Model selection

Due to the growing interest of ecologists in species distribution modeling, a large range of modeling techniques is now available (Reiss et al. 2011, Guillera-Arroita et al. 2015, Qiao et al. 2015). Running the most appropriate model involves selecting the best modeling technique for the data under analysis and also involves considering the scientific objectives to be addressed (Reiss et al. 2011, Qiao et al. 2015).

Here, we compared several modeling techniques using the 'biomod2' library in R v.3.3.0 (Thuiller et al. 2016) and tested the performance of these approaches with regards to the chronological addition of new data and the transferability performance of models between areas. Several models were generated with an increasing number of occurrence data (see Fig. S1 in Supplement 1 at www.int-res.com/articles/suppl/m594p149_supp.pdf). The best modeling techniques were then compared with each other using a non-random cross-validation procedure (Fig. S2; Wenger & Olden 2012) in order to determine the approach with the best accuracy in transferability performances (Randin et al. 2006, Wenger & Olden 2012).

Results showed high performance and stability values for random forest (RF) and boosted regression trees (BRT) in our case study (see Supplement 1). However, BRT performed better in transferability than did RF (Heikkinen et al. 2012). Previous works have shown that RF does not deal correctly with missing values and patchy data sets (Breiman 2001, Barbet-Massin et al. 2012, Qiao et al. 2015; see Table S1 in Supplement 1 for a review). Therefore, BRT was chosen in the present work to generate the analyses.

BRT calibration was completed using the 'gbm' R package (Elith et al. 2008, Ridgeway 2015). The 3 main parameters (learning rate [lr], tree complexity [tc], bag fraction [bg]) were selected using the method developed by Elith et al. (2008) to determine the combination of values that would minimize the predicted deviance of the models (Elith & Leathwick 2014). The parameters were finally set at lr = 0.0001, tc = 2 and bf = 0.75.

Following Barbet-Massin et al. (2012), we sampled the same number of background data as the number of presence data available for computing BRT models. Considering the low number of presence data points available, 100 model replicates (i.e. background sampling) were generated for each analysis. Finally, to correct for data aggregation in space, presence duplicates were removed when present in the same 10 km resolution pixel.

Model performance was assessed by measuring the area under the receiver operating curve (AUC) of each model replicate using the 'dismo' R library (Hijmans et al. 2016). AUC expresses the relationship between model sensitivity and the commission error (1 − specificity), where sensitivity corresponds to the number of presence pixels correctly predicted as present, and specificity is the number of absence pixels correctly predicted as absent (Fielding & Bell 1997). The use of the AUC to evaluate SDM performance has been debated (Lobo et al. 2008, Peterson et al. 2008), but the AUC remains the most appropriate metric for presence-background models since values remain stable with low-prevalence data sets and are not sensitive to threshold effects (Hand 2009, van Proosdij et al. 2016). Following the recommendation of Jiménez-Valverde (2012), we used the AUC to estimate the robustness of the models but not for direct comparisons between models that were generated for different species, on different study areas or with different training samples.

Correcting for sampling bias

The data collected during the various scientific cruises over the Kerguelen Plateau over the last 145 yr present conspicuous spatial heterogeneities. The resulting biases can generate an unequal number of records in different sectors of the study area and heterogeneous patterns in record distribution. Such heterogeneities can increase the risk of overestimating the contribution of environmental conditions to the models in the most frequently sampled areas (Araújo & Guisan 2006).

Table 1. Environmental descriptors selected for species distribution models. *indicates that environmental layers were available for the following time periods: 1955–2012, 1955–1964, 1965–1974, 1975–1994 and 2005–2012. Minimum and maximum values are shown for the period 1955–2012. Spatial resolution of layers: 10 km grid-cell pixels. Spatial extent: 46–56° S, 63–81° E

Environmental descriptors	Units	Description	Min. value	Max. value	Source
Depth	m	Bathymetric grid around the Kerguelen Plateau	−4977.0000	−1.0000	This study. Derived from the Biogeographic Atlas of the Southern Ocean (De Broyer et al. 2014)
Sea surface mean temperature*	°C	Mean sea surface temperature	3.0566	7.6223	World Ocean Atlas (2013)
Sea surface temperature amplitude*	°C	Amplitude between mean summer and mean winter sea surface temperature	−3.3036	−1.4108	World Ocean Atlas (2013)
Seafloor mean temperature*	°C	Mean seafloor temperature	−0.2978	4.6422	This study. Derived from World Ocean Atlas (2013) sea surface temperature layers
Seafloor temperature amplitude*	°C	Amplitude between mean summer and mean winter seafloor temperature	−2.5757	0.8867	This study. Derived from World Ocean Atlas (2013) sea surface temperature layers
Sea surface mean salinity*	PSS	Mean sea surface salinity	33.6849	33.8251	World Ocean Atlas (2013)
Sea surface salinity amplitude*	PSS	Amplitude between mean summer and mean winter sea surface salinity	−0.0859	0.3165	World Ocean Atlas (2013)
Seafloor salinity amplitude*	PSS	Amplitude between mean summer and mean winter seafloor salinity	− 169	0.0937	This study. Derived from World Ocean Atlas (2013) sea surface salinity layers
Mean surface chl *a*	mg m^{-3}	Surface chlorophyll *a* concentration. Summer mean over 2002–2009	0.1358	2.7324	MODIS AQUA (NASA) 2010
Sediments	Categorical	Sediment features	14 categories		McCoy (1991), updated by H. J. Griffiths (unpubl. data)
Geomorphology	Categorical	Geomorphologic features	27 categories		ATLAS ETOPO2 2014 (Douglass et al. 2014)
Slope	Degrees	Bathymetric slope	4.8229×10^{-5}	0.1547	Biogeographic Atlas of the Southern Ocean (De Broyer et al. 2014)
Mean seafloor oxygen concentration	ml l^{-1}	Mean seafloor oxygen concentration over 1955–2012	4.0080	7.6223	This study. Derived from World Ocean Atlas (2013) sea surface oxygen concentration layers

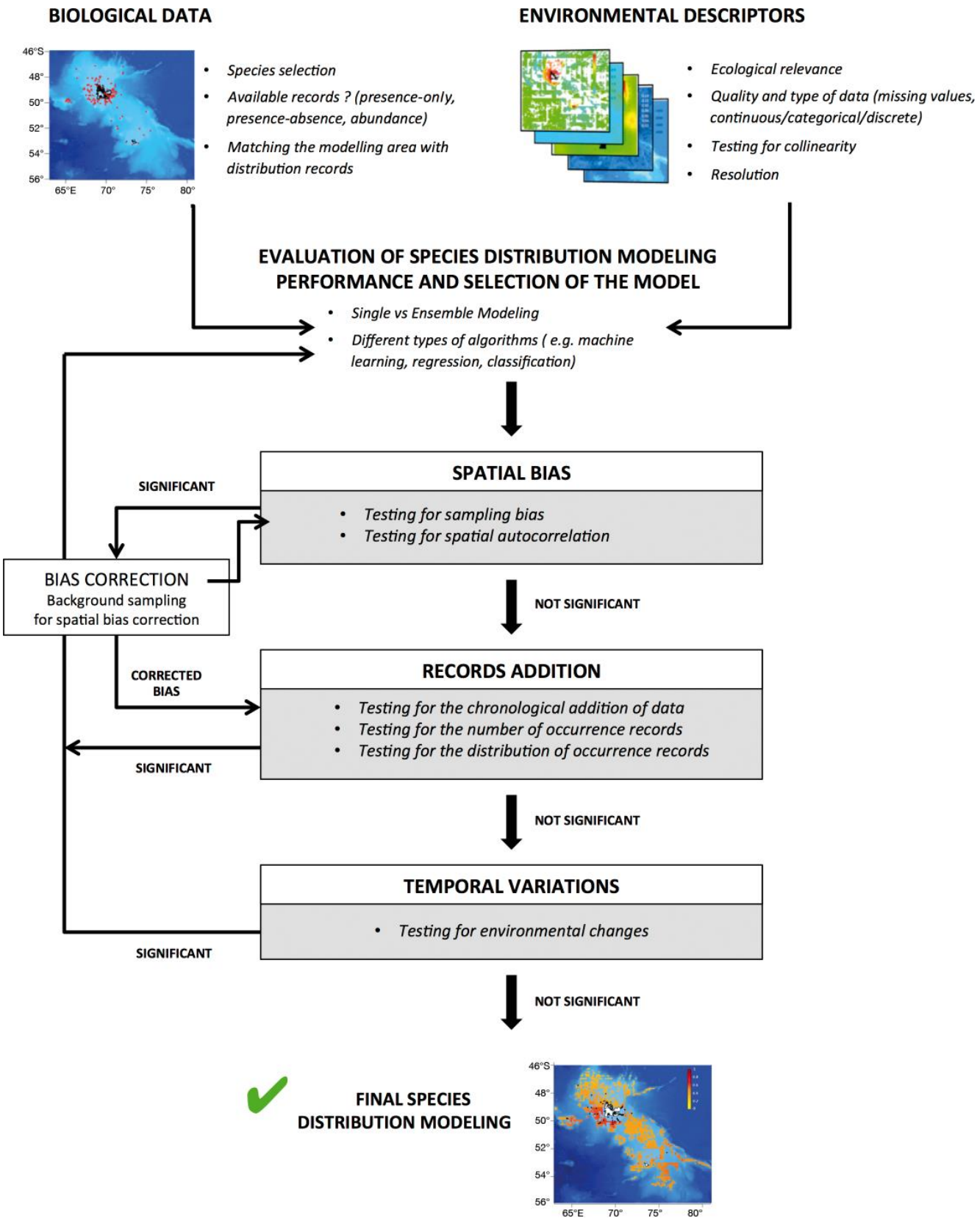


Fig. 2. Tests and procedures carried out in the present work. Arrows indicate the stepwise procedure with statistical validation leading either to the following step or correction/stepback requirements

The effect of spatial heterogeneities on the quality of distribution models was tested using a null model approach. The first null model (null model #1) was generated by sampling presence data at random within the total set of sites that were visited during the different campaigns, whether echinoid specimens were collected at these sites or not (see Fig. 3). Because absence data were not available, this approach allowed us to assess the weight of sampling bias in the models. If a sampling bias is significant, null model #1 is expected to produce distribution maps with higher suitability values in the most frequently sampled areas (Merckx et al. 2011).

A second null model (null model #2) was built by simulating presence data sampled at random over the entire study area. Null model #2 was expected to produce distribution maps of equal suitability over the entire study area. If sampling is spatially biased, we expect that null model #1 would deviate from null model #2 (Raes & ter Steege 2007).

The 2 null models were generated for the 4 selected species. The number of presence-only data used in the models was contained between the number of data points collected from the MD04 campaign until the PROTEKER campaign, between 1974 and 2015, which corresponds to periods of high sampling effort (Fig. 1B). In each null model, 100 replicates were produced. Time-averaged environmental descriptors (1955–2012) were used for the analysis.

To correct for sampling bias when null models #1 and #2 significantly differed from each other, we used the methodology proposed by Phillips et al. (2009), which has been shown to improve modeling performance (Phillips et al. 2009, Aguirre-Gutiérrez et al. 2013). A grid layer was built using a kernel density estimation (KDE) to represent spatial sampling bias. The layer was calculated from the map of visited sites. The estimated proportion of presence-only data present in each pixel was determined using the 'kde2d' function of the 'MASS' R package (Venables & Ripley 2002). Background data were sampled according to the weighting scheme of the KDE layer, to reduce discrepancies between presence-only records and background data (Phillips et al. 2009, Barbet-Massin et al. 2012). In order to test for the efficiency of model correction based on the KDE, Pearson's r correlation was computed between pixel values of the KDE layer (the proxy for sampling effort) and the predicted probabilities of models after the KDE correction.

Spatial heterogeneities in data collection can also generate spatial autocorrelation (SAC) between presence records, which can violate model calibration assumptions and affect model accuracy with incorrect parameter estimations (Segurado et al. 2006, Dormann 2007, Crase et al. 2012). Several approaches have been developed to take SAC into account in SDMs (see Crase et al. 2012 for a review). They consist of including an additional term in the models (the auto-covariate) which represents the influence of neighboring records on modeling predictions. The significance of SAC was tested using the Moran *I* autocorrelation index computed on model residuals (Luoto et al. 2005, Crase et al. 2012) for both original and corrected models. Models were built using time-averaged environmental descriptors (1955–2012).

Testing for the effect of the chronological addition of new records on model performance

Our data set consisted of presence-only data collected during various scientific cruises with distinct sampling protocols, which may alter the performance of the models (Fig. 1). To test for model reliability, we separately analysed the influence of (1) the chronological addition of presence records, (2) data number alone and (3) sampling patterns (the distribution of data in space). The analyses were performed for *A. cordatus*, *C. nutrix* and *S. diadema*; not enough data were available for *B. antarcticus*. We used time-averaged environmental descriptors (1955–2012) to generate the models.

To test for the potential effect of the chronological addition of new data on model performance, we followed the protocol proposed by Aguiar et al. (2015). The data set was split into distinct subsets corresponding to main periods of sampling effort (1975, including Marion Dufresne campaigns; 1993, including ANARE campaigns; 2010, including POKER II campaign; 2015, including PROTEKER campaigns). New presence data were progressively added to the models, following the chronological collection of new records. The influence of the chronological addition of data was assessed by measuring the correlation between models using Schoener's *D* statistic. Schoener's *D* is a correlation metric adapted to the study of niche similarities (Warren et al. 2008, Rödder & Engler 2011). It evaluates the similarity of pixel values between 2 distribution grids. A *D* value of 0 means that the 2 maps are perfectly different, and a *D* value of 1 means that maps are perfectly similar. Values were computed using the 'niche.overlap' function of the 'ENMeval' R package (Muscarella et al. 2014).

The significance of correlations was tested following a null model protocol, using 100 replicates, pairwise-

compared using the Schoener's *D* statistic (Raes & ter Steege 2007, Warren et al. 2008, Ficetola et al. 2009).

The distinct effect of data addition and sampling patterns were tested separately. To test for the effect of data addition alone, models were built by sampling an increasing number of presence data at random in the total area for *A. cordatus* (n = 54, 76, 95), *C. nutrix* (n = 46, 54, 106, 114) and *S. diadema* (n = 54, 66, 98). These thresholds correspond to the number of presence-only data used in the chronological addition analysis.

Finally, to test for the effect of sampling patterns, different models were produced by sampling presence data at random either within a subset of real data collected along transects (MD03 campaign) or within a subset of real data collected at random (POKER II, PROTEKER campaigns). All models were compared with each other.

Testing for the effect of temporal variations on model performance

To test for the effect of environmental shifts on the models, different distribution models were generated using distinct environmental descriptors for 4 periods (1955–1964; 1965–1974; 1975–1994; 2005–2012) and the complete set of presence data available. Similarities between models were measured using Schoener's *D* statistic.

RESULTS

Environmental shifts

Mean sea surface temperature and amplitude, mean seafloor temperature and amplitude and mean sea surface salinity and amplitude all differed significantly among all studied decades ($p < 0.001$). Only seafloor temperature amplitude did not significantly differ between the time periods 2005–2012 and 1955–1964. These results indicate that significant environmental shifts occurred during the studied time period, and this may induce important variations in the models since the data set extends over 145 yr.

Spatial bias

Null model #1 predicted higher suitability values in areas with the most intense sampling effort, corresponding to the northern part of the Kerguelen Plateau and the vicinity of the Kerguelen archipelago (Fig. 3A). In contrast, null model #2 predicted medium suitability values over the entire Kerguelen Plateau because presence data were sampled randomly in the area (Fig. 3B). The difference between null models #1 and #2 was significant for the 4 species (Fig. 3), showing that sampling bias has a significant impact on model outputs, which will overestimate environment suitability in areas with the highest number of sampling sites if no correction is applied.

Correlation between visited areas and predicted probability distribution decreased in models built with the KDE-correction compared to non-corrected models (Table 2), showing that the correction is efficient at reducing the influence of sampling bias on modeling performance. However, the correction proved less efficient in models of the coastal and narrow niche species *Abatus cordatus*, for which correlation values after the KDE correction remained high (r = 0.44) (Table 2).

SAC was significant for non-corrected models (Moran index, $I_{min} = 0.05$, $I_{max} = 0.16$) but values were not significant in corrected models ($I_{min} = 0.04$, $I_{max} = 0.06$), except for *A. cordatus* (see Table S2 and Fig. S3 in Supplement 2). This shows that the KDE procedure corrected for SAC in 3 of the 4 studied species.

Chronological addition of new records

The different models built with a chronological addition of new data showed high AUC values (mean ± SD) for *Ctenocidaris nutrix* and *A. cordatus* ($0.814 \pm 0.018 < AUC_{C.nutrix} < 0.883 \pm 0.024$ and $0.908 \pm 0.023 < AUC_{A.cordatus} < 0.909 \pm 0.018$ respectively), demonstrating the relevance of all models (Fig. 4, see Fig. S4 in Supplement 3). For these 2 species, Schoener's *D* correlation values were high (mean ± SD; $D_{A.cordatus} = 0.978 \pm 0.023$, $D_{C.nutrix} = 0.968 \pm 0.020$) and significant, showing that the models were similar to each other. (see Table S3 in Supplement 3)

In contrast, models generated for *Sterechinus diadema* significantly differed from each other with lower Schoener's *D* statistics ($D_{S.diadema} = 0.932 \pm 0.036$) (see Fig. S5 in Supplement 3). Therefore, the chronological addition of new data has contrasting impacts on model outputs in the studied species, which may be explained by the sensitivity of models to data addition and to sampling patterns.

Data addition and sampling patterns

Comparison of models produced with an increasing number of data points presents high and signifi-

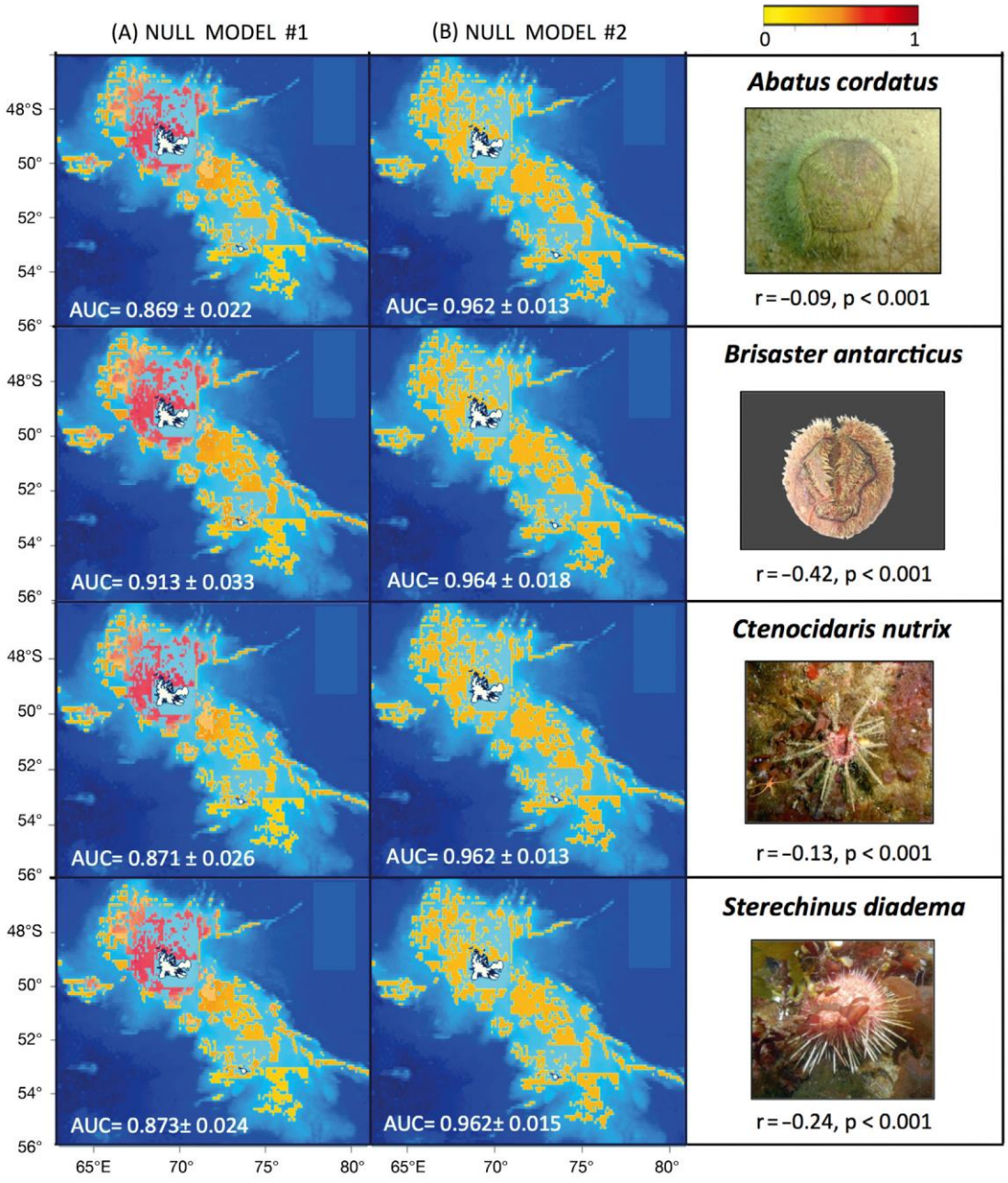


Fig. 3. (A) Null model #1 and (B) null model #2 for the different species under study. Mean (±SD) area under the receiver operating curve (AUC) values are given for the 100 replicates. Comparisons between models compiled with Pearson's r correlation values and associated probabilities (colour bar)

cant Schoener's D values (minimum $D = 0.979 \pm 0.031$ for *S. diadema*, maximum $D = 0.985 \pm 0.020$ for *C. nutrix*), showing that model outputs did not vary significantly with increasing data in our case study (Table 3).

To test for the influence of sampling patterns, models built using subsets with contrasting distribution patterns (radial versus random patterns) were compared. Schoener's D statistics measured between these 2 types of models presented low values, suggesting a significant influence of sampling pattern on model output (Table 3).

Environmental change and model performance

The different models generated with contrasting environmental descriptors were highly similar, as

Table 2. Pearson's r correlation of pixel values between the kernel density estimation (KDE) layer and the predicted probability of each species model. Statistic probabilities are all < 0.05

	Before KDE correction	After KDE correction
Abatus cordatus	0.72	0.44
Brisaster antarcticus	0.60	–0.17
Ctenocidaris nutrix	0.80	0.11
Sterechinus diadema	0.61	0.20

shown by high Schoener's *D* and low standard deviation values (*D* = 0.981 ± 0.005). This proves that environmental shifts have no significant impact on model outputs. In addition, the respective contributions of environmental descriptors to models did not vary significantly among periods for the 4 species. However, *A. cordatus* seems to be less impacted by environmental shifts than the other species (Fig. 5).

Finally, the contribution of time-averaged environmental descriptors over the total studied period (1955–2012) differed from contributions computed for each decade separately (Fig. 5).

Final species distribution models

Sampling bias analyses and model corrections showed that reliable distribution models can be built for *C. nutrix* only; this was the only data set in which spatial and temporal heterogeneities did not impact prediction performances significantly. A final, reliable model was produced for *C. nutrix* over the Kerguelen Plateau (Fig. 6).

DISCUSSION

Data scarcity and heterogeneity

First research surveys of the Kerguelen Plateau date back to the oceanographic campaign of the HMS Challenger in 1872. One and a half centuries later, our knowledge of benthic species distribution on the Kerguelen Plateau has significantly increased, but remains patchy (Koubbi et al. 2016). As in most parts of the Southern Ocean, modeling species distributions on the Kerguelen Plateau faces significant limitations due to gaps and heterogeneities in the data (Guillaumot et al. 2016). Such limitations can seriously limit the relevance of modeling procedures, which are re-

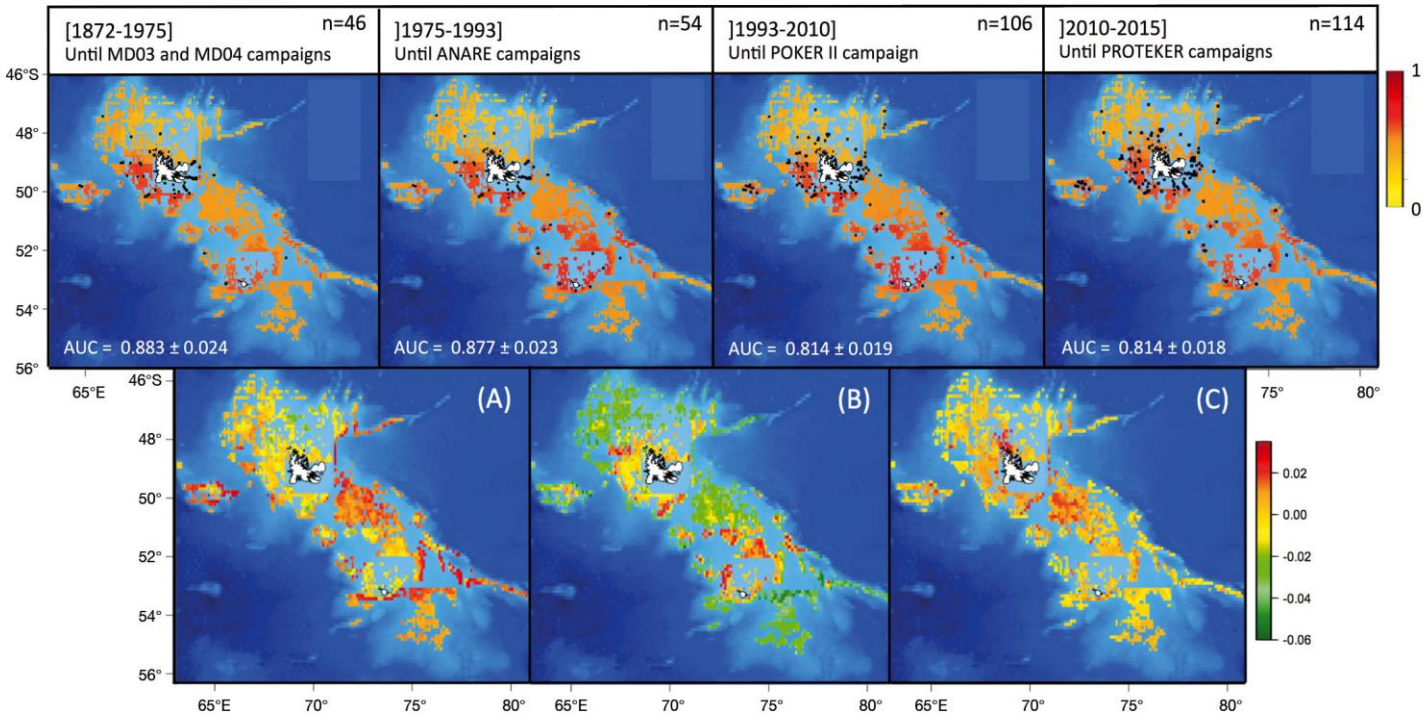


Fig. 4. First row: distribution models of *Ctenocidaris nutrix* for 4 periods, with increasing number of presence data points to build the model (averaged maps of 100 model replicates). Colour bar: probabilities of distribution predicted by the model (between 0 and 1). Second row: difference in probability distribution between (A) n = 54 and n = 46, (B) n = 106 and n = 54 and (C) n = 114 and n = 106. Colour bar represent differences in distribution probabilities between maps

Table 3. Influence of data addition and sampling patterns on models for *Abatus cordatus*, *Ctenocidaris nutrix* and *Sterechinus diadema*. Data addition: mean (±SD) Schoener's *D* and associated p-value computed between models (100 replicates) produced for the different species with n = 54, 76, 95, n = 46, 54, 106, 114 and n = 54, 66, 98 occurrences randomly sampled from the total dataset. Sampling pattern: Schoener's *D* and associated p-value computed between models (100 replicates) produced with subsets contrasting in data distribution patterns (transect versus random sampling)

Species	Data addition		Sampling pattern	
	D_{obs}	p	D_{obs}	p
Abatus cordatus	0.981 ± 0.025	<0.05	–	
Ctenocidaris nutrix	0.985 ± 0.020	<0.05	0.941 ± 0.030	0.147
Sterechinus diadema	0.979 ± 0.031	<0.05	0.842 ± 0.040	0.941

quired by environmental managers for conservation purposes (Féral et al. 2016, Koubbi et al. 2016). In the present work, we followed a step-by-step protocol to assess, quantify and correct the potential effects of data scarcity and heterogeneity on SDMs, a critical issue when considering the growing interest for modeling approaches in Antarctic and sub-Antarctic regions (Gutt et al. 2012). Our results demonstrate that such approaches can prove feasible and reliable in certain case studies, when data quality and sampling bias can be tested and corrected.

Coping with spatial and temporal bias in presence-only datasets

Spatial bias and SAC

Building SDMs for remote and little-accessed regions often requires the use of spatially biased data sets conditioned by sampling caveats. Because parts of these regions that are the most easily accessed aggregate most of the available presence data, more weight is given to the most frequently sampled sites, and thus model performance is reduced (Phillips et al. 2009). In the present work, a significant difference was measured between the 2 null models (generated by selecting presence data either from visited stations only or from random sites over the total investigated area), highlighting the strong heterogeneity of

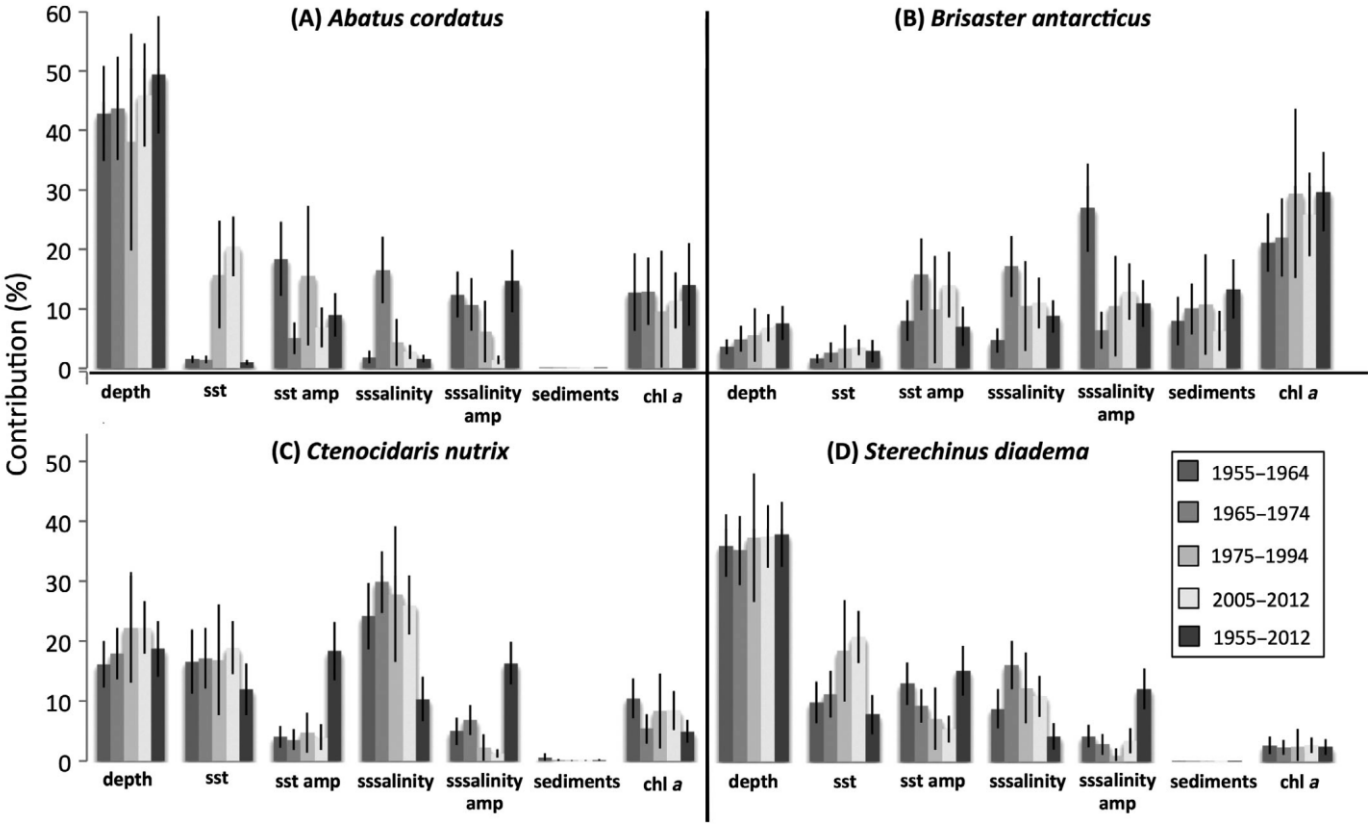


Fig. 5. Mean (±SD) contributions of environmental descriptors to the models for the 4 time periods and species under study. sst: sea surface temperature; sst amp: sea surface temperature amplitude; sssalinity: sea surface salinity; sst amp: sea surface salinity amplitude; chl a: chlorophyll *a* (see Guillaumot et al. 2016 for details)

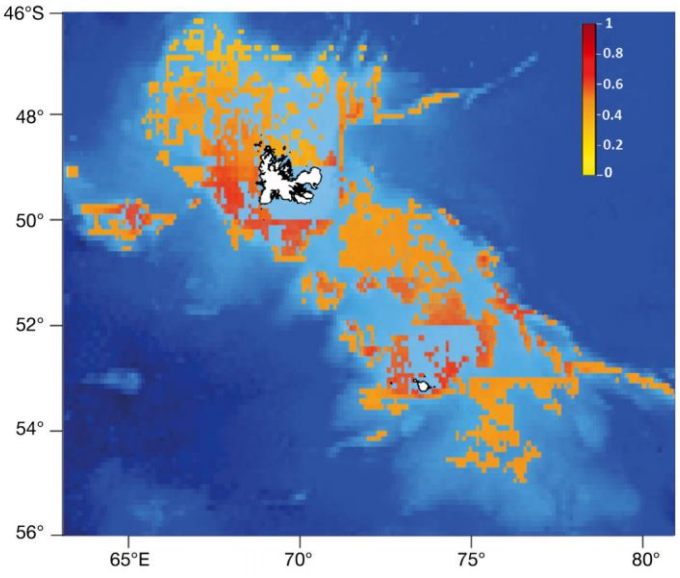


Fig. 6. Species distribution model generated for *Ctenocidaris nutrix* using all presence-only data available and present environmental descriptors (2005–2012). Mean (±SD) area under the receiver operating curve (AUC) = 0.813 ± 0.02

sampling effort with more data collected in the northern part of the Kerguelen Plateau and in shallow coastal areas.

The significant SAC values that were computed from model residuals also reveal the impact of sampling bias. The significance of SAC on uncorrected model residuals can be partly explained by the relative accumulation and high density of presence data in shallow areas of the Kerguelen Plateau, where species presence probability is over-predicted. One could argue that SAC analysis does not apply to SDMs, as species presence proximities must be considered in the environmental niche space, not in the geography. However, in the present study, the difference between null models constitutes operational evidence of the impact of sample clumping on model outputs, which is also revealed by significant SAC values.

To correct for sampling bias, we used a background-based correction method (Phillips et al. 2009) that was previously used in studies based on presence-only and limited data sets (Mateo et al. 2010, Pokharel et al. 2016, Phillips et al. 2017). These methods allowed us to reduce the effect of sample spatial bias on modeling performance by weighting background records according to sampling patterns. In the present study, the correction was proven to be efficient to correct both for the influence of uneven sampling effort on predicted distributions (Table 2) and for SAC on all SDMs except for models of *Abatus cordatus*. *A. cordatus* is a coastal, shallow marine species that was mainly sampled in the northern part of the Kerguelen Plateau. Species presence records are strongly conditioned by the location of the most intense sampling efforts. This is in line with previous studies that highlighted the difficulties of modeling the distribution of narrow-niche species with low prevalence distribution (i.e. corresponding to the proportion of the area where presence records are located) (Barbet-Massin et al. 2012, Qiao et al. 2015). In small presence-only datasets, the methodologies used to correct for spatial bias are not as efficient for narrow-niche species as for broader-niche species. Reducing the extent of distribution modeling of narrow-niche species to the boundaries of their environmental limits could prove a good alternative.

Influence of record addition

The chronological addition of new data had a limited impact on certain model outputs, as demonstrated by high similarities between the chronological models generated for *A. cordatus* and *Ctenocidaris nutrix*. In contrast, chronological models of *Sterechinus diadema* differed significantly from each other. A detailed analysis of data increments proved that the increasing number of presences had no impact on modeling performance, which is not in line with previous works (Stockwell & Peterson 2002, Wisz et al. 2008). However, these results can be altered by our incomplete knowledge of full species distributions due to sampling bias and the limited number of data sets available (Hernandez et al. 2006, Bean et al. 2012). With *S. diadema*, differences between the chronological models were due to contrasted spatial patterns between data sets (transects versus random patterns).

Historical data and environmental change

Significant environmental shifts were measured for the descriptors analysed between 1955 and 2012 over the Kerguelen Plateau (i.e. mean sea surface temperature and amplitude, mean surface salinity and amplitude). However, for all species, distribution models built for each decade were highly similar to each other. These results confirm that temporal heterogeneities in data sets do not necessarily impact the robustness of the models, because spe-

cies preferences for their environment may be wider than the magnitude of changes in time. Working with both present and historical data to improve the completeness of occurrence records proved reliable when assuming that species niche and distribution have not significantly changed during the studied time period.

Between 1955 and 2012, the respective contributions of temperature and salinity to the models did not vary over the range of within-decade variation for *B. antarcticus*, *C. nutrix* or *S. diadema*; variations between decades were more marked in models produced for *A. cordatus*. This near-shore species is found in shallow waters of the Kerguelen and Heard islands, where environmental descriptors include many no-data pixels (Guillaumot et al. 2016). Consequently, the varying contributions of temperature and salinity to the models of *A. cordatus* between decades cannot be attributed with certainty to the effect of environmental change, but rather to modeling limitations.

Sea surface temperature and salinity amplitudes contributed significantly to the models, contributing more than the averaged parameters (i.e. *A. cordatus* and *B. antarcticus*; Fig. 5). This is in line with the results of Bradie & Leung (2017), who tested for the contribution of several environmental descriptors across a wide panel of taxa. They showed the importance of including seasonal means and extremes in models to further depict species distributions, considering their stronger relationships with species niche width and ecological traits (i.e. growth and survival; see Franklin 2009).

Using time-averaged descriptors over the entire period (1955–2012) may be considered the best approach to produce representative models, independent of short-term environmental variations. Unexpectedly, our results showed that for all species, contributions of time-averaged descriptors to the models were much more different than all differences between decadal descriptors (Fig. 5). This suggests that using time-averaged descriptors for long time periods does not necessarily improve model reliability compared to using descriptors averaged over shorter time periods. This also highlights the importance of the descriptor selection in modeling procedures, a critical issue for improving model performance as already stressed in previous studies (Bradie & Leung 2017). This is particularly relevant for certain regions of the Southern Ocean, such as the Western Antarctic Peninsula, which has experienced among the most significant environmental changes in the world's oceans during the last decades (Turner et al. 2014).

Influence of species niche width in modeling performances

Among the 4 studied species, *A. cordatus* has the narrowest ecological niche and most restricted distribution in the vicinity of coastal areas of the Kerguelen and Heard archipelagoes. Such limited geographic and environmental distributions compared to the total extent of the studied area implies that similar environmental conditions prevail in geographically close occurrence sites. This induces a strong SAC pattern that explains the difficulties encountered when correcting for spatial bias compared to other species models. Moreover, the limited environmental variability between coastal sampling sites of the different oceanographic surveys can also explain the absence of a data-addition effect on modeling performances for *A. cordatus*.

In contrast, *C. nutrix* and *S. diadema* have wider ecological niches than *A. cordatus* (Fig. 1). For these 2 species, record data are more widely distributed and show contrasting sampling patterns (i.e. transect-like versus random patterns) that were shown to influence modeling performance in *S. diadema* only (Table 3). This can be explained by the higher number of presence records available for *C. nutrix* (n = 114 and 98 for *C. nutrix* and *S. diadema* respectively) that allowed a more complete survey of *C. nutrix* distribution. Finally, only the *C. nutrix* data set contained the quality and number of occurrence records that fulfilled all methodological requirements to produce a reliable distribution model.

Considering species niche width in order to cope with spatial and temporal bias in SDMs is important, as already shown by Tessarolo et al. (2014) who studied the influence of survey designs on the performance of distribution models for endemic species with narrow ecological niches. They concluded that survey designs have a low impact on models in comparison with the effect of niche width, number of data points and type of modeling technique used. However, they did not generate any analysis of species with broad ecological niches as a comparison. Our results are also in line with other modeling studies in which distribution models of species with broad niches were the least stable (Reiss et al. 2011, Guo et al. 2015, Qiao et al. 2015, Ranc et al. 2017).

CONCLUSIONS

The use of SDMs has gained importance during the last decades, providing complementary information

for environmental managers. Modeling results can help interpolate species distributions, identify the potential drivers of a species' distribution and predict the potential effects of environmental changes on habitat suitability. However, modeling species distributions over vast and remote marine areas like the Southern Ocean using poor and heterogeneous data sets remains challenging, and improvement of biological and environmental data sets is still required.

In the present study, we showed that reliable SDMs can be produced in such areas as long as the amount and quality of data allow testing and correcting for the effects of biases. Using historical data requires proper environmental descriptors for modeling the effect of environmental changes on species distributions. Using time-averaged predictors over long time periods can generate unrealistic models.

Model selection is also crucial at this stage and the statistical performance of models is not the only criteria to be considered. Modeling procedures must be chosen with regards to the scientific issues that are being addressed. Two procedures (BRT and RF) performed best in our case study, but one of them (BRT) proved to be more relevant because it dealt better with transferability and data patchiness.

Modeling species distributions in data-poor areas poses the practical problem of the minimum number of presence-only data points required to run reliable models, although this is not the only or most critical issue. The number of occurrence records must be high enough for testing model robustness and reliability. In regions with limited access, sampling effort may be heterogeneous, which influences model performance. We showed that sampling bias can be corrected, but the efficiency of the correction depends on species niche width, with narrow-niche species models being more troublesome to correct. In our study, *A. cordatus* is a species limited to shallow coastal areas, which implies a strong correlation between species occurrence and sampling patterns. Restricting the model to a more reduced area could allow for correction of spatial bias and improve modeling performance.

There is also a crucial need for improving the quality of data sets (Kennicutt et al. 2014) and running more accurate models to better tackle conservation issues (Rodríguez et al. 2007, Guisan et al. 2013). For the time being, producing uncertainty maps can be an alternative (Rocchini et al. 2011, Tessarolo et al. 2014) and can provide additional information to environmental managers and stakeholders (Addison et al. 2013, Guisan et al. 2013).

Model reliability and performance also depend on the interaction between data set completeness and a species' intrinsic ecological properties. Hence, we showed that the type and width of ecological niches are important to consider, with the distribution of narrow-niche species being easier to model and less sensitive to incomplete data sets (Guo et al. 2015, Ranc et al. 2017). However, narrow niches usually imply that species are distributed over small areas, for which distribution models will be highly sensitive to extrapolations.

Our protocol showed that reliable SDMs can be produced when enough data are available and data set bias can be tested and corrected. In the present study, only one SDM (*C. nutrix*) could be corrected for spatial and temporal heterogeneities to generate reliable distribution predictions. However, our results stress the need to consider methodological issues when modeling species distributions based on poor and spatially biased data sets, and should contribute to bringing new insights and enhancing modeling performance in future studies.

Acknowledgements. This work is a contribution to the IPEV program PROTEKER funded by the French polar institute (IPEV program no.1044) and contribution no. 21 to the vERSO project (www.versoproject.be) funded by the Belgian Science Policy Office (BELSPO, contract no. BR/132/A1/vERSO). We thank the 19 scientific cruises for the collection of the data used to realise this work (Table 1, Guillaumot et al. 2016, see also Supplement 3).

LITERATURE CITED

Addison PF, Rumpff L, Bau SS, Carey JM and others (2013) Practical solutions for making models indispensable in conservation decision-making. Divers Distrib 19:490–502

Aguiar LMDS, Rosa ROL, Jones G, Machado RB (2015) Effect of chronological addition of records to species distribution maps: the case of *Tonatia saurophila maresi* (Chiroptera, Phyllostomidae) in South America. Austral Ecol 40:836–844

Aguirre-Gutiérrez J, Carvalheiro LG, Polce C, van Loon EE, Raes N, Reemer M, Biesmeijer JC (2013) Fit-for-purpose: species distribution model performance depends on evaluation criteria–Dutch hoverflies as a case study. PLOS ONE 8:e63708

Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. J Biogeogr 33:1677–1688

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: How, where and how many? Methods Ecol Evol 3: 327–338

Bradie J, Leung B (2017) A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. J Biogeogr 44:1344–1361

Breiman L (2001) Random forests. Mach Learn 45:5–32

Brotons L, Thuiller W, Araújo MB, Hirzel AH (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27: 437–448

Costa GC, Nogueira C, Machado RB, Colli GR (2010) Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. Biodivers Conserv 19:883–899

Crase B, Liedloff AC, Wintle BA (2012) A new method for dealing with residual spatial autocorrelation in species distribution models. Ecography 35:879–888

Danis B, Van de Putte A, Renaudier S, Griffiths H (2013) Connecting biodiversity data during the IPY: the path towards e-polar science. In: Verde C, di Priso G (eds) Adaptation and evolution in marine environments, Vol 2. Springer, Berlin, p 21–32

David B, Choné T, Mooi R, De Ridder C (2005) Antarctic Echinoidea. Synopses of the Antarctic benthos, Vol 10. Koeltz Scientific, Königstein

De Broyer C, Koubbi P, Griffiths H, Grant A and others (2014) Biogeographic atlas of the Southern Ocean. Scientific Committee on Antarctic Research, Cambridge

Dormann CF (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Glob Ecol Biogeogr 16:129–138

Dormann CF, Elith J, Bacher S, Buchmann C and others (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46

Douglass LL, Turner J, Grantham HS, Kaiser S, Constable A and others (2014) A hierarchical classification of benthic biodiversity and assessment of protected areas in the Southern Ocean. PLOS ONE 9:e100551

Duque-Lazo J, Van Gils HAMJ, Groen TA, Navarro-Cerrillo RM (2016) Transferability of species distribution models: the case of *Phytophthora cinnamomi* in southwest Spain and southwest Australia. Ecol Modell 320:62–70

Elith J, Leathwick J (2014) Boosted regression trees for ecological modeling. www.lcis.com.tw/paper_store/paper_store/brt(5)-2015115131033846.pdf

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802–813

Féral JP, Saucède T, Poulin E, Marschal C and others (2016) PROTEKER: implementation of a submarine observatory at the Kerguelen Islands (Southern Ocean). Underwat Technol 34:3–10

Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. J Appl Ecol 43:393–404

Ficetola GF, Thuiller W, Padoa-Schioppa E (2009) From introduction to the establishment of alien species: bioclimatic differences between presence and reproduction localities in the slider turtle. Divers Distrib 15:108–116

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24:38–49

Franklin J (2009) Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge

Greathead C, González-Irusta JM, Clarke J, Boulcott P, Blackadder L, Weetman A, Wright PJ (2015) Environmental requirements for three sea pen species: relevance to distribution and conservation. ICES J Mar Sci 72: 576–586

Griffiths HJ, Danis B, Clarke A (2011) Quantifying Antarctic marine biodiversity: the SCAR-MarBIN data portal. Deep Sea Res II 58:18–29

Guillaumot C, Martin A, Fabri-Ruiz S, Eléaume M, Saucède T (2016) Echinoids of the Kerguelen Plateau: occurrence data and environmental setting for past, present, and future species distribution modelling. ZooKeys 630:1–17

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A and others (2015) Is my species distribution model fit for purpose? Matching data and models to applications. Glob Ecol Biogeogr 24:276–292

Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I and others (2013) Predicting species distributions for conservation decisions. Ecol Lett 16:1424–1435

Guo C, Lek S, Ye S, Li W, Liu J, Li Z (2015) Uncertainty in ensemble modelling of large-scale species distribution: effects from species characteristics and model techniques. Ecol Modell 306:67–75

Gutt J, Zurell D, Bracegridle T, Cheung W and others (2012) Correlative and dynamic species distribution modelling for ecological predictions in the Antarctic: a cross-disciplinary concept. Polar Res 31:11091

Gutt J, Barnes D, Lockhart SJ, van de Putte A (2013) Antarctic macrobenthic communities: a compilation of circumpolar information. Nat Conserv 4:1–13

Hand DJ (2009) Measuring classifier performance: a coherent alternative to the area under the ROC curve. Mach Learn 77:103–123

Heikkinen RK, Marmion M, Luoto M (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography 35:276–288

Hernandez PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773–785

Hijmans RJ, Phillips S, Leathwick J, Elith J (2016) dismo: species distribution modeling. R package version 1.1-1. https://CRAN.R-project.org/package=dismo

Hortal J, Lobo JM, Jiménez-Valverde A (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. Conserv Biol 21:853–863

Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos 117:847–858

Jackson AL, Inger R, Parnell AC, Bearhop S (2011) Comparing isotopic niche widths among and within communities: SIBER–stable isotope Bayesian ellipses in R. J Anim Ecol 80:595–602

Jiménez-Valverde A (2012) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Glob Ecol Biogeogr 21:498–507

Kennicutt MC, Chown SL, Cassano JJ, Liggett D and others (2014) Six priorities for Antarctic science. Nature 512: 23–25

Koubbi P, Mignard C, Causse R, Da Silva O and others (2016) Ecoregionalisation of the Kerguelen and Crozet islands oceanic zone. Part I: Introduction and Kerguelen oceanic zone. Commission for the Conservation of Antarctic Marine Living Resources Report, Working Group on Ecosystem Monitoring and Management No.16/43 CCAMLR, Bologna

Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17:145–151

Loiselle BA, Jørgensen PM, Consiglio T, Jiménez I, Blake JG, Lohmann LG, Montiel OM (2008) Predicting species distributions from herbarium collections: Does climate bias in collection sampling influence model outcomes? J Biogeogr 35:105–116

Luoto M, Pöyry J, Heikkinen RK, Saarinen K (2005) Uncertainty of bioclimate envelope models based on the geographical distribution of species. Glob Ecol Biogeogr 14: 575–584

Mateo RG, Croat TB, Felicísimo AM, Muñoz J (2010) Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections.

Divers Distrib 16:84–94
McCoy FW (1991) Southern Ocean sediments; circum-Antarctic to 30°S. In: Hayes DE (ed) Marine geological and geophysical atlas of the circum-Antarctic to 30° S. Antarctic Research Series, Vol 54. American Geophysical Union, Washington, DC, p 37–46
McCune JL (2016) Species distribution models predict rare species occurrences despite significant effects of landscape context. J Appl Ecol 53:1871–1879
Merckx B, Steyaert M, Vanreusel A, Vincx M, Vanaverbeke J (2011) Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. Ecol Modell 222:588–597
Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, Anderson RP (2014) ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for ecological niche models. Methods Ecol Evol 5:1198–1205
Naimi B, Hamm NA, Groen TA, Skidmore AK, Toxopeus AG (2014) Where is positional uncertainty a problem for species distribution modelling? Ecography 37:191–203
Newbold T (2010) Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Prog Phys Geogr 34: 3–22
Peterson AT, Pape M, Soberón J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Modell 213:63–72
Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Modell 190:231–259
Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19:181–197
Phillips ND, Reid N, Thys T, Harrod C, Payne N and others (2017) Applying species distribution modelling to a data poor, pelagic fish complex: the ocean sunfishes. J Biogeogr 44:2176–2187
Pierrat B, Saucède T, Laffont R, De Ridder C and others (2012) Large-scale distribution analysis of Antarctic echinoids using ecological niche modelling. Mar Ecol Prog Ser 463:215–230
Pokharel KP, Ludwig T, Storch I (2016) Predicting potential distribution of poorly known species with small database: the case of four horned antelope *Tetracerus quadricornis* on the Indian subcontinent. Ecol Evol 6:2297–2307
Qiao H, Soberón J, Peterson AT (2015) No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. Methods Ecol Evol 6:1126–1136
Raes N, ter Steege H (2007) A null-model for significance testing of presence-only species distribution models. Ecography 30:727–736
Ranc N, Santini L, Rondinini C, Boitani L, Poitevin F, Angerbjörn, Maiorano L (2017) Performance tradeoffs in target-group bias correction for species distribution models. Ecography 40:1076–1087
Randin CF, Dirnböck T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in space? J Biogeogr 33: 1689–1703
Reiss H, Cunze S, König K, Neumann H, Kröncke I (2011) Species distribution modelling of marine benthos: a North Sea case study. Mar Ecol Prog Ser 442:71–86
Reiss H, Birchenough S, Borja A, Buhl-Mortensen L and others (2015) Benthos distribution modelling and its relevance for marine ecosystem management. ICES J Mar Sci 72:297–315
Reutter BA, Helfer V, Hirzel AH, Vogel P (2003) Modelling habitat suitability using museum collections: an example with three sympatric *Apodemus* species from the Alps. J Biogeogr 30:581–590
Ridgeway G (2015) gbm: generalized boosted regression models. R package version 2.1.1. https://CRAN.R-project.org/package=gbm
Robinson LM, Elith J, Hobday AJ, Pearson RG, Kendall BE, Possingham HP, Richardson AJ (2011) Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Glob Ecol Biogeogr 20:789–802
Rocchini D, Hortal J, Lengyel S, Lobo JM and others (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. Prog Phys Geogr 35:211–226
Rödder D, Engler JO (2011) Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. Glob Ecol Biogeogr 20:915–927
Rodríguez JP, Brotons L, Bustamante J, Seoane J (2007) The application of predictive modelling of species distribution to biodiversity conservation. Divers Distrib 13:243–251
Segurado PA, Araújo MB, Kunin WE (2006) Consequences of spatial autocorrelation for niche-based models. J Appl Ecol 43:433–444
Stockwell DR, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. Ecol Modell 148: 1–13
Tessarolo G, Rangel TF, Araújo MB, Hortal J (2014) Uncertainty associated with survey design in species distribution models. Divers Distrib 20:1258–1269
Thuiller W, Georges D, Engler R, Breiner F (2016) biomod2: ensemble platform for species distribution modeling. R package version 3.3-7. https://CRAN.R-project.org/package=biomod2
Turner J, Barrand NE, Bracegirdle TJ, Convey P and others (2014) Antarctic climate change and the environment: an update. Polar Rec 50:237–259
Van de Putte AP, Griffiths HJ, Raymond B, Danis B (2014) Data and mapping. In: De Broyer C, Koubbi P, Griffiths HJ, Raymond B and others (eds) Biogeographic atlas of the Southern Ocean. Scientific Committee on Antarctic Research, Cambridge, p 14–15
van Proosdij ASJ, Sosef MSM, Wieringa JJ, Raes N (2016) Minimum required number of specimen records to develop accurate species distribution models. Ecography 39:542–552
Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York, NY
Vierod AD, Guinotte JM, Davies AJ (2014) Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. Deep Sea Res II 99:6–18
Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62:2868–2883
Wenger SJ, Olden JD (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol Evol 3:260–267
Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A (2008) Effects of sample size on the performance of species distribution models. Divers Distrib 14:763–773
World Ocean Atlas (2013) v2. National Centers for Environmental Information, NOAA. https://www.nodc.noaa.gov/OC5/woa13/woa13data.html

Editorial responsibility: Jean-Sébastien Lauzon-Guay, Dartmouth, Nova Scotia, Canada

Submitted: July 27, 2017; Accepted: March 1, 2018
Proofs received from author(s): April 21, 2018

The following supplements accompany the article

Methods for improving species distribution models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau

Charlène Guillaumot*, Alexis Martin, Marc Eléaume, Thomas Saucède

*Corresponding author: charleneguillaumot21@gmail.com

Marine Ecology Progress Series 594: 149–164 (2018)

Supplement 1: Evaluation and choice of the model

In the present work, we ran ensemble models as a decision tool to select algorithms that are the most appropriate to the type of data to be analysed (Scales et al. 2016). The performance of 10 different algorithms was compared using the default parametrization settings proposed in the 'biomod2' R package (see Thuiller et al. (2016) for calibration details and Marmion et al. (2009) for modeling documentation). The compared algorithms include Artificial Neural Network (ANN), Boosted Regression Trees (BRT), Classification Tree Analysis (CTA), Flexible Discriminant Analysis (FDA), Generalized Additive Model (GAM), Generalized Linear Models (GLM), Multivariate Adaptive Regression Splines (MARS), Maximum Entropy (MaxEnt), Random Forest (RF), and Surface Range Envelope (SRE).
Two analyses were realised to compare the respective performance of the models. First, for each algorithm, AUC values of 100 model replicates were computed. Models were performed using all occurrence data available for the species *Ctenocidaris nutrix* and *Sterechinus diadema* only (Fig S1A, S1C) because there were not enough data to perform the analysis for *Abatus cordatus* and *Brisaster antarcticus*.
In a second step, standard deviation of the 100 replicates were compared between models as the number of data was progressively increased between runs to represent the improvement of sampling effort through time (Fig S1B, S1D).
Presence-only records associated to non-informative environmental data (NA/, no data values) were removed as required to perform the *biomod2* analysis. Occurrence duplicates located on one single 0.1° grid cell were removed to reduce spatial weighting. 200 pseudo-absences were selected to perform the analysis.

Results show that Boosted Regression Trees (BRT) and Random Forest (RF) are the algorithms that perform best to model the distribution of *C. nutrix* and *S. diadema* (Fig. S1), with relatively stable (SD < 0.025) and high AUC values varying between [0.976,1] and [0.994,1] respectively of the analysis that studies data addition. Unexpectedly, algorithms previously shown to be well suited to presence-only data and small datasets (e.g. SRE or MaxEnt, see Araújo and Peterson 2012, Yackulic et al. 2013) did not perform well in our case study. Low performances of SRE have already been reported (Elith et al. 2006). The low number of pseudo-absences used to calibrate the model could explain the low performance of MaxEnt (Barbet-Massin et al. 2012, Phillips and Dudik 2008).

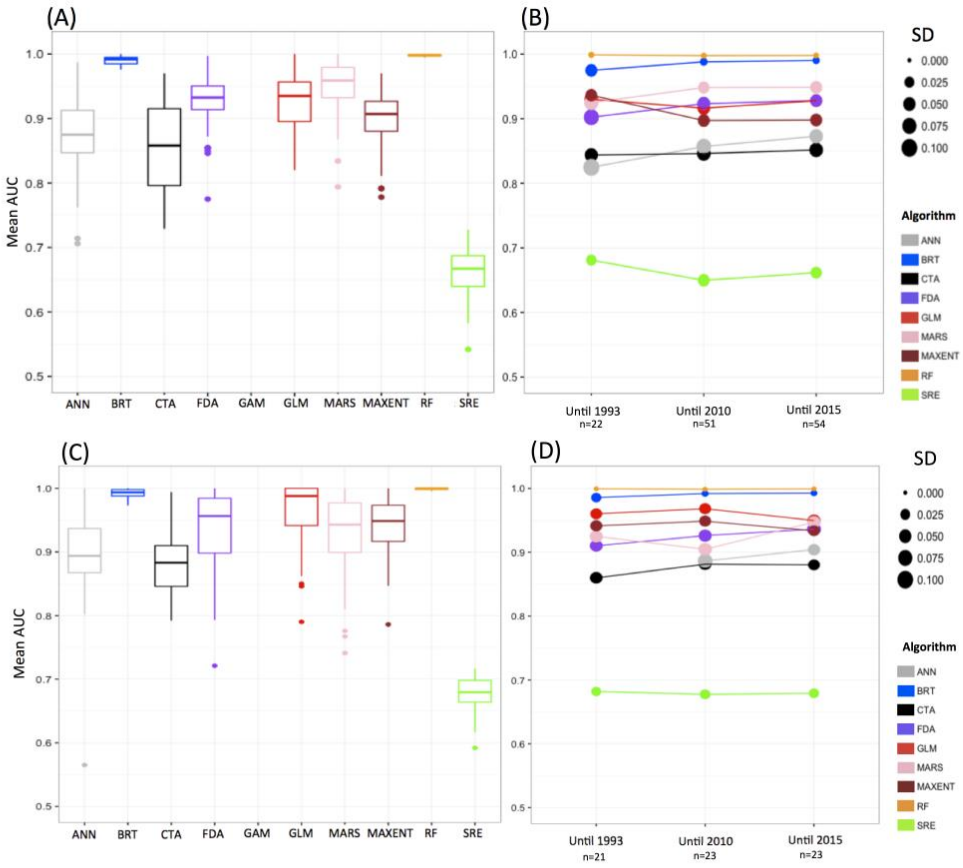


Figure S1: Compared performances of the different models for the species *Ctenocidaris nutrix* (A-B) and *Sterechinus diadema* (C-D). (A, C) Mean AUC values of model replicates for each algorithm. (B, D) Variation of mean AUC values and Standard Deviation (SD) of model replicates with data addition (n=22, n=51 or n=54 occurrences for *C. nutrix* and n=21, n= 23 and n=23 for *S. diadema*). For each analysis, 200 background data were randomly sampled in the studied area. Environmental descriptors correspond to [1955-2012].

The respective performance of BRT and RF (Fig. S1) was tested for spatial transferability following a non-random three-fold cross-validation procedure (Fig. S2, Wenger and Olden 2012). Model transferability is defined as the "extrapolative accuracy" of a model that is, the model ability to extrapolate in space and time (Randin et al. 2006, Wenger and Olden 2012). Three models were computed simultaneously using three different subsets of occurrences for *C. nutrix* (Fig. S2) alternatively used as training and test data (50 replicates). The three averaged models were compared with each others using the Schoener's D similarity index. D mean and standard deviation values were computed for all comparisons. All analyses were performed using time-averaged environmental parameters for the total period under study [1955-2012]. We considered that the most similar the distribution maps are the better the transferability performance is (Fig. S2).

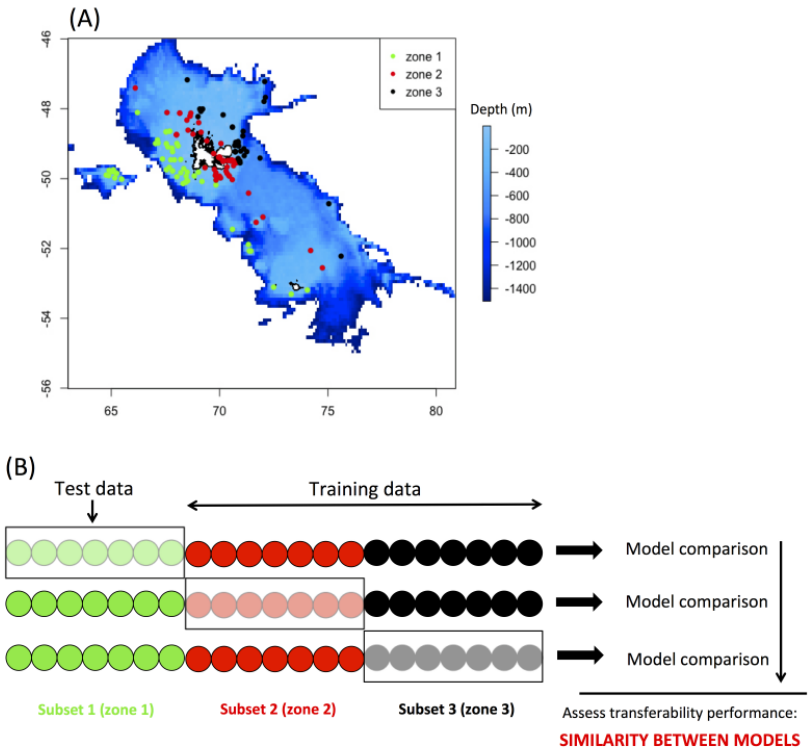


Figure S2: (A) Map showing the distribution of presence-only data in the three subsets defined for the cross-validation procedure. (B) Non-random three-fold cross-validation procedure performed to test for the transferability performance of models. Zones 1, 2 and 3 refere to (A).

Comparison between maps shows higher similarity values between the different models run with BRT (Schoener's $\bar{D}$=0.867± 0.034) than with RF ($\bar{D}$=0.761± 0.036), which highlights that BRT performs best for spatial transferability. Because transferability performance is a central criterion of model selection in our study (Araújo and Guisan 2006, Wenger and Olden 2012), BRT was selected for the further analyses. This result is in line with previous studies that highlight the high performance of BRT for prediction (Elith and Graham 2009, Guo et al. 2015) and transferability (Heikkinen et al. 2012, Wenger and Olden 2012, Crimmins et al. 2013) while RF has been shown to generate geographically restricted models with high accuracy (Guo et al. 2015, Qiao et al. 2015, Beaumont et al. 2016).

Table S1: Overall comparison between Random Forest, Boosted Regression Tree (BRT) and MaxEnt respective performances with reference works.

	Random Forest	BRT	MaxEnt
DATASET			
• *Type of data*	Presence-only/ presence-absence	Presence-only/ presence-absence	Presence-only/ presence-absence
• *Missing biological values*	Interpolation required first (Breiman 2001)	allowed	allowed
• *Categorical descriptors*	Biased if different categorical levels (Duan et al. 2014)	allowed	allowed
Robustness to spatially biased data	**More sensible than BRT to patchy patterns** (Marmion et al. 2009, Barbet-Massin et al. 2012) More adapted to bias correction methods than BRT (Barbet-Massin et al. 2012)		Not adapted (Royle et al. 2012) Unstable predictions (this study)
Overall modelling performance	**High performance and interpolation accuracy** (Wenger and Olden 2012, Guo et al. 2015) Biological responses often unrealistic (Beaumont et al. 2016)	Medium performance (Qiao et al. 2015) Performed better than RF in previous works on benthic marine species (Reiss et al. 2011)	High performance even with complex environmental interactions (Elith et al. 2011)
• *Transferability performance*	**Poor** (Wenger and Olden 2012, Crimmins et al. 2013).	**Good** (Heikkinen et al. 2012)	One of the highest (Heikkinen et al. 2012, Duque-Lazo et al. 2016)
• *Extrapolation performance*	**Not suitable** (Qiao et al. 2015, Beaumont et al. 2016) due to **overfitting** (Wenger and Olden 2012, Aguirre-Gutiérrez et al. 2013)	Good (Heikkinen et al. 2012). **High prediction performance** (Elith et al. 2006, Elith and Graham 2009, Guo et al. 2015)	Perform worse than BRT (this study) Tend to overpredict (Duan et al. 2014)
Required computation time	Long (Elith and Graham 2009, García-Callejas and Araújo 2016)	Medium (this study)	Medium (this study)

611

Supplement 2: Spatial autocorrelation (SAC)

Table S2: Moran I SAC index computed from mean residuals of the 100 model replicates and the associated significance for each species before and after spatial bias correction.

	Before correction		After correction	
	I_{obs}	p-value	I_{obs}	p-value
Abatus cordatus	0.16	1.19e-9	0.06	5.85e-4
Brisaster antarcticus	0.05	0.04	0.04	0.08
Ctenocidaris nutrix	0.07	7.37e-8	0.01	0.17
Sterechinus diadema	0.06	3.90e-3	0.02	0.13

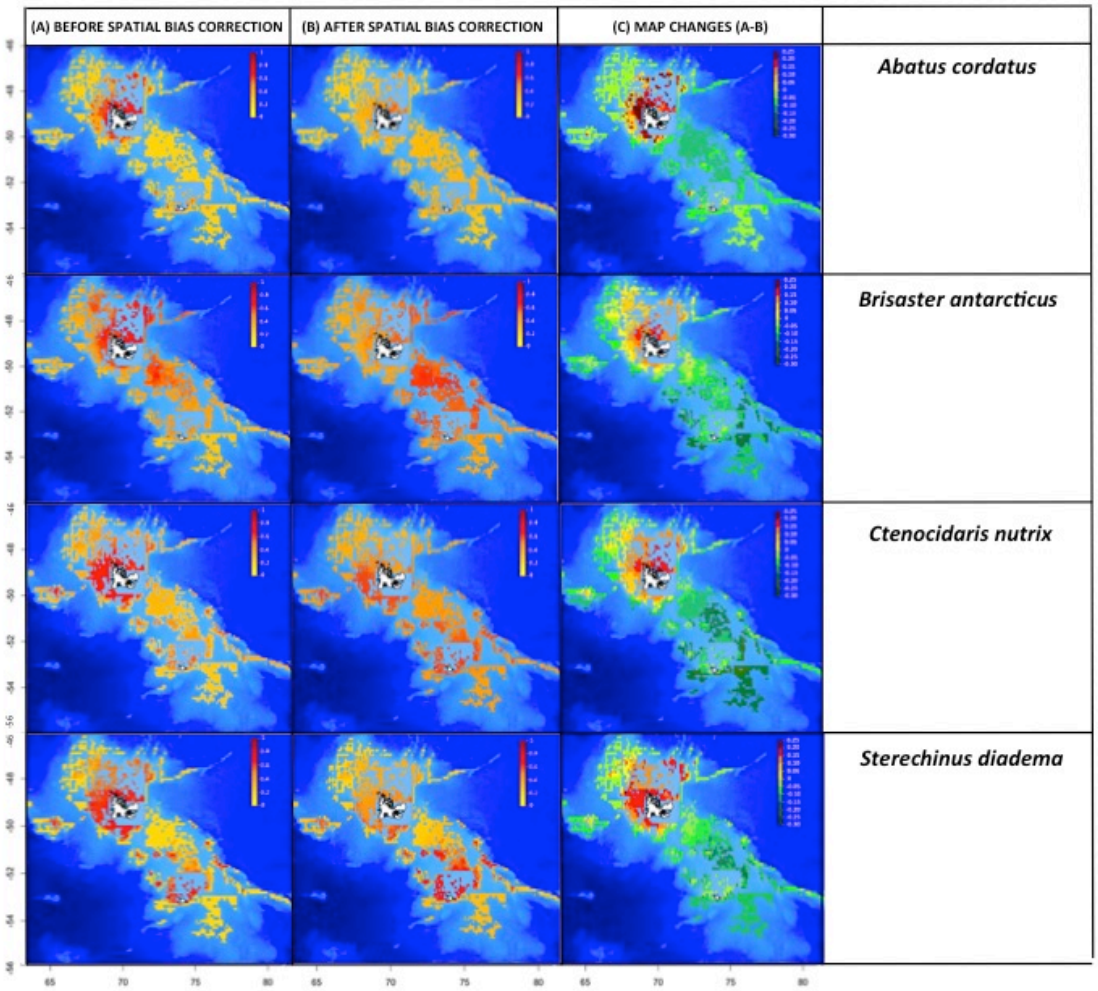


Figure S3: Maps showing species distribution models computed before and after correcting for spatial bias by background sampling.

Supplement 3: Testing the influence of chronological addition of occurrences

We thank the 19 scientific cruises for the collection of the data used to realise this work (Table 1, Guillaumot et al. 2016). We thank the master, the crew and the scientific team of the FV "Austral" that collected, sorted, and made available for studies the benthic samples of the POKER II (2010) cruise. We are grateful to the leader of the cruise, Pr. Guy Duhamel (MNHN) and Echinodermata curators Nadia Améziane and Marc Eléaume for giving us the opportunity to study POKER II sea urchins. Work at sea was supported by the Terres Australes et Antarctiques Françaises (TAAF), the Syndicat des Armateurs Réunionnais de Palangriers Congélateurs (SARPC), the Direction des Pêches Maritimes et de l'Aquaculture, Ministère de l'Agriculture et de l'Alimentation (DPMA), the Réserve Naturelle of TAAF, and the Muséum national d'Histoire naturelle, Paris.

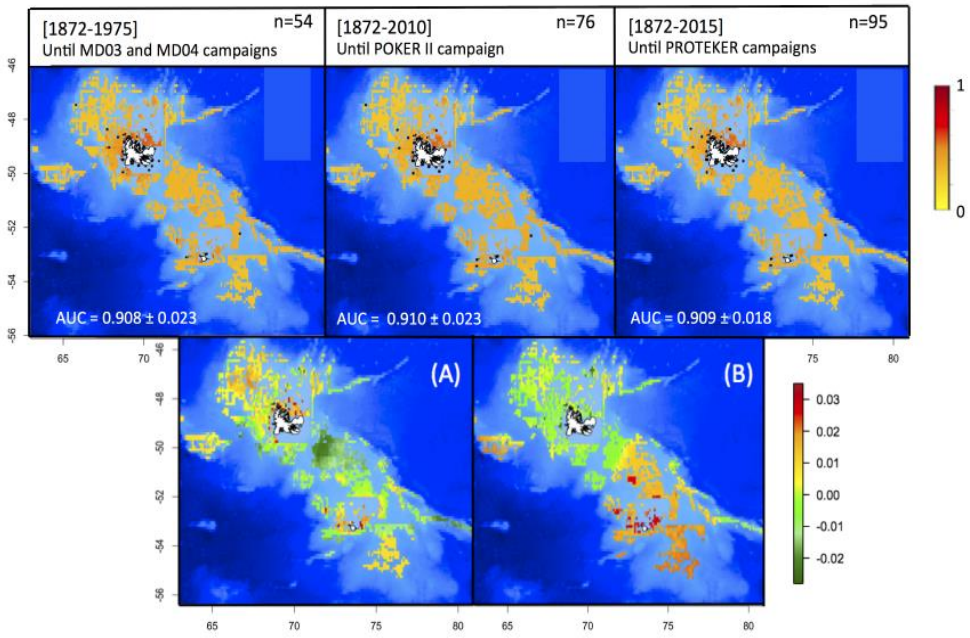


Figure S4: First row: distribution models of *Abatus cordatus* with increasing number of occurrences. Averaged maps of 100 model replicates. Second row: (A) Difference in probability distribution between n=76 and n=54, (B) between n=95 and n=76.

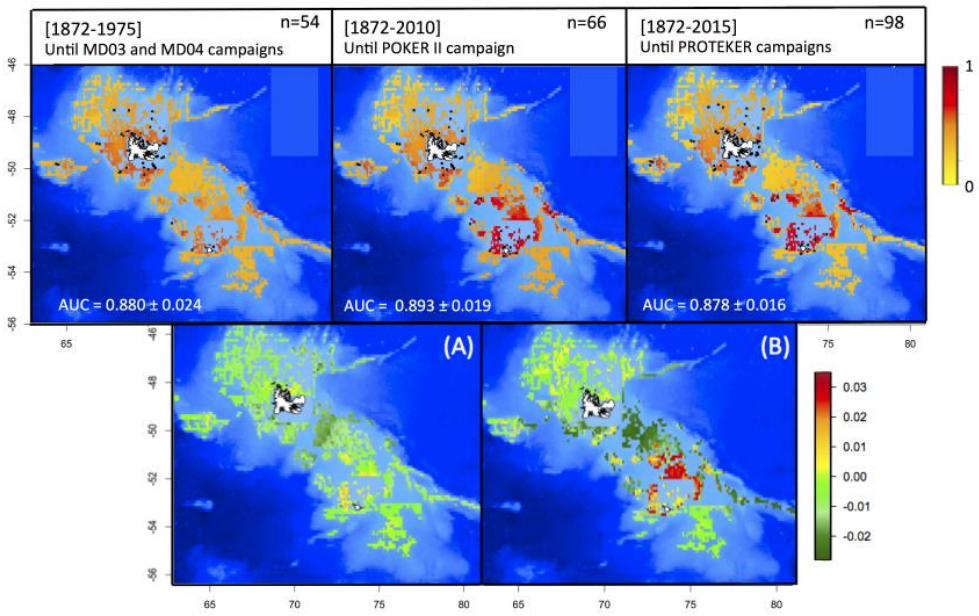


Figure S5: First row: distribution models of *Sterechinus diadema* with increasing number of occurrences. Averaged maps of 100 model replicates. Second row: (A) Difference in probability distribution between n=66 and n=54, (B) between n=98 and n=66.

Table S3: Effect of chronological addition of new data on model performance. Comparison between distribution maps. Upper diagonal: Schoener‘s D correlation between averaged maps. Lower diagonal: significance p-value of the associated Schoener‘s D correlation.

Abatus cordatus	←1975] n=54	-	←2010] n=76	←2015] n=95
←1975] n=54	-	-	0.972±0.025	0.980±0.021
-	-	-	-	-
←2010] n=76	0.002	-	-	0.981±0.023
←2015] n=95	0	-	0	-
Ctenocidaris nutrix	←1975] n=46	←1993] n=54	←2010] n=106	←2015] n=114
←1975] n=46	-	0.964±0.026	0.969±0.020	0.967±0.020
←1993] n=54	0.017	-	0.960±0.020	0.961±0.020
←2010] n=106	0.005	0.037	-	0.988±0.013
←2015] n=114	0.010	0.028	0	-
Sterechinus diadema	←1975] n=54	-	←2010] n=66	←2015] n=98
←1975] n=54	-	-	0.930±0.030	0.928±0.037
-	-	-	-	-
←2010] n=66	0.369	-	-	0.937±0.042
←2015] n=98	0.411	-	0.262	-

References

Aguirre-Gutiérrez J, Carvalheiro LG, Polce C, van Loon EE and others (2013) Fit-for-purpose: species distribution model performance depends on evaluation criteria–Dutch hoverflies as a case study. PloS one. 8: e63708.

Araújo MB and Guisan A (2006) Five (or so) challenges for species distribution modelling. J. Biogeogr. 33: 1677–1688.

Araújo MB and Peterson AT (2012) Uses and misuses of bioclimatic envelope modeling. Ecology. 93: 1527–1539.

Barbet-Massin M, Jiguet F, Albert C and Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Method. Ecol. Evol.* 3: 327–338.

Beaumont LJ, Graham E, Duursma DE, Wilson PD and others (2016) Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges? Ecol. Model. 342: 135–146.

Breiman L (2001) Random forests. Machine learning. 45: 5–32.

Crimmins SM, Dobrowski SZ and Mynsberge AR (2013) Evaluating ensemble forecasts of plant species distributions under climate change. Ecol. Model. 266: 126–130.

Duan RY, Kong XQ, Huang MY, Fan WY and Wang ZG (2014) The predictive performance and stability of six species distribution models. PloS one. 9: e112764.

Duque-Lazo J, Van Gils HAMJ, Groen TA and Navarro-Cerrillo RM (2016) Transferability of species distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. Ecol. Model. 320: 62–70.

Elith JH, Graham R, Anderson M, Dudík S and others (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography. 29: 129–151.

Elith J and Graham CH (2009) Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. Ecography. 32: 66–77.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE and Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17: 43–57.

García-Callejas D and Araújo MB (2016) The effects of model and data complexity on predictions from species distributions models. Ecol. Model. 326: 4–12.

Guo C, Lek S, Ye S, Li W, Liu J and Li Z (2015) Uncertainty in ensemble modelling of large-scale species distribution: Effects from species characteristics and model techniques. Ecol. Model. 306: 67–75.

Heikkinen RK, Marmion M and Luoto M (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography. 35: 276–288.

Hijmans RJ, Phillips S, Leathwick J and Elith J (2016) dismo: Species Distribution Modeling. R package version 1.1-1. https://CRAN.R-project.org/package=dismo.

Marmion M, Luoto M, Heikkinen RK and Thuiller W (2009) The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecol. Model. 220: 3512–3520.

Merow C, Smith M and Silander J (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography. 36: 1058–1069.

Phillips S and Dudík M (2008) Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. Ecography. 31: 161–175.

Qiao H, Soberón J and Peterson AT (2015) No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. Method. Ecol. Evol. 6: 1126–1136.

Randin CF, Dirnböck T, Dullinger S, Zimmermann NE and others (2006) Are niche-based species distribution models transferable in space? J. Biogeogr. 33: 1689–1703.

Reiss H, Cunze S, König K, Neumann H and Kröncke I (2011) Species distribution modelling of marine benthos: a North Sea case study. Mar. Ecol. Prog. Ser. 442: 71–86.

Royle JA, Chandler RB, Yackulic C and Nichols JD (2012) Likelihood analysis of species occurrence probability from presence- only data for modelling species distributions. Method. Ecol. Evol. 3: 545–554.

Scales KL, Miller PI, Ingram SN, Hazen EL, Bograd SJ and Phillips RA (2016). Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models. Divers. Distrib. 22: 212-224.

Thuiller W, Georges D, Engler R and Breiner F (2016) biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7. https://CRAN.R-project.org/package=biomod2.

Yackulic C, Chandler R, Zipkin E, Royle J and others (2013) Presence-only modelling using MaXent: when can we trust the inferences? Meth. Ecol. Evol. 4: 236-243.

Wenger SJ and Olden JD (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Method. Ecol. Evol. 3: 260–267.

Can the Patagonian crab *Halicarcinus planatus* (Fabricius, 1775) reach Antarctic coasts? Study of the dispersal potential of its larvae using a Lagrangian approach.

López-Farrán Zambra[1,2,3], Dulière Valérie[4], Guillaumot Charlène[5,6] *et al.*

[1] LEM-Laboratorio de Ecología Molecular, Instituto de Ecología y Biodiversidad, Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile.
[2] Research Center Dynamics of High Latitude Marine, Ecosystem (Fondap-IDEAL), Universidad Austral de Chile, Valdivia, Chile
[3] LEMAS-Laboratorio de Ecología de Macroalgas Antárticas y Sub antárticas, Universidad de Magallanes, Punta Arenas, Chile
[4] Royal Belgian Institute Natural Sciences, Brussels, Belgium
[5] Laboratoire de Biologie marine CP160/15, Université Libre de Bruxelles, Bruxelles, Belgium.
[6] Biogéosciences, UMR 6282 CNRS, Université Bourgogne Franche-Comté, Dijon, France.

Work in progress, first results.

1. FOREWORDS AND OBJECTIVES

In February 2010, an ovigerous female of *Halicarcinus planatus* (Fabricius, 1775) (Brachyura, Hymenosomatidae) was found alive in shallow subtidal waters of Deception Island (WAP; Aronson et al. 2014). This small crab (carapace width up to 15 mm and 20 mm for female and male, respectively, in Punta Arenas; Fig. 1) is an opportunistic feeder (Boschi et al. 1969), commonly found sheltered under rocks in the intertidal and subtidal zones, in between holdfasts of the giant kelp *Macrocystis pyrifera* or sheltered in hydrozoans and mussel colonies (Richer De Forges 1977, Chuang and Ng 1994, Vinuesa and Ferrari 2008).

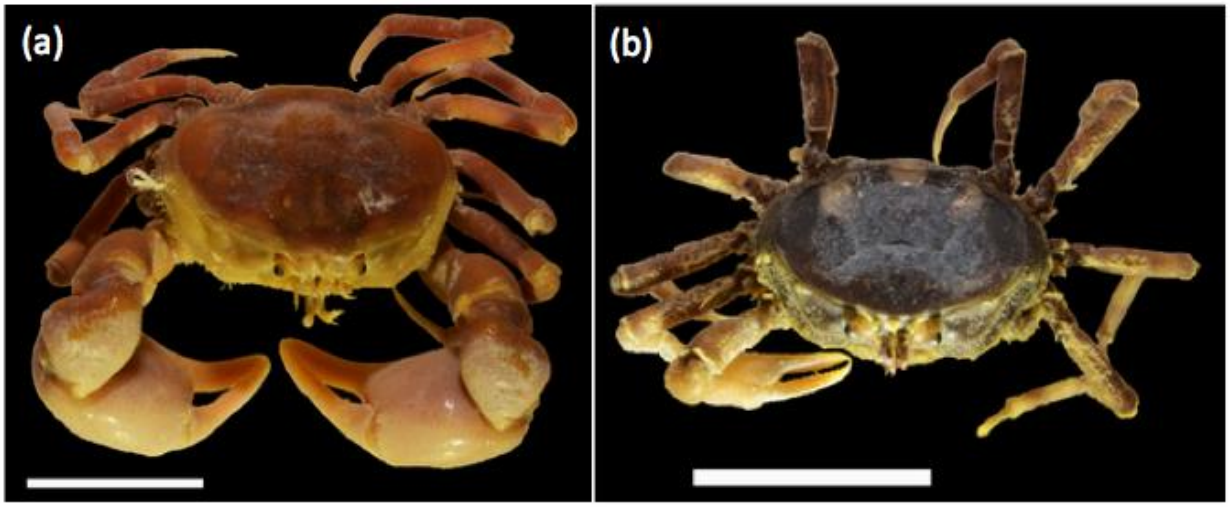


Figure 1. Male (a) and female (b) specimens of *Halicarcinus planatus* (Fabricius, 1775) collected in the Magellan Strait. Scale: 1 cm. Photos credit to C. Ceroni and K.Gerard.

The crab is commonly found in the Southern tip of South America, including Tierra del Fuego, Cape Horn and Diego Ramirez Islands and in some sub-Antarctic Islands (Prince Edward and Marion Islands, Crozet and Kerguelen Islands, Falkland Islands, Macquarie Island, Auckland Islands in New-Zealand and Campbell Island archipelago) up to 36°S latitude (Fig. 2) (Boschi et al. 1969, Melrose 1975, Richer De Forges 1977, Griffiths et al. 2013, Aronson et al. 2014). *Halicarcinus planatus* mainly inhabits shallow environments, although it has also been reported from lower intertidal down to 270 m (Garth 1958, Vinuesa 2005, Griffiths et al. 2013, Varisco et al. 2016).

Halicarcinus planatus has a strong dispersal potential mediated by an extended planktonic larval stage (Richer De Forges 1977, Diez and Lovrich 2010, Ferrari et al. 2011), lasting between 45 and 60 days (at temperatures of 11-13°C and 8°C in laboratory respectively) prior to benthic settlement (Boschi et al. 1969, Diez and Lovrich 2010). The species can either spawn between April and May (e.g. in the Beagle Channel), or in the end of the austral winter between August and December (e.g. in the Kerguelen Islands) (Diez and Lovrich, 2010, Diez et al. 2012, Vinuesa and Ferrari 2008).
According to Richer de Forges (1977) the larval hatching occurs between October and November and the larvae can be found in the epipelagic plankton in austral summer, i.e. from November to March.

Halicarcinus planatus has the physiological capacity to withstand low temperatures. Indeed, while most decapod taxa exposed to cold waters experience increased magnesium ion concentration in the hemolymph ($[Mg^{2+}]_{HL}$), reducing metabolic rates and aerobic activity, potentially leading to death (Frederich et al. 2001, Thatje et al. 2005, Aronson et al. 2007, Diez and Lovrich 2010), *H. planatus* has the capacity to overcome these issues by reducing $[Mg^{2+}]_{HL}$ (Frederich et al. 2001) providing capacity for survival in cold waters like the Kerguelen Islands, where winter seawater temperatures range between +1.1 and +3.0°C (Féral et al. 2019).

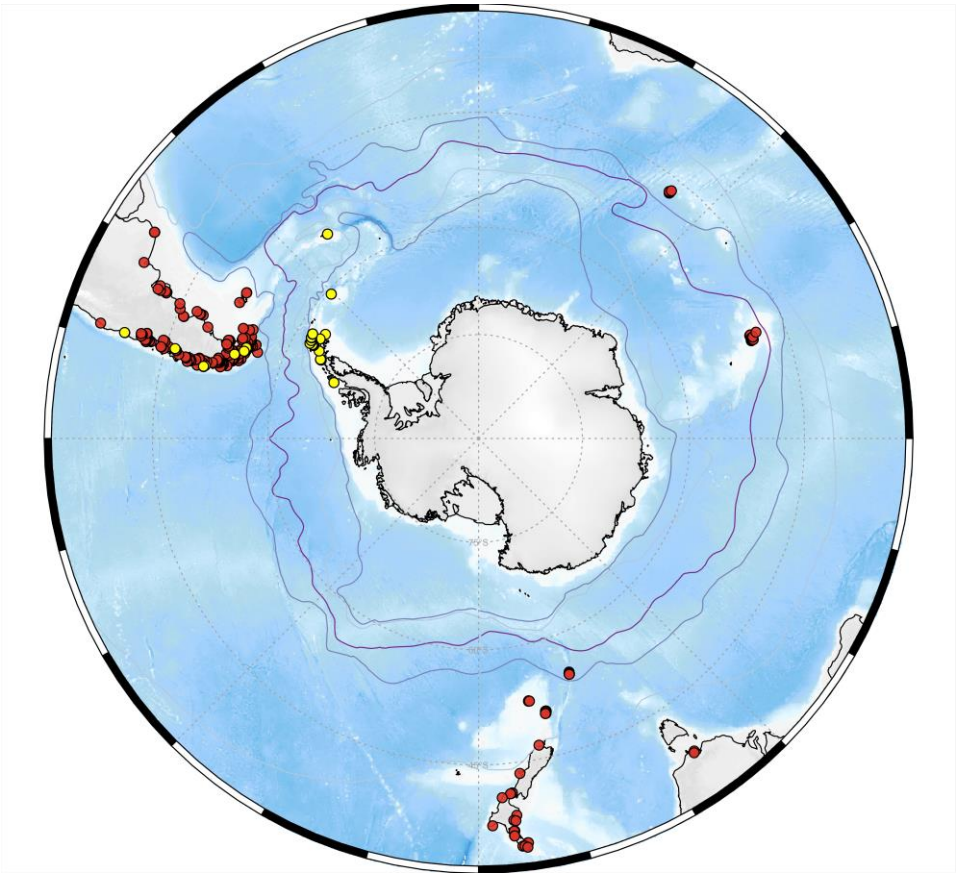


Figure 2. Presence (red dots) and absence (yellow dots) records of *Halicarcinus planatus* in the Southern Ocean, collected during different sampling expeditions carried out between 2015 and 2019 (PROTEKER 1, 4, 5 and 6, INACH ECA 53, 54 and 55), obtained from collaborators, and retrieved from IOBIS and GBIF databases, and from the scientific literature (López-Farrán et al. 2020).

OBJECTIVES

Following the recent discovery of a living specimen of *H. planatus* in Deception Island, the aim of this study is to use a dispersal model (Lagrangian approach) to evaluate whether crab propagules can reach Antarctic coasts from their actual presence locations. Simulations would help assess whether the natural transport via currents would be responsible of its presence in Deception Island.

2. MATERIAL AND METHODS

Lagrangian model settings

The Lagrangian particle model used in this study is based on the model described in Dulière et al. (2013) and made available as a module of the free and open-source aquatic modelling system COHERENS v2 (Luyten 2011).

Particles are transported under advective and diffusive processes. The classical fourth-order Runge-Kutta method is used to estimate horizontal transport. The diffusive velocities are obtained from random walk theory with constant horizontal and vertical diffusion coefficients of 10 and 0.0001m^2.s^{-1}, respectively. The same diffusion coefficient values are used as in Young et al. (2014) and are equivalent to values observed in the Southern Ocean (empirical values or commonly accepted by modellers; Sheen et al. 2013, Watson et al. 2013). A bouncing condition is used for particles reaching the sea surface or seabed, and particles that leave the model domain through the ocean open boundary are assumed to have left the region. Stranding is not allowed, so when a

particle reaches a dry cell, its position is set to its previous position at sea. The Lagrangian module is used off-line with a computation time step of 5 minutes.
The hydrodynamic conditions used to force the Lagrangian model are based on the 2008-2016 PHY_001_024 datasets produced by the high-resolution global analysis and forecasting system, provided by Mercator Ocean (Law Shune et al. 2019). These products contain daily mean fields of sea surface elevation and horizontal ocean currents. In addition, they also contain sea ice information (*i.e*. concentration, thickness and velocity), sea water potential temperature, sea water salinity and ocean mixed layer thickness. These datasets have been generated with NEMO 3.1 and LIM2 EVP models forced with 3-hourly atmospheric forcing from ECMWF (European Centre for Medium-Range Weather Forecasts, https://www.ecmwf.int/). Daily averaged model products are made available after interpolation from the native model grid to a global standard Arakawa C grid of 1/12° horizontal resolution and 50 fixed vertical levels (from 0 to 5,000 m). The quality of the Global high-resolution products has been assessed in Lellouche et al. (2019). 3D vertical ocean currents are estimated from the divergence in the horizontal velocity from the PHY_001_024 forcing fields, assuming null surface and bottom vertical velocity.
The model grid was built from a sub-sample of the global grid of the hydrodynamic forcing field from latitude 36°S down to the South Pole. The horizontal resolution of 1/12° (~8km) was kept and the 50 vertical levels have been adapted to 50 sigma levels for the COHERENS system.

Biological assumptions used in the Lagrangian model

Particles were assumed to drift during three months, corresponding to the maximal duration of the larval stage. They were launched from presence locations points (Fig. 2). Particles were limited to 200 m depth during their journey. The propagules were assumed to passively drift (without swimming capacity). No nycthemeral behaviour was considered. Particles were released from August to November and the drift was studied until the end of February.

Studied years.

The model was launched over different years, in order to assess possible contrasts in propagule dispersal trajectories under the influence of contrasting climatic regimes. Noteworthy, the intensity of the Southern Annular Mode (SAM, Limpasuvan and Hartmann 1999), has been shown to be strongly and linearly teleconnected to the phase of El Niño Southern Oscillation (Carvalho et al. 2005, L'Heureux and Thompson 2006, Ciasto and Thompson 2008). A negative SAM with strong El Niño episodes is characterised by warmer temperatures and stronger westerly winds. In contrast, years with strong positive SAM and with La Niña episodes present weaker westerly winds and a dryer and colder atmosphere (Nicolas et al. 2017). The model was therefore considered for 2009-2010 ('normal' SAM= -0.54), 2015-2016 (positive SAM: +1.77) and 2016-2017 (negative SAM: -1.02). SAM values are available at http://www.nerc-bas.ac.uk/public/icd/gjma/newsam.1957.2007.seas.txt and in (Doddridge and Marshall 2017).

3. RESULTS

Results suggest that the crab larvae, within their 3-month drift in the water column, cannot reach Antarctic coasts independently of the different climatic regimes (Fig. 3). The Antarctic Circumpolar Current (ACC) constitutes a physical barrier that prevents propagules from crossing the Southern Ocean (Fig. 4).
The simulations also highlights the influence of the ACC to disperse the particles eastward (Fig. 5) and explain the genetic connectivity in between the different populations of *Halicarcinus planatus* (Fig. 2).

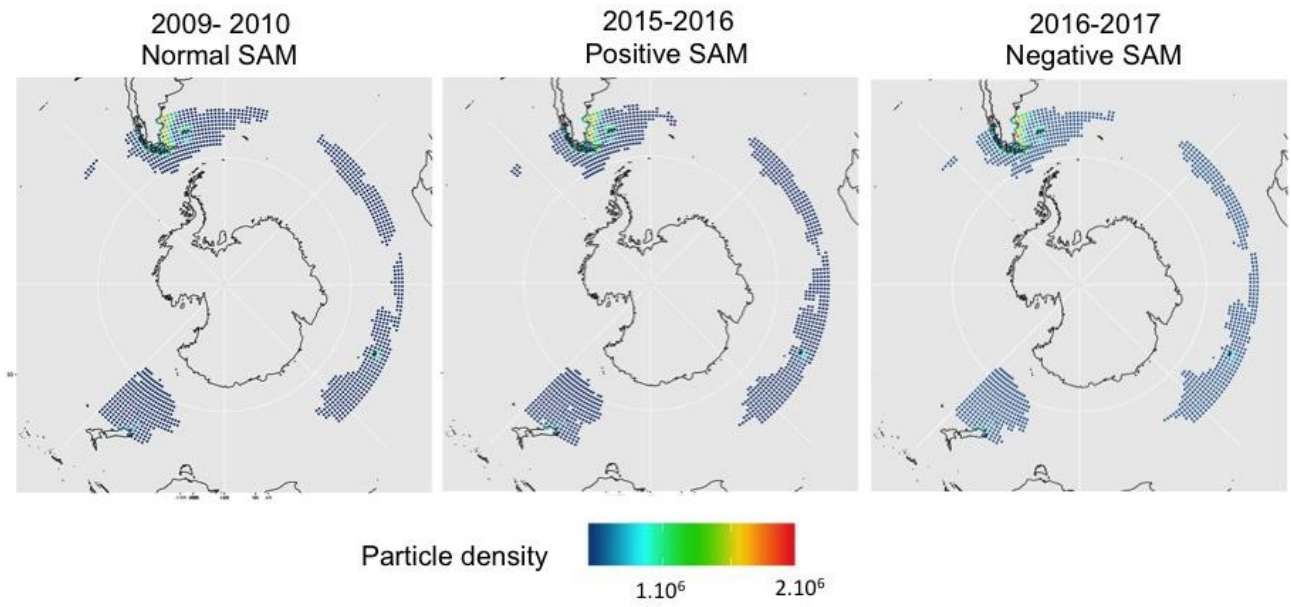


Figure 3. Maps with particle densities predicted for the different climatic scenarios. Average densities from August to February. Particles of 3-month age, located between 0 to 200 m depth.

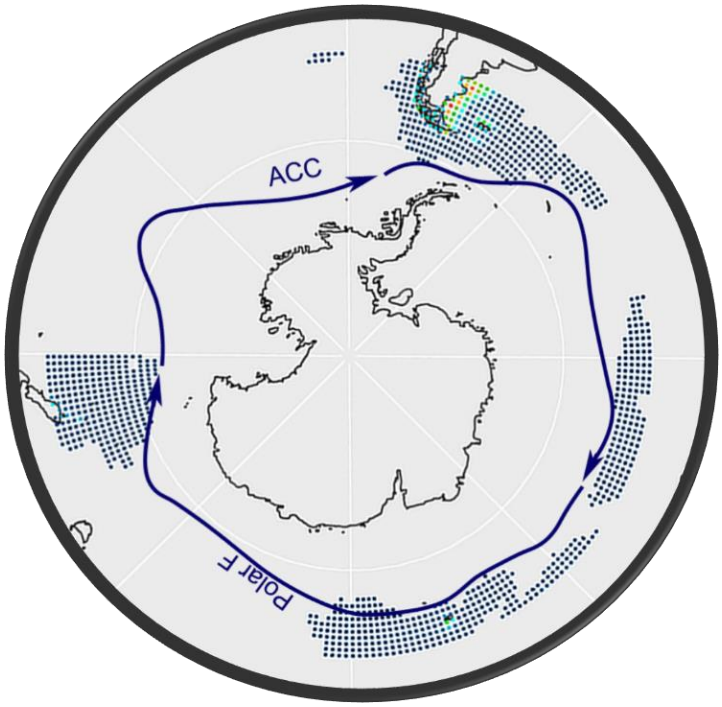


Figure 4. Illustration of the link between model results and the position of the Antarctic Circumpolar Current (ACC). The ACC constitutes a strong physical barrier that prevents propagules from reaching Antarctic coasts and disperse them eastward.

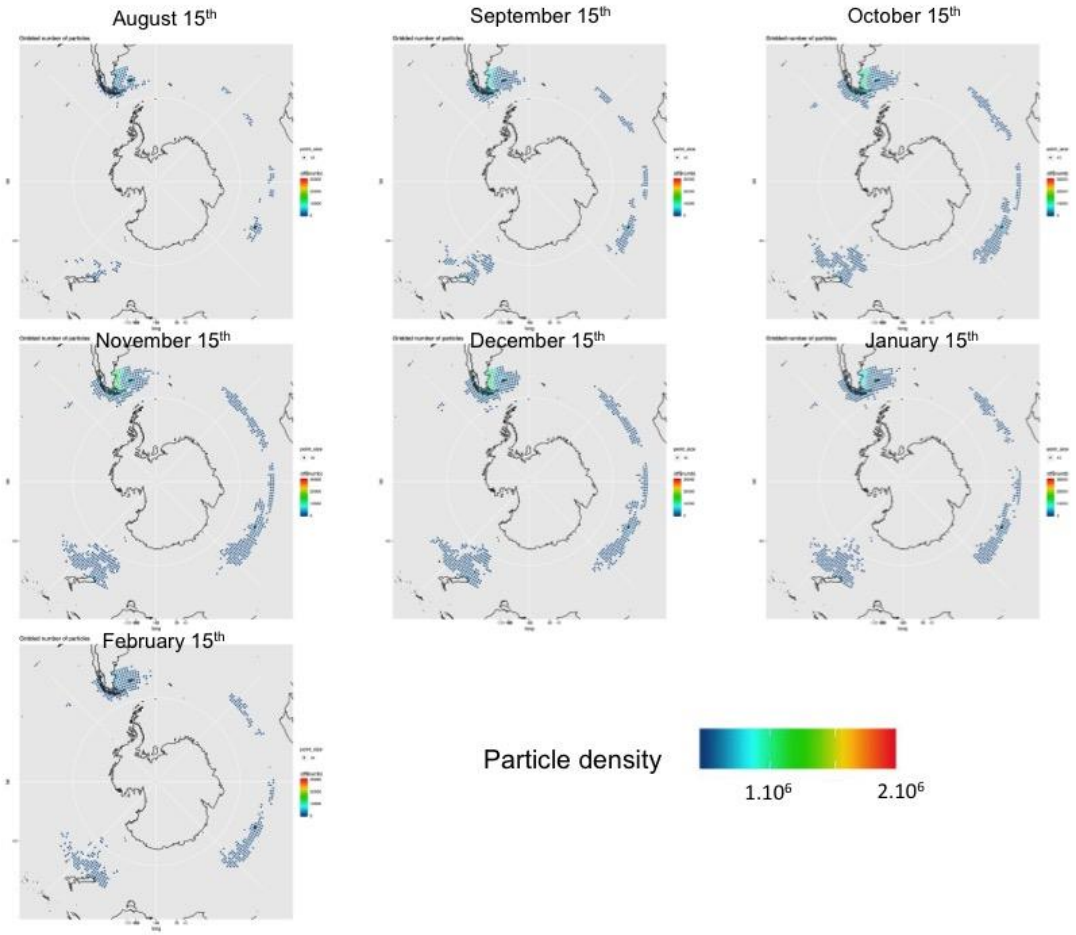


Figure 5. Snapshots of particle densities on the 15th of each month (from August to February 2009-2010) for particles younger than 3 months old and at depth between 0 and 200 m.

REFERENCES

Aronson, R.B., Thatje, S., Clarke, A., Peck, L.S., Blake, D.B., Wilga, C.D. & Seibel, B.A. (2007). Climate change and invasibility of the Antarctic benthos. *Annual Review of Ecology, Evolution and Systematics, 38*, 129-154.

Aronson, R.B., Frederich, M., Price, R. & Thatje, S. (2014). Prospects for the return of shell-crushing crabs to Antarctica. *Journal of Biogeography, 42*(1), 1–7.

Boschi, E.E., Scelzo, M.A. & Goldstein, B. (1969). Desarrollo larval del cangrejo, *Halicarcinus planatus* (Fabricius) (Crustacea, Decapoda, Hymenosomidae), en el laboratorio, con observaciones sobre la distribución de la especie. *Bulletin of Marine Science, 19*(1), 225-242.

Carvalho, L.M., Jones, C. & Ambrizzi, T. (2005) Opposite phases of the Antarctic oscillation and relationships with intraseasonal to interannual activity in the tropics during the austral summer. *Journal of Climate,* 18, 702–718.

Chuang, C.T. & Ng, P.K. (1994). The ecology and biology of Southeast Asian false spider crabs (Crustacea: Decapoda: Brachyura: Hymenosomatidae). *Hydrobiologia*, 285(1–3), 85–92.

Ciasto, L.M. & Thompson, D. (2008). Observations of large-scale ocean–atmosphere interaction in the Southern Hemisphere. *Journal of Climate*, 21, 1244–1259.

Diez, M.J. & Lovrich, G.A. (2010). Reproductive biology of the crab *Halicarcinus planatus* (Brachyura, Hymenosomatidae) in sub-Antarctic waters. *Polar Biology*, *33*(3), 389-401.

Doddridge, E.W. & Marshall, J. (2017). Modulation of the seasonal cycle of Antarctic sea ice extent related to the Southern Annular Mode. *Geophysical Research Letters*, *44*(19), 9761-9768.

Dulière, V., Ovidio F. & Legrand S. (2013) Development of an Integrated Software for Forecasting the Impacts of Accidental Oil Pollution- OSERIT. Final Report. Brussels: Belgian Science Policy, 68 pp. (Research Programme Science for a Sustainable Development)
http://www.belspo.be/belspo/SSD/science/Reports/OSERIT_FinRep_AD.pdf

Diez, M.J., Spivak, E.D., Anger, K.,& Lovrich, G.A. (2012). Seasonal variations in size, biomass, and elemental composition (CHN) of Halicarcinus planatus (Brachyura: Hymenosomatidae) larvae from the Beagle Channel, Southern South America. *Journal of Crustacean Biology*, *32*(4), 575-581.

Féral, J.P., Poulin, E., González-Wevar, C.A., Améziane, N., Guillaumot, C., Develay, E. & Saucède, T. (2019). Long-term monitoring of coastal benthic habitats in the Kerguelen Islands: a legacy of decades of marine biology research. In: Welsford, D., Dell J., Duhamel G. (Eds), The Kerguelen Plateau: marine ecosystem and fisheries. Proceedings of the Second Symposium. Australian Antarctic Division, Kingston, Tasmania, Australia, pp. 383-402.

Ferrari, R., Malcolm, H., Neilson, J., Lucieer, V., Jordan, A., Ingleton, T., ... & Hill, N. (2018). Integrating distribution models and habitat classification maps into marine protected area planning. *Estuarine, Coastal and Shelf Science*, 212, 40-50.

Frederich, M., Sartoris, J. & Pörtner, H.O. (2001). Distribution patterns of decapod crustaceans in polar areas: A result of magnesium regulation? *Polar Biology*, 24(10), 719–723.

Garth, J.S. (1958). Brachyura of the Pacific Coast of America. *Allan Hancock Pacific Expeditions*, 21, i–xii, 1–804.

Griffiths, H.J., Whittle, R.J., Roberts, S.J., Belchier, M. & Linse, K. (2013). Antarctic crabs: invasion or endurance? *PLoS One*, *8*(7).

Law Chune, S., Nouel, L., Fernandez, E. & Derval, C. (2019) Product Manual for the Global Ocean Sea Physical Analysis and Forecasting Products GLOBAL_ANALYSIS_FORECAST_PHY_001_024. 30pp.

L'Heureux, M.L. & Thompson, W.J. (2006). Observed relationships between the El Niño–Southern Oscillation and the extratropical zonal-mean circulation. *Journal of Climate*, 19, 276–287.

Lellouche, J-M., Legalloudec, O., Regnier, C., Levier, B., Greiner, E. & Drevillon, M. (2019). Quality Information Document For Global Sea Physical Analysis and Forecasting Product GLOBAL_ANALYSIS_FORECAST_PHY_001_024. 91pp.

Limpasuvan, V. & Hartmann, D.L. (1999). Eddies and the annular modes of climate variability. *Geophysical Research Letters*, 26, 3133–313. https://climatedataguide.ucar.edu/

López-Farrán, Z., Gerard, K., Saucède, T., Brickle, P., Waters, J., González-Wevar, C., Naretto, J., Poulin, E., Rosenfeld, S., Ojeda, J. & Ceroni C. (2020). Records of *Halicarcinus planatus* at sub-Antarctic regions. Version 1.2. Universidad de Magallanes. Occurrence dataset https://doi.org/10.15468/qnqnn2.

Luyten, P. (2011). COHERENS --- A Coupled Hydrodynamical-Ecological Model for Regional and Shelf Seas: User Documentation. Version 2.0. RBINS-MUMM Report, Royal Belgian Institute of Natural Sciences, 1202pp.

Melrose, M.J. (1975). The marine fauna of New Zealand: family Hymenosomatidae (Crustacea, Decapoda, Brachyura) (Vol. 34). Wellington, N.Z.

Nicolas, J.P., Vogelmann, A.M., Scott, R.C., Wilson, A.B., Cadeddu, M.P., Bromwich, D.H., ... & Powers, H.H. (2017). January 2016 extensive summer melt in West Antarctica favoured by strong El Niño. *Nature Communications*, *8*, 15799.

Richer de Forges, B. (1977). Étude du crabe des îles Kerguelen: *Halicarcinus planatus* (Fabricius). *Comité National Français des Recherches Antarctiques*, *42*, 71-133.

Sheen, K.L., Brearley, J.A., Naveira Garabato, A.C., Smeed, D.A., Waterman, S., Ledwell, J.R. ... & Watson, A.J. (2013). Rates and mechanisms of turbulent dissipation and mixing in the Southern Ocean: Results from the Diapycnal and Isopycnal Mixing Experiment in the Southern Ocean (DIMES). *Journal of Geophysical Research: Oceans*, *118*(6), 2774-2792.

Thatje, S., Anger, K., Calcagno, J.A., Lovrich, G.A., Pörtner, H.O. & Arntz, W.E. (2005). Challenging the cold: crabs reconquer the Antarctic. *Ecology*, *86*(3), 619-625.

Varisco, M., Colombo, J., Isola, T. & Vinuesa, J. (2016). Growth and maturity of the spider crab *Halicarcinus planatus* (Brachyura: Hymenosomatidae) females in the southwestern Atlantic Ocean. Can these parameters be influenced by the population sex ratio? *Marine Biology Research*, *12*(6), 647-655.

Vinuesa, J.H. (2005). Distribución de crustáceos decápodos y estomatópodos del golfo San Jorge, Argentina. *Revista de biología marina y oceanografía*, *40*(1), 7-21.

Vinuesa, J.H. & Ferrari, L. (2008). Reproduction of *Halicarcinus planatus* (crustacea, decapoda, hymenosomatidae) in the Deseado River estuary, southwestern Atlantic Ocean. *Marine Biology*, *154*(2), 345-351.

Watson, A.J., Ledwell, J.R., Messias, M.J., King, B.A., Mackay, N., Meredith, M.P., ... & Garabato, A.C. N. (2013). Rapid cross-density ocean mixing at mid-depths in the Drake Passage measured by tracer release. *Nature*, *501*(7467), 408-411.

Young, E.F.,Thorpe, S.E., Banglawala, N. & Murphy E.J. (2014), Variability in transport pathways on and around the South Georgia shelf, Southern Ocean: Implications for recruitment and retention, *Journal of Geophysical Research Oceans, 119*, 241–252.

Title: Modelling the response of Antarctic marine species to environmental changes. Methods, applications and limitations.

Keywords: ecological modelling; species distribution models; physiological models; lagrangian dispersal models; Southern Ocean ; marine benthic species.

Abstract: Among tools that are used to fill knowledge gaps on natural systems, ecological modelling has been widely applied during the last two decades. Ecological models are simple representations of a complex reality. They allow to highlight environmental drivers of species ecological niche and better understand species responses to environmental changes. However, applying models to Southern Ocean benthic organisms raises several methodological challenges. Species presence datasets are often aggregated in time and space nearby research stations or along main sailing routes. Data are often limited in number to correctly describe species occupied space and physiology. Finally, environmental datasets are not precise enough to accurately represent the complexity of marine habitats. **Can we thus generate performant and accurate models at the scale of the Southern Ocean ? What are the limits of such approaches ? How could we improve methods to build more relevant models ?** In this PhD thesis, three different model categories have been studied and their performance evaluated. (1) Mechanistic physiological models (Dynamic Energy Budget models, DEB) simulate how the abiotic environment influences individual metabolism and represent the species fundamental niche. (2) Species distribution models (SDMs) predict species distribution probability by studying the relationship between species presences and the environment. They represent the species realised niche. (3) Dispersal lagrangian models predict the drift of propagules in water masses. Results show that physiological models can be developed for marine Southern Ocean species to simulate the metabolic variations in link with the environment and predict population dynamics. However, more data are necessary to highlight detailed physiological contrasts between populations and to accurately evaluate models. Results obtained for SDMs suggest that models generated at the scale of the Southern Ocean and future simulations are not relevant, given the lack of data available to characterise species occupied space, the lack of precision and accuracy of future climate scenarios and the impossibility to evaluate models. Moreover, model extrapolate on a large proportion of the projected area. Adding information on species physiological limits (observations, results from experiments, physiological model outputs) was shown to reduce extrapolation and to improve the capacity of models to estimate the species realised niche. Spatial aggregation of occurrence data, which influenced model predictions and evaluation was also succefully corrected. Finally, dispersal models showed an interesting potential to highlight the role of geographic barriers or conversely of spatial connectivity and also the link between species distribution, physiology and phylogeny history. This PhD thesis provides several methodological advice, annoted codes and tutorials to help implement future modelling works applied to Southern Ocean marine species.

Titre : Modéliser la réponse des espèces antarctiques aux changements environnementaux. Méthodes, applications et limites.

Mots-clés : modélisation écologique ; modèles de distribution d'espèces ; modèles physiologiques ; modèles de dispersion lagrangiens ; océan Austral ; espèces marines benthiques.

Résumé : Parmi les outils qui permettent de mieux comprendre les systèmes naturels, la modélisation écologique a connu un essor particulièrement important depuis une vingtaine d'années. Les modèles écologiques, représentation simplifiée d'une réalité complexe, permettent de mettre en avant les facteurs environnementaux qui déterminent la niche écologique des espèces et de mieux comprendre leur réponse aux changements de l'environnement. Dans le cas des faunes marines antarctiques, la modélisation écologique fait face à plusieurs défis méthodologiques. Les jeux de données de présence des espèces sont très souvent agrégés dans le temps et dans l'espace, à proximité des stations de recherche. Ces données sont souvent trop peu nombreuses pour caractériser l'espace environnemental occupé par les espèces ainsi que leur physiologie. Enfin, les jeux de données environnementales manquent encore de précision pour finement représenter la complexité des habitats marins. **Dans ces conditions, est-il possible de générer des modèles performants et justes à l'échelle de l'océan Austral ? Quelles sont les approches possibles et leurs limites ? Comment améliorer les méthodes afin de générer de meilleurs modèles ?** Au cours de ce travail de thèse, trois types de modèles ont été étudiés et leurs performances évaluées. (1) Les modèles physiologiques de type DEB (Dynamic Energy Budget) simulent la manière dont l'environnement abiotique influe sur le métabolisme des individus et proposent une représentation de la niche fondamentale des espèces. (2) Les modèles de distribution d'espèces (SDMs pour Species Distribution Models) prédisent la probabilité de distribution des espèces en étudiant la relation spatiale entre données de présence et environnement. Ils proposent une représentation de la niche réalisée des espèces. Enfin (3), les modèles de dispersion de type lagrangien prédisent le mouvement de propagules dans les masses d'eau. Les résultats montrent que les modèles physiologiques réussissent à simuler les variations métaboliques des espèces antarctiques en fonction de l'environnement et à prédire les dynamiques de populations. Cependant, davantage de données sont nécessaires pour pouvoir caractériser finement les différences physiologiques entre populations et évaluer correctement les modèles. Les résultats obtenus pour les SDMs montrent que les modèles générés à l'échelle de l'océan Austral et leurs prédictions futures ne sont pas fiables du fait du manque de données disponibles pour caractériser l'espace occupé par les espèces, du manque de précision des scénarios climatiques futurs et de l'impossibilité d'évaluer les modèles. De plus, les modèles extrapolent sur une très grande proportion de l'espace projeté. L'apport d'information complémentaire sur les limites physiologiques des espèces (observations, résultats d'expériences, sorties de modèles physiologiques) permet de réduire l'extrapolation et d'augmenter la capacité des modèles à décrire la niche réalisée des espèces. L'agrégation spatiale des données, qui influençait les prédictions et l'évaluation des modèles a également pu être corrigée. Enfin, les modèles de dispersion ont montré un potentiel intéressant pour révéler le rôle des barrières géographiques ou à l'inverse, la connectivité spatiale, mais également le lien existant entre distribution, physiologie et histoire phylogénétique des espèces. Ce travail de thèse propose de nombreux conseils et fournit des codes annotés parfois sous forme de tutoriels, afin de constituer une aide utile aux futurs travaux de modélisation sur les espèces marines antarctiques.