



HAL
open science

Approches quantitatives pour comprendre et prédire l'écologie, la distribution et la biodiversité des habitats benthiques dans l'Anthropocène

Clément Violet

► **To cite this version:**

Clément Violet. Approches quantitatives pour comprendre et prédire l'écologie, la distribution et la biodiversité des habitats benthiques dans l'Anthropocène. Biodiversité et Ecologie. Université de Bretagne occidentale - Brest, 2023. Français. NNT : 2023BRES0092 . tel-04611787

HAL Id: tel-04611787

<https://theses.hal.science/tel-04611787>

Submitted on 14 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat de

L'UNIVERSITE DE BRETAGNE OCCIDENTALE

École Doctorale N° 598
Sciences de la Mer et du Littoral
Spécialité : *Ecologie Marine*

Par

Clément VIOLET

**Approches quantitatives pour comprendre et prédire l'écologie,
la distribution et la biodiversité des habitats benthiques dans
l'Anthropocène**

Thèse présentée et soutenue à Plouzané, le 18 décembre 2023
Unité de recherche : Ifremer - DYNECO-LEBCO

Rapporteurs avant soutenance :

Stéphane DRAY Directeur de Recherche, Université Lyon 1
David MOUILLOT Professeur, Université de Montpellier

Composition du Jury :

Présidente :	Emmanuelle CAM	Professeure, Université de Bretagne Occidentale
Examineurs :	Sakina-Dorothée AYATA	Maîtresse de conférences, Sorbonne Université
	Thibaut DE BETTIGNIES	Chargé de mission, PatriNat
	Stéphane DRAY	Directeur de Recherche, Université Lyon 1
	David MOUILLOT	Professeur, Université de Montpellier
Directeur de thèse :	Stanislas DUBOIS	Cadre de Recherche, Ifremer

Invités :

Co-directeur de thèse :	Martin MARZLOFF	Cadre de Recherche, Ifremer
Encadrant scientifique :	Aurélien BOYÉ	Cadre de Recherche, Ifremer
Encadrant scientifique :	Mathieu CHEVALIER	Cadre de Recherche, Ifremer

Remerciements

Il y a un peu plus de vingt ans, qui aurait pu imaginer que le petit garçon que j'étais, blotti dans sa serviette de plage, terrifié à la seule pensée de fouler ce sable répugnant, souillé par les laisses de mer et infesté de ces immondes amphipodes bondissant de toute part, se destinerait à rédiger une thèse en écologie marine ? Ces quelques lignes ne pourront exprimer toute la gratitude que j'ai envers vous, vous qui m'avez permis d'arriver jusqu'au bout de ce périple.

Je souhaite tout d'abord exprimer ma sincère gratitude à Sakina-Dorothee Ayata, Emmanuelle Cam, Thibaut de Bettignies d'avoir accepté de faire partie de ce jury de thèse. Je souhaite remercier tout particulièrement Stéphane Dray et David Mouillot pour avoir accepté d'en être les rapporteurs.

Je remercie également Anik Brind'Amour, Philippe Cugier, Clément Garcia Cédric Gaucherel, Olivier Gauthier pour avoir eu la gentillesse de faire partie de mon CSI. Merci infiniment pour votre temps et votre écoute attentive. Merci pour vos conseils tout au long de cette thèse, qui en ont grandement amélioré la qualité.

Merci à Stanislas Dubois d'avoir dirigé cette thèse et pour l'écoute et tout le temps que tu m'as accordé durant toute cette thèse.

Je souhaite exprimer ma profonde gratitude à mes directeurs de thèse, débutant par toi, Martin, pour avoir accepté de m'accueillir sous ton aile bienveillante dès le Master. Mes remerciements s'étendent également à Aurélien et Mathieu pour leur participation essentielle dans cette formidable aventure qu'a constituée l'encadrement de ma thèse. Votre curiosité intellectuelle insatiable et votre rigueur m'ont constamment poussé vers l'excellence, enrichissant indéniablement mon parcours. La gentillesse et la bienveillance que vous avez tous trois manifestées à mon égard ont grandement contribué au plaisir que j'éprouvais à me rendre au laboratoire chaque jour.

Je tiens à exprimer ma profonde gratitude à tous ceux qui ont contribué et qui contribuent

encore au programme de suivi REBENT. Sans votre travail assidu et votre dévouement, cette thèse n'aurait pas vu le jour. I would also like to extend heartfelt thanks to Graham Edgar, Lizzi Oh, and Rick Stuart-Smith for your invaluable assistance throughout my PhD journey. Your dedication, alongside the tremendous efforts of all the volunteers involved in the Reef Life Survey Program, has been instrumental in the progression and success of my research. Your contributions have not only enriched my work but also underscored the profound impact of collaborative endeavour.

Je tiens à exprimer toute ma reconnaissance à Olivier Gauthier et Emmanuelle Cam qui m'ont généreusement donné l'opportunité d'enseigner durant ces trois années. Chaque heure passée en compagnie des étudiants a été pour moi une source de joie intense, me poussant à approfondir avec rigueur les sujets que j'avais la responsabilité de leur transmettre.

Ma passion pour l'enseignement, je la dois en grande partie à toi, Olivier. Tes cours n'ont pas été seulement instructifs ; ils ont été une véritable révélation. Cependant, cette vocation ne s'est pas nourrie de ta seule influence, aussi significative soit-elle (et c'est bien dommage !). Je suis également redevable à Gauthier Schaal et Jacques Grall, qui m'ont montré qu'enseigner pouvait se faire dans une ambiance chaleureuse et décontractée. Enfin, mes sincères remerciements vont à Christophe et David. Votre approche innovante et particulière de l'enseignement a profondément transformé ma conception de la pédagogie, et pour cela, ma gratitude est infinie.

Je souhaite du plus profond de mon coeur remercier Emilien, Touria, Aline B., Aurélien B., Antoine, Mathieu, Céline, Philippe, Amelia, Gabin, Stan, Justine, Aline G., Pierre-Olivier, Martin, Flavia, Aurélien T. et Mick' pour faire du LEBCO et de sa salle de pause un endroit si chaleureux, merci infiniment pour ces fous rires.

Un immense merci à Véronique pour avoir aidé durant ces trois années à naviguer dans les différents arcanes administratifs d'Ifremer et de m'avoir aidé à préparer mes différentes missions.

Un énorme merci à tous les doctorants (et les pièces rapportées !) de l'Open Space pour cette ambiance incroyable qui nous booste chaque jour ! Une pensée spéciale pour ceux qui ont déjà pris leur envol vers de nouvelles aventures : Bastien, Kévin, Lyndsay, Marion, Lou, et Alex. Léa, dont la gentillesse et le rire ont le pouvoir magique de réchauffer l'Open Space, même lors des matinées frisquettes d'automne, d'hiver, de printemps ou d'été. Un p'tit clin d'oeil à notre Dinde Supérieure, Laure, doyenne incontestée de notre petit univers ; chacune de nos

discussions a été un vrai kif ! Et toi, Mathisse, tu es simplement la meilleure (et pas seulement pour le café brûlant qui m'accueille chaque matin). Impossible de résumer ici tous ces moments épiques vécus ensemble, que ce soit en installant du carrelage et des moulures dans le bureau de Mathieu, lors des sessions cours de surf ou lors de nos innombrables soirées bars. En parlant bar toi, Bastouille, avec tes exploits légendaires, tu restes indétronable ! Nos discussions Tour de France et les étapes dans le Pays-Basque resteront gravés dans ma mémoire. Un immense merci également, pour ces moments "oxygénation" ces dernières semaines, j'en avais grandement besoin. Mathieu, tu es sans conteste LA 'pièce rapportée' de cet Open Space, merci pour ces discussions et ces fous rires en pause ou à la cantine qui resteront gravés dans ma mémoire ! Chonchon, un merci tout spécial pour ton super boulot durant ton M2, essentiel pour ma thèse, et merci pour tous ces memes politiques croustillants sur Twitter, quelle rigolade ! Merci, Irene, d'avoir enduré mon espagnol hésitant sans jamais perdre patience, c'est un exploit ! Quant aux petits nouveaux, Lucas, Maeva, et Chloé, votre aventure ne fait que commencer, mais je suis déjà ravi d'avoir croisé votre route. Je suis persuadé que vous allez faire un excellent job. Savourez chaque instant, malgré les défis que la thèse représente, ça en vaut la peine ! Un immense merci à tous pour ces moments inoubliables !

Un phénoménal big-up à mon crew, la team "A Fond la Fonte" : Manon, Marine, Thomas, Mathisse, Laure, Aurélien, Mathieu ! J'avoue, mon assiduité brillait peut-être plus à la cantine qu'à la salle de sport. Mais qu'importe, chaque déjeuner passé ensemble était épique. Merci pour ces instants mémorables !

Je tiens à saisir cette opportunité pour exprimer ma profonde gratitude envers tous mes amis qui ont jalonné mon parcours depuis mon arrivée à Brest. Un merci tout particulier à Elodie, Elodie, Pierre, Thibault, François, Anthony, Mathilde, Angelina, Maurane, Christophe et David pour tous ces précieux moments partagés. Ma reconnaissance s'étend également aux amis rencontrés au cours de ce voyage académique, notamment Lyndsay, Léa, Lou, Kevin, Lucas, Pierre-Léo, Aurélien, Gabin, Mathieu et Noémie, Romain, Fred, Gaspard et Claire, qui ont chacun illuminé ce voyage.

Merci au Tortuga, aux Fauvettes et maintenant à Gourmand, Mais Pas Que (a.k.a. l'After Work) d'avoir été les partenaires officiels de ma thèse.

Je souhaite rendre un hommage tout particulier à Masha, mon tigre domestique, qui a fait de moi son humain de compagnie.

Merci infiniment à ma famille, et surtout à mes parents, vous méritez amplement toute ma gratitude pour votre soutien constant tout au long de mes études, une fondation solide qui m'a porté jusqu'à cet instant où je rédige ces quelques phrases. Sans vous, rien de tout cela n'aurait été possible et je vous en suis infiniment reconnaissant. Une petite mention spéciale à toi Guillaume, pour avoir partagé ces années inoubliables à Brest avec moi, affichant avec fierté les couleurs du Stade Brestois 29. Ensemble, nous avons bravé la pluie, le vent, la neige, ou même les trois à la fois, dans l'arène légendaire du stade Francis-Le Blé. Merci frangin !

Je souhaite également remercier ma belle famille. Gilles, Dominique, Gwen (sans oublier le p'tit Maël), un merci du fond du cœur pour votre accueil chaleureux, ainsi que pour le soutien indéfectible et l'affection que vous m'avez généreusement offerte au fil des années.

Gaëlle, ton amour et ton soutien inconditionnel ont été mon refuge et ma force durant cette thèse. Ces années n'auraient pas eu la même saveur sans ta présence constante et rassurante à mes côtés. Je sais que les derniers mois ont été rudes pour toi. Sans toi, ce travail de thèse n'aurait pas seulement été plus difficile — il aurait été impossible. Je t'aime.

Ces remerciements sont déjà beaucoup trop longs, mais sache que je te remercie du fond du cœur toi, toi que je n'ai malheureusement pas eu la place de citer, mais qui a croisé ma route pendant cette aventure universitaire et l'a rendu unique. Merci infiniment.

*I don't know anything, but I do know that
everything is interesting if you go into it
deeply enough.*

RICHARD FEYNMAN

Table des matières

1	Introduction Générale	19
1.1	Écosystèmes Côtiers	21
1.1.1	Écosystèmes Côtiers : Biodiversité et Enjeux Humains	21
1.1.2	Contribution des Écosystèmes Côtiers aux Services Écosystémiques	22
1.1.3	Les Écosystèmes Côtiers face aux Pressions Anthropiques	23
1.2	Ecologie des Communautés	27
1.2.1	Mesurer et Interpréter la Diversité des Espèces dans une Communauté	27
1.2.2	La niche écologique : un cadre explicatif de la diversité des communautés	28
1.2.3	Influence des Espèces Ingénieures sur la Structure des Communautés	30
1.2.4	La Disparition des Habitats Biogéniques	31
1.2.5	Dynamique des Écosystèmes : Résistance, Résilience et Changement de Régime	32
1.3	De l'Observation <i>In Situ</i> à la Modélisation	34
1.3.1	Impacts de la Révolution Digitale sur la Recherche en Écologie .	35
1.3.2	Des Méthodes d'Ordination aux Modèles de Distribution d'Espèces : Une Exploration des Approches Quantitatives en Écologie	37
1.4	Contexte & Objectifs de cette Thèse	43
	Bibliographie	47
2	Essential ingredients in Joint Species Distribution Models	59
	Préambule	59
2.1	Abstract	61
2.2	Introduction	62
2.3	Materials & Methods	64
2.3.1	<i>Hierarchical Modelling of Species Community (HMSC)</i>	65

TABLE DES MATIÈRES

2.3.2	Datasets	65
2.3.3	Comparison of alternative model structures	67
2.3.4	Model fitting	67
2.3.5	Assessing model performance and interpretability	68
2.4	Results	72
2.4.1	Model Fit & Predictive power	72
2.4.2	Variance partitioning	74
2.4.3	Species niche estimated	76
2.4.4	Exploring the residual correlation	78
2.5	Discussion	79
	Bibliographie	83
	Appendix A - Data Sources and Descriptions of the Datasets	88
	Appendix B - Model Convergence	92
	Appendix C - Complementary Results	106
3	From local seafloor imagery to global patterns in benthic habitats : contribution of citizen science to habitat classification across latitudes	121
	Préambule	121
3.1	Abstract	123
3.2	Introduction	124
3.3	Materials & Methods	127
3.3.1	Data	128
3.4	Clustering pipeline	130
3.4.1	Dimension reduction - <i>UMAP</i>	130
3.4.2	Clustering - <i>HDBSCAN</i>	131
3.4.3	Evaluation of the clustering output	132
3.4.4	Interpretation of the clusters	132
3.4.5	Spatio-temporal distribution of benthic habitat states	133
3.5	Results	133
3.6	Discussion	142
	Bibliographie	146
	Appendix A - <i>RLS</i> Classification schema	154
	Appendix B - Clustering pipeline results	158
	Appendix C - Interpretation of the uncovered clusters	177

Appendix D - Complementary results	194
4 Patterns of Co-occurrence and Spatial Richness in Marine Habitats : Insights from Large-Scale Modelling	201
Préambule	201
4.1 Abstract	202
4.2 Introduction	203
4.3 Materials & Methods	205
4.3.1 Datasets	205
4.3.2 Statistical analyses	208
4.4 Results	210
4.4.1 Model Performance & Predictors importance	210
4.4.2 Habitat states patterns	210
4.4.3 Spatial distribution	213
4.5 Discussion	216
4.5.1 Explanatory and predictive power of the model	216
4.5.2 Influence of biophysical factors and anthropogenic pressures	218
4.5.3 Co-occurrence patterns	219
4.5.4 Richness Spatial Patterns	220
4.6 Conclusion	221
Bibliographie	221
Appendix A - Datasets	229
Appendix B - Model fitting & Evaluation	235
Appendix C - Model Explanation	266
5 Discussion Générale et Perspectives	283
5.1 Améliorer les Capacités Prédictives des Modèles de Distribution d'Espèces	283
5.2 Modélisation des Niches Environnementales : Entre Attentes et Réalités . .	287
5.3 Analyse des états d'habitats marins à l'échelle du globe : avancées et perspectives	289
5.4 Implications des travaux entrepris dans cette thèse	289
Bibliographie	291

Table des figures

1.1	Illustration du continuum terre-mer et des différents habitats marins qu'abritent les écosystèmes côtiers. De gauche à droite : Pré salé, Mangrove, Herbier marin, Récif de bivalves, Récif corallien, Forêt de laminaires. Illustration WERNBERG et al. (2023)	21
1.2	Illustration de services écosystémiques fournis par les océans. a. Cueillette de <i>Laminaria digitata</i> en mer d'Iroise. b. Cérémonie du <i>Paddle-out</i> . c. Forêt de laminaires et de fucoïdes contribuant à différents services de régulation. © Image : a. Agrocean. b. Tyler Rock, Freesurf Magazine. c. John Turnbull	24
1.3	Anomalies de températures et vagues de chaleurs extrêmes identifiées au large des côtes dans l'est de l'Atlantique le 17 juin 2023. a. Anomalies de températures de surface supérieure à 2°C dans l'ensemble de l'Atlantique Est avec localement des anomalies comprises en +4 et +6°C au large des côtes irlandaises. b. Catégorisation des vagues de chaleur dans l'Atlantique Est, avec des vagues de chaleur allant du niveau 1 "vague de chaleur modérée" (i.e golfe de Gascogne) jusqu'au niveau 5 "au-delà de l'extrême" au large de l'Irlande. © Infographie : Scott DUNCAN (2023)	26
1.4	Niche écologique d'une espèce marine virtuelle n'utilisant que trois facteurs environnementaux ou ressources.	30
1.5	Différentes réponses qu'un écosystème peut présenter face à des pressions. Chaque courbe représente l'ensemble des points d'équilibre potentiel d'un écosystème. Si ces graphes permettent de développer des visions théoriques des dynamiques écologiques, ils sous-entendent une simplification de la nature intrinsèquement changeante des écosystèmes. a. Réponse linéaire b. Réponse avec seuil sans hystérésis c. Réponse unimodale d. Réponse avec seuil et hystérésis. Dans ce cas, les courbes en pointillés reflètent des points d'équilibre instable. Extrait de SPAKE et al. (2022)	33

1.6	Schéma décrivant la réponse théorique d'un écosystème à une pression. a. L'écosystème présente une forte résistance à la pression subite. b. L'écosystème présente une résistance moyenne à la pression, mais une forte résilience. c. L'écosystème présente à la fois une faible résistance et résilience	34
1.7	Processeur Intel 4004 commercialisé en 1971. © Photographie : Thomas NGUYEN (2016)	36
1.8	Schéma du principe de fonctionnement d'une méthode de réduction de dimension basée sur les graphes de voisinages. a. Calcul d'un graphe de voisinage dans l'espace multidimensionnel originel des données. b. Apprentissage d'une représentation de ce graphe dans un espace de plus faible dimension. c. Projection des observations dans l'espace en plus faible dimension. Adapté de MCINNES et al. (2018)	39
1.9	Stratégies de modélisations de distribution d'espèces. FERRIER et GUISAN (2006)	42
1.10	Schéma illustrant les interactions majeures entre les habitats benthiques, la faune qui y est associée et les conditions environnementales. Cadre de pensée dans lequel s'inscrit cette thèse. Adapté d'après MARZLOFF (2022)	44
1.11	Schéma illustrant les travaux réalisés dans le cadre de ce travail de thèse.	46
2.1	Study workflow. A. <i>Hierarchical Model of Species Community (HMSC)</i> structure, incorporating environmental variables, phylogeny, and species-specific traits. B. Data pre-processing : partitioning community data into train and test datasets, estimating phylogenetic distance (using taxonomic classification), and dimension reduction of species-trait matrix using fuzzy-PCA. C. Summary of the four alternative model structures fitted to presence/absence and abundance data : the <i>Benchmark</i> , Phylogeny, Traits & Phylogeny models only consider the targeted polychaetes, while the Whole Community model includes additional species potentially interacting with the target assemblage. Random effects for sampling year, site, and habitat were included in all models.	70

2.2 Change in explanatory and predictive power of three model structures (yellow for *Phylogeny (Ph)*, red for *Traits and phylogeny (TrPh)*, and green for *Whole community (WhC)* models) relative to the *Benchmark model (Bench; purple)*. Changes are expressed as percentages relative to the benchmark fitted on presence/absence (top row) or abundance (bottom row) data. Points above the zero line indicate performance gain. 73

2.3 Comparison of model performances with regards to their ability to predict community structures when fitted with abundance data for the train (purple) and test (yellow) dataset. Left : differences in pairwise dissimilarities estimated on the observed and the predicted assemblages. Centre : differences in species richness between observed and predicted assemblages. Right : differences in total abundance between observed and predicted assemblages. 75

2.4 Change in explained variance related to environmental predictors (left column) and random effects (right column) for three alternative model structures (yellow for *Phylogeny (Ph)*, red for *Traits and phylogeny (TrPh)*, and green for *Whole community (WhC)* models) relative to the *Benchmark model (Bench; purple)*. Percentage changes were computed relative to the *Benchmark model* fitted with presence/absence (top panels) or abundance (bottom panels) data. Positive values indicate an increase in the proportion of variance explained by the focal model compared to the *Benchmark model*. See Figure S23 and S24 for the raw percentages, expressed as percentages of explained variance or total amount of variance respectively. . . . 76

2.5 Number (y-axis) and proportion (computed across all coefficients for each model, as indicated above individual bars) of response curves (i.e. one for each species-predictor combination according to the typology (nine shapes highlighted by the black curve in each panel) defined by RIGAL et al. (2020). Results are presented for the different model structures 77

2.6 (A) Comparison of residual species-species correlations associated with the three random effects estimated by the *Whole Community Model* (*WhC*; y-axis) and the *Benchmark model* (*Bench*; x-axis) fitted on abundance data. The colour scale highlights the density of points in each scatter plot. (B) Distribution of the δ index characterizing change in sign (negative values indicate sign change) and magnitude (higher absolute values indicate higher numerical difference) between residual correlations estimated by the *WhC* model and the *Bench* model adjusted with abundance data for the three random effects (Habitat, Site, Year). 79

3.1 Comparaison de différents algorithmes de groupement utilisés en écologie (en colonne; à savoir k-means, Ward, groupement lien complet ou lien simple, DBSCAN ou HDBSCAN) sur différents jeux de données simulés en deux dimensions (en ligne). Les jeux de données des deux premières lignes contiennent deux groupes, les trois derniers en contiennent trois. . . 123

3.2 Two-dimensional *UMAP* embedding of the benthic cover data of the 6,554 *RLS* transects. Each point corresponds to an *RLS* transect, coloured according to membership for the selected *UMAP-HDBSCAN* pipeline. Black dots represent points classified as noise (n=1464). The 17 clusters can be interpreted as follows (see Fig. 3.3 and S20-36) : 1. *Foliose brown algae* (n=148) 2. *Filamentous algae* (n=208) 3. *Other Sessile invertebrates* (n=185) 4. *Foliose red algae* (n=123) 5. *Seagrass* (n=83) 6. *Soft coral and gorgonians* (n=98) 7. *Bushy fucooids* (n=577) 8. *Canopy forming algae* (n=894) 9. *Unconsolidated substrate* (n=151) 10. *Crustose coralline and turf algae* (n=286) 11. *Green calcified algae* (n=166) 12. *Bare substrates* (n=329) 13. *Crustose coralline algae* (n=409) 14. *Sand* (n=220) 15. *Branching coral* (n=110) 16. *Turf and sand* (n=207) 17. *Turf algae* (n=897) . . 134

3.3 a. *HDBSCAN* condensed clustering tree of the *UMAP* 2D embedding b. Heatmap of the mean substrate coverage (rounded to the nearest integer) for each cluster identified by the *UMAP-HDBSCAN* pipeline. 137

3.4 Violin plot of the absolute latitudinal distribution of the different hard cluster solutions. 138

TABLE DES FIGURES

3.5 a. Spatial distribution of reef surveys from the Reef Life Survey database used for analyses. b. Map of dominant clusters in each MEOW ecoregion. Dominant clusters were determined as the greatest count of transect labels in each ecoregion. c. Spatial distribution of the proportion of transects classified as noise in each ecoregion. d. Gini-Simpson diversity index calculated by the occurrence of clusters in each ecoregion of the world 140

3.6 Well-sampled ecoregions in Australia, with high number of transects and temporal replications between 2008 and 2021. b. Distribution of 44 sites surveyed between 2008 and 2021 in the Cape Howe region ; Colour coding indicates cluster identity. Dots are jittered along the y-axis. c. Number and proportion of transects in the different clusters for the period 2008-2013 in the four ecoregions highlighted in a. d. The same analysis for the period 2014-2021. 141

3.7 Two-dimensional UMAP embedding of the 5,090 clustered *RLS* transects. Each picture is representative of its cluster. The arrows indicate potential transitions as identified in JOUFFRAY et al. (2015) and CORNWALL et al. (2023). 143

4.1 Heatmap of the relative contribution of variables to the Random Forest explanatory power. The variables are colour coded according to the type of predictor (blue : abiotic conditions, green : fishing pressure, orange : other anthropogenic pressures). The relative contribution of each variable for a given state is expressed relative to the sum of all variable's contributions for that state. 211

4.2 Partial Dependence Plots of the a. Habitat states richness b. probability of each of the 14 habitat states against the three most important variables identified previously for these habitat states, namely : Sea Surface Temperature, Fishing PC1 (Fishing pressure increases as Fishing PC1 increases) and Marine Heatwaves Anomaly. For all plots, the X-axis is expressed in the original unit of the variables. For graphical representation, we only plot here the [0% ; 90%] values of Fishing PC1, excluding the few positive extreme values. 212

4.3 a. Distribution of the number of possible habitat states per transects, the numbers above the bar represent the number of transects predicted to have this number of habitat states. b. Relative frequency of habitat states as a function of predicted number habitat states in one transect. c. Heatmap of the frequency of co-occurrence of each habitat state. Each line of the heatmap has been standardised by the total number of co-occurrences with the focal habitat state. 214

4.4 a. Minimal, b. median and c. maximal richness (i.e. number of predicted reef states) per spatial block 215

5.1 Comparaison entre les quatre structures de modèle alternatives étudiées dans le premier chapitre de cette thèse. Le graphique présente pour chaque espèce (le long de l'axe des x) la variance qui est expliquée (le long de l'axe y) par (1) les variables environnementales (Environnement) et (2) les trois effets aléatoires. Les résultats sont présentés pour les modèles ajustés aux données d'abondance (à gauche) et de présence/absence (à droite). Les espèces sont classées par ordre décroissant de variance expliquée par l'environnement pour le modèle de référence. 288

Liste des tableaux

2.1	Multi-assessment framework providing a list of metrics to assess, interpret or compare jSDMs across different ecological facets at the species, community or overall level.	71
3.1	Description of the 24 categories used in this study to overall capture the diversity of habitat types sampled by <i>Reef Life Surveys</i> worldwide. The 50 original <i>RLS</i> categories were lumped into these 24 categories that represent ecologically-consistent groups associated with different levels of structural complexity.	128
4.1	Environmental variables extracted from <i>Bio-Oracle</i> and anthropogenic variables extracted from <i>Ocean Health Index</i> . See ASSIS et al. (2018) for a detailed description of the <i>Bio-Oracle</i> variables and their acquisition method and Table S1 to S3 in HALPERN et al. (2019) for detailed description and acquisition method of anthropogenic variables	207

Introduction Générale

L'écologie est une science multidisciplinaire, qui a subi des développements rapides au cours des dernières décennies et dont le but est de comprendre la contribution relative des facteurs biotiques et abiotiques sur la dynamique spatio-temporelle de la biodiversité (REECE et al., 2012). Ces patrons de biodiversité peuvent s'étudier à des échelles différentes, qui peuvent être grandes comme les forêts tropicales ou la mer (CHAPIN et al., 2011), ou à des échelles plus locales, par exemple le microbiome intestinal (FOSTER et al., 2017). La biodiversité peut se mesurer à différents niveaux d'organisation allant de l'individu à l'écosystème en passant par les espèces, les populations ou les communautés. Différentes méthodes existent pour décrire et comprendre les patrons de biodiversité et ont donné lieu à la création de nouvelles sous-disciplines comme l'écologie expérimentale, l'écologie des communautés ou l'écologie quantitative.

Les premiers naturalistes qui contribuèrent au développement de l'écologie ont tout d'abord adopté une approche descriptive en voyageant à travers le monde et en collectant des spécimens végétaux et animaux comme l'on fait LINNAEUS (1789), HUMBOLDT et BONPLAND (1805), DARWIN (1839) et bien d'autres encore. Pendant leurs voyages, ces pionniers de l'écologie ont été fascinés par les différences dans le nombre d'espèces pouvant coexister dans différentes parties de globe. Ces patrons de diversité observés ont généré un ensemble de questions macroécologiques auxquelles il était difficile de répondre à cause du manque de données à l'échelle du globe et de moyens de les analyser. Ainsi, bien que des hypothèses furent développées rapidement pour expliquer les variations de biodiversité observée, ces dernières étaient difficilement testables. Pendant longtemps, les écologues ne purent s'intéresser qu'à des échelles spatiales et temporelles restreintes et tester des hypothèses sur les patrons de distribution des espèces à des échelles locales. Grâce aux développements technologiques permettant de collecter toujours plus de don-

nées à des résolutions toujours plus fines et à des échelles spatio-temporelles toujours plus importantes, par exemple avec les apports de la télédétection ou les développements de la biologie moléculaire (EDGAR et al., 2016) et l’apport d’outils numériques et de modélisation (LEGENDRE & LEGENDRE, 2012; GUISAN et al., 2017), les chercheurs ont pu développer de nouvelles hypothèses et théories pour expliquer les patrons de diversité observée sur notre planète (voir par exemple HUBBELL (2001), VELLEND (2017) ou bien encore LEIBOLD et CHASE (2018)).

Cette thèse s’inscrit dans un des sous-domaines de l’écologie qui est celle de l’écologie des communautés. L’écologie des communautés s’intéresse aux communautés, c’est-à-dire aux espèces vivant interagissant entre elles de manière directe ou indirecte, même endroit et au même moment (VELLEND, 2017). Ce champ de recherche vise à identifier les règles générales qui régissent les patrons de diversité, l’abondance et la composition en espèces des communautés (VELLEND, 2010).

Lors de l’émergence de ce champ disciplinaire, les chercheurs se sont intéressés tout d’abord à étudier les règles d’assemblage des espèces autour du prisme des interactions biotiques et particulièrement trophiques ou bien en se concentrant sur les facteurs abiotiques et les gradients environnementaux (WHITTAKER, 1975). Ces études se firent tout d’abord sur des échelles spatiales limitées (WHITTAKER, 1960). A l’ère du numérique, la collecte massive de données à de plus grandes échelles spatiales et l’application des outils de l’écologie quantitative nous permettent de mieux prendre en compte l’importance relative de ces différents facteurs (GASTON, 2000). Dans le cadre de cette thèse, et en m’appuyant sur des jeux de données collectées à plus ou moins larges échelles spatiales et temporelles, je vais essayer de comprendre en appliquant certains des outils, des techniques et méthodes de l’écologie quantitative, comment les habitats benthiques biogéniques structurent la diversité des communautés des écosystèmes marins côtiers et comment ils répondent face aux changements environnementaux et anthropiques, afin de mieux appréhender le futur des écosystèmes marins côtiers face aux changements globaux.

1.1 Écosystèmes Côtiers

1.1.1 Écosystèmes Côtiers : Biodiversité et Enjeux Humains

À la croisée des milieux terrestres et marins, les écosystèmes côtiers sont des écosystèmes abritant une grande biodiversité (MCLEAN et al., 2001). Cette biodiversité est supportée par la variété d'habitats que comprennent les écosystèmes côtiers (Fig. 1.1). Côté terre, on trouve des zones urbaines, industrielles, mais également des espaces naturels comme les falaises ou les dunes (BURKE et al., 2000). Côté mer, ces écosystèmes comprennent des milieux intertidaux tels que les plages, les estuaires et les lagons, ainsi que des milieux subtidaux comme les forêts de laminaires ou les récifs coralliens (BURKE et al., 2000), qui sont parmi les milieux abritant une des plus grandes biodiversités et productivités au monde (WERNBERG et al., 2023).

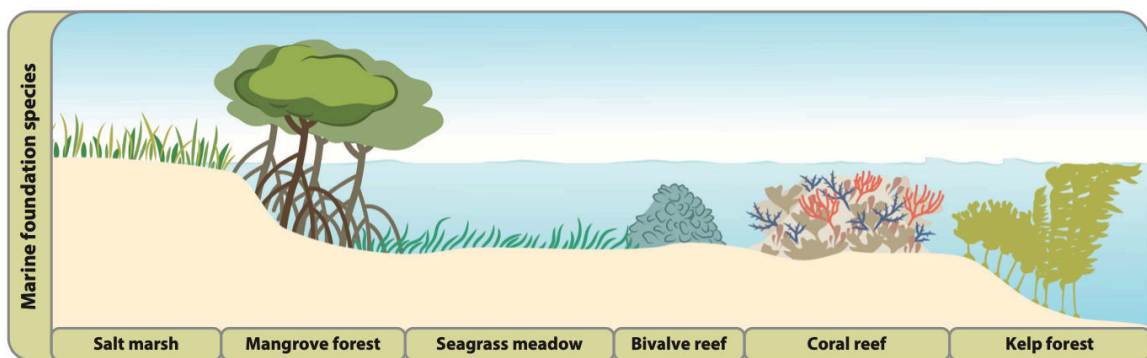


Figure 1.1 – Illustration du continuum terre-mer et des différents habitats marins qu'abritent les écosystèmes côtiers. De gauche à droite : Pré salé, Mangrove, Herbier marin, Récif de bivalves, Récif corallien, Forêt de laminaires. Illustration WERNBERG et al. (2023)

Parallèlement à leur importance écologique, les régions côtières jouent un rôle crucial dans la dynamique des populations humaines. En effet, 65% des mégaloilles de la planète sont situées sur les côtes (BLACKBURN et al., 2019). Actuellement, près de 40% de la population, soit 2,8 milliards d'individus, vivent dans un rayon de 100 km autour des côtes (MAUL & DUEDALL, 2019). Si ce pourcentage devrait rester stable jusqu'en 2035, la population côtière devrait néanmoins augmenter de 355 millions d'habitants au cours des quinze prochaines années (MAUL & DUEDALL, 2019). Cette concentration démographique illustre l'importance cruciale des zones côtières pour le développement socio-économique dans les décennies à venir, mais cette concentration démographique implique également des pressions importantes sur les écosystèmes côtiers.

En somme, les écosystèmes côtiers, en plus d'être d'une forte richesse écologique, sont au cœur d'enjeux démographiques et socio-économiques. Leur préservation et leur gestion durable sont essentielles, non seulement au regard de la biodiversité qu'ils abritent, mais aussi pour l'avenir des populations humaines qui en dépendent.

1.1.2 Contribution des Écosystèmes Côtiers aux Services Écosystémiques

Notre planète est recouverte à 71% par les océans ; or les zones côtières définies par la limite du plateau continental ne représentent qu'environ 9% de la superficie des océans (HARRIS et al., 2014). Cependant, cette zone abriterait y abriterait selon les estimations, près de 91 % de la biodiversité (COSTELLO & CHAUDHARY, 2017). Cette forte biodiversité soutient un éventail de services écosystémiques essentiels aux populations humaines. Ces services peuvent être classés en différentes catégories : les services de régulation, les services d'approvisionnement et les services culturels (REID et al. (2005) ; Fig. 1.2).

Les services d'approvisionnement sont les plus simples à définir. Ce sont les services écosystémiques qui fournissent des ressources à l'être humain comme de la nourriture ou bien des matériaux. Par exemple, les petites pêcheries côtières capturent environ 25 millions de tonnes de poissons par an (COCHRANE et al., 2023) et 90% des espèces ciblées par la pêche (FAO, 2022). De même, les pêcheries de crustacés et de coquillages ainsi que leurs élevages représentent des secteurs commerciaux importants avec respectivement, près de 17 millions et 24 millions de tonnes (FAO, 2022). Les écosystèmes côtiers constituent également un réservoir important en termes de ressources médicinales. De nombreux organismes tels que les algues, oursins ou hippocampes entrent dans la composition de médicament : par exemple, dans celle de nombreuses médecines traditionnelles (KUMARAVEL et al., 2012 ; THURSTAN et al., 2018 ; SIBIYA et al., 2021), mais également en médecine moderne avec par exemple l'isolement de composés anticancéreux (SCHWARTSMANN et al., 2001), antiviraux (YASUHARA-BELL & LU, 2010), antibiotiques (BENOIST et al., 2020) ou antidouleurs (MILJANICH, 2004).

Les écosystèmes marins côtiers offrent un nombre de services immatériels, souvent encapsulés dans la terminologie des services culturels, qui englobent une gamme diversifiée de bénéfices que l'humanité tire directement ou indirectement de ces écosystèmes. L'un des aspects les plus manifestes réside dans l'esthétisme des paysages marins, qui inspirent l'art, la littérature et d'autres expressions culturelles (GHERMANDI et al., 2010). Parallèlement,

la biodiversité et la beauté naturelle de ces régions stimulent des activités récréationnelles et constituent un moteur essentiel du tourisme côtier. Les écosystèmes côtiers sont aussi au centre des cultures et des spiritualités qu'entretiennent les populations côtières avec leur environnement marin. En effet, pour de nombreuses populations, la mer n'est pas simplement un élément géographique, mais un pilier de leur patrimoine culturel et de leur histoire collective (GHERMANDI et al., 2010).

Les services de régulation sont eux moins visibles, puisqu'ils rassemblent tous les processus qui permettent de maintenir notre environnement dans des conditions stables. Par exemple, les écosystèmes côtiers participent à la régulation de la qualité de l'eau, en piégeant des polluants dans les écosystèmes estuariens (BARBIER et al., 2011) ou en protégeant le trait de côte des vagues et de l'érosion à l'aide des récifs coralliens, des prés-salés ou bien des forêts de mangroves (BARBIER et al., 2011; HARRIS et al., 2018). Les herbiers marins et les forêts de laminaires offrent également d'autres services en permettant notamment de moduler les effets du changement climatique global en agissant comme des puits de carbones (FOURQUREAN et al., 2012; FILBEE-DEXTER & WERNBERG, 2018).

La valeur économique des services produits par les écosystèmes est toutefois dépendante de leur état de santé (SUTTON et al., 2016). Or plus de 75% des écosystèmes terrestres et 66% des écosystèmes marins sont dégradés (IPBES, 2019). Cette dégradation conduit à une réduction drastique de la capacité des écosystèmes à fournir des services essentiels et à soutenir les activités économiques qui en dépendent (IPBES, 2019). Ainsi, la préservation et la restauration des écosystèmes deviennent non seulement un impératif écologique et économique, mais sont aussi essentielles pour le bien-être et la santé des populations humaines. Assurer leur durabilité est primordial pour garantir la pérennité des bénéfices socio-économiques qu'ils procurent à l'humanité (IPBES, 2019).

1.1.3 Les Écosystèmes Côtiers face aux Pressions Anthropiques

Le rôle majeur de l'espèce humaine dans la transformation de notre planète n'est maintenant plus à démontrer. Notre impact est tellement significatif que les géologues proposent de désigner notre époque actuelle comme l'Anthropocène (SUBRAMANIAN, 2019). Ce nouvel âge géologique est symbolisé par la libération massive de CO₂ dans l'atmosphère. Plus de 555 milliards de tonnes depuis la première révolution industrielle ont été relâchées (LEWIS & MASLIN, 2015), entraînant un changement climatique global (CALVIN et al., 2023). Face à

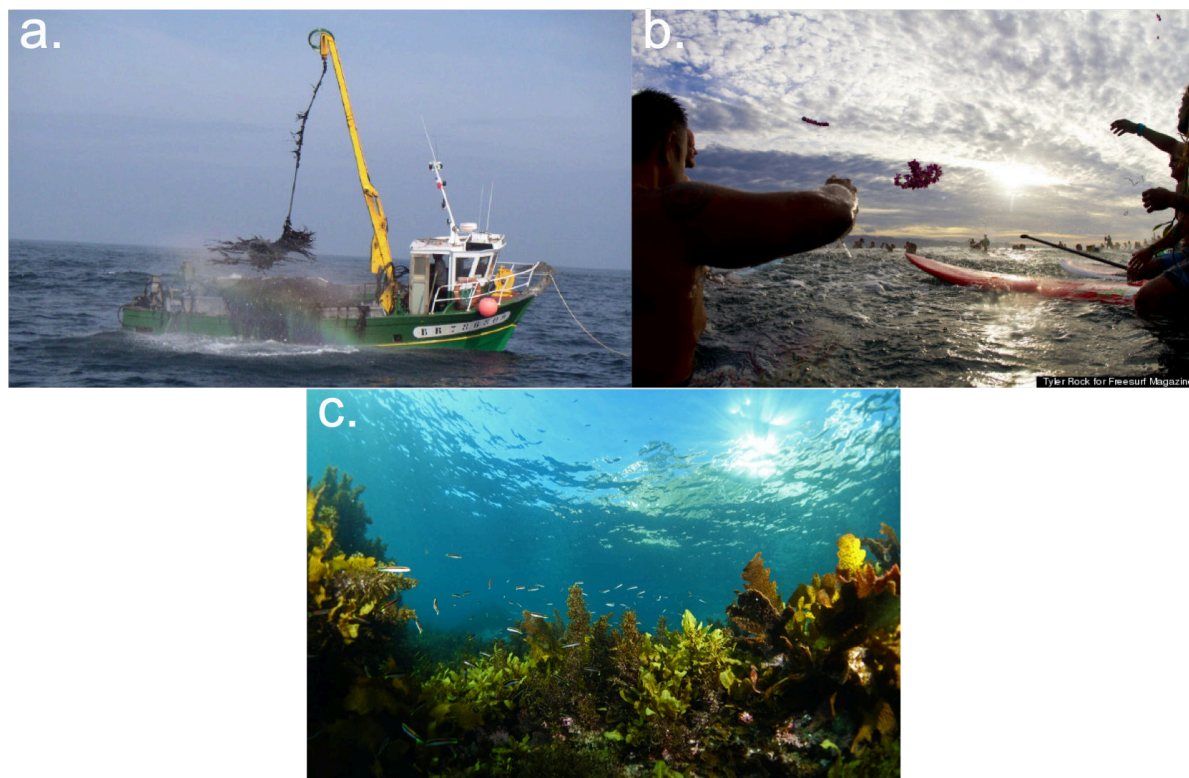


Figure 1.2 – *Illustration de services écosystémiques fournis par les océans. a. Cueillette de Laminaria digitata en mer d'Iroise. b. Cérémonie du Paddle-out. c. Forêt de laminaires et de fucoïdes contribuant à différents services de régulation. © Image : a. Agrocean. b. Tyler Rock, Freesurf Magazine. c. John Turnbull*

ce bouleversement, la biosphère, et notamment les écosystèmes côtiers, se retrouve grandement perturbée. Les modifications que nous observons aujourd'hui, liées aux activités humaines (CALVIN et al., 2023), peuvent être classées en deux types d'impacts : directs et indirects.

La plupart des pressions exercées sur les écosystèmes côtiers découlent de l'augmentation de la densité de population côtière. La densification de la population humaine le long des côtes entraîne de nombreuses perturbations, par exemple des ruissellements urbains qui sont source de différentes pollutions que ce soit aux métaux lourds ou en polluants organiques ou plastiques (TODD et al., 2019). A ces pollutions facilement identifiables s'ajoutent celles liées au bruit ou à la lumière (TODD et al., 2019). Ces dernières peuvent modifier le comportement de nombreuses espèces (TODD et al., 2019) et entraîner des modifications plus ou moins profondes des communautés. Par exemple, les efflorescences

d'algues causées par l'eutrophisation peuvent entraîner la mort de nombreux organismes comme des poissons, crustacés ou bien mollusques (TODD et al., 2019). Ces efflorescences ont également un fort impact économique sur les zones littorales souvent lié entre autres à une diminution de l'activité touristique, à des impacts sur la santé humaine et sur les pêcheries (DOBRICIC et al., 2016).

L'accroissement des activités humaines liées à l'usage de l'espace côtier entraîne également d'autres menaces, par exemple la prolifération d'espèces non indigènes envahissantes, souvent introduites par des navires en transit dans différents ports (HARDIMAN & BURGİN, 2010). Attachées aux coques des navires, ou contenue dans l'eau des ballastes de plus gros navires, des espèces non indigènes animales comme végétales ainsi que de nouveaux pathogènes peuvent être introduits dans de nouvelles zones géographiques (HARDIMAN & BURGİN, 2010).

D'autres impacts anthropiques directs viennent de l'exploitation des ressources des espaces maritimes côtiers. Ainsi, les ressources halieutiques sont largement exploitées : près de 65% des espèces commerciales sont exploitées à leur niveau maximal alors que 35% des espèces sont surexploitées (FAO, 2022). L'extraction d'autres ressources comme le sable, utilisé pour la construction et extrait via des opérations de dragage, impacte également fortement les écosystèmes côtiers, en particulier les écosystèmes benthiques (BOYD et al., 2003; BOYD et al., 2005; ERFTEMEIJER & ROBIN LEWIS, 2006). Au-delà de ces impacts directs, les opérations de dragage peuvent modifier la bathymétrie et les courants (BRAY, 2008), remettent en suspensions des sédiments dans la colonne d'eau (@ ERFTEMEIJER & ROBIN LEWIS, 2006), et relarguent un certain nombre de polluants (FILHO et al., 2004).

L'impact anthropique indirect sur les écosystèmes est, quant à lui, principalement dû au changement climatique global, lié aux émissions de gaz à effet de serre (CO₂ et CH₄ en particulier), avec un réchauffement de l'océan compris entre 0.68 et 1.01 °C depuis l'ère préindustrielle (FOX-KEMPER et al., 2021). Conjointement avec l'augmentation globale des températures moyennes de l'eau de surface, on observe une augmentation en fréquence, en durée et en intensité d'événements climatiques marins extrêmes, tels que les vagues de chaleur marine (OLIVER et al., 2018) ou les tempêtes. Ces vagues de chaleur sont l'équivalent des canicules terrestres et se produisent tant en zones tempérées que tropicales (HOBDAY et al. (2016); Fig. 1.3). Elles ont, par exemple, entraîné des vagues de blanchissement des

coraux et une augmentation de leur mortalité autour de Moorea en 2019 (WYATT et al., 2023) ou encore la disparition de forêts de laminaires le long de la côte ouest-australienne (WERNBERG et al., 2016). Les prédictions du *GIEC* font état d'une augmentation de 4 à 8 fois de la fréquence des vagues de chaleur en fonction des scénarios de réduction des émissions de gaz à effet de serre (FOX-KEMPER et al., 2021). Ainsi l'impact de ces canicules marines sur les écosystèmes marins devrait augmenter dans les années à venir (OLIVER et al., 2018 ; SMITH et al., 2023).

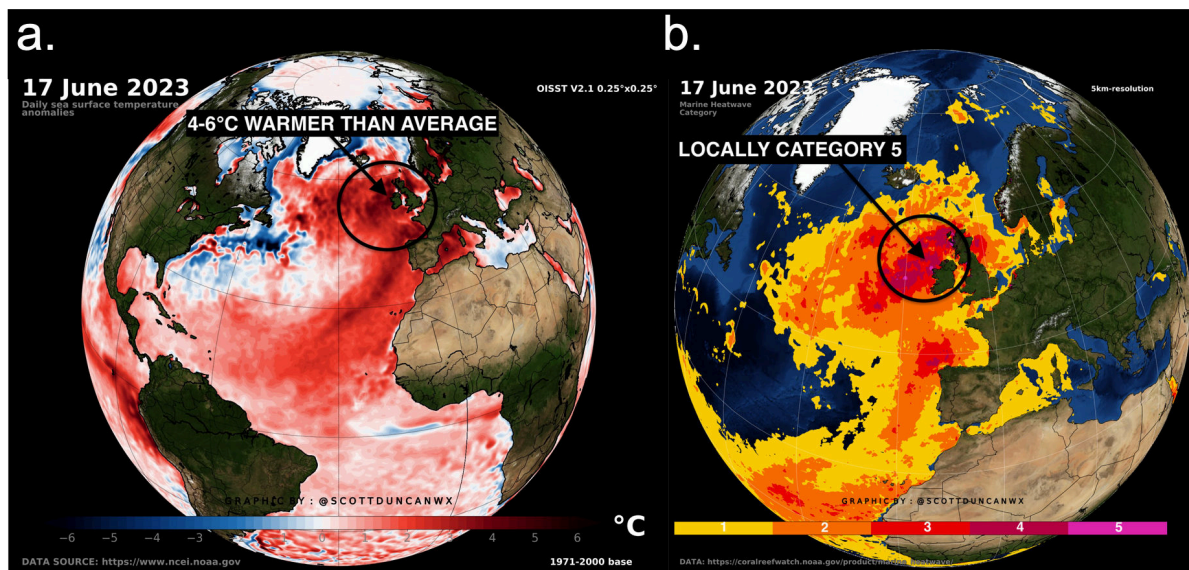


Figure 1.3 – Anomalies de températures et vagues de chaleurs extrêmes identifiées au large des côtes dans l’est de l’Atlantique le 17 juin 2023. a. Anomalies de températures de surface supérieure à 2°C dans l’ensemble de l’Atlantique Est avec localement des anomalies comprises en +4 et +6°C au large des côtes irlandaises. b. Catégorisation des vagues de chaleur dans l’Atlantique Est, avec des vagues de chaleur allant du niveau 1 “vague de chaleur modérée” (i.e golfe de Gascogne) jusqu’au niveau 5 “au-delà de l’extrême” au large de l’Irlande. © Infographie : Scott DUNCAN (2023)

Les activités anthropiques ont également induit des perturbations considérables dans la chimie des océans. Les bassins océaniques agissent comme des puits de carbone en transformant le CO₂ en acide carbonique via le processus de dissolution. Cette dissolution entraîne une réduction du pH marin (CANADELL et al., 2021). Depuis l’ère préindustrielle, on note une diminution du pH océanique, oscillant initialement au-dessus de 8.1 pour se rapprocher de 8 actuellement soit une acidification de 21% (CANADELL et al., 2021). Cette acidification a des répercussions notables sur les écosystèmes côtiers. Notamment, elle affecte la capacité des organismes tels que les algues coralligènes, les coraux et les

bivalves à synthétiser leurs structures calcaires (CHAN & CONNOLLY, 2013; KUFFNER et al., 2008; ZHAO et al., 2017). La modification de la composition chimique des océans peut se traduire également par une diminution de la concentration en oxygène (SCHMIDTKO et al., 2017; BREITBURG et al., 2018). Les zones hypoxiques ont des conséquences majeures pour les organismes côtiers : une teneur en oxygène plus faible compromet la croissance et les fonctions reproductives des organismes les moins tolérants avec des conséquences sur les communautés et le fonctionnement des écosystèmes (LIMBURG et al., 2020).

Les écosystèmes côtiers subissent une augmentation constante de ces multiples pressions (IPBES, 2019). Ces pressions opèrent à différentes échelles spatiales et temporelles et génèrent des effets complexes souvent cumulatifs, ce qui complique la compréhension et surtout la prédiction de leurs impacts. L'évolution récente des écosystèmes terrestres et marins (CARMONA et al., 2021), dans le contexte du changement climatique global, met en lumière la nécessité de mise en place de mesures de préservation et de stratégies de gestion durable. Sans interventions, ces écosystèmes pourraient atteindre un seuil de dégradation au-delà duquel leur restauration deviendrait très limitée, voire impossible (PALMER & STEWART, 2020).

Les perturbations subies par les écosystèmes affectent directement les communautés d'espèces qu'ils abritent. Pour appréhender pleinement les conséquences de ces changements sur la structure et le fonctionnement des écosystèmes côtiers, il est essentiel de saisir comment les communautés y maintiennent leur équilibre. La section suivante se penchera sur l'écologie des communautés, mettant en lumière l'importance des espèces animales et végétales pour la résilience des écosystèmes côtiers face aux pressions environnementales et anthropiques.

1.2 Ecologie des Communautés

1.2.1 Mesurer et Interpréter la Diversité des Espèces dans une Communauté

Les communautés écologiques, définies comme l'ensemble des espèces coexistant dans une zone géographique spécifique, sont caractérisées par leur diversité spécifique et l'abondance respective de chaque espèce (VELLEND, 2010). Au niveau local, la diversité d'une communauté, appelée diversité α (WHITTAKER, 1960), se reflète par exemple, dans le

nombre d'espèces présentes (c.-à-d. la richesse spécifique). Lors de la comparaison des communautés entre plusieurs sites, c'est-à-dire en évaluant la variabilité des communautés en termes de composition et de structure (par exemple, les patrons de dominance en termes d'abondance), les chercheurs évaluent la diversité β (WHITTAKER, 1972). Une faible diversité β signifie que les sites ont des communautés similaires, tandis qu'une valeur élevée indique des différences marquées entre les communautés. Étudier la diversité β est une manière pour les chercheurs d'appréhender les changements environnementaux en cours, car ces changements n'affectent pas nécessairement le nombre d'espèces présentes dans une zone de manière consistante. En revanche, les effets de ces changements sont visibles à toutes les échelles via la composition des communautés (DORNELAS et al., 2023).

Mesurer la diversité spécifique, ne permet pas d'appréhender l'ensemble des facettes de la biodiversité. La biodiversité est caractérisée par une multitude de facettes différentes, dont la diversité des lignées évolutives (c.-à-d. diversité phylogénétique), ou bien la diversité des traits fonctionnels des espèces (c.-à-d. diversité fonctionnelle) (BAGOUSSE-PINGUET et al., 2019). La diversité fonctionnelle est un aspect important de la biodiversité à mesurer, puisque l'étude des variations de cet aspect de la biodiversité permet de mieux comprendre l'effet des perturbations sur le fonctionnement global des écosystèmes (MOUILLOT et al., 2013).

L'écologie des communautés, champ disciplinaire dans lequel s'inscrivent ces travaux de thèse, vise donc à comprendre les facteurs régissant la diversité des communautés, tout en cherchant à en établir des principes universels. Ainsi, pour comprendre ces facteurs, il est important de décrire les changements de biodiversité à différentes échelles spatiales et temporelles, et via différentes méthodes pour comprendre la biodiversité (DORNELAS et al., 2023).

1.2.2 La niche écologique : un cadre explicatif de la diversité des communautés

Alors que les mesures de diversité offrent des indications précieuses sur la variabilité et la similarité des communautés d'un écosystème, elles ne fournissent pas, à elles seules, un cadre explicatif des mécanismes sous-jacents. Pour comprendre les facteurs qui façonnent ces patrons de diversité, il est essentiel de se tourner vers la théorie des niches. Cette dernière nous offre des outils et des perspectives pour appréhender les rôles écologiques et

les besoins environnementaux des espèces, permettant ainsi d'expliquer les raisons pour lesquelles certaines espèces coexistent tandis que d'autres sont exclues.

Joseph Grinnell dans son article fondateur a décrit le concept de niche écologique (GRINNELL, 1917). La niche écologique grinnellienne peut être définie comme l'ensemble des conditions environnementales qui permettent aux espèces de persister et de se reproduire (GRINNELL, 1917). Quelques années plus tard, Elton étendit le concept de Grinnell en définissant la niche écologique comme "la place d'un animal dans l'environnement abiotique et sa relation aux sources de nourriture et à ses ennemies" (ELTON, 1927). Aujourd'hui, la définition de niche écologique la plus utilisée en écologie moderne est la définition de Hutchinson qui décrit la niche écologique comme un système de coordonnées qui permettent de décrire un hypervolume à n-dimension (HUTCHINSON (1957); Fig. 1.4). Cet espace multidimensionnel compte une dimension pour chaque ressource utilisée par l'espèce en question, ainsi qu'une dimension supplémentaire pour chaque espèce interagissant avec elle (HUTCHINSON, 1957). Hutchinson décrivit plus précisément deux types de niches : la niche fondamentale, l'hypervolume qui décrit l'ensemble des conditions favorables dans lesquelles une espèce peut exister, et la niche réalisée, qui est un sous-ensemble de la niche fondamentale contrainte par les interactions biotiques et la disponibilité environnementale (p. ex. contrainte de dispersion). La niche réalisée, accessible aux écologues par le biais de mesures *in situ*, est souvent perçue comme tronquée, car les complexités inhérentes à la mesure exhaustive de toutes les dimensions de la niche sur le terrain est difficile (Fig. 1.4; CHEVALIER et al. (2022)).

Le concept de niche est utile aux écologues, car il permet de mesurer un optimum écologique (c.-à-d. les conditions où la *fitness* de l'espèce est maximal), d'estimer la largeur de la niche (c.-à-d. l'ensemble des conditions environnementales dans lesquelles une espèce peut être trouvée), de mesurer le chevauchement des niches de deux espèces (c.-à-d. dans quelle mesure deux espèces peuvent être trouvées dans des conditions similaires) (CHASE & LEIBOLD, 2003).

Ainsi, la théorie des niches avance l'idée que la coexistence de nombreuses espèces en un lieu spécifique résulte du fait que ces espèces occupent des niches écologiques distinctes, présentant ainsi un faible chevauchement. Ce constat conduit à une diversité α élevée pour ce site. En ce qui concerne la diversité β , lorsque deux sites possèdent des conditions environnementales semblables et que les niches disponibles sont comparables entre elles,

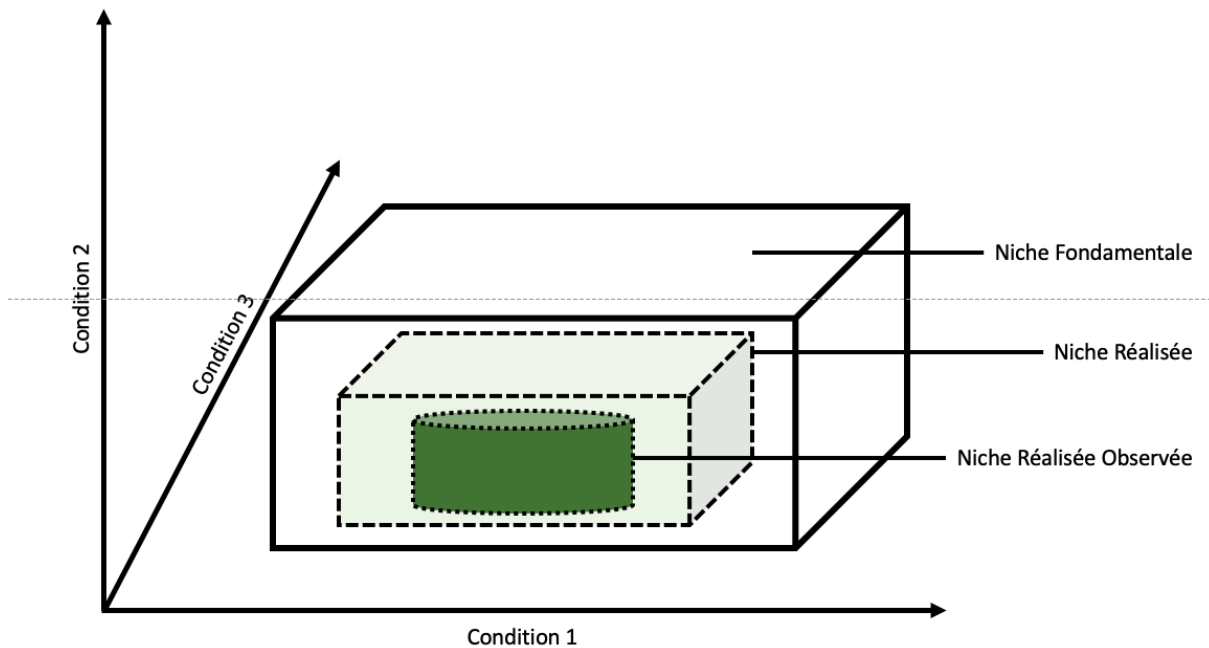


Figure 1.4 – Niche écologique d'une espèce marine virtuelle n'utilisant que trois facteurs environnementaux ou ressources.

la diversité α sera faible. À l'opposé, si les conditions environnementales diffèrent entre les deux sites, la diversification potentielle des niches entre ces habitats pourrait entraîner une diversité β élevée.

1.2.3 Influence des Espèces Ingénieures sur la Structure des Communautés

Au sein d'une communauté, toutes les espèces qui la composent n'ont pas la même influence sur sa structure et ses fonctions. Parmi l'ensemble des espèces qui ont un rôle particulier pour structurer la communauté, il est important dans le cadre de cette thèse de mentionner les espèces ingénieures. Une espèce ingénieure est une espèce qui de par ses structures, sa présence ou son comportement modifie la structure et le fonctionnement de son habitat (JONES et al., 1994).

Il existe plusieurs catégories d'espèces ingénieures. Parmi ces catégories, il est possible de distinguer les espèces fondatrices (DAYTON, 1972) qui jouent un rôle crucial dans la structuration des communautés, car leurs fortes abondances ou biomasses vont modifier

profondément l'environnement (ELLISON, 2019). Leur importance est d'autant plus marquée lorsque l'on reconnaît que bon nombre d'espèces fondatrices sont responsables de la création d'habitats biogéniques (THOMSEN et al., 2010; THOMSEN et al., 2018). Ces habitats biogéniques, issus de la production biologique des organismes vivants, ouvrent dans l'environnement de nouvelles niches écologiques, en modifiant les ressources disponibles, influençant les flux de matière et d'énergie, et peuvent abriter une multitude d'autres organismes, induisant ainsi une augmentation de la biodiversité locale (BRUNO et al., 2003; SUNDAY et al., 2017; ROMERO et al., 2015).

En somme, les habitats biogéniques et les espèces qui les forment jouent un rôle clé dans la réponse présente et future des écosystèmes marins côtiers face aux changements globaux et impacts anthropiques (SUNDAY et al., 2017; HARLEY et al., 2006; BULLERI et al., 2018). Leur étude est donc critique d'un point de vue de la gestion et de la préservation des écosystèmes, mais elle permet aussi de répondre à des questions d'écologie fondamentale. Mieux comprendre le fonctionnement et la réponse de ces habitats peut ainsi fournir des éclairages quant aux processus qui promeuvent la biodiversité (ROMERO et al., 2015), aux mécanismes de rétroaction qui stabilisent les écosystèmes (THOMSEN et al., 2010), ainsi que sur les liens entre biodiversité et fonctionnement des écosystèmes (HASTINGS et al., 2007).

1.2.4 La Disparition des Habitats Biogéniques

Les habitats biogéniques jouent un rôle essentiel dans la génération et le maintien de la biodiversité et peuvent également servir de sentinelles face aux changements environnementaux (ROCA et al., 2016; NUGUES & ROBERTS, 2003; FREDERICQ et al., 2019). Leur sensibilité à une multitude de pressions, qu'elles soient d'origine anthropique ou naturelle, en fait des indicateurs des changements environnementaux. Par exemple, un déclin observé dans un herbier marin peut signaler une dégradation de la qualité des eaux (ROCA et al., 2016). De la même manière, les coraux, avec leur sensibilité au stress thermique, sont des témoins des effets du changement climatique (HOEGH-GULDBERG, 1999). Ainsi, la surveillance de l'état de santé de ces habitats biogéniques peut nous fournir des informations vitales sur les changements plus larges qui se produisent dans l'écosystème.

Les habitats biogéniques sont de plus en plus menacés par les activités humaines (WERNBERG et al., 2023). En quarante ans, la grande barrière de corail, le plus grand récif corallien au monde, a perdu près de la moitié de sa superficie (HUGHES et al., 2015)

et les herbiers marins ont perdu dans le même temps plus de 30% de leur superficie mondiale (WAYCOTT et al., 2009 ; DUNIC et al., 2021). Les causes de la diminution de superficie de ces habitats biogéniques sont multifactorielles et reposent sur l'ensemble des pressions directes et indirectes décrites précédemment. Dans les chapitres suivants, les termes *ecosystem engineer*, *biogenic habitat* et *foundation species* sont ainsi utilisés indistinctement.

Cependant, cette disparition n'est pas immédiate, les habitats peuvent passer par plusieurs stades de dégradation différents appelés dans la suite de cette thèse "état d'habitat". Par exemple, DONOVAN et al. (2018) identifie 5 états d'habitats de récifs coralliens avec des degrés de biomasse de poissons, de couverture en corail et de macroalgues différentes témoignant de différents degrés de dégradation (c.-à-d. un état non dégradé, trois états de transitions et un état dégradé).

Les conséquences de la dégradation et de la disparition des habitats biogéniques sont difficiles à anticiper du fait de leur importance au sein des écosystèmes. La perte de ces habitats entraîne une cascade d'événements dans l'écosystème. Sans ces habitats pour servir de zones de frayère, de nourriture et de protection, de nombreuses espèces se retrouvent vulnérables (p. ex. HUGHES et al. (2009)). Le déclin des habitats biogéniques est ainsi devenu l'un des principaux moteurs du déclin de la biodiversité (IPBES, 2019), bousculant la stabilité des écosystèmes tout entiers. Ainsi, l'étude des habitats biogéniques offre également des perspectives intéressantes pour le développement de programmes pour sauvegarder les écosystèmes marins (LOH et al., 2019). Il est donc important de mieux comprendre leur écologie et les impacts qu'ils ont sur leurs communautés associées, ainsi que de caractériser leurs états et leurs distributions à l'échelle globale. Caractériser leurs états et leurs distributions présents et futurs nécessite entre autres de développer des outils pour prédire les changements d'état que subissent ou risque de subir ces habitats face aux changements environnementaux (WERNBERG et al., 2023). Cela nécessite notamment de comprendre les dynamiques complexes que peuvent montrer ces écosystèmes face à des pressions multiples et variées (NYSTRÖM et al., 2012).

1.2.5 Dynamique des Écosystèmes : Résistance, Résilience et Changement de Régime

Les communautés écologiques sont des systèmes dynamiques, continuellement influencés par une multitude de perturbations. Ces dernières peuvent être qualifiées de "pulse" quand

ponctuelles sur de courtes durées, ou de “press” quand leurs impacts se font ressentir sur une période plus longue (BENDER et al., 1984; RYO et al., 2019). Leur amplitude peut varier (JENTSCH & WHITE, 2019), entraînant des impacts soit progressifs, soit soudains sur les écosystèmes et leurs communautés (Fig. 1.5; SPAKE et al. (2022)).

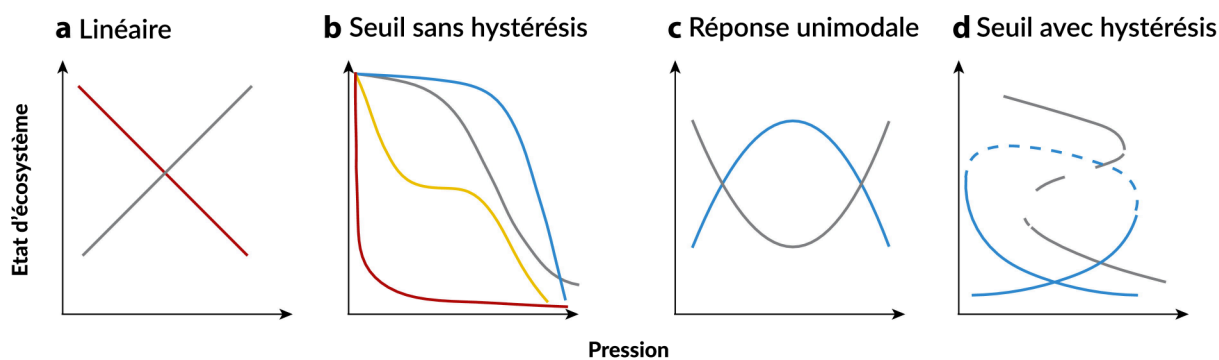


Figure 1.5 – Différentes réponses qu’un écosystème peut présenter face à des pressions. Chaque courbe représente l’ensemble des points d’équilibre potentiel d’un écosystème. Si ces graphes permettent de développer des visions théoriques des dynamiques écologiques, ils sous-entendent une simplification de la nature intrinsèquement changeante des écosystèmes. a. Réponse linéaire b. Réponse avec seuil sans hystérésis c. Réponse unimodale d. Réponse avec seuil et hystérésis. Dans ce cas, les courbes en pointillés reflètent des points d’équilibre instable. Extrait de SPAKE et al. (2022)

Face à ces perturbations, les communautés manifestent des propriétés de résistance et de résilience (Fig. 1.6). La résistance caractérise la capacité d’un système à maintenir sa structure et son fonctionnement en présence de perturbations. Par exemple, dans les écosystèmes marins, les mangroves démontrent une forte résistance aux tempêtes tropicales grâce à leur système racinaire complexe (KAZEMI et al., 2018). La résilience, en revanche, fait référence à la capacité d’un système à retourner à son état antérieur (supposément à l’équilibre) suite à une perturbation. Il convient de noter que différents écosystèmes, en fonction de la composition de leurs communautés et de leurs propriétés intrinsèques, manifestent des réponses variées aux perturbations.

Dans les cas où les pressions environnementales dépassent un certain seuil, les écosystèmes peuvent subir des changements de régime écologiques (SCHEFFER et al., 2001), passant à un nouvel état écologique persistant, souvent considéré comme moins favorable (NYSTRÖM et al., 2012; ROCHA et al., 2015). Ces transitions peuvent être linéaires ou non-linéaires (SCHEFFER et al., 2015), et peuvent présenter des seuils au-delà desquels la restauration à l’état initial est ardue, voire impossible (c.-à-d. phénomène d’hysté-

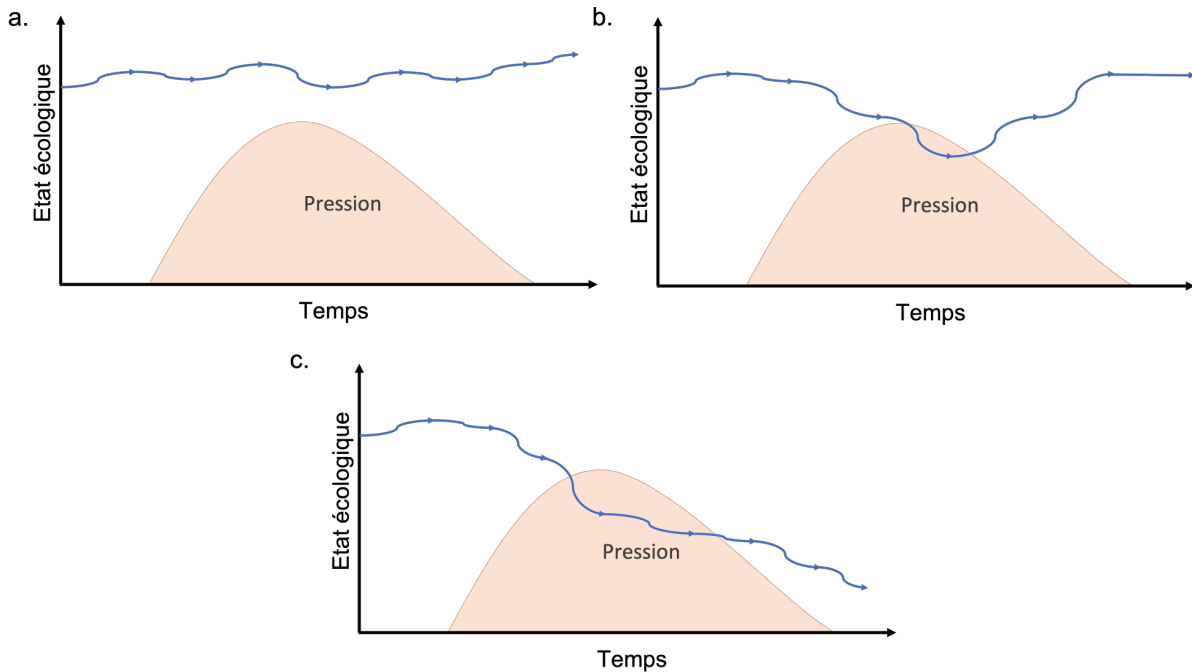


Figure 1.6 – Schéma décrivant la réponse théorique d'un écosystème à une pression. a. L'écosystème présente une forte résistance à la pression subite. b. L'écosystème présente une résistance moyenne à la pression, mais une forte résilience. c. L'écosystème présente à la fois une faible résistance et résilience

sis, Fig. 1.5 d; MARZLOFF et al. (2016); JOHNSON et al. (2017); ROCHA et al. (2018)). Ces changements de régime ont été observés dans de nombreux écosystèmes : de la forêt à la savane (STAYER et al., 2011), de la toundra aux forêts boréales (SCHEFFER et al., 2012), des récifs coralliens aux récifs dominés par les macroalgues (MCMANUS & POLSENBERG, 2004), ou bien encore l'effondrement des pêcheries (GÅRDMARK et al., 2015).

Comprendre les mécanismes sous-jacents à ces changements de régime est essentiel pour élaborer des stratégies de gestion et de conservation efficaces. En raison de la complexité et de l'interconnectivité des facteurs impliqués, la prédiction et la prévention de ces changements restent des défis considérables pour la recherche écologique contemporaine.

1.3 De l'Observation *In Situ* à la Modélisation

Une première approche pour étudier les mécanismes sous-jacents aux changements de biodiversité réside dans l'expérimentation *in situ* ou en mésocosmes. Ces approches sont

particulièrement utilisées en écologie pour décrire des systèmes simplifiés à des échelles spatio-temporelles réduites et composées de quelques espèces (WITMAN et al., 2015). Ainsi, ces approches expérimentales ignorent généralement les processus à plus larges échelles et les processus régionaux qui régissent les patrons de biodiversités et biaisent notre compréhension des phénomènes écologiques menant aux communautés observées (LEIBOLD et al., 2017). Augmenter les échelles des expérimentations serait donc une solution à ce problème, mais effectuer des manipulations expérimentales pour contrôler la biodiversité à l'échelle de systèmes plus complexes (p. ex. à l'échelle d'un écosystème ou d'un biome) est tout simplement irréalisable (PETERSEN et al., 2009).

L'utilisation de la modélisation apparaît comme une alternative judicieuse pour surpasser ces problèmes d'échelles. Au cours du temps, de nombreuses approches de modélisations ont été développées et sont utilisées en écologie (GEARY et al., 2020). Parmi l'ensemble des approches, certaines nécessitent une description complète des mécanismes écologiques en jeu et permettent par exemple d'étudier le comportement d'un écosystème sous différents scénarios, d'autres à partir de données parcellaires permettent d'inférer par exemple les niches environnementales réalisées de diverses espèces et enfin, les modèles conceptuels permettent de formaliser ou de développer des cadres théoriques du fonctionnement des écosystèmes avec très peu de données (voir Fig. 1 dans GEARY et al. (2020) pour les différents cas d'usage des différents types de modèles et leurs besoins respectifs en données). De par leur diversité, les modèles permettent d'inférer des processus aux échelles pertinentes pour comprendre les règles qui régissent les patrons de diversités observés dans les écosystèmes (LEIBOLD et al., 2017). Ainsi, dans le contexte global de l'écologie des communautés, la modélisation est un outil de choix pour mieux comprendre le fonctionnement des écosystèmes dans leur ensemble (EDGAR et al., 2016).

1.3.1 Impacts de la Révolution Digitale sur la Recherche en Écologique

Les évolutions rapides en écologie et dans d'autres domaines de la recherche, tant d'un point de vue théorique qu'appliquées, ont été rendues possibles par les développements grandissant de l'électronique et de l'informatique. Un des premiers changements technologiques majeurs dans les années 1970 concerne la sortie grand public du microprocesseur comme l'Intel 4004 (Fig. 1.7). Les microprocesseurs ont révolutionné le monde de l'informatique puisque pour la première fois l'ensemble des composants électroniques d'un processeur furent miniaturisés pour tenir dans un seul boîtier, engendrant un gain de

place considérable, augmentant la vitesse de calcul et une diminution de la consommation énergétique. Ce changement technologique a permis la démocratisation de l'informatique et l'apparition de l'ordinateur personnel, rendant l'informatique plus accessible au monde universitaire, et notamment aux écologues.

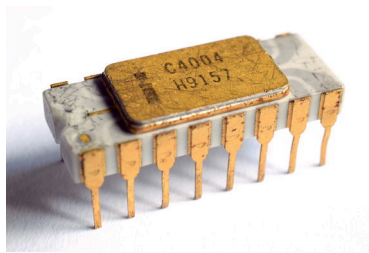


Figure 1.7 – *Processeur Intel 4004 commercialisé en 1971. © Photographie : Thomas NGUYEN (2016)*

Dans le même temps, les développements théoriques (p. ex. théorie des niches (HUTCHINSON, 1957), théorie de la biogéographie insulaire (MACARTHUR & WILSON, 1967), théorie des communautés écologiques (VELLEND, 2017), théorie des métacommunautés (LEIBOLD & CHASE, 2018)) ont fait passer progressivement l'écologie des communautés d'une science descriptive à une science hypothético-déductive où différentes hypothèses ont pu être testées à l'aide de modèles statistiques, univariées et/ou multivariées (LEGENDRE & LEGENDRE, 2012). Grâce à la puissance de calcul apportée par l'essor de l'informatique, de nouvelles méthodes d'analyses de données ont été mises au point et c'est ainsi qu'en 1979 un ouvrage en deux tomes regroupant l'ensemble des méthodologies statistiques multidimensionnelles adaptées aux données écologiques a vu le jour (LEGENDRE & LEGENDRE, 1979b ; LEGENDRE & LEGENDRE, 1979a).

L'essor de l'électronique et de l'informatique a également permis d'apporter de nouveaux outils aux écologues pour l'observation des écosystèmes de la planète. Ainsi, depuis les années 1960 et le début du programme Landsat (WULDER et al., 2019), le nombre de programmes spatiaux pour l'observation de la Terre à l'aide de satellites, n'a cessé d'augmenter et il est aujourd'hui possible de dénombrer plusieurs centaines de satellites d'observation actuellement en orbite (TATEM et al., 2008 ; ANDERSON, 2018). L'essor de l'internet à la fin des années 1990 a également marqué un nouveau tournant pour les sciences en général. Ce réseau informatique a permis de connecter les chercheurs du monde entier, facilitant la collaboration et la diffusion des connaissances dans le monde entier. Un bon exemple de projet collaboratif à grande échelle concerne les plateformes *Global Biodiver-*

sity Information Facility (GBIF) ou Ocean Biodiversity Information System (OBIS), qui ont pour but de collecter des données sur l'occurrence et la distribution des espèces pour contrôler l'état de la biodiversité à large échelle. Cet essor du numérique est également corrélé avec celui des programmes de suivis de la biodiversité à long terme (BUCKLEY et al., 2021) et des programmes de suivis de sciences participatives tels que *eBird* (WOOD et al., 2011), *iNaturalist* (SHEM UNGER & DUMAIS, 2021) ou bien encore le programme *Reef Life Survey* (EDGAR & STUART-SMITH, 2014).

L'écologie rentre donc dans le monde du big data, présentant de nouvelles perspectives et problématiques aux chercheurs (FARLEY et al., 2018). Face à cette prolifération de données, l'écologie des communautés doit désormais adopter des outils numériques sophistiqués issus des sciences des données pour identifier des motifs sous-jacents ou pour tester de nouvelles hypothèses. Ainsi, il est d'une importance croissante de former les nouvelles générations d'écologues aux outils numériques et aux méthodologies de la science des données. Grâce à ces connaissances, ils pourraient faire émerger de nouvelles techniques et de nouveaux paradigmes pour mieux comprendre et décrire la biodiversité dans le contexte du changement global.

1.3.2 Des Méthodes d'Ordination aux Modèles de Distribution d'Espèces : Une Exploration des Approches Quantitatives en Écologie

Comme établi précédemment, les communautés se caractérisent par leur dynamisme intrinsèque, rendant leur étude complexe. Face à cette complexité, une panoplie d'outils numériques et mathématiques a vu le jour. Les sections suivantes fourniront une présentation de certains de ces outils, en particulier ceux utilisés dans cette thèse, à savoir : (1) les méthodes d'ordination, essentielles pour déchiffrer les patrons écologiques ; (2) les méthodes de groupements, conçues pour détecter des discontinuités au sein de données écologiques ; enfin (3) les modèles de distribution d'espèces, qui se basent sur la théorie des niches pour expliquer et prédire la distribution spatiale des espèces.

Ordinations

Les ordinations classiquement utilisées en écologie numérique telles l'*Analyse en Composante Principale (PCA)*, l'*Analyse de Correspondance (CA)* ou le *Positionnement multidimensionnel (MDS)* sont des techniques d'analyse multidimensionnelles qui reposent

sur l'idée de trouver une façon de factoriser la matrice de composition d'espèces d'une certaine façon en fonction de l'espace des dimensions souhaitées (Euclidien pour la *PCA*, Chi-carré pour la *CA*, et divers selon la fonction de distance appliquée pour la *MDS*) pour résumer les informations contenues dans la matrice de communauté (LEGENDRE & LEGENDRE, 2012; UDELL et al., 2015).

De nombreuses avancées ont été faites dans le champ de recherche active qu'est celui de la réduction de la dimensionnalité des données. De nouvelles méthodes basées sur les graphes de voisinage (Fig. 1.8) ont vu le jour ces dernières années comme *Isomap* (TENENBAUM et al., 2000), *t-SNE* (van der MAATEN & HINTON, 2008) ou bien encore *UMAP* (MCINNES et al., 2020). Le principe de l'ensemble de ces méthodes est de construire des graphes dans l'espace multidimensionnel originel des données en connectant les points "proches" les uns aux autres par des arêtes. L'idée générale est de déterminer quels sont les points (et donc les communautés) qui sont le plus semblables (proches dans l'espace multidimensionnel) ou dissemblables (éloignés dans l'espace multidimensionnel). Les méthodes basées sur des graphes de voisinage projettent les points dans un espace en plus faible dimension, tout en cherchant à préserver les relations de voisinage entre les points décrits par le graphe dans l'espace multidimensionnel. Pendant des années, ces méthodes numériquement plus gourmandes en ressources et/ou sans package facilement utilisable étaient rarement utilisées par les écologues. Cependant, ces méthodes commencent à faire leur apparition dans la boîte à outils des écologues des communautés (ROBERTS, 2020; MILOŠEVIĆ et al., 2022).

Méthodes de Groupement en Ecologie

D'une manière générale, l'être humain cherche à catégoriser les objets qui le côtoient afin de mieux les décrire. Avant de les nommer, il nous est nécessaire de les regrouper en catégories distinctes, qui ne sont rien d'autre que des abstractions utiles à notre appréhension du monde. Ainsi, il faut pouvoir développer des définitions, des heuristiques ou bien encore des méthodes pour identifier des groupes distincts dans un milieu qui, bien que parfois discret, est généralement continu (LEGENDRE & LEGENDRE, 2012).

L'usage des méthodes de groupement reste très fréquent en écologie des communautés. Ce type de méthodes permet entre autres de définir des groupes d'espèces ayant les mêmes traits (MOUILLOT et al., 2021), ou partageant les mêmes patrons spatiaux de distribution (PANG et al., 2023), d'identifier des zones similaires par rapport à leur composition

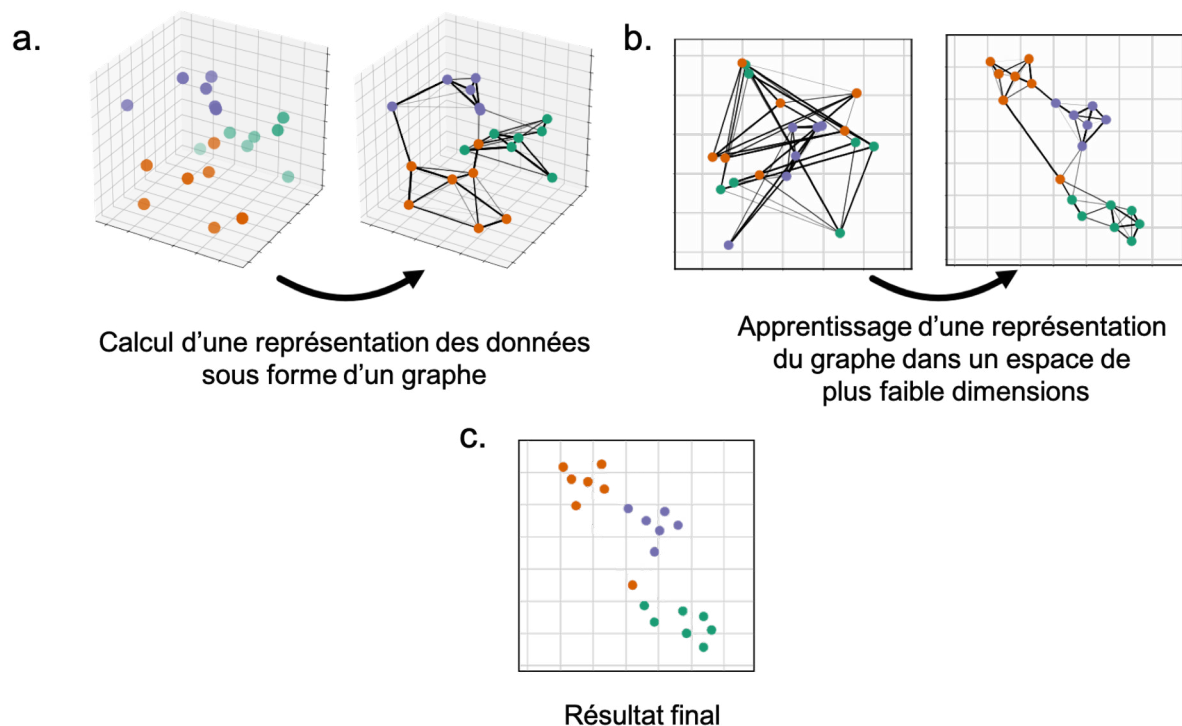


Figure 1.8 – Schéma du principe de fonctionnement d'une méthode de réduction de dimension basée sur les graphes de voisinages. a. Calcul d'un graphe de voisinage dans l'espace multidimensionnel original des données. b. Apprentissage d'une représentation de ce graphe dans un espace de plus faible dimension. c. Projection des observations dans l'espace en plus faible dimension. Adapté de MCINNES et al. (2018)

spécifique (LEGENDRE & LEGENDRE, 2012), d'identifier des écorégions dans l'océan global (SONNEWALD et al., 2020) ou bien encore d'identifier des patrons spatiaux dans des réseaux écologiques (OHLSSON & EKLÖF, 2020).

Parmi les méthodes fréquemment utilisées en écologie, il est possible de citer entre autres les méthodes hiérarchiques, comme celle de Ward ou du groupement à liens simples ou complets, et non hiérarchiques comme le *k-means* (LEGENDRE & LEGENDRE, 2012). L'avantage de ces méthodes est leur facilité de compréhension. Cependant, le groupement hiérarchique à liens simples est sensible au bruit contenu dans les données (LEGENDRE & LEGENDRE, 2012) alors que les autres méthodes ont pour contraintes de n'identifier que des groupes sphériques.

Pour contourner ces deux problèmes, d'autres méthodes ont été développées en sciences des données. Ces algorithmes de groupement issu d'une famille appelée "*Density-Based*

Clustering” imposent une autre contrainte aux groupes à identifier : ces derniers doivent contenir un certain nombre d’observations avant de pouvoir être déclarés comme étant un cluster à part entière (c’est-à-dire qu’ils doivent être denses) tout en permettant à des observations trop éloignées de ces groupes de n’être assignées à aucun d’entre eux (observations considérées alors comme bruit ; KRIEGEL et al. (2011)). Ainsi, la contrainte de la forme des groupes est relaxée tout en permettant de définir des groupes dépourvus d’observations bruitées. Bien que certaines de ces méthodes aient été développées dès la fin du 20ème siècle (ESTER et al., 1996 ; ANKERST et al., 1999), elles n’ont été utilisées que très récemment en écologie dues à leur complexité en coût de calcul et leur difficulté à paramétrer (voir OHLSSON et EKLÖF (2020) et SONNEWALD et al. (2020)).

Modèles de Distribution d’Espèces

Les modèles de distribution d’espèces sont des outils largement utilisés en écologie pour comprendre, prédire et cartographier la distribution spatiale d’une espèce en fonction des conditions environnementales (GUISAN & ZIMMERMANN, 2000). Enracinés dans la théorie des niches, ces modèles reposent sur l’identification des relations entre les occurrences connues (ou l’abondance) d’une espèce et un ensemble de variables environnementales à ces localités, qui décrivent alors la niche réalisée de l’espèce (GUISAN et al., 2017).

Il est possible de classer les modèles de distribution d’espèces en deux catégories : les modèles dits “mécanistiques” et les modèles dits “corrélatifs”. Les modèles “mécanistiques” reposent sur une représentation explicite des mécanismes associés aux variations de biodiversité (p. ex. augmentation du métabolisme de l’espèce en condition de stress) (KEARNEY et al., 2010), alors que les modèles “corrélatifs” s’appuient sur les corrélations observées entre l’environnement et les occurrences/abondance des espèces étudiées sans mettre en évidence les mécanismes sous-jacents (BRISCOE et al., 2023).

Les modèles mécanistiques modélisent la distribution des espèces ou des communautés (HARFOOT et al., 2014), à l’aide d’équations différentielles qui intègrent explicitement les processus physiologiques de réponses aux variations environnementales (KEARNEY & PORTER, 2009). Cependant, l’une des critiques principales de la modélisation mécanistique réside dans sa difficulté à être appliquée pour les espèces, communautés ou écosystèmes peu étudiés, du fait de la nécessité de traduire des hypothèses écologiques en équations (EVANS et al., 2015) et de l’acquisition de données physiologiques sur les espèces étudiées (GANDRA et al., 2015).

Les modèles corrélatifs quant à eux sont bien plus utilisés en écologie des communautés pour modéliser la distribution des espèces (*Species Distribution Models* ou *SDM*, en anglais). Basés sur des modèles statistiques tels que les régressions linéaires, ces modèles trouvent des relations corrélatives entre les données environnementales et les données d'occurrence (ou d'abondance) d'une espèce (GUISAN et al., 2017). Toutefois, les relations observées entre les occurrences et les données environnementales sont souvent non-linéaires et peuvent être mieux modélisées grâce à des modèles additifs généralisés (GUISAN et al., 2002), ou à l'aide de modèles de *Machine Learning* (ELITH et al., 2006). Cette variété d'approches de modélisation s'est enrichie avec l'apport de nouvelles données qui ont permis de mieux caractériser la niche écologique des espèces en prenant en compte leurs traits (POLLOCK et al., 2012) ou bien les relations phylogénétiques entre différentes espèces (IVES & HELMUS, 2011). L'ensemble de ces techniques de modélisation ont permis de faire des progrès majeurs en écologie notamment en permettant de mieux comprendre la niche écologique réalisée des espèces, en créant des cartes de distribution d'espèces menacées pour aider à la mise en place de plans de gestion (FRANKLIN, 2013), en prédisant les zones d'expansion d'espèces envahissantes (MAINALI et al., 2015), ou en étudiant les changements de distribution d'espèces et des patrons de diversités associés selon différents scénarios de changement climatique (ARAÚJO et al., 2019).

Néanmoins, cette approche ne permet pas de modéliser l'ensemble d'une communauté puisque, en considérant les espèces comme indépendantes les unes des autres, elle ne prend pas en compte la nature multivariée des données en écologie (WARTON et al., 2015; OVASKAINEN et al., 2017). Pour pallier à cet écueil, FERRIER et GUISAN (2006) proposèrent plusieurs approches : (1) “*grouper, puis prédire*”, (2) “*prédire puis grouper*” et enfin (3) “*prédire et grouper simultanément*” (Fig. 1.9).

Pour résumer les différentes approches proposées par FERRIER et GUISAN (2006), la première consiste à utiliser des techniques de classification ou d'ordination (voir Sec. 1.3.2) pour d'abord caractériser différents types de communautés d'espèces avant d'ensuite prédire leurs distributions et leurs déterminants. La seconde approche consiste à modéliser séparément chaque espèce de la communauté, puis les prédictions de chaque modèle sont agrégées par des méthodes de classification ou d'ordination. Une variante de cette approche consiste à modéliser chaque espèce indépendamment des autres puis à compiler les prédictions de ces modèles pour prédire les assemblages d'espèces présents, sans considérer les potentielles interactions interspécifiques (*stacked-SDM* en anglais; POLLOCK et al.

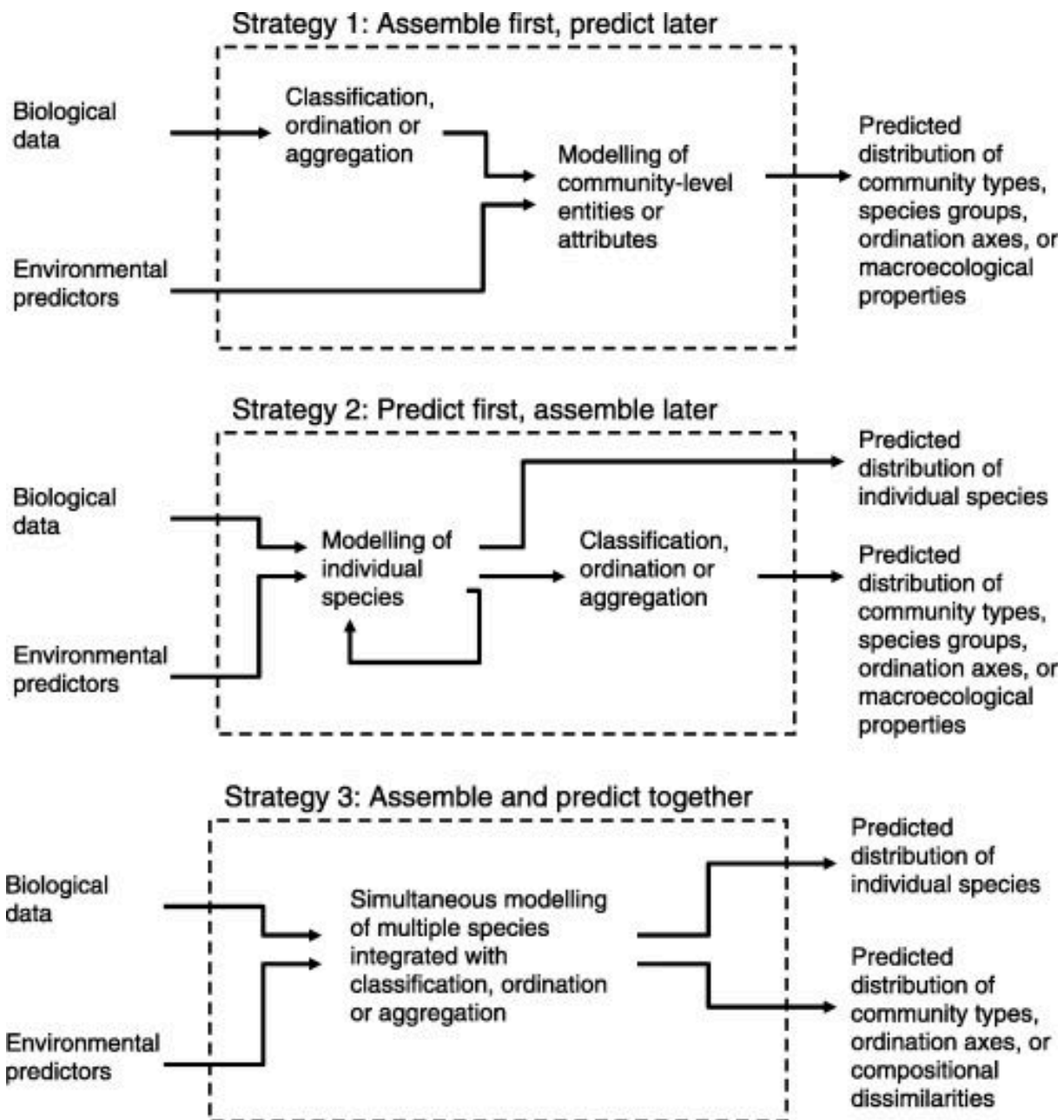


Figure 1.9 – Stratégies de modélisations de distribution d'espèces. FERRIER et GUIBAN (2006)

(2020)). Enfin, la dernière approche propose de prédire toutes les espèces et de les agréger en une communauté en une seule étape via des approches de modélisation multidimensionnelles, qui prennent en compte les patrons de co-occurrence entre espèces (WARTON et al., 2015). Chacune de ces approches présente des forces et des faiblesses différentes (D'AMEN et al., 2017; POLLOCK et al., 2020; NORBERG et al., 2019).

Ainsi l’approche “*prédire et grouper simultanément*” est selon FERRIER et GUIBAN (2006) la plus polyvalente, permettant de prédire la distribution de chaque espèce individuellement tout en capturant de manière adéquate les propriétés macroécologiques des communautés. Plus difficile à mettre en place, cette approche a récemment connu un regain d’intérêt grâce au développement de nouvelles approches pour estimer la distribution jointe de toutes les espèces d’une communauté à l’aide d’outils statistiques accessibles aux écologues (WARTON et al., 2015 ; HUI, 2016 ; OVASKAINEN et al., 2017 ; NIKU et al., 2019 ; CHIQUET et al., 2021 ; PICHLER et al., 2020). Ces *Joint Species Distribution Models (jSDM)* se proposent de modéliser, à l’aide d’une distribution multidimensionnelle (c.-à-d. normale multivariée), toutes les espèces d’une communauté en fonction de différents prédicteurs, tout en prenant en compte l’effet de prédicteurs potentiels (c.-à-d. non inclus dans le modèle) grâce à un système de variables latentes (WARTON et al., 2015).

Si cette approche “*prédire et grouper simultanément*” paraît prometteuse, son implémentation n’est pas sans défi (POGGIATO et al., 2021 ; D’AMEN et al., 2017). D’autres approches méthodologiques ont aussi des atouts qu’il convient d’examiner, notamment lorsqu’il s’agit de faire le lien avec la gestion des écosystèmes (D’AMEN et al., 2017 ; POLLOCK et al., 2020). L’approche “*grouper, puis prédire*” a par exemple été utilisée pour explorer les déterminants à large échelle des patrons biogéographiques en milieu marin (BELANGER et al., 2012). Dans ce contexte, cette thèse s’appuiera sur une diversité de méthodes complémentaires pour répondre aux objectifs suivants.

1.4 Contexte & Objectifs de cette Thèse

Les habitats biogéniques marins, dont les forêts d’algues, les herbiers marins et les récifs coralliens, représentent des piliers fondamentaux pour la sauvegarde de la biodiversité marine et le bon fonctionnement des écosystèmes côtiers (IPBES, 2019). Malheureusement, leur intégrité est mise à mal par une myriade de pressions, qu’elles soient d’origine anthropique ou naturelle (IPBES, 2019). Afin de garantir la préservation de la biodiversité marine et la pérennité des fonctions écologiques essentielles des zones côtières, une compréhension approfondie de ces habitats est impérative. Cette compréhension passe par la caractérisation des états d’habitats et des facteurs qui régissent leur distribution, mais aussi par une meilleure description de leur dynamique et des interrelations complexes que ces habitats biogéniques entretiennent avec la biodiversité qu’ils abritent (Fig. 1.10).

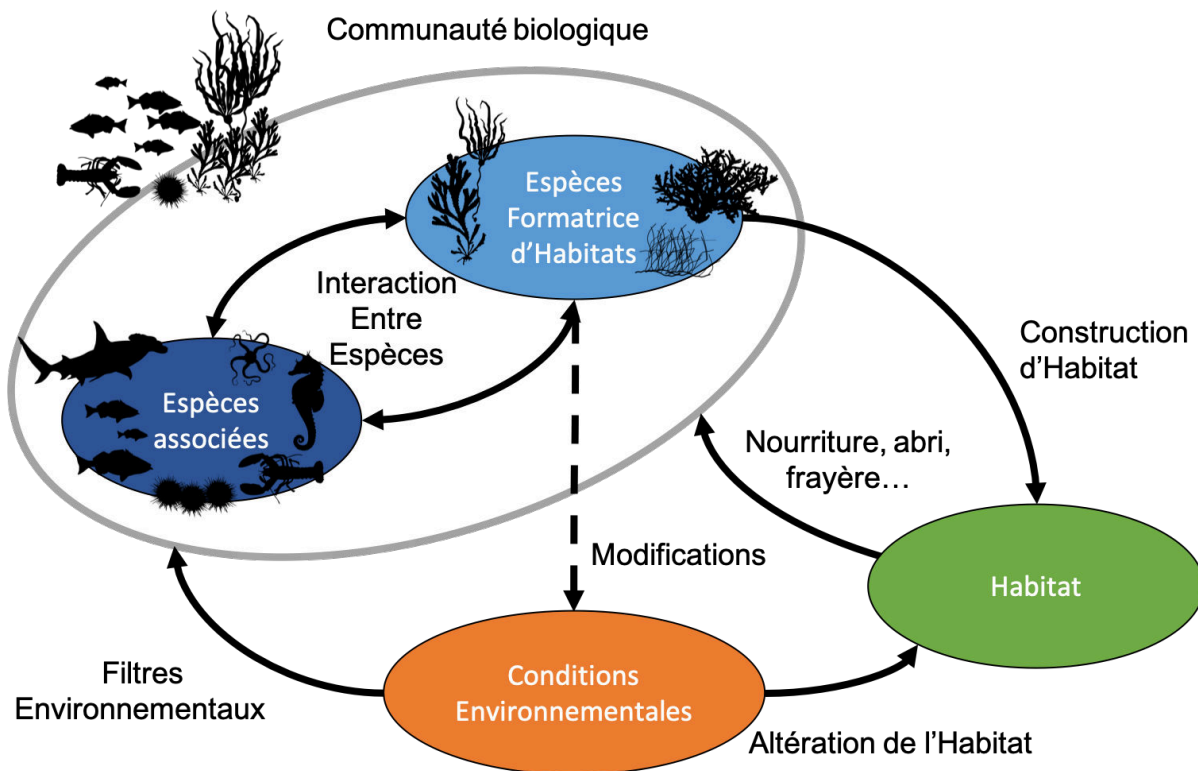


Figure 1.10 – Schéma illustrant les interactions majeures entre les habitats benthiques, la faune qui y est associée et les conditions environnementales. Cadre de pensée dans lequel s'inscrit cette thèse. Adapté d'après MARZLOFF (2022)

L'ambition première de cette thèse est d'améliorer notre compréhension globale du fonctionnement des habitats biogéniques côtiers. Elle vise à caractériser le rôle des habitats biogéniques et de leurs différents états sur la biodiversité environnante, mais aussi à éclairer notre vision des transitions potentielles d'habitats au sein des environnements benthiques côtiers (Fig. 1.10). Pour mener à bien ces objectifs, ce travail de thèse a été divisé en trois chapitres qui utilisent deux approches distinctes (Fig. 1.11).

Dans le premier chapitre, j'ai testé l'approche "prédire et grouper simultanément", via l'usage de *jSDM*, pour comprendre les règles d'assemblage d'espèces et les patrons de biodiversité associés à deux habitats caractéristiques des écosystèmes benthiques côtiers de Bretagne. En particulier, l'objectif était de comprendre et prédire les réponses des communautés de macroinvertébrés associées à ces habitats le long de gradients environnementaux. Pour ce faire, j'ai développé une méthodologie d'évaluation des performances des différentes prédictions réalisables par un *jSDM* que j'ai appliquée à un jeu de données

issu de séries temporelles de huit ans couvrant l'ensemble du littoral breton (rebe.fr). Grâce à cela, j'ai pu identifier les avantages et les limites de différentes structures de *jSDM* qui intègrent différentes sources d'information (phylogénie, trait, information sur la communauté accompagnatrice) quant à leur capacité à expliquer et prédire des patrons de biodiversité locaux et leur déploiement à large échelle.

Dans les chapitres deux et trois, j'ai testé une approche se basant sur les états d'habitats biogéniques comme proxy pour étudier la biodiversité, dans une démarche "*grouper puis, prédire*". Dans le second chapitre, j'ai ainsi cherché à créer une typologie d'habitats benthiques biogéniques en me basant sur des données de couverture d'habitats, collectées grâce à un programme de sciences participatives déployé à l'échelle du globe (reeflifesurvey.com). J'ai développé une nouvelle méthodologie de *Machine Learning* pour classifier ces données d'observation en alliant une méthode de réduction de dimensions basée sur les graphes de voisinage et un algorithme de groupement basé sur la densité. A l'aide d'outils d'interprétation de modèles de *Machine Learning*, j'ai pu comprendre quelle était la composition de ces types d'habitats, étudier leur distribution spatiale et temporelle et leur diversité à l'échelle globale.

Dans le troisième chapitre, j'ai qualifié les niches environnementales des états d'habitats biogéniques découverts dans le chapitre précédent par une approche de modèles de distribution d'espèces. J'ai identifié l'influence relative de différentes pressions anthropiques et variables environnementales sur la distribution de ces états d'habitats. Les résultats de ces travaux ont montré que plusieurs états d'habitats ont des niches environnementales similaires. Ces travaux ont permis ainsi de mettre en évidence des zones et des conditions environnementales où l'apparition de différents états d'habitats est plus probable.

Objectifs

Prédire et grouper simultanément

Grouper, puis prédire

Chapitre 1
Comprendre et prédire la diversité benthique de deux habitats côtiers

Décrire	Comprendre	Prédire
	+	+

Chapitre 2
Identification d'états d'habitats côtiers à l'échelle du globe

Décrire	Comprendre	Prédire
+		

Chapitre 3
Prédictibilité et déterminants des états d'habitats à une échelle nationale

Décrire	Comprendre	Prédire
	+	+

Données

Emprise Spatiotemporelle



Bretagne | Rebert (2008-2014)

Habitats




Herbiers et sédiments intertidaux

Objet d'étude




Communautés de Macroinvertébrés benthiques

Emprise Spatiotemporelle

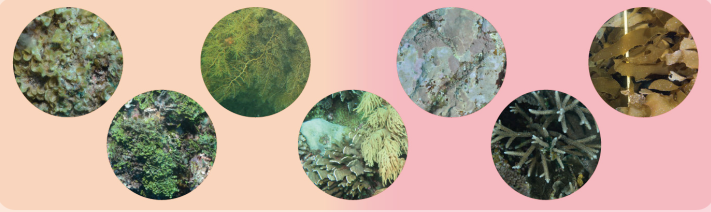


Monde entier ← Reef Life Survey (2008-2021) →




Australie

Habitats récifaux




Objet d'étude

Pourcentage de Couverture d'Habitats



Objet d'étude

Etats d'Habitats



Méthodes utilisées

Joint Species Distribution Models
HMSC

Groupement
UMAP + HDBSCAN

Species Distribution Models
Random Forest

Aspect étudié

- Niches environnementales
- Diversité α et β
- Prédicibilité des espèces

- Etats d'habitats

- Niches environnementales
- Distribution spatiale
- Prédicibilité des états d'habitats

Figure 1.11 – Schéma illustrant les travaux réalisés dans le cadre de ce travail de thèse.

Bibliographie

- ANDERSON, C. B., (2018), Biodiversity monitoring, earth observations and the ecology of scale, *Ecology Letters*, 2110, 1572-1585.
- ANKERST, M., BREUNIG, M. M., KRIEGEL, H.-P., & SANDER, J., (1999), OPTICS : Ordering Points to Identify the Clustering Structure, *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, 49-60.
- ARAÚJO, M. B., ANDERSON, R. P., BARBOSA, A. M., BEALE, C. M., DORMANN, C. F., EARLY, R., GARCIA, R. A., GUISAN, A., MAIORANO, L., NAIMI, B., O'HARA, R. B., ZIMMERMANN, N. E., & RAHBEK, C., (2019), Standards for distribution models in biodiversity assessments, *Science Advances*, 51, eaat4858.
- BAGOUSSE-PINGUET, Y. L., SOLIVERES, S., GROSS, N., TORICES, R., BERDUGO, M., & MAESTRE, F. T., (2019), Phylogenetic, functional, and taxonomic richness have both positive and negative effects on ecosystem multifunctionality, *Proceedings of the National Academy of Sciences*, 11617, 8419-8424.
- BARBIER, E. B., HACKER, S. D., KENNEDY, C., KOCH, E. W., STIER, A. C., & SILLIMAN, B. R., (2011), The value of estuarine and coastal ecosystem services, *Ecological Monographs*, 812, 169-193.
- BELANGER, C. L., JABLONSKI, D., ROY, K., BERKE, S. K., KRUG, A. Z., & VALENTINE, J. W., (2012), Global environmental predictors of benthic marine biogeographic structure, *Proceedings of the National Academy of Sciences*, 10935, 14046-14051.
- BENDER, E. A., CASE, T. J., & GILPIN, M. E., (1984), Perturbation Experiments in Community Ecology : Theory and Practice, *Ecology*, 651, 1-13.
- BENOIST, L., HOUYVET, B., HENRY, J., CORRE, E., ZANUTTINI, B., & ZATYLNÝ-GAUDIN, C., (2020), In-Depth In Silico Search for Cuttlefish (*Sepia officinalis*) Antimicrobial Peptides Following Bacterial Challenge of Haemocytes, *Marine Drugs*, 189, 439.
- BLACKBURN, S., PELLING, M., & MARQUES, C., (2019), Chapter 38 - Megacities and the Coast : Global Context and Scope for Transformation, In E. WOLANSKI, J. W. DAY, M. ELLIOTT & R. RAMACHANDRAN (Éd.), *Coasts and Estuaries* (p. 661-669), Elsevier.
- BOYD, S., LIMPENNY, D., REES, H., & COOPER, K., (2005), The effects of marine sand and gravel extraction on the macrobenthos at a commercial dredging site (results 6 years post-dredging), *ICES Journal of Marine Science*, 622, 145-162.
- BOYD, S., LIMPENNY, D., REES, H., COOPER, K., & CAMPBELL, S., (2003), Preliminary observations of the effects of dredging intensity on the re-colonisation of dredged sediments off the southeast coast of England (Area 222), *Estuarine, Coastal and Shelf Science*, 571, 209-223.
- BRAY, R. N., (2008), *Environmental aspects of dredging* (1^{re} éd.), CRC Press.
- BREITBURG, D., LEVIN, L. A., OSCHLIES, A., GRÉGOIRE, M., CHAVEZ, F. P., CONLEY, D. J., GARÇON, V., GILBERT, D., GUTIÉRREZ, D., ISENSEE, K., JACINTO, G. S., LIMBURG, K. E., MONTES, I., NAQVI, S. W. A., PITCHER, G. C., RABALAIS, N. N., ROMAN, M. R., ROSE, K. A., SEIBEL, B. A., ... ZHANG, J., (2018), Declining oxygen in the global ocean and coastal waters, *Science*, 3596371, eaam7240.

- BRISCOE, N. J., MORRIS, S. D., MATHEWSON, P. D., BUCKLEY, L. B., JUSUP, M., LEVY, O., MACLEAN, I. M. D., PINCEBOURDE, S., RIDDELL, E. A., ROBERTS, J. A., SCHOUTEN, R., SEARS, M. W., & KEARNEY, M. R., (2023), Mechanistic forecasts of species responses to climate change : The promise of biophysical ecology, *Global Change Biology*, 296, 1451-1470.
- BRUNO, J. F., STACHOWICZ, J. J., & BERTNESS, M. D., (2003), Inclusion of facilitation into ecological theory, *Trends in Ecology & Evolution*, 183, 119-125.
- BUCKLEY, H. L., DAY, N. J., LEAR, G., & CASE, B. S., (2021), Changes in the analysis of temporal community dynamics data : a 29-year literature review, *PeerJ*, 9, e11250.
- BULLERI, F., ERIKSSON, B. K., QUEIRÓS, A., AIROLDI, L., ARENAS, F., ARVANITIDIS, C., BOUMA, T. J., CROWE, T. P., DAVOULT, D., GUIZIEN, K., IVEŠA, L., JENKINS, S. R., MICHALET, R., OLABARRIA, C., PROCACCINI, G., SERRÃO, E. A., WAHL, M., & BENEDETTI-CECCHI, L., (2018), Harnessing positive species interactions as a tool against climate-driven loss of coastal biodiversity, *PLOS Biology*, 169, 1-19.
- BURKE, L., KURA, Y., KASSEM, K., REVENGA, C., SPALDING, M., & McALLISTER, D., (2000), Pilot analysis of global ecosystems : coastal ecosystems, *World Resources Institute*.
- CALVIN, K., DASGUPTA, D., KRINNER, G., MUKHERJI, A., THORNE, P. W., TRISOS, C., ROMERO, J., ALDUNCE, P., BARRETT, K., BLANCO, G., CHEUNG, W. W., CONNORS, S., DENTON, F., DIONGUE-NIANG, A., DODMAN, D., GARSCHAGEN, M., GEDEN, O., HAYWARD, B., JONES, C., ... HA, M., (2023, juillet), *IPCC, 2023 : Climate Change 2023 : Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* (rapp. tech.).
- CANADELL, J., MONTEIRO, P., COSTA, M., COTRIM DA CUNHA, L., COX, P., ELISEEV, A., HENSON, S., ISHII, M., JACCARD, S., KOVEN, C., LOHILA, A., PATRA, P., PIAO, S., ROGELJ, J., SYAMPUNGANI, S., ZAEHLE, S., & ZICKFELD, K., (2021), Global Carbon and other Biogeochemical Cycles and Feedbacks, In V. MASSON-DELMOTTE, P. ZHAI, A. PIRANI, S. CONNORS, C. PÉAN, S. BERGER, N. CAUD, Y. CHEN, L. GOLDFARB, M. GOMIS, M. HUANG, K. LEITZELL, E. LONNOY, J. MATTHEWS, T. MAYCOCK, T. WATERFIELD, O. YELEKÇI, R. YU & B. ZHOU (Éd.), *Climate Change 2021 : The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 673-816), Cambridge University Press.
- CARMONA, C. P., TAMME, R., PÄRTEL, M., de BELLO, F., BROSE, S., CAPDEVILA, P., GONZÁLEZ-M., R., GONZÁLEZ-SUÁREZ, M., SALGUERO-GÓMEZ, R., VÁSQUEZ-VALDERRAMA, M., & TOUSSAINT, A., (2021), Erosion of global functional diversity across the tree of life, *Science Advances*, 713, eabf2675.
- CHAN, N. C. S., & CONNOLLY, S. R., (2013), Sensitivity of coral calcification to ocean acidification : a meta-analysis, *Global Change Biology*, 191, 282-290.
- CHAPIN, F. S., MATSON, P. A., & VITOUSEK, P. M., (2011), *The Ecosystem Concept*, Springer New York.
- CHASE, J. M., & LEIBOLD, M. A., (2003, juillet), *Ecological Niches : Linking Classical and Contemporary Approaches*, University of Chicago Press.

- CHEVALIER, M., ZARZO-ARIAS, A., GUÉLAT, J., MATEO, R. G., & GUISAN, A., (2022), Accounting for niche truncation to improve spatial and temporal predictions of species distributions, *Frontiers in Ecology and Evolution*, 10.
- CHIQUET, J., MARIADASSOU, M., & ROBIN, S., (2021), The Poisson-Lognormal Model as a Versatile Framework for the Joint Analysis of Species Abundances, *Frontiers in Ecology and Evolution*, 9.
- COCHRANE, K. L., BAHRI, T., DABBADIE, L., FERNANDEZ- REGUERA, D., KALIKOSKI, D. C., MA, X., & VANNUCCINI, S., (2023), The impact of climate change on coastal fisheries and aquaculture. In *Reference Module in Earth Systems and Environmental Sciences*, Elsevier.
- COSTELLO, M. J., & CHAUDHARY, C., (2017), Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation, *Current Biology*, 2711, R511-R527.
- D'AMEN, M., RAHBK, C., ZIMMERMANN, N. E., & GUISAN, A., (2017), Spatial predictions at the community level : from current approaches to future frameworks, *Biological Reviews*, 921, 169-187.
- DARWIN, C., (1839), *Journal of Researches Into the Geology and Natural History : Of the Various Countries Visited by HMS Beagle, Under the Command of Cpt. Fitzroy, from 1832 to 1836*, Colburn.
- DAYTON, P. K., (1972), Toward an understanding of community resilience and the potential effects of enrichments to the benthos at McMurdo Sound, Antarctica, *Proceedings of the colloquium on conservation problems in Antarctica*, 81-96.
- DOBRICIC, S., POZZOLI, L., SANSEVERINO, I., CONDUTO, D., & LETTIERI, T., (2016), *Algal bloom and its economic impact*, European Commission ; Joint Research Centre.
- DONOVAN, M. K., FRIEDLANDER, A. M., LECKY, J., JOUFFRAY, J.-B., WILLIAMS, G. J., WEDDING, L. M., CROWDER, L. B., ERICKSON, A. L., GRAHAM, N. A. J., GOVE, J. M., KAPPEL, C. V., KARR, K., KITTINGER, J. N., NORSTRÖM, A. V., NYSTRÖM, M., OLESON, K. L. L., STAMOULIS, K. A., WHITE, C., WILLIAMS, I. D., & SELKOE, K. A., (2018), Combining fish and benthic communities into multiple regimes reveals complex reef dynamics, *Scientific Reports*, 81, 16943.
- DORNELAS, M., CHASE, J. M., GOTELLI, N. J., MAGURRAN, A. E., MCGILL, B. J., ANTÃO, L. H., BLOWES, S. A., DASKALOVA, G. N., LEUNG, B., MARTINS, I. S., MOYES, F., MYERS-SMITH, I. H., THOMAS, C. D., & VELLEND, M., (2023), Looking back on biodiversity change : lessons for the road ahead, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3781881, 20220199.
- DUNCAN, S., (2023). <https://twitter.com/ScottDuncanWX/status/1670833932986732544>
- DUNIC, J. C., BROWN, C. J., CONNOLLY, R. M., TURSCHWELL, M. P., & CÔTÉ, I. M., (2021), Long-term declines and recovery of meadow area across the world's seagrass bioregions, *Global Change Biology*, 2717, 4096-4109.
- EDGAR, G. J., BATES, A. E., BIRD, T. J., JONES, A. H., KININMONTH, S., STUART-SMITH, R. D., & WEBB, T. J., (2016), New Approaches to Marine Conservation Through the Scaling Up of Ecological Data, *Annual Review of Marine Science*, 81, 435-461.
- EDGAR, G. J., & STUART-SMITH, R. D., (2014), Systematic global assessment of reef fish communities by the Reef Life Survey program, *Scientific Data*, 11, 140007.

- ELITH, J., H. GRAHAM, C., P. ANDERSON, R., DUDÍK, M., FERRIER, S., GUISAN, A., J. HIJMANS, R., HUETTMANN, F., R. LEATHWICK, J., LEHMANN, A., LI, J., G. LOHMANN, L., A. LOISELLE, B., MANION, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., MCC. M. OVERTON, J., TOWNSEND PETERSON, A., ... E. ZIMMERMANN, N., (2006), Novel methods improve prediction of species' distributions from occurrence data, *Ecography*, 292, 129-151.
- ELLISON, A. M., (2019), Foundation Species, Non-trophic Interactions, and the Value of Being Common, *iScience*, 13, 254-268.
- ELTON, C., (1927), *Animal ecology*, University of Chicago Press.
- ERFTEMEIJER, P. L., & ROBIN LEWIS, R. R., (2006), Environmental impacts of dredging on seagrasses : A review, *Marine Pollution Bulletin*, 5212, 1553-1572.
- ESTER, M., KRIEGEL, H.-P., SANDER, J., & XU, X., (1996), A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226-231.
- EVANS, T. G., DIAMOND, S. E., & KELLY, M. W., (2015), Mechanistic species distribution modelling as a link between physiology and conservation, *Conservation Physiology*, 31, cov056.
- FAO, (2022), Fishery and Aquaculture Statistics 2022/FAO annuaire. Statistiques des pêches et de l'aquaculture 2022/FAO anuario, *FAO yearbook*.
- FARLEY, S. S., DAWSON, A., GORING, S. J., & WILLIAMS, J. W., (2018), Situating Ecology as a Big-Data Science : Current Advances, Challenges, and Solutions, *BioScience*, 688, 563-576.
- FERRIER, S., & GUISAN, A., (2006), Spatial modelling of biodiversity at the community level, *Journal of Applied Ecology*, 433, 393-404.
- FILBEE-DEXTER, K., & WERNBERG, T., (2018), Rise of Turfs : A New Battlefield for Globally Declining Kelp Forests, *BioScience*, 682, 64-76.
- FILHO, G. M. A., CREED, J. C., ANDRADE, L. R., & PFEIFFER, W. C., (2004), Metal accumulation by *Halodule wrightii* populations, *Aquatic Botany*, 804, 241-251.
- FOSTER, K. R., SCHLUTER, J., COYTE, K. Z., & RAKOFF-NAHOUM, S., (2017), The evolution of the host microbiome as an ecosystem on a leash, *Nature*, 5487665, 43-51.
- FOURQUREAN, J. W., DUARTE, C. M., KENNEDY, H., MARBÀ, N., HOLMER, M., MATEO, M. A., APOSTOLAKI, E. T., KENDRICK, G. A., KRAUSE-JENSEN, D., MCGLATHERY, K. J., & SERRANO, O., (2012), Seagrass ecosystems as a globally significant carbon stock, *Nature Geoscience*, 57, 505-509.
- FOX-KEMPER, B., HEWITT, H., XIAO, C., AÐALGEIRSDÓTTIR, G., DRIJFHOUT, S., EDWARDS, T., GOLLEDGE, N., HEMER, M., KOPP, R., KRINNER, G., MIX, A., NOTZ, D., NOWICKI, S., NURHATI, I., RUIZ, L., SALLÉE, J.-B., SLANGEN, A., & YU, Y., (2021), Ocean, Cryosphere and Sea Level Change, In V. MASSON-DELMOTTE, P. ZHAI, A. PIRANI, S. CONNORS, C. PÉAN, S. BERGER, N. CAUD, Y. CHEN, L. GOLDFARB, M. GOMIS, M. HUANG, K. LEITZELL, E. LONNOY, J. MATTHEWS, T. MAYCOCK, T. WATERFIELD, O. YELEKÇI, R. YU & B. ZHOU (Éd.), *Climate Change 2021 : The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 1211-1362), Cambridge University Press.
- FRANKLIN, J., (2013), Species distribution models in conservation biogeography : developments and challenges, *Diversity and Distributions*, 1910, 1217-1223.

- FREDERICQ, S., KRAYESKY-SELF, S., SAUVAGE, T., RICHARDS, J., KITTLE, R., ARAKAKI, N., HICKERSON, E., & SCHMIDT, W. E., (2019), The Critical Importance of Rhodoliths in the Life Cycle Completion of Both Macro- and Microalgae, and as Holobionts for the Establishment and Maintenance of Marine Biodiversity, *Frontiers in Marine Science*, 5.
- GANDRA, M., SEABRA, R., & LIMA, F. P., (2015), A low-cost, versatile data logging system for ecological applications, *Limnology and Oceanography : Methods*, 133, e10012.
- GÅRDMARK, A., CASINI, M., HUSS, M., van LEEUWEN, A., HJELM, J., PERSSON, L., & de ROOS, A. M., (2015), Regime shifts in exploited marine food webs : detecting mechanisms underlying alternative stable states using size-structured community dynamics theory, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130262.
- GASTON, K. J., (2000), Global patterns in biodiversity, *Nature*, 4056783, 220-227.
- GEARY, W. L., BODE, M., DOHERTY, T. S., FULTON, E. A., NIMMO, D. G., TULLOCH, A. I. T., TULLOCH, V. J. D., & RITCHIE, E. G., (2020), A guide to ecosystem models and their environmental applications, *Nature Ecology & Evolution*, 411, 1459-1471.
- GHERMANDI, A., NUNES, P. A. L. D., PORTELA, R., NALINI, R., & TEELUCKSINGH, S. S., (2010), Recreational, Cultural and Aesthetic Services from Estuarine and Coastal Ecosystems, *SSRN Electronic Journal*.
- GRINNELL, J., (1917), The Niche-Relationships of the California Thrasher, *The Auk*, 344, 427-433.
- GUISAN, A., EDWARDS, T. C., & HASTIE, T., (2002), Generalized linear and generalized additive models in studies of species distributions : setting the scene, *Ecological Modelling*, 1572, 89-100.
- GUISAN, A., THUILLER, W., & ZIMMERMANN, N. E., (2017), *Habitat Suitability and Distribution Models : With Applications in R*, Cambridge University Press.
- GUISAN, A., & ZIMMERMANN, N. E., (2000), Predictive habitat distribution models in ecology, *Ecological Modelling*, 1352, 147-186.
- HARDIMAN, N., & BURGIN, S., (2010), Recreational impacts on the fauna of Australian coastal marine ecosystems, *Journal of Environmental Management*, 9111, 2096-2108.
- HARFOOT, M. B. J., NEWBOLD, T., TITTENSOR, D. P., EMMOTT, S., HUTTON, J., LYUTSAREV, V., SMITH, M. J., SCHARLEMANN, J. P. W., & PURVES, D. W., (2014), Emergent Global Patterns of Ecosystem Structure and Function from a Mechanistic General Ecosystem Model, *PLoS Biology*, 124, 1-24.
- HARLEY, C. D. G., RANDALL HUGHES, A., HULTGREN, K. M., MINER, B. G., SORTE, C. J. B., THORNBUR, C. S., RODRIGUEZ, L. F., TOMANEK, L., & WILLIAMS, S. L., (2006), The impacts of climate change in coastal marine systems, *Ecology Letters*, 92, 228-241.
- HARRIS, D. L., ROVERE, A., CASELLA, E., POWER, H., CANAVESIO, R., COLLIN, A., POMEROY, A., WEBSTER, J. M., & PARRAVICINI, V., (2018), Coral reef structural complexity provides important coastal protection from waves under rising sea levels, *Science Advances*, 42, eaao4350.
- HARRIS, P., MACMILLAN-LAWLER, M., RUPP, J., & BAKER, E., (2014), Geomorphology of the oceans, *Marine Geology*, 352, 4-24.
- HASTINGS, A., BYERS, J. E., CROOKS, J. A., CUDDINGTON, K., JONES, C. G., LAMBRINOS, J. G., TALLEY, T. S., & WILSON, W. G., (2007), Ecosystem engineering in space and time, *Ecology Letters*, 102, 153-164.

- HOBDAY, A. J., ALEXANDER, L. V., PERKINS, S. E., SMALE, D. A., STRAUB, S. C., OLIVER, E. C., BENTHUYSEN, J. A., BURROWS, M. T., DONAT, M. G., FENG, M., HOLBROOK, N. J., MOORE, P. J., SCANNELL, H. A., SEN GUPTA, A., & WERNBERG, T., (2016), A hierarchical approach to defining marine heatwaves, *Progress in Oceanography*, 141, 227-238.
- HOEGH-GULDBERG, O., (1999), Climate change, coral bleaching and the future of the world's coral reefs, *Marine and freshwater research*, 508, 839-866.
- HUBBELL, S. P., (2001), *The unified neutral theory of biodiversity and biogeography*, Princeton University Press.
- HUGHES, A. R., WILLIAMS, S. L., DUARTE, C. M., HECK JR, K. L., & WAYCOTT, M., (2009), Associations of concern : declining seagrasses and threatened dependent species, *Frontiers in Ecology and the Environment*, 75, 242-246.
- HUGHES, T. P., DAY, J. C., & BRODIE, J., (2015), Securing the future of the Great Barrier Reef, *Nature Climate Change*, 56, 508-511.
- HUI, F. K., (2016), boral – Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r, *Methods in Ecology and Evolution*, 76, 744-750.
- HUMBOLDT, A. v., & BONPLAND, A., (1805), *Essai sur la géographie des plantes : accompagné d'un tableau physique des régions équinoxiales, fondé sur des mesures exécutées, depuis le dixième degré de latitude boréale jusqu'au dixième degré de latitude australe, pendant les années 1799, 1800, 1801, 1802 et 1803*, A Paris, Chez Levrault, Schoell et compagnie, libraires, XIII-1805.
- HUTCHINSON, G. E., (1957), Concluding Remarks, *Cold Spring Harbor Symposia on Quantitative Biology*, 220, 415-427.
- IPBES, (2019), *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (E. BRONDIZIO, S. DIAZ, J. SETTELE & H. T. a. NGO, Éd.).
- IVES, A. R., & HELMUS, M. R., (2011), Generalized linear mixed models for phylogenetic analyses of community structure, *Ecological Monographs*, 813, 511-525.
- JENTSCH, A., & WHITE, P., (2019), A theory of pulse dynamics and disturbance in ecology, *Ecology*, 1007, e02734.
- JOHNSON, C. R., CHABOT, R. H., MARZLOFF, M. P., & WOTHERSPOON, S., (2017), Knowing when (not) to attempt ecological restoration, *Restoration Ecology*, 251, 140-147.
- JONES, C. G., LAWTON, J. H., & SHACHAK, M., (1994), Organisms as Ecosystem Engineers, *Oikos*, 693, 373.
- KAZEMI, A., EVANS, H. B., VAN DE RIET, K., CASTILLO, L., & CURET, O. M., (2018), Hydrodynamics of Mangrove Roots and its Applications in Coastal Protection, *OCEANS 2018 MTS/IEEE Charleston*, 1-8.
- KEARNEY, M., & PORTER, W., (2009), Mechanistic niche modelling : combining physiological and spatial data to predict species' ranges, *Ecology Letters*, 124, 334-350.
- 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*, 33, 203-213.

- KRIEGEL, H.-P., KRÖGER, P., SANDER, J., & ZIMEK, A., (2011), Density-based clustering, *WIREs Data Mining and Knowledge Discovery*, 13, 231-240.
- KUFFNER, I. B., ANDERSSON, A. J., JOKIEL, P. L., RODGERS, K. S., & MACKENZIE, F. T., (2008), Decreased abundance of crustose coralline algae due to ocean acidification, *Nature Geoscience*, 12, 114-117.
- KUMARAVEL, K., RAVICHANDRAN, S., BALASUBRAMANIAN, T., & SONNESCHEIN, L., (2012), Seahorses – A source of traditional medicine, *Natural Product Research*, 2624, 2330-2334.
- LEGENDTRE, L., & LEGENDRE, P., (1979a), La structure des données écologiques, In MASSON (Éd.), *Ecologie Numérique* (T. 2), Les Presses de l'Université du Québec.
- LEGENDTRE, L., & LEGENDRE, P., (1979b), Le traitement multiple des données écologiques, In MASSON (Éd.), *Ecologie Numérique* (T. 1), Les Presses de l'Université du Québec.
- LEGENDTRE, P., & LEGENDRE, L., (2012), *Numerical Ecology* (3rd), Elsevier.
- LEIBOLD, M. A., & CHASE, J. M., (2018, décembre), *Metacommunity Ecology, Volume 59*, Princeton University Press.
- LEIBOLD, M. A., CHASE, J. M., & ERNEST, S. K. M., (2017), Community assembly and the functioning of ecosystems : how metacommunity processes alter ecosystems attributes, *Ecology*, 984, 909-919.
- LEWIS, S. L., & MASLIN, M. A., (2015), Defining the Anthropocene, *Nature*, 5197542, 171-180.
- LIMBURG, K. E., BREITBURG, D., SWANEY, D. P., & JACINTO, G., (2020), Ocean Deoxygenation : A Primer, *One Earth*, 21, 24-29.
- LINNAEUS, C., (1789), *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species ; cum characteribus, differentiis, synonymis, locis* (12^e éd., T. 1), Laurentius Salvius.
- LOH, T.-L., ARCHER, S. K., & DUNHAM, A., (2019), Monitoring program design for data-limited marine biogenic habitats : A structured approach, *Ecology and Evolution*, 912, 7346-7359.
- MACARTHUR, R. H., & WILSON, E. O., (1967), *The theory of island biogeography*, Princeton university press.
- MAINALI, K. P., WARREN, D. L., DHILEEPAN, K., MCCONNACHIE, A., STRATHIE, L., HASSAN, G., KARKI, D., SHRESTHA, B. B., & PARMESAN, C., (2015), Projecting future expansion of invasive species : comparing and improving methodologies for species distribution modeling, *Global Change Biology*, 2112, 4464-4480.
- MARZLOFF, M. P., (2022), Schéma conceptuel du projet de recherche ANR TRIDENT.
- MARZLOFF, M. P., LITTLE, L. R., & JOHNSON, C. R., (2016), Building Resilience Against Climate-Driven Shifts in a Temperate Reef System : Staying Away from Context-Dependent Ecological Thresholds, *Ecosystems*, 191, 1-15.
- MAUL, G. A., & DUEDALL, I. W., (2019), Demography of Coastal Populations, In C. W. FINKL & C. MAKOWSKI (Éd.), Springer International Publishing.
- MCINNES, L., HEALY, J., & MELVILLE, J., (2020), UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction, *arXiv*.
- MCINNES, L., HEALY, J., SAUL, N., & GROSSBERGER, L., (2018), UMAP : Uniform Manifold Approximation and Projection, *The Journal of Open Source Software*, 329, 861.
- MCLEAN, R. F., TSYBAN, A., BURKETT, V., CODIGNOTTO, J. O., FORBES, D. L., MIMURA, N., BEAMISH, R. J., & ITTEKKOT, V., (2001), Coastal Zones and Marine Ecosystems, In J. J. MCCARTHY, O. F.

- CANZIANI, N. A., LEARY, D. J., DOKKEN & K. S. WHITE (Éd.), *Climate change 2001 : impacts, adaptation, and vulnerability : contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change* (p. 343-379, T. 2), Cambridge University Press.
- MCMANUS, J. W., & POLSENERG, J. F., (2004), Coral–algal phase shifts on coral reefs : Ecological and environmental aspects, *Progress in Oceanography*, 602, 263-279.
- MILJANICH, G. P., (2004), Ziconotide : Neuronal Calcium Channel Blocker for Treating Severe Chronic Pain, *Current Medicinal Chemistry*, 1123, 3029-3040.
- MILOŠEVIĆ, D., MEDEIROS, A. S., STOJKOVIĆ PIPERAC, M., CVIJANOVIĆ, D., SOININEN, J., MILOSAVLJEVIĆ, A., & PREDIĆ, B., (2022), The application of Uniform Manifold Approximation and Projection (UMAP) for unconstrained ordination and classification of biological indicators in aquatic ecology, *Science of The Total Environment*, 815, 152365.
- MOUILLOT, D., GRAHAM, N. A., VILLÉGER, S., MASON, N. W., & BELLWOOD, D. R., (2013), A functional approach reveals community responses to disturbances, *Trends in Ecology & Evolution*, 283, 167-177.
- MOUILLOT, D., LOISEAU, N., GRENIÉ, M., ALGAR, A. C., ALLEGRA, M., CADOTTE, M. W., CASAJUS, N., DENELLE, P., GUÉGUEN, M., MAIRE, A., MAITNER, B., MCGILL, B. J., MCLEAN, M., MOUQUET, N., MUNOZ, F., THUILLER, W., VILLÉGER, S., VIOLLE, C., & AUBER, A., (2021), The dimensionality and structure of species trait spaces, *Ecology Letters*, 249, 1988-2009.
- NGUYEN, T., (2016), *Processeur Intel C4004 avec traces grises*. https://commons.wikimedia.org/wiki/File:Intel_C4004.jpg
- NIKU, J., HUI, F. K. C., TASKINEN, S., & WARTON, D. I., (2019), gllvm : Fast analysis of multivariate abundance data with generalized linear latent variable models in r, *Methods in Ecology and Evolution*, 1012, 2173-2182.
- NORBERG, A., ABREGO, N., BLANCHET, F. G., ADLER, F. R., ANDERSON, B. J., ANTILA, J., ARAÚJO, M. B., DALLAS, T., DUNSON, D., ELITH, J., FOSTER, S. D., FOX, R., FRANKLIN, J., GODSOE, W., GUIGAN, A., O'HARA, B., HILL, N. A., HOLT, R. D., HUI, F. K., ... OVASKAINEN, O., (2019), A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels, *Ecological Monographs*, e01370.
- NUGUES, M. M., & ROBERTS, C. M., (2003), Partial mortality in massive reef corals as an indicator of sediment stress on coral reefs, *Marine Pollution Bulletin*, 463, 314-323.
- NYSTRÖM, M., NORSTRÖM, A. V., BLENCKNER, T., de la TORRE-CASTRO, M., EKLÖF, J. S., FOLKE, C., ÖSTERBLOM, H., STENECK, R. S., THYRESSON, M., & TROELL, M., (2012), Confronting Feedbacks of Degraded Marine Ecosystems, *Ecosystems*, 155, 695-710.
- OHLSSON, M., & EKLÖF, A., (2020), Spatial resolution and location impact group structure in a marine food web, *Ecology Letters*, 2310, 1451-1459.
- OLIVER, E. C. J., DONAT, M. G., BURROWS, M. T., MOORE, P. J., SMALE, D. A., ALEXANDER, L. V., BENTHUYSEN, J. A., FENG, M., SEN GUPTA, A., HOBDAI, A. J., HOLBROOK, N. J., PERKINS-KIRKPATRICK, S. E., SCANNELL, H. A., STRAUB, S. C., & WERNBERG, T., (2018), Longer and more frequent marine heatwaves over the past century, *Nature Communications*, 91, 1324.

- OVASKAINEN, O., TIKHONOV, G., NORBERG, A., GUILLAUME BLANCHET, F., DUAN, L., DUNSON, D., ROSLIN, T., & ABREGO, N., (2017), How to make more out of community data? A conceptual framework and its implementation as models and software, *Ecology Letters*, 205, 561-576.
- PALMER, M. A., & STEWART, G. A., (2020), Ecosystem restoration is risky ... but we can change that, *One Earth*, 36, 661-664.
- PANG, S. E. H., SLIK, J. W. F., ZURELL, D., & WEBB, E. L., (2023), The clustering of spatially associated species unravels patterns in tropical tree species distributions, *Ecosphere*, 146, e4589.
- PETERSEN, J., KEMP, W., KENNEDY, V., DENNISON, W., & KANGAS, P., (2009), Enclosed Experimental Ecosystems and Scale : Tools for Understanding and Managing Coastal Ecosystems, In J. E. PETERSEN, V. S. KENNEDY, W. C. DENNISON & W. M. KEMP (Éd.), Springer US.
- PICHLER, M., BOREUX, V., KLEIN, A.-M., SCHLEUNING, M., & HARTIG, F., (2020), Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks, *Methods in Ecology and Evolution*, 112, 281-293.
- POGGIATO, G., MÜNKEMÜLLER, T., BYSTROVA, D., ARBEL, J., CLARK, J. S., & THUILLER, W., (2021), On the Interpretations of Joint Modeling in Community Ecology, *Trends in Ecology & Evolution*.
- POLLOCK, L. J., MORRIS, W. K., & VESK, P. A., (2012), The role of functional traits in species distributions revealed through a hierarchical model, *Ecography*, 358, 716-725.
- POLLOCK, L. J., O'CONNOR, L. M., MOKANY, K., ROSAUER, D. F., TALLUTO, M. V., & THUILLER, W., (2020), Protecting Biodiversity (in All Its Complexity) : New Models and Methods, *Trends in Ecology & Evolution*, 3512, 1119-1128.
- REECE, J., URRY, L., CAIN, M., WASSERMAN, S., MINORSKY, P., & JACKSON, R., (2012), *Biologie de Campbell* (9^e éd.), ERPI.
- REID, W., MOONEY, H., CROPPER, A., CAPISTRANO, D., CARPENTER, S., CHOPRA, K., DASGUPTA, P., DIETZ, T., DURAIAPPAH, A., HASSAN, R., KASPERSON, R., LEEMANS, R., MAY, R., MCMICHAEL, A., PINGALI, P., SAMPER, C., SCHOLES, R., WATSON, R., ZAKRI, A., ... ZUREK, M., (2005), *Ecosystems and human well-being - Synthesis : A Report of the Millennium Ecosystem Assessment*, Island Press.
- ROBERTS, D. W., (2020), Comparison of distance-based and model-based ordinations, *Ecology*, 1011, e02908.
- ROCA, G., ALCOVERRO, T., KRAUSE-JENSEN, D., BALSBY, T., VAN KATWIJK, M., MARBÀ, N., SANTOS, R., ARTHUR, R., MASCARÓ, O., FERNÁNDEZ-TORQUEMADA, Y., PÉREZ, M., DUARTE, C., & ROMERO, J., (2016), Response of seagrass indicators to shifts in environmental stressors : A global review and management synthesis, *Ecological Indicators*, 63, 310-323.
- ROCHA, J., YLETYINEN, J., BIGGS, R., BLENCKNER, T., & PETERSON, G., (2015), Marine regime shifts : drivers and impacts on ecosystems services, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130273.
- ROCHA, J. C., PETERSON, G., BODIN, Ö., & LEVIN, S., (2018), Cascading regime shifts within and across scales, *Science*, 3626421, 1379-1383.
- ROMERO, G. Q., GONÇALVES-SOUZA, T., VIEIRA, C., & KORICHEVA, J., (2015), Ecosystem engineering effects on species diversity across ecosystems : a meta-analysis, *Biological Reviews*, 903, 877-890.

- RYO, M., AGUILAR-TRIGUEROS, C. A., PINEK, L., MULLER, L. A., & RILLIG, M. C., (2019), Basic Principles of Temporal Dynamics, *Trends in Ecology & Evolution*, 348, 723-733.
- SCHEFFER, M., CARPENTER, S. R., DAKOS, V., & van NES, E. H., (2015), Generic Indicators of Ecological Resilience : Inferring the Chance of a Critical Transition, *Annual Review of Ecology, Evolution, and Systematics*, 461, 145-167.
- SCHEFFER, M., CARPENTER, S., FOLEY, J. A., FOLKE, C., & WALKER, B., (2001), Catastrophic shifts in ecosystems, *Nature*, 4136856, 591-596.
- SCHEFFER, M., HIROTA, M., HOLMGREN, M., NES, E. H. V., & CHAPIN, F. S., (2012), Thresholds for boreal biome transitions, *Proceedings of the National Academy of Sciences*, 10952, 21384-21389.
- SCHMIDTKO, S., STRAMMA, L., & VISBECK, M., (2017), Decline in global oceanic oxygen content during the past five decades, *Nature*, 5427641, 335-339.
- SCHWARTSMANN, G., DA ROCHA, A. B., BERLINCK, R. G., & JIMENO, J., (2001), Marine organisms as a source of new anticancer agents, *The Lancet Oncology*, 24, 221-225.
- SHEM UNGER, A. T., Mark Rollins, & DUMAIS, H., (2021), iNaturalist as an engaging tool for identifying organisms in outdoor activities, *Journal of Biological Education*, 555, 537-547.
- SIBIYA, A., JEYAVANI, J., SIVAKAMAVALLI, J., RAVI, C., DIVYA, M., & VASEEHARAN, B., (2021), Bioactive compounds from various types of sea urchin and their therapeutic effects — A review, *Regional Studies in Marine Science*, 44, 101760.
- SMITH, K. E., BURROWS, M. T., HOBDAV, A. J., KING, N. G., MOORE, P. J., SEN GUPTA, A., THOMSEN, M. S., WERNBERG, T., & SMALE, D. A., (2023), Biological Impacts of Marine Heatwaves, *Annual Review of Marine Science*, 151, 119-145.
- SONNEWALD, M., DUTKIEWICZ, S., HILL, C., & FORGET, G., (2020), Elucidating ecological complexity : Unsupervised learning determines global marine eco-provinces, *Science Advances*, 622.
- SPAKE, R., BARAJAS-BARBOSA, M. P., BLOWES, S. A., BOWLER, D. E., CALLAGHAN, C. T., GARBOWSKI, M., JURBURG, S. D., van KLINK, R., KORELL, L., LADOUCEUR, E., ROZZI, R., VIANA, D. S., XU, W.-B., & CHASE, J. M., (2022), Detecting Thresholds of Ecological Change in the Anthropocene, *Annual Review of Environment and Resources*, 471.
- STAVER, A. C., ARCHIBALD, S., & LEVIN, S., (2011), Tree cover in sub-Saharan Africa : Rainfall and fire constrain forest and savanna as alternative stable states, *Ecology*, 925, 1063-1072.
- SUBRAMANIAN, M., (2019), Anthropocene now : influential panel votes to recognize Earth's new epoch, *Nature*.
- SUNDAY, J. M., FABRICIUS, K. E., KROEKER, K. J., ANDERSON, K. M., BROWN, N. E., BARRY, J. P., CONNELL, S. D., DUPONT, S., GAYLORD, B., HALL-SPENCER, J. M., KLINGER, T., MILAZZO, M., MUNDAY, P. L., RUSSELL, B. D., SANFORD, E., THIYAGARAJAN, V., VAUGHAN, M. L. H., WIDDICOMBE, S., & HARLEY, C. D. G., (2017), Ocean acidification can mediate biodiversity shifts by changing biogenic habitat, *Nature Climate Change*, 71, 81-85.
- SUTTON, P. C., ANDERSON, S. J., COSTANZA, R., & KUBISZEWSKI, I., (2016), The ecological economics of land degradation : Impacts on ecosystem service values, *Ecological Economics*, 129, 182-192.
- TATEM, A. J., GOETZ, S. J., & HAY, S. I., (2008), Fifty years of earth observation satellites : Views from above have lead to countless advances on the ground in both scientific knowledge and daily life, *American scientist*, 965, 390.

- TENENBAUM, J. B., de SILVA, V., & LANGFORD, J. C., (2000), A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, 2905500, 2319-2323.
- THOMSEN, M. S., ALTIERI, A. H., ANGELINI, C., BISHOP, M. J., GRIBBEN, P. E., LEAR, G., HE, Q., SCHIEL, D. R., SILLIMAN, B. R., SOUTH, P. M., WATSON, D. M., WERNBERG, T., & ZOTZ, G., (2018), Secondary foundation species enhance biodiversity, *Nature Ecology & Evolution*, 24, 634-639.
- THOMSEN, M. S., WERNBERG, T., ALTIERI, A., TUYA, F., GULBRANSEN, D., MCGLATHERY, K. J., HOLMER, M., & SILLIMAN, B. R., (2010), Habitat Cascades : The Conceptual Context and Global Relevance of Facilitation Cascades via Habitat Formation and Modification, *Integrative and Comparative Biology*, 502, 158-175.
- THURSTAN, R. H., BRITTAIN, Z., JONES, D. S., CAMERON, E., DEARNALEY, J., & BELLGROVE, A., (2018), Aboriginal uses of seaweeds in temperate Australia : an archival assessment, *Journal of Applied Phycology*, 303, 1821-1832.
- TODD, P. A., HEERY, E. C., LOKE, L. H. L., THURSTAN, R. H., KOTZE, D. J., & SWAN, C., (2019), Towards an urban marine ecology : characterizing the drivers, patterns and processes of marine ecosystems in coastal cities, *Oikos*, 1289, 1215-1242.
- UDELL, M., HORN, C., ZADEH, R., & BOYD, S., (2015), Generalized Low Rank Models, *arXiv*.
- van der MAATEN, L., & HINTON, G., (2008), Visualizing Data using t-SNE, *Journal of Machine Learning Research*, 986, 2579-2605.
- VELLEND, M., (2010), Conceptual Synthesis in Community Ecology, *The Quarterly Review of Biology*, 852, 183-206.
- VELLEND, M., (2017, 2023-10-13), *The Theory of Ecological Communities*, Princeton University Press.
- WARTON, D. I., BLANCHET, F. G., O'HARA, R. B., OVASKAINEN, O., TASKINEN, S., WALKER, S. C., & HUI, F. K., (2015), So many variables : joint modeling in community ecology, *Trends in Ecology & Evolution*, 3012, 766-779.
- WAYCOTT, M., DUARTE, C. M., CARRUTHERS, T. J. B., ORTH, R. J., DENNISON, W. C., OLYARNIK, S., CALLADINE, A., FOURQUREAN, J. W., HECK, K. L., HUGHES, A. R., KENDRICK, G. A., KENWORTHY, W. J., SHORT, F. T., & WILLIAMS, S. L., (2009), Accelerating loss of seagrasses across the globe threatens coastal ecosystems, *Proceedings of the National Academy of Sciences*, 10630, 12377-12381.
- WERNBERG, T., BENNETT, S., BABCOCK, R. C., de BETTIGNIES, T., CURE, K., DEPCZYNSKI, M., DUFOIS, F., FROMONT, J., FULTON, C. J., HOVEY, R. K., HARVEY, E. S., HOLMES, T. H., KENDRICK, G. A., RADFORD, B., SANTANA-GARCON, J., SAUNDERS, B. J., SMALE, D. A., THOMSEN, M. S., TUCKETT, C. A., ... WILSON, S., (2016), Climate-driven regime shift of a temperate marine ecosystem, *Science*, 3536295, 169-172.
- WERNBERG, T., S. THOMSEN, M., K. BAUM, J., J. BISHOP, M., F. BRUNO, J., A. COLEMAN, M., FILBEE-DEXTER, K., GAGNON, K., HE, Q., MURDIYARSO, D., ROGERS, K., R. SILLIMAN, B., A. SMALE, D., STARKO, S., & A. VANDERKLIFT, M., (2023), Impacts of Climate Change on Marine Foundation Species, *Annual Review of Marine Science*, 161.
- WHITTAKER, R. H., (1972), Evolution and Measurement of Species Diversity, *Taxon*, 212-3, 213-251.

- WHITTAKER, R. H., (1960), Vegetation of the Siskiyou Mountains, Oregon and California, *Ecological Monographs*, 303, 279-338.
- WHITTAKER, R. H., (1975), *Communities and ecosystems*. (2^e éd.), Macmillan.
- WITMAN, J. D., LAMB, R. W., & BYRNES, J. E. K., (2015), Towards an integration of scale and complexity in marine ecology, *Ecological Monographs*, 854, 475-504.
- WOOD, C., SULLIVAN, B., ILIFF, M., FINK, D., & KELLING, S., (2011), eBird : Engaging Birders in Science and Conservation, *PLOS Biology*, 912, 1-5.
- WULDER, M. A., LOVELAND, T. R., ROY, D. P., CRAWFORD, C. J., MASEK, J. G., WOODCOCK, C. E., ALLEN, R. G., ANDERSON, M. C., BELWARD, A. S., COHEN, W. B., DWYER, J., ERB, A., GAO, F., GRIFFITHS, P., HELDER, D., HERMOSILLA, T., HIPPLE, J. D., HOSTERT, P., HUGHES, M. J., ... ZHU, Z., (2019), Current status of Landsat program, science, and applications, *Remote Sensing of Environment*, 225, 127-147.
- WYATT, A. S. J., LEICHTER, J. J., WASHBURN, L., KUI, L., EDMUNDS, P. J., & BURGESS, S. C., (2023), Hidden heatwaves and severe coral bleaching linked to mesoscale eddies and thermocline dynamics, *Nature Communications*, 141, 25.
- YASUHARA-BELL, J., & LU, Y., (2010), Marine compounds and their antiviral activities, *Antiviral Research*, 863, 231-240.
- ZHAO, L., SCHÖNE, B. R., MERTZ-KRAUS, R., & YANG, F., (2017), Insights from sodium into the impacts of elevated pCO₂ and temperature on bivalve shell formation, *Journal of Experimental Marine Biology and Ecology*, 486, 148-154.

Essential ingredients in Joint Species Distribution Models

Préambule

Les modèles de distribution d'espèces (*Species Distribution Models, SDM*) ont récemment bénéficié d'avancées méthodologiques significatives, en particulier grâce à l'adoption de modèles statistiques multivariés, désignés sous le terme de *jSDM* (*Joint Species Distribution Models*; WARTON et al. (2015)). Ces modèles se révèlent prometteurs pour élucider de nombreuses questions en écologie des communautés. Ils permettent de comprendre l'effet de l'environnement, des traits fonctionnels, de l'histoire évolutive sur les espèces étudiées, les raisons leur présence dans une zone donnée, et enfin la structure de la communauté étudiée (WARTON et al., 2015; OVASKAINEN et al., 2017b). Ces méthodologies ont suscité un intérêt croissant au sein de la communauté écologique pour diverses raisons. En premier lieu, elles semblent particulièrement adaptées pour décrypter une multitude de processus écologiques (voir Table 1 dans OVASKAINEN et al. (2017b)). En second lieu, elles semblent manifester une capacité prédictive supérieure en comparaison avec les *SDM* (NORBERG et al., 2019), et se distinguent par leur capacité à inférer des interactions interspécifiques (WARTON et al., 2015; OVASKAINEN et al., 2017b; OVASKAINEN et al., 2017a).

Néanmoins, certaines incertitudes subsistent quant à l'application optimale de ces modèles. La flexibilité intrinsèque de ce cadre méthodologique suscite des interrogations sur la stratégie à appliquer : pour optimiser l'inférence des processus écologiques d'une part, et la performance prédictive d'autre part. Des questions restent en suspens, quant à l'inclusion par exemple, de données sur les traits fonctionnels des espèces de la communauté étudiée, difficile à collecter, et de leur impact sur le pouvoir explicatif et prédictif des mo-

dèles. Si leur capacité à améliorer la précision prédictive a été largement démontrée sur des ensembles de données de présence/absence, notamment dans les études de NORBERG et al. (2019) et WILKINSON et al. (2021), la performance de ces modèles quand ils sont ajustés à des données d’abondance reste, à notre connaissance, insuffisamment explorée dans la littérature. Certains travaux récents ont exprimé des réserves quant à la capacité de ces méthodes à déduire efficacement les réseaux d’interactions interspécifiques à partir de données de présence/absence (BLANCHET et al., 2020; VIOLET, 2020; POGGIATO et al., 2021).

Il est/semble nécessaire d’évaluer la sensibilité de l’approche par *jSDM*, avant d’envisager d’utiliser ce genre de modèles pour réaliser des inférences sur les processus écologiques et des prédictions qui visent à déterminer l’importance des habitats biogéniques dans la structuration des communautés benthiques en utilisant des ensembles de données d’abondance collectées à grande échelle, par exemple celles issues du programme *Reef Life Survey (RLS)*. Cette évaluation a été réalisée en développant une méthodologie d’évaluation des performances des *jSDMs* et en l’appliquant à un jeu de données collectées par le programme de recherche *REBENT*. Dans le cadre de ce chapitre, nous nous sommes intéressés à un sous-ensemble du jeu de données du *REBENT*, avec un focus sur la communauté de polychètes intertidale inféodée à deux habitats différents : les sédiments nus et des herbiers intertidaux dont les processus écologiques ont déjà été étudiés le long de la côte bretonne (BOYÉ et al., 2017; BOYÉ et al., 2019). L’objectif est de vérifier si étendre l’approche *jSDM* à une échelle globale telle que celle présentée par le *RLS* serait pertinent pour quantifier l’influence des habitats biogéniques sur la structure des communautés faunistiques. Ceci afin de mieux prédire et comprendre les conséquences de la disparition des habitats biogéniques.

Ce chapitre de thèse a déjà fait l’objet d’une communication orale “*Effects of including community-level or species-specific data on the performances of a Joint Species Distribution Model : a case study with species-rich marine benthic communities*”. International Statistical Ecology Conference 2022. Une publication a été soumise, et envoyée en revue par deux fois en quasiment un an, au journal *Methods in Ecology and Evolution* a été récemment refusée. C’est pourquoi ce chapitre de thèse devrait ainsi être prochainement soumis pour publication au *Journal of Applied Ecology*.

2.1 Abstract

1. *Joint Species Distribution Models (jSDM)* are increasingly used to explain and predict biodiversity patterns. *jSDMs* account for species co-occurrence patterns and can include phylogeny or functional traits to better capture the processes shaping communities. Yet, several factors may alter the interpretability and predictive ability of *jSDMs* : missing abiotic predictors, omitting ecologically-important species, or increasing the number of model parameters by adding phylogeny and/or trait information.
2. We developed a novel framework to comprehensively assess the interpretability, explanatory and predictive power of *jSDMs* at both species and community levels. We compared performances of four alternative model formulations : (1) a *Benchmark jSDM* with only abiotic predictors and residual co-occurrence patterns, (2) a *jSDM* adding phylogeny to the *Benchmark*, (3) a *jSDM* adding traits to model 2, and (4) the *Benchmark jSDM* with additional non-target species that are not of direct interest but potentially interact with the target assemblage. Models were fitted on both presence/absence and abundance data for 99 target polychaete species sampled in two coastal habitats over 500km and 8 years, along with information on 179 non-target species and traits data for the target species.
3. For both presence/absence and abundance data, explanatory power was good for all models but their interpretability and predictive power varied. Relative to the *Benchmark* model, predictive errors on species abundances decreased by 95% or 53%, when including non-target species, or phylogeny, respectively. These differences across models relate to changes in both species-environment relationships and residual co-occurrence patterns. While considering trait data did not improve explanatory or predictive power, it facilitated interpretation of trait-mediated species response to environmental gradients.
4. This study demonstrates that any *jSDM* formulation comes with tradeoffs between either explaining or predicting the occurrence or abundance of species. Hence, it highlights the need to compare alternative model formulations using the original and comprehensive assessment framework proposed in this study. Overall, this work contributes to a better understanding of *jSDMs*' performances across multiple facets and provides insights and tools for model selection based on specific objectives and available data.

2.2 Introduction

Community ecology aims at describing, explaining, and predicting changes in communities (TREDENNICK et al., 2021). Understanding the processes that determine species distribution is a prerequisite to characterize and predict community structure and associated ecological dynamics, which is critical to inform effective management or restoration measures in a rapidly changing world (DIETZE et al., 2018; BRUDVIG & CATANO, 2022). *Joint Species Distribution Models (jSDM)* are particularly well-suited tools to address these challenges, whether to characterize the processes that shape observed communities (WARTON et al., 2015; OVASKAINEN et al., 2017a), or to predict future changes in species assemblages (NORBERG et al., 2019; POLLOCK et al., 2020).

jSDMs are multivariate (i.e. multi-species) extensions of *Species Distribution Models (SDMs)*, which have been broadly applied over the past decades - across all terrestrial and marine realms - to understand and predict both species occurrences (ELITH et al., 2006; NORBERG et al., 2019) and species abundances (HOWARD et al., 2014; WALDOCK et al., 2022) using a set of covariates (e.g. climatic variables). Relative to *SDMs*, *jSDMs* explanatory power can benefit from accounting for species assembly rules (OVASKAINEN et al., 2017b). In particular, relative to single-species *SDMs* that only consider the abiotic niche of species (i.e. the Grinnellian niche), *jSDM* can theoretically also account for interspecific interactions (i.e. the Eltonian niche).

In *jSDMs*, the variability in community composition not explained by covariates is captured by a residual covariance matrix representing species co-occurrence patterns potentially representing biotic interactions (OVASKAINEN et al., 2017b). This feature is highly attractive to ecologists because it provides a way to disentangle the relative influence of abiotic and biotic processes on biodiversity patterns (GODSOE et al., 2017) while also improving model's predictive power (GIANNINI et al., 2013; STANICZENKO et al., 2017). However, in practice, inferring and interpreting residual co-occurrence patterns using *jSDMs* remains challenging for several reasons (BLANCHET et al., 2020; HOLT, 2020).

First, while *jSDMs* have been applied to a large number of species presence/absence datasets (NORBERG et al., 2019; WILKINSON et al., 2019; WILKINSON et al., 2021), simulation studies showed that inferred co-occurrence networks do not necessarily provide evidence for species interactions (DORMANN et al., 2018; BLANCHET et al., 2020) but only capture spatial and temporal associations between species (KEIL et al., 2021). Some authors speculated

that *jSDMs* applied to abundance data - instead of presence/absence data - could provide a better proxy for biotic interactions (BLANCHET et al., 2020; MOMAL et al., 2020). Accordingly, *jSDM* have increasingly been applied to abundance data (HUI, 2016; OVASKAINEN et al., 2017b; CHIQUET et al., 2021). While challenges related to modelling abundance data was recently explored in the context of species distribution modelling (WALDOCK et al., 2022), the predictive and the explanatory power of *jSDM* fitted to abundance data remains relatively untested compared to presence/absence data (NORBERG et al., 2019; WILKINSON et al., 2021).

Second, regardless of the type of data considered (i.e. presence/absence or abundance), several factors may limit or affect the interpretability and predictive ability of *jSDM*. For instance, co-occurrence patterns estimated in *jSDM* are affected by unaccounted environmental variables implying that *jSDMs* cannot fully separate the environmental and the biotic niche of species (BLANCHET et al., 2020; POGGIATO et al., 2021). Beyond missing environmental predictors, accounting for extra species that can influence the target community (e.g. competitors) is key to improve *jSDMs*' inference and predictions. However, because many ecological studies only focus on particular taxonomic groups (POLLOCK et al., 2014; HÄKKILÄ et al., 2018) and disregard non-target taxa, co-occurrence patterns estimated from *jSDMs* are almost always skewed by missing ecological actors (MOMAL et al., 2021). How this bias affects the predictive ability of *jSDM* remains untested.

Finally, similar to *SDMs*, *jSDMs* can theoretically be extended to include additional sources of information about modelled species (NIKU et al., 2019; OVASKAINEN et al., 2017b). For instance, accounting for phylogenetic relationships between species (IVES & HELMUS, 2011) or for the link between functional traits and environmental responses (POLLOCK et al., 2012) can improve both the explanatory and the predictive powers of *SDMs* (MORALES-CASTILLA et al., 2017; VESK et al., 2021). These findings support the hypothesis that similar species, in terms of traits and/or recent evolutionary history, usually share similar environmental preferences. While inclusion of species-specific information in *jSDMs* should yield similar benefits (OVASKAINEN et al., 2017b), the relative influence of additional sources of information on their interpretability and predictive power remains untested (NORBERG et al., 2019; WILKINSON et al., 2019; ABREGO et al., 2022).

Overall, many practical questions remain concerning the application of *jSDMs* to ecological community monitoring data, in particular related to inclusion of additional sources

of information within the models. While some comparative assessments of *jSDM* performance exists (e.g. NORBERG et al. (2019); WILKINSON et al. (2021)), including some comparison of the benefit of trait and phylogenetic data in some phyla (e.g. ABREGO et al. (2022)), there has been no formal assessment of the relative importance of species-specific information (trait and/or phylogeny) compared to the role of missing species. Furthermore, comparative assessments have rarely been performed on both presence/absence and abundance data. To a few exceptions (WALDOCK et al., 2022), most assessments were made considering presence/absence data (NORBERG et al., 2019; WILKINSON et al., 2019) and mostly focused on predictive power (NORBERG et al., 2019; WILKINSON et al., 2019), hence disregarding the interpretability/explanatory aspects of the models (TREDENNICK et al., 2021). Yet, *jSDMs* are increasingly fitted on abundance data (BRIMACOMBE et al., 2021) and used for explanatory purposes (ABREGO et al., 2017). Hence, there is a mismatch between current understanding of *jSDMs* performance and their application by ecologists. In practice, most *jSDM* applications consider a single model structure and do not explore the effects of including additional sources of information. Perhaps this shortcoming relates to the high-dimensionality of *jSDMs* which makes their comparison challenging.

In this study, we developed a multi-faceted assessment framework to evaluate the extent to which alternative parameterization of *jSDM* can lead to a better interpretability or predictability at the species and community levels. To illustrate its usefulness, we applied this general framework to a case study presenting typical features of community ecology datasets. Specifically, by comparing predictions obtained from a *Benchmark* model excluding additional sources of information (i.e. a classical *jSDM*), we tested the effect of (1) including phylogeny alone and in combination with trait data, (2) incorporating monitoring information related to non-target species and (3) considering abundance instead of presence/absence data. We hypothesized that all these sources of information should improve *jSDM* predictive and explanatory powers, but did not assume a priori that a given modeling strategy would lead to greater improvements in model performances.

2.3 Materials & Methods

We used the *HMSC* (*Hierarchical Modeling of Species Communities*) framework applied to a long-term monitoring dataset. The following subsections sequentially describe our workflow (as illustrated in Fig. 2.1) : (1) the *HMSC* framework, (2) the data used in this

study, (3) data splitting between train and test datasets to assess the explanatory and predictive powers of models, respectively, (4) the rationales for the suite of alternative models considered and, (5) a multi-faceted framework developed to assess tradeoffs in jSDMs' performances in relation to different study purposes.

2.3.1 Hierarchical Modelling of Species Community (HMSC)

HMSC is a multivariate hierarchical generalized linear mixed model using Bayesian inference (OVASKAINEN & ABREGO, 2020). It has two parts : one for fixed effects and another for random effects. The fixed part models the species' realized niche, where each dimension of the niche is a covariate (e.g. temperature ; Fig. 2.1). Including trait data can improve species niche estimates by accounting for trait-environment relationships, where species with similar traits are expected to respond similarly along environmental gradients (OVASKAINEN et al., 2017b). Including phylogenetic data can help capture residual ecological information not included in the available trait data, as phylogenetically-close species tend to share similar traits and niches (WIENS et al., 2010). Alongside traits, phylogeny can improve niche estimates for rare species by borrowing information from similar species (OVASKAINEN et al., 2017b ; OVASKAINEN et al., 2017a ; OVASKAINEN & ABREGO, 2020). The random part of *HMSC* relies on latent variables, i.e. covariates that capture residual variance, including missing environmental features or biotic interactions (OVASKAINEN et al., 2017b ; OVASKAINEN et al., 2017a ; OVASKAINEN & ABREGO, 2020). The H matrix (site loadings) estimates missing covariate values, while the Λ matrix (species loadings) represents species' response to these missing covariates (Fig. 2.1).

2.3.2 Datasets

Faunistic data

The *REBENT* program (rebent.ifremer.fr) is an ongoing monitoring of benthic macrofauna across multiple stations along Brittany's coastline (Western France). Here, we used data from BOYÉ et al. (2019), covering 23 sites and two intertidal soft-bottom habitats : bare sands and seagrass meadows (Fig. S1) where infaunal communities were monitored yearly using the same protocol between 2006 and 2014. A detailed description of the sampling methodology is provided in (BOYÉ et al., 2017 ; BOYÉ et al., 2019). At each site, three faunal samples (0.03 m² cores) were taken at each of three fixed sampling points distributed 200 m apart. These samples were then pooled together to estimate abundances at the

site level. For each sampling event, individuals were identified to the lowest taxonomic level possible (mostly species level ; for simplicity we hereafter refer to “species”).

Overall, across a total of 375 sampling units (i.e. unique combination of years, sites and habitats), 152,583 individuals belonging to 519 species were collected and identified. To avoid convergence issues and poor model inference, we filtered out species occurring less than four times (across the 180 samples used as train set, see below), resulting in the removal of 241 species. The remaining 278 species included 99 polychaete species (the targeted assemblage) and 179 non-target species of bivalves, molluscs, and amphipoda, which may predate or compete with polychaetes (GRALL et al., 2006 ; JANKOWSKA et al., 2019). We chose to focus on polychaetes as this taxonomic group exhibits diverse lifestyles (JUMARS et al., 2015), can be used to monitor the health of benthic habitats (GIANGRANDE et al., 2005), and because trait data and ecological information were available from previous studies (BOYÉ et al., 2019).

Traits and phylogeny data

Traits data were retrieved from (BOYÉ et al., 2019) for the 99 polychaete species present in the train set (see below). The 11 fuzzy-coded traits available (see BOYÉ et al. (2019) for details) were synthesized using a fuzzy-PCA, with the *fpca* function from the *ade4* R package (THIOULOUSE et al., 2018). The first three axes, which account for 59% of the total variance of the trait matrix, were included in the model as synthetic traits data (Fig. 2.1). The first axis distinguishes mobile predatory species from sessile microphages ; the second axis differentiates semelparous species from iteroparous species ; and, the third axis separates burrowers from tube-dwellers (Fig. S2).

Phylogeny was not available, hence we followed common practices (OVASKAINEN & ABREGO, 2020) and retrieved the taxonomic classification of these 99 polychaetes through the WoRMS database (www.marinespecies.org ; january 2020) and used this information as a proxy for phylogenetic relationships (Fig. 2.1 ; OVASKAINEN et ABREGO (2020)). Phylogenetic distances were then estimated using the *ape* R package (PARADIS & SCHLIEP, 2019).

Environmental data

Following BOYÉ et al. (2019) (see Appendix A for details about data sources), we selected seven environmental variables to characterise the ecological niche of each species. These seven variables quantify different components of coastal environmental variability including hydrology (sea water temperature, salinity and current velocity), sedimentology (mud and organic matter content), substrate heterogeneity (Trask index) and local wave exposure (fetch). For each of these seven variables, the first- and second-degree polynomials were computed to account for non-linear responses.

2.3.3 Comparison of alternative model structures

The first model (denoted *Bench*) only considers data for the 99 target polychaete species and the 7 environmental covariates (Fig. 2.1). The second model (denoted *Ph*) adds phylogenetic data to the *Bench* model (Fig. 2.1). The third model (denoted *TrPh*) adds traits data to the *Ph* model. The fourth model (denoted *WhC*) has the same structure as the *Bench* model but includes data on the whole community (278 species, including 179 additional non-target species; Fig. 2.1). *WhC* excludes traits (unavailable for the non-target taxa) and phylogenetic data for faster computation. Each model was fitted twice, either with presence/absence or abundance data, using probit and lognormal Poisson distributions respectively. All models include the same random effects (Fig. 2.1) : temporal, spatial (sites), and habitats (bare vs seagrass).

2.3.4 Model fitting

We estimated model parameters by running 5 chains using a MCMC sampling algorithm over 375,000 iterations. The first 125,000 iterations were discarded as burn-in while the remaining were thinned every 250 iterations yielding 1,000 posterior samples per chain for an overall total of 5,000 posterior samples for each parameter. We assessed convergence for each model parameter using both potential scale reduction factor (GELMAN & RUBIN, 1992) and effective sample size as reported in supplementary materials (Appendix B). All models were fitted using the DATARMOR supercomputing facility.

2.3.5 Assessing model performance and interpretability

For independent assessment of models' predictive performance, the dataset was split into a train and a test set, instead of using strict cross-validation procedure that would have considerably increase the computational burden (see also NORBERG et al. (2019)). The train dataset consisted of 180 sampling units (21 sites; one or two habitats, and six to nine years per site; Fig. S1). The test set comprised 35 sampling units covering a 9-year period at two specific sites with both seagrass and bare sand habitats. These sites were chosen as representative of both regional macrofaunal species diversity (all the species observed in the test set are also observed in the train set) and mean environmental conditions (which limits model extrapolation outside of the trained parameter space; Fig. S3-S4; BOYÉ et al. (2017); BOYÉ et al. (2022); TOUMI et al. (2023)).

To assess *jSDM*'s performance, we used a set of complementary metrics to evaluate both their explanatory and predictive abilities on the train and test dataset, respectively (Table 1). AUC and RMSE, calculated only for the 99 target species (i.e. polychaetes) even for the *WhC* model that includes a total of 278 species, were used to assess overall and species-level performance for presence/absence and abundance models, respectively. Relationships between observed and predicted mean species abundances across all sites were also visualized for abundance models.

Along with the raw AUC and RMSE values, we also visualized and quantified changes relative to the *Bench* model for the *Ph*, *TrPh* and *WhC* models. For abundance models, we computed the overall relative change in mean RMSE across species as :

$$\text{Relative change} = \frac{\text{mean}(\text{RMSE}_{\text{focal model}}) - \text{mean}(\text{RMSE}_{\text{Bench}})}{\text{mean}(\text{RMSE}_{\text{Bench}})} \times 100 \quad (2.1)$$

AUC and RMSE only partially capture model accuracy at the community scale (Table 1). To explore this aspect, we focused on differences between predicted and observed assemblage richness and total abundances (for abundance models). We also compared observed and predicted Sørensen (for presence/absence) and Bray-Curtis (abundance) pairwise-dissimilarity matrices to explore how well β -diversity patterns were reproduced by the models. For these three metrics, we computed relative change for both the train

and test datasets between mean predicted and mean observed values as follows :

$$\text{Relative change} = \frac{\text{mean}(\text{Metric}_{\text{predict}}) - \text{mean}(\text{Metric}_{\text{obs}})}{\text{mean}(\text{Metric}_{\text{obs}})} \times 100 \quad (2.2)$$

where “Metric” is a community-based measure (e.g. species richness, total abundance, dissimilarity matrices) estimated from model predictions or observations at the sample level (i.e. unique combination of site, habitat and year ; or, pairs of samples for dissimilarity). To evaluate model interpretability, we calculated the amount of explained variance per species and the proportion that can be attributed to environmental covariates (fixed effects) and random effects. We compared the overall relative change in the proportion of variance explained by the covariates and by the random effects for the *Ph*, *TrPh* and *WhC* relative to the *Bench* model (by comparing mean values across species similarly to Eq. 2.1). We also propose a novel way of exploring species-environment relationships (Table 1) by classifying the response curves estimated from the different models based on their shapes, considering both their direction (decline, null, or increase) and their acceleration (decelerated, constant, or accelerated) (RIGAL et al., 2020). Finally, we compared the residual co-occurrence patterns associated with each random effect of the *Bench* model with those of the best performing model (*WhC*). We quantified differences in magnitude and sign of residual species-species correlations using the following index :

$$\delta = |\text{corr}_{\text{best model}} - \text{corr}_{\text{Benchmark}}| \times \text{sign}(\text{corr}_{\text{best model}} \times \text{corr}_{\text{Benchmark}}) \quad (2.3)$$

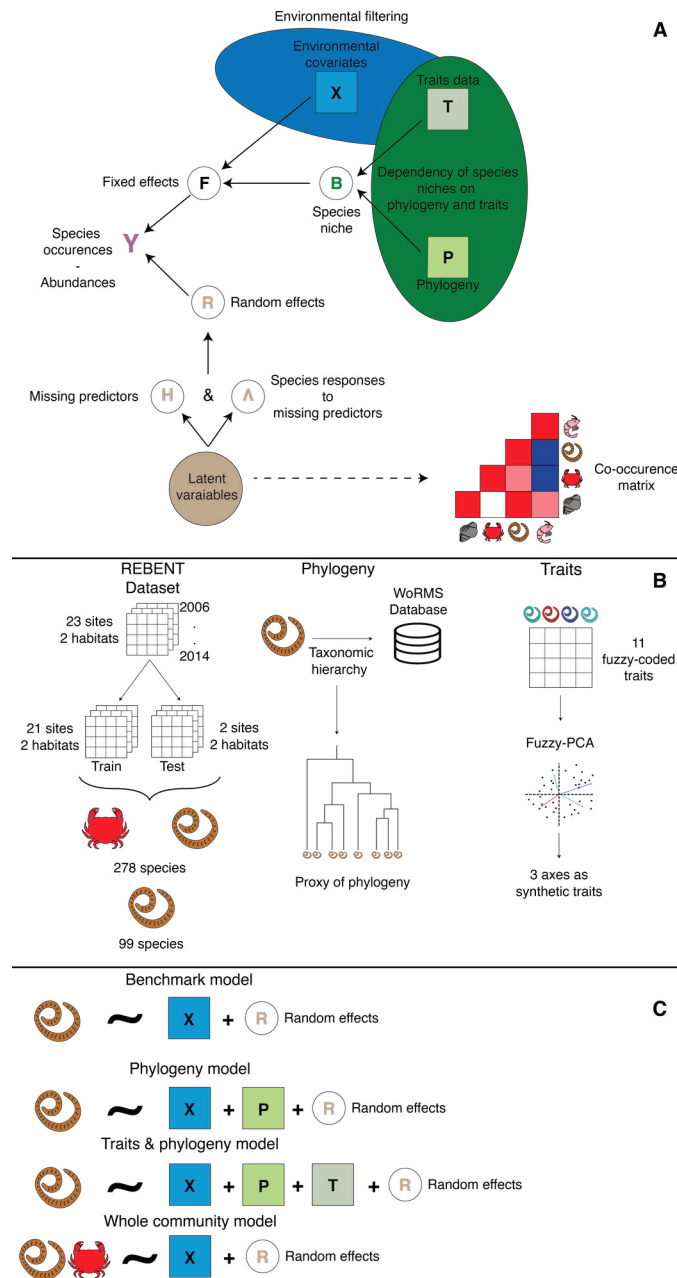


Figure 2.1 – Study workflow. *A*. Hierarchical Model of Species Community (HMSC) structure, incorporating environmental variables, phylogeny, and species-specific traits. *B*. Data pre-processing : partitioning community data into train and test datasets, estimating phylogenetic distance (using taxonomic classification), and dimension reduction of species-trait matrix using fuzzy-PCA. *C*. Summary of the four alternative model structures fitted to presence/absence and abundance data : the Benchmark, Phylogeny, Traits & Phylogeny models only consider the targeted polychaetes, while the Whole Community model includes additional species potentially interacting with the target assemblage. Random effects for sampling year, site, and habitat were included in all models.

Table 2.1 – Multi-assessment framework providing a list of useful metrics to assess, interpret or compare *j*SDMs across different ecological facets (rows) at the species, community or overall level. *Italicized metrics are used in this study.*

Model outputs		Example of derived-metrics for model interpretation	Example of derived-metrics for model evaluation		Example of performance measures to assess the explanatory/predictive power of models ¹	
			Presence/Absence	Abundance	Presence/Absence	Abundance
Species level	Abundance, occurrence probability, environmental coefficients	Variable importance (e.g. LIME, SHAP ²), Heatmap of environmental coefficients, <i>Response curves</i> ³ , <i>Variance partitioning</i>	Number of Presence/Absence, Proportion of occupied sites	Total abundance, site-specific abundance	<i>AUC</i> , Kappa, F1-Score	<i>RMSE</i> , MAE, R2, Correlation between predicted and observed values
α -diversity	Site-specific community composition	Diversity index (e.g. Shannon entropy, Simpson-Gini index), <i>Total abundance</i> , <i>Total richness</i> , Proportion of rare/abundant species <i>Pairwise dissimilarity</i> (e.g. Jaccard/Bray-Curtis) ^{4, 5}	Total Beta diversity, Turnover, Nestedness, Local Contribution to Beta Diversity (LCBD), Species Contribution to Beta Diversity (SCBD)		<i>Differences between predicted and observed values</i> , RMSE, MAE, R2, Correlations (e.g. Kendall, Pearson) between observed and predicted alpha or beta diversity indices	
β -diversity						
Overall assessment (all sites)	Regional community composition	Diversity index (e.g. Shannon entropy, Simpson-Gini index), Total abundance, Total richness, Proportion of rare/abundant species	Comparison with observed or reconstructed networks (expert-based or estimated e.g. based on trophic analyses), using e.g. correlations, <i>residual correlation index</i> (δ) ⁶		Average over all species : <i>AUC</i> , Kappa, F1-Score	Average over all species : <i>RMSE</i> , MAE, R2, Correlation <i>between predicted and observed values</i>
	Residual correlation matrix	Co-occurrence network analysis (e.g centrality, number of degrees)				
	Trait-based regression coefficients	<i>Traits-environment response curves</i> , Heatmap of traits-environment coefficients				

¹ All performance measures can theoretically be compared between models. For instance, we here measured differences between models using a measure of relative change in RMSE or AUC relative to the Bench model using Eq. 1. Other measures could be correlations between model predictions.

² See RYO et al. (2021)

³ To ease model comparison and interpretation, we propose to summarize the information contained in species response curves using the framework initially proposed by RIGAL et al. (2020) for classifying species temporal trajectories based on their trend, acceleration, direction and velocity. Applied to regression coefficients, it allows to classify the response of species to each environmental variable into several shapes that are easy to interpret, to link with ecological theory, and to compare across models.

⁴ For *j*SDM assessment, pairwise dissimilarities can be computed on the observed site-by-species matrix and on the predicted one. Comparing these values (e.g. through correlation analysis or simply through differences) will inform on how well the model reproduces/predict beta diversity patterns. Alternatively, pairwise dissimilarities can be computed between the observed taxa composition of a sample and its predicted one. These dissimilarities then become a metric to assess model performance based on species-composition predictions.

⁵ For *j*SDMs comparisons, pairwise dissimilarities computed between the observed taxa composition of a sample and its predicted one can be compared across models (e.g. through correlations) to assess to what extent differences between predicted and observed taxa composition are congruent across different models. Alternatively, comparing correlations between pairwise dissimilarities computed on the observed site-by-species matrix and on the predicted one will inform on which model best predict beta diversity patterns.

⁶ Species interaction networks can be reconstructed under certain conditions using the residual correlation matrices estimated by *j*SDM (see MOMAL et al. (2020)). The comparison between these reconstructed interaction networks and already known interaction networks (based on trophic data, experimental data, expert knowledge or qualitative information on species interactions) can serve as a means of model validation.

⁷ Comparing modelled species trait-environment responses (e.g., signs, shape of response curves) with expected responses (e.g. from theory, experiments or expert knowledge) can also serve to validate qualitatively the models.

2.4 Results

Both MCMC convergence and effective sample size of the different *jSDMs* were satisfactory (see Appendix D).

2.4.1 Model Fit & Predictive power

Species level

Presence/absence models showed excellent explanatory power with mean AUCs above 0.9 on the train dataset, but lower predictive power with mean AUCs around 0.66 on the test set (Fig. S17). Both explanatory (mean AUC between 0.92 and 0.93) and predictive (mean AUC between 0.64 and 0.66) power were overall similar across models (Fig. 2.2, Fig. S17). Within the target species assemblage, predictions improved for 41 species and worsened for 36 species (out of the 99 target species, which implies marginal changes for the remaining 22) in the *WhC* model relative to the benchmark. In comparison in the *Ph* or the *TrPh* models, predictions only improved for 26 and 27 species, respectively, and worsened for 49 and 48 species, respectively.

Abundance models also showed a satisfactory explanatory power with a mean RMSE close to nine for all models, given a mean abundance in the train dataset of 307.31 ± 583.58 (mean \pm sd). Overall, all models underpredicted species abundances (Fig. S18-19). While explanatory power was similar across models, larger variations were observed for predictive power. The *Bench* model had a mean RMSE of 126.67 (for a mean abundance in the test dataset of 700.57 ± 818.66 ; Fig S17). The *Ph* model performed better (mean RMSE of 62.23; -50.87% compared to the *Bench*; Fig. S17) whereas the *TrPh* model did worse (mean RMSE of 139.21; +9.90%; Fig. S17). The best model was the *WhC* with a mean RMSE of 6.59 (-94.80% compared to the *Bench*, Fig. S17). Out of the 99 target species, the *WhC* model predictions improved for 57 species but declined for 15 species relative to the *Bench*. Conversely, performance gain for the *Ph* and *TrPh* models were poor relative to the *Bench*, as predictions improved for 38 and 31 species, respectively, but declined for 40 and 46 species, respectively.

We further investigated this gain in predictive power of the *WhC* model fitted to abundance data by examining the relationships between changes in predictive power and the occurrence or abundance of the species. On the test set, performance of the *WhC* mo-

del most improved relative to the *Bench* model for rare species (correlation with average species abundance : Kendall's $\tau = 0.12$, p-value < 0.05 ; Fig. S20). However, we found no patterns between change in RMSE relative to the *Bench* model and proportion of presence (Kendall's $\tau = 0.12$, p-value = 0.09 ; Fig S21).

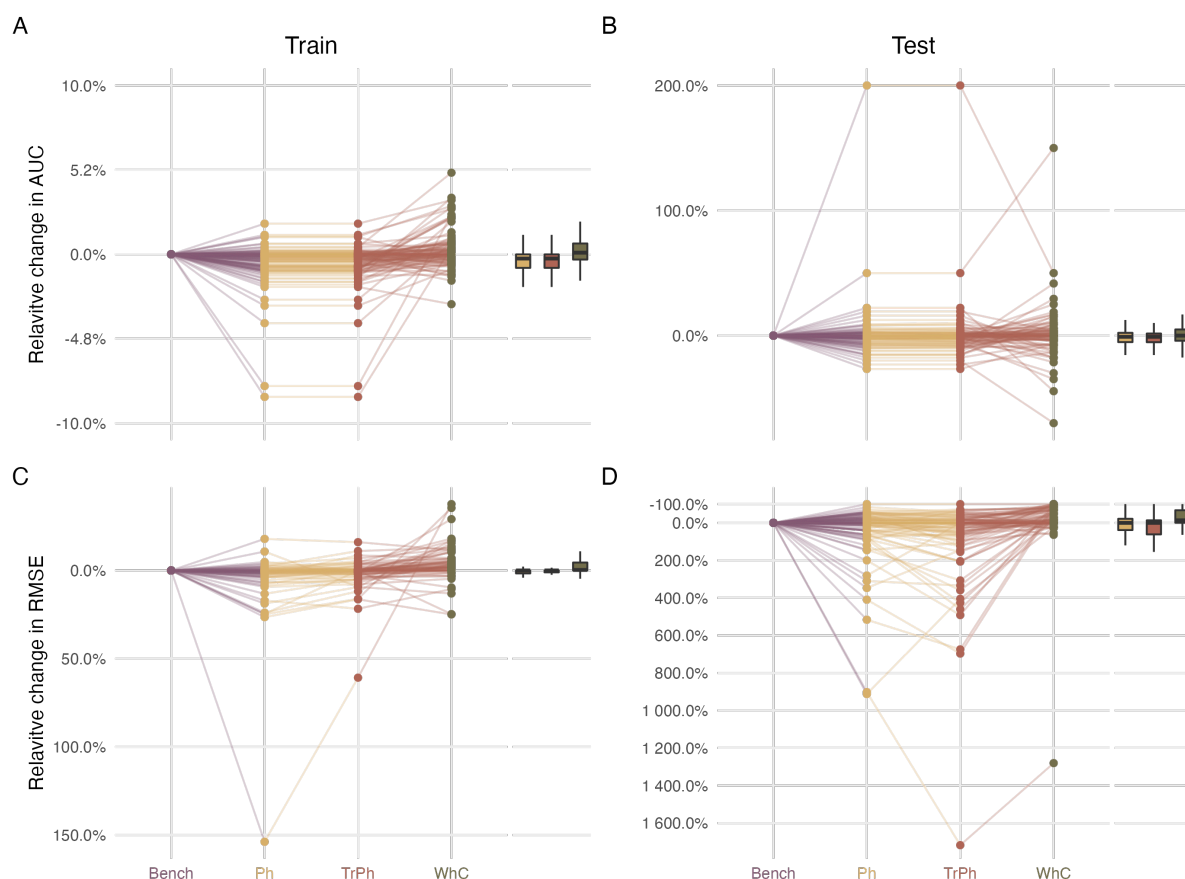


Figure 2.2 – Change in explanatory and predictive power of three model structures (yellow for Phylogeny (Ph), red for Traits and phylogeny (TrPh), and green for Whole community (WhC) models) relative to the Benchmark model (Bench ; purple). Changes are expressed as percentages relative to the benchmark fitted on presence/absence (top row) or abundance (bottom row) data. Points above the zero line indicate performance gain.

Community level

In terms of alpha diversity, the *Bench*, the *Ph* and *TrPh* models fitted on abundance data all underpredicted mean species richness of the train set by 4 species on average (-29.2% compared to observed data ; Fig. 2.3). In contrast, the *WhC* model overpredicted mean richness by 11 species on average (+80% compared with observed data). Similar

results were found on the test dataset with the *Bench*, *Ph* and *TrPh* models underpredicting richness by 5 species (-24.9 %) on average whereas the *WhC* model overpredicted richness by 7 species (+35.8% compared with observed data). Similar results were found for models fitted on presence/absence data (Fig. S22).

All models overall underpredicted mean total abundance relative to the train dataset (Fig. 2.3), by 153 individuals for the *Bench* model (-49.8% compared to observed data) and by 159 and 155 individuals (-51.7% and -50.4%) for the *Ph* and *TrPh* models, respectively. The *WhC* model only underpredicted total abundance by 22 individuals (-7.12% compared to observed data). Relative to the test dataset, the *Bench*, the *Ph* and the *TrPh* models overpredicted mean total abundance by 1642 (+234% compared to observations), 465 (+66.3%), and 1969 individuals (+281%), respectively. In contrast, the *WhC* model underpredicted mean total abundance of the test data samples by 404 individuals on average (-57.6%).

Mean beta diversity patterns in the train dataset were overall well captured by all models fitted on abundance or presence/absence data (Fig. 2.3). Observed dissimilarities were slightly overpredicted by all abundance models : by 0.057 for the *Bench* (+7.3% compared with observed data), 0.050 for the *Ph* (+6.4%), 0.054 for the *TrPh* (+6.9%) and 0.070 for the *WhC* models (+8.9%). Differences for presence/absence models were of similar order but all models underpredicted mean pairwise dissimilarities between samples (Fig. S22). On the test dataset, beta diversity patterns were rather poorly captured by the models fitted on abundance data. The *Bench* model overpredicted the pairwise dissimilarities by 0.364 on average (+67.1% compared with observed data), the *Ph* model by 0.365 (+67.4%), the *TrPh* model by 0.375 (+69.1%) and the *WhC* model by 0.338 (+62.4%). Similar results were observed for presence/absence models with slightly smaller overpredictions (Fig. S22).

2.4.2 Variance partitioning

The mean amount of total variance explained across the 99 polychaetes varied between 21 and 23% for models fitted with presence/absence data and between 18 and 30% for abundance-based models (Fig. S23). For all models, environmental variables, rather than random effects, accounted for most (more than 68% \pm 18% ; mean \pm sd) of the explained variance (Fig. S24). However, a larger part of variance is explained by random effects in

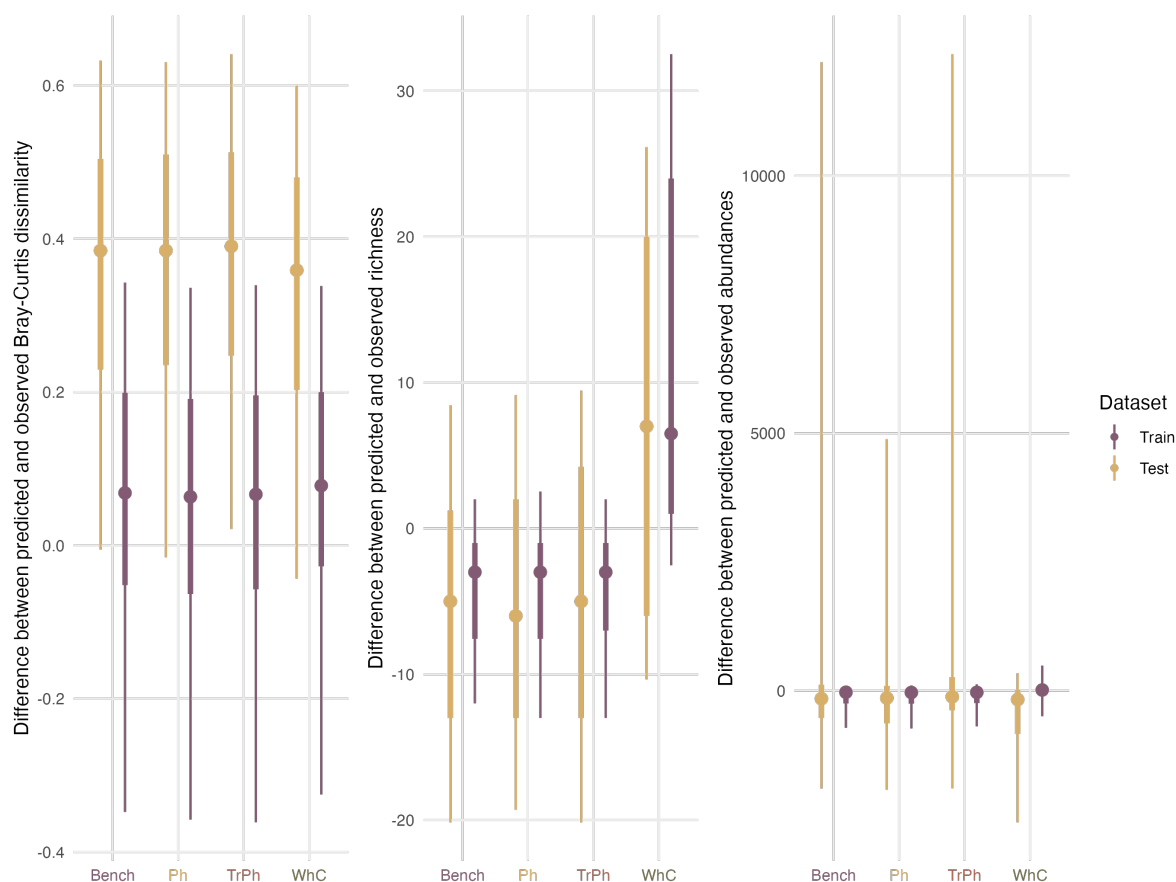


Figure 2.3 – Comparison of model performances with regards to their ability to predict community structures when fitted with abundance data for the train (purple) and test (yellow) dataset. Left : differences in pairwise dissimilarities estimated on the observed and the predicted assemblages. Centre : differences in species richness between observed and predicted assemblages. Right : differences in total abundance between observed and predicted assemblages.

the *WhC* model compared to the other models, including the *Bench* (Fig. S24). Compared to the *Bench* model fitted with abundance data, the relative change in the part of variance explained by random effects across the 99 species decreased by 17.00% for the *Ph* model, 10.90% for the *TrPh* model and increased by 224% for the *WhC* model (Fig. 2.4). Similar results with smaller relative changes were obtained across presence/absence models (Fig. 2.4; Fig. S23-24).

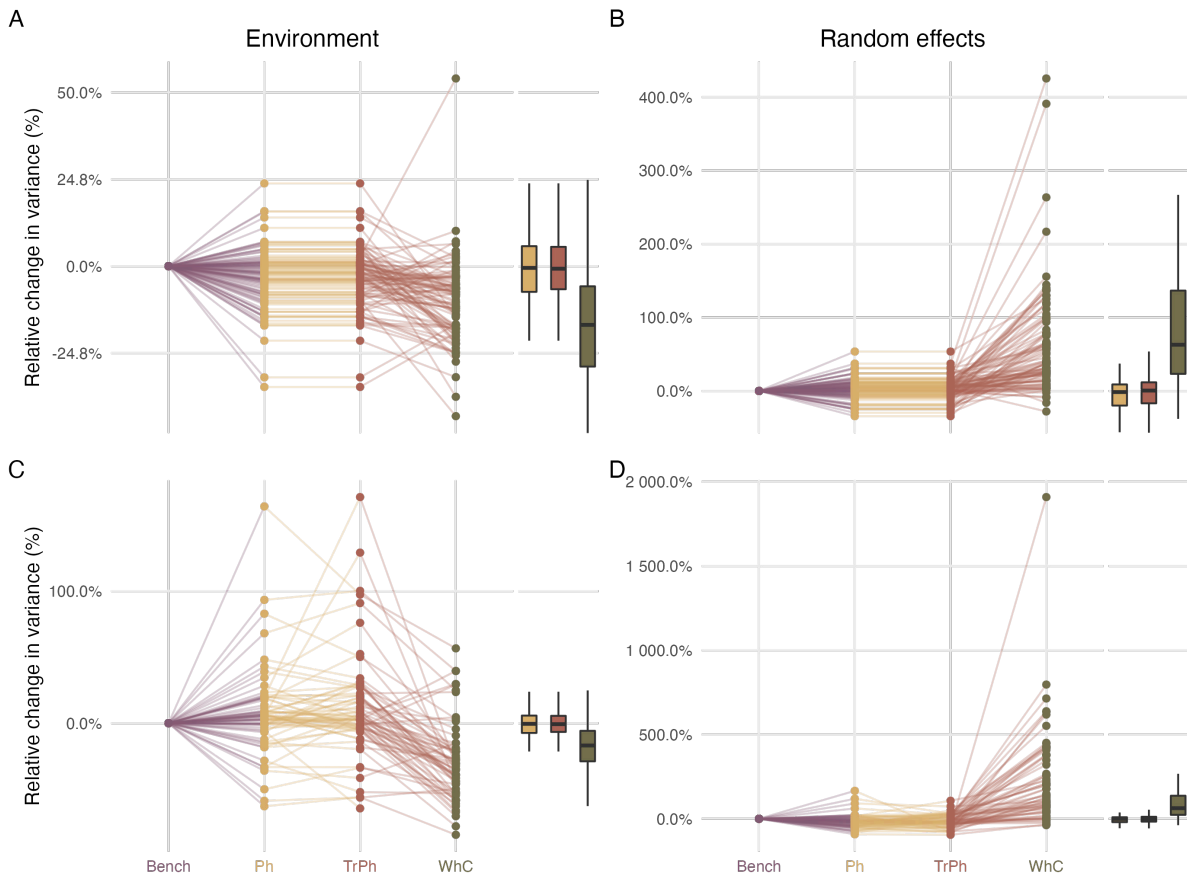


Figure 2.4 – Change in explained variance related to environmental predictors (left column) and random effects (right column) for three alternative model structures (yellow for Phylogeny (Ph), red for Traits and phylogeny (TrPh), and green for Whole community (WhC) models) relative to the Benchmark model (Bench; purple). Percentage changes were computed relative to the Benchmark model fitted with presence/absence (top panels) or abundance (bottom panels) data. Positive values indicate an increase in the proportion of variance explained by the focal model compared to the Benchmark model. See Figure S23 and S24 for the raw percentages, expressed as percentages of explained variance or total amount of variance respectively.

2.4.3 Species niche estimated

For abundance models, the large majority (>60%) of flat response curves indicated a lack of meaningful species-environment relationships (Fig. 2.5). This proportion reached 83% for the *WhC* model. The prevalence of flat relationships did not appear to be related to convergence issues (Fig. S15-16) or to be driven by a specific covariate (Fig. S25). Convex or concave response curves were rare in abundance models. Significant relationships primarily included constant or accelerated declines, representing approximately 10%

and 15% of response curves in the *Bench*, *TrPh*, and *Ph* models (Fig. 2.5). In the *WhC* model, these percentages decreased to 7% and 6%, respectively (Fig. 2.5). Similar findings were observed for presence/absence models (Fig. S26 ; Fig. S27).

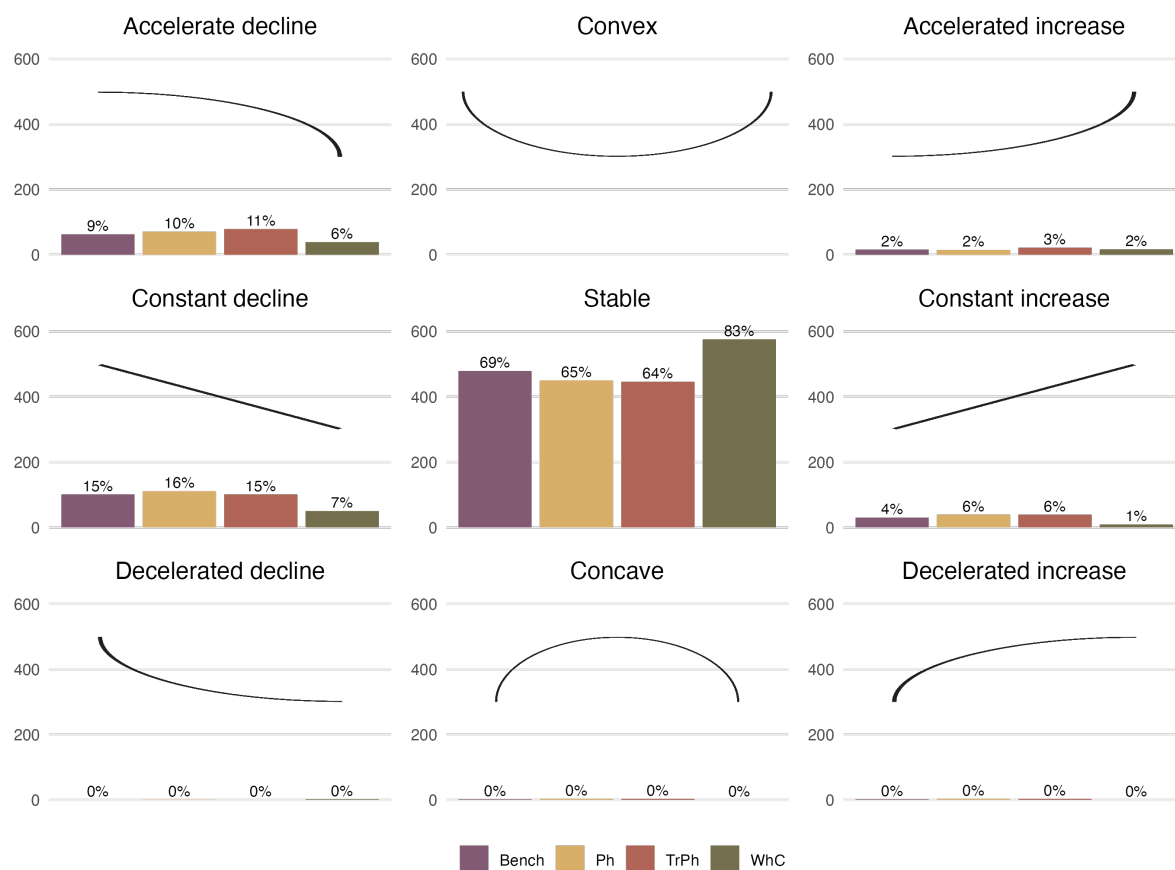


Figure 2.5 – Number (*y*-axis) and proportion (computed across all coefficients for each model, as indicated above individual bars) of response curves (*i.e.* one for each species-predictor combination according to the typology (nine shapes highlighted by the black curve in each panel) defined by RIGAL *et al.* (2020). Results are presented for the different model structures

Both abundance and presence/absence *TrPh* models (which include species functional traits) reveal some meaningful trait-environment relationships between the first fuzzy-PCA axis and the seven environmental predictors. This suggests that the occurrence of certain traits is likely favored (or hindered) under certain environmental conditions (Fig. S28). For instance, mobile predatory species showed larger declines in abundance as fetch increases than sessile suspensivores (Fig. S28). Moreover, increase in organic matter concentration and decrease in current velocities were associated with higher abundances

of suspensive feeders.

2.4.4 Exploring the residual correlation

Residual species-species correlations were compared between the *Bench* model and the *WhC* model, only for the 99 target species, using both presence/absence (Fig S29) and abundance data (Fig. 2.6). We only focus this comparison on the *WhC* model (rather than other models) because of its higher predictive performance and higher proportion of explained variance by random effects (Fig. 2.4). Residual correlations estimated from both models were highly correlated, both for presence/absence and abundance data (Fig. 2.6 and Fig. S29). However, agreement between models varied across different random effects from a moderate correlation between residuals associated with the Habitat random effects ($R^2 = 0.57$) or with the Site random effects ($R^2 = 0.64$), to a high correlation between residuals related to the Year random effects ($R^2 = 0.95$). The δ index main modal distribution, which is centered on zero, confirms an overall agreement between residual correlations estimated from both models in relation to the Year random effects with a marginal proportion of sign changes (0.45% of sign changes related to correlation greater than 0.25; Fig. 2.6 B) only related to low species-species residual correlations (<0.25 ; Fig. 2.6 A and Fig. S29). In contrast, the δ index highlights inconsistencies in both magnitude and signs changes between residuals associated with the Habitat and the Site (12.2% and 9.11% of sign changes related to correlation greater than 0.25) random effects. Similar results were obtained for presence/absence models (Fig. S29).

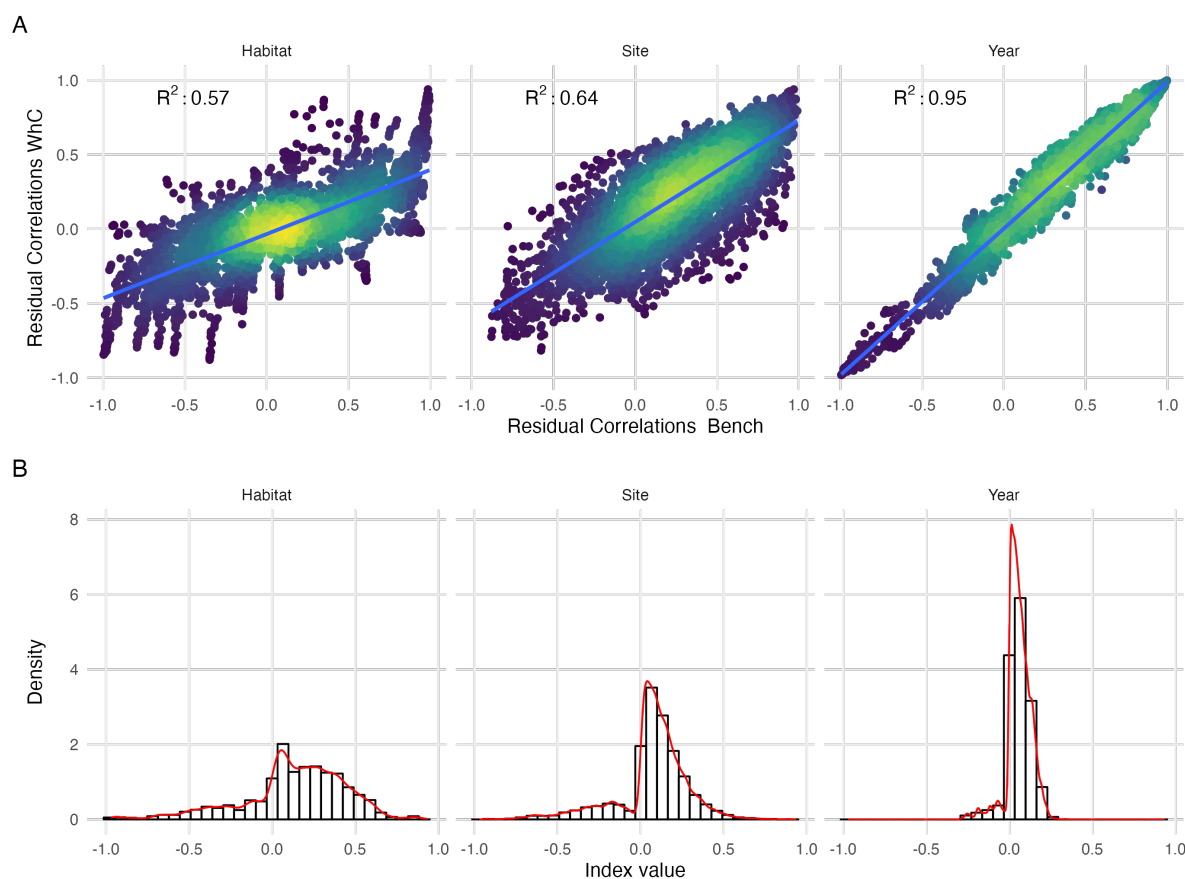


Figure 2.6 – (A) Comparison of residual species-species correlations associated with the three random effects estimated by the Whole Community Model (WhC ; y -axis) and the Benchmark model (Bench ; x -axis) fitted on abundance data. The colour scale highlights the density of points in each scatter plot. (B) Distribution of the δ index characterizing change in sign (negative values indicate sign change) and magnitude (higher absolute values indicate higher numerical difference) between residual correlations estimated by the WhC model and the Bench model adjusted with abundance data for the three random effects (Habitat, Site, Year).

2.5 Discussion

Case studies in community ecology typically rely on partial and heterogeneous observations (POLLOCK et al., 2020) but also on incomplete knowledge of target species ecological features (e.g. traits, phylogeny ; TYLER et al. (2012)). This study investigated how $jSDM$ performance varies depending on the type of information included (i.e. phylogeny, traits or data on non-target species) using a multi-assessment framework (spanning interpretability, inference and prediction, for both species- and community-level metrics, Table 1) enabling a thorough evaluation of model performance.

We found that *jSDMs*' performance, in particular predictive power of abundance models at the species level, mostly increased when including information related to the 179 non-target species sampled alongside with the 99 polychaetes of interest. However, improvement in species-level predictions does not directly translate into enhanced performance at the community level. The *WhC* model did not improve estimates of beta diversity or total abundance relative to the other models and largely overpredicted species richness, as previously suggested (ZURELL et al., 2018). Given *HMSC* hierarchical structure (POGGIATO et al., 2021), inclusion of monitoring data related to other species likely improves model performance for the target assemblage by capturing relevant drivers that are not explicitly considered. For instance, it can help describe target species' realized niches by accounting for ecological processes related to environmental conditions (including trait-mediated responses) or biotic interactions that are not explicitly captured otherwise (OVASKAINEN et al., 2017b). In our case, main differences between residual correlations estimated by the *Benchmark model* and the *Whole community* model relate to spatial random effects (i.e. site and habitat). In contrast, the temporal random effect yielded similar residual co-occurrences in both models. This suggests that including non-target species in our case, mostly helped capture spatial variability in species associations across sites and habitats.

Importantly, while we show that including non-target species can improve predictive performance, in particular for rare species, benefits of accounting for non-target species might vary depending on robustness of non-target species monitoring data (e.g. detection issues), their role within the ecosystem (e.g. engineer species are likely more influential on local communities than rare transient species), or processes shaping the target assemblage (if influence of abiotic factors dominates, then adding other species will have marginal consequences on model performance). While the list of additional species to consider can be prioritized based on existing knowledge in well-studied ecosystems, such information is often unavailable. Furthermore, a specific investigation, that might rely on simulated datasets to overcome limitations related to real world datasets (DIRENZO et al., 2022), would be required to determine specific criteria, as well as optimal number of non-target species to include. While species communities and assemblages are largely defined arbitrarily (STROUD et al., 2015), a systematic assessment of *jSDM* performance as increasing the number and types (for instance based on their functional or trophic roles) of non-target species would be valuable to optimise model performance for the species of management interests.

jSDMs have already been used to model the distribution of a wide variety of species ranging from micro-organisms (MINARD et al., 2019) to megafauna (BRIMACOMBE et al., 2021) inhabiting many different ecosystems. Here, while we studied assemblages associated with two specific coastal habitats, i.e. seagrass and sand, that have original characteristics as they are located at the land-sea interface (BOYÉ et al., 2019), our case study reflects typical aspects of applied ecological research. These include issues related to data limitation and availability but also typical features of ecological communities (e.g. prevalence of rare and transient species; MAGURRAN et HENDERSON (2003)); SNELL TAYLOR et al. (2018)). Valuable insights on trait-environment relationships are scarce in our study, which reflects how contributions of functional ecology in *jSDMs* are likely limited by trait data quality and availability (TYLER et al., 2012; de JUAN et al., 2022). For instance, we found an interaction between trophic modalities (i.e. microphagous versus macrophagous diet) and fetch (Fig. S15), indicating that organisms that filter on small particles are less likely to occur in wave-exposed sites where high levels of sediment resuspension can block their filtering systems (MANNING et al., 2014). Yet, the limited number of informative trait-environment relationships or species-environment relationships either suggest that neutral processes may shape polychaete assemblages (BOYÉ et al., 2019); or rather highlight a mismatch between trait data, environmental data, and the ecological processes at play (de JUAN et al., 2022). For instance, environmental variables only capture mean climatological conditions, but fail to quantify variability in the coastal environment, such as extreme events and seasonal or annual variability. Likewise, the list of available fuzzy-coded traits only partially captures species capacity to adapt to environmental variability (de JUAN et al., 2022). Thus, effectiveness of inclusion of traits in *jSDMs* is likely to be limited, or to rely on effort to collect relevant trait information. In our case, while including traits does not improve model predictive power, it somehow enhances our understanding of species responses along environmental gradients. Hence, if the goal is not prediction but inference (TREDENNICK et al., 2021), including traits and proxies of phylogeny can facilitate *jSDM* interpretation.

This paper lays out an original framework to systematically compare multiple facets of alternative *jSDM* formulations (i.e. including phylogeny, traits or additional species) on model interpretability, explanatory and predictive power (Table 1). Using a set of complementary metrics, we specifically assess performance of alternative model formulations fitted to presence-absence or abundance data at the species and community levels. Our framework goes beyond existing guidelines proposed to assess the performance of *jSDM*

fitted on presence-absence data (WILKINSON et al., 2021) or that focus on the predictive power of abundance-based models (e.g. WALDOCK et al. (2022)). It specifically compares the performance (both explanatory and predictive) and interpretability of alternative models' formulations accounting for the multiple and high-dimensional components that are typical of *jSDMs*, namely : (1) species and community level predictions including alpha and beta diversity metrics and ranking of predictions according to species prevalence/abundance ; (2) species-environment relationships where we transposed the framework initially developed for time series by RIGAL et al. (2020) into an effective tool to classify response curves according to 9 categories across high numbers of species (e.g. 99 in our case study) ; (3) trait-environment relationships ; and (4) residual species-species correlations associated with random effects thanks to a new index that summarizes both changes in the sign and magnitude of the residual correlations.

Overall, our results provide new insights into the most appropriate strategies for *jSDM* fitting, according to modelling objectives (TREDENNICK et al., 2021) and available data. While the four models considered had similar explanatory power, adding extra information to standard *jSDMs* that only consider abiotic predictors can prove useful in cases. For instance, adding monitoring data for other non-target species can substantially increase model predictive power by modifying inferred species-environment relationships and residual correlation matrices. Similarly, adding traits or phylogeny can improve model interpretability. Future studies will be key to consolidate our findings on simulated case studies (ZURELL et al., 2010 ; DIRENZO et al., 2022), or across contrasted ecosystems, for instance dominated either by environmental filtering, or by competitive processes. Generalizing this approach across ecosystems will further help prioritize data collection effort on the long term. For this purpose, we recommend using a multi-model inference framework similar to the one used in this study to systematically assess trade-offs associated with alternative *jSDMs* formulations.

Bibliographie

- ABREGO, N., BÄSSLER, C., CHRISTENSEN, M., & HEILMANN-CLAUSEN, J., (2022), Traits and phylogenies modulate the environmental responses of wood-inhabiting fungal communities across spatial scales, *Journal of Ecology*, *1104*, 784-798.
- ABREGO, N., DUNSON, D., HALME, P., SALCEDO, I., & OVASKAINEN, O., (2017), Wood-inhabiting fungi with tight associations with other species have declined as a response to forest management, *Oikos*, *1262*.
- BLANCHET, F. G., CAZELLES, K., & GRAVEL, D., (2020), Co-occurrence is not evidence of ecological interactions, *Ecology Letters*, *237*, 1050-1063.
- BOYÉ, A., GAUTHIER, O., BECHELER, R., LE GARREC, V., HILY, C., MAGUER, M., & GRALL, J., (2022), Drivers and limits of phenotypic responses in vulnerable seagrass populations : *Zostera marina* in the intertidal, *Journal of Ecology*, *1101*, 144-161.
- BOYÉ, A., LEGENDRE, P., GRALL, J., & GAUTHIER, O., (2017), Constancy despite variability : Local and regional macrofaunal diversity in intertidal seagrass beds, *Journal of Sea Research*, *130*, 107-122.
- BOYÉ, A., THIÉBAUT, É., GRALL, J., LEGENDRE, P., BROUDIN, C., HOUBIN, C., GARREC, V. L., MAGUER, M., DROUAL, G., & GAUTHIER, O., (2019), Trait-based approach to monitoring marine benthic data along 500 km of coastline, *Diversity and Distributions*, *2512*, 1879-1896.
- BRIMACOMBE, C., BODNER, K., & FORTIN, M.-J., (2021), Inferred seasonal interaction rewiring of a freshwater stream fish network, *Ecography*, *442*, 219-230.
- BRUDVIG, L. A., & CATANO, C. P., (2022), Prediction and uncertainty in restoration science, *Restoration Ecology*, e13380.
- CHIQUET, J., MARIADASSOU, M., & ROBIN, S., (2021), The Poisson-Lognormal Model as a Versatile Framework for the Joint Analysis of Species Abundances, *Frontiers in Ecology and Evolution*, *9*.
- de JUAN, S., BREMNER, J., HEWITT, J., TÖRNROOS, A., MANGANO, M. C., THRUSH, S., & HINZ, H., (2022), Biological traits approaches in benthic marine ecology : Dead ends and new paths, *Ecology and Evolution*, *126*, e9001.
- DIETZE, M. C., FOX, A., BECK-JOHNSON, L. M., BETANCOURT, J. L., HOOTEN, M. B., JARNEVICH, C. S., KEITT, T. H., KENNEY, M. A., LANEY, C. M., LARSEN, L. G., LOESCHER, H. W., LUNCH, C. K., PIJANOWSKI, B. C., RANDERSON, J. T., READ, E. K., TREDENNICK, A. T., VARGAS, R., WEATHERS, K. C., & WHITE, E. P., (2018), Iterative near-term ecological forecasting : Needs, opportunities, and challenges, *Proceedings of the National Academy of Sciences*, *1157*, 1424-1432.
- DIRENZO, G. V., HANKS, E., & MILLER, D. A. W., (2022), A practical guide to understanding and validating complex models using data simulations, *Methods in Ecology and Evolution*.
- DORMANN, C. F., BOBROWSKI, M., DEHLING, D. M., HARRIS, D. J., HARTIG, F., LISCHKE, H., MORETTI, M. D., PAGEL, J., PINKERT, S., SCHLEUNING, M., SCHMIDT, S. I., SHEPPARD, C. S., STEINBAUER, M. J., ZEUSS, D., & KRAAN, C., (2018), Biotic interactions in species distribution modelling : 10 questions to guide interpretation and avoid false conclusions, *Global Ecology and Biogeography*, *279*, 1004-1016.
- ELITH, J., H. GRAHAM, C., P. ANDERSON, R., DUDÍK, M., FERRIER, S., GUISAN, A., J. HIJMANS, R., HUETTMANN, F., R. LEATHWICK, J., LEHMANN, A., LI, J., G. LOHMANN, L., A. LOISELLE, B.,

- MANION, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., MCC. M. OVERTON, J., TOWNSEND PETERSON, A., ... E. ZIMMERMANN, N., (2006), Novel methods improve prediction of species' distributions from occurrence data, *Ecography*, 292, 129-151.
- GELMAN, A., & RUBIN, D. B., (1992), Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, 74, 457-472.
- GIANGRANDE, A., LICCIANO, M., & MUSCO, L., (2005), Polychaetes as environmental indicators revisited, *Marine Pollution Bulletin*, 5011, 1153-1162.
- GIANNINI, T. C., CHAPMAN, D. S., SARAIVA, A. M., ALVES-DOS-SANTOS, I., & BIESMEIJER, J. C., (2013), Improving species distribution models using biotic interactions : a case study of parasites, pollinators and plants, *Ecography*, 366, 649-656.
- GODSOE, W., FRANKLIN, J., & BLANCHET, F. G., (2017), Effects of biotic interactions on modeled species' distribution can be masked by environmental gradients, *Ecology and Evolution*, 72, 654-664.
- GRALL, J., LE LOC'H, F., GUYONNET, B., & RIERA, P., (2006), Community structure and food web based on stable isotopes (^{15}N and ^{13}C) analysis of a North Eastern Atlantic maerl bed, *Journal of Experimental Marine Biology and Ecology*, 3381, 1-15.
- HÄKKILÄ, M., ABREGO, N., OVASKAINEN, O., & MÖNKKÖNEN, M., (2018), Habitat quality is more important than matrix quality for bird communities in protected areas, *Ecology and Evolution*, 88, 4019-4030.
- HOLT, R. D., (2020), Some thoughts about the challenge of inferring ecological interactions from spatial data., *Biodiversity Informatics*, 151, 61-66.
- 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*, 56, 506-513.
- HUI, F. K., (2016), boral – Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r, *Methods in Ecology and Evolution*, 76, 744-750.
- IVES, A. R., & HELMUS, M. R., (2011), Generalized linear mixed models for phylogenetic analyses of community structure, *Ecological Monographs*, 813, 511-525.
- JANKOWSKA, E., MICHEL, L. N., LEPOINT, G., & WŁODARSKA-KOWALCZUK, M., (2019), Stabilizing effects of seagrass meadows on coastal water benthic food webs, *Journal of Experimental Marine Biology and Ecology*, 510, 54-63.
- JUMARS, P. A., DORGAN, K. M., & LINDSAY, S. M., (2015), Diet of Worms Emended : An Update of Polychaete Feeding Guilds, *Annual Review of Marine Science*, 71, 497-520.
- KEIL, P., WIEGAND, T., TÓTH, A. B., MCGLINN, D. J., & CHASE, J. M., (2021), Measurement and analysis of interspecific spatial associations as a facet of biodiversity, *Ecological Monographs*, 913, e01452.
- MAGURRAN, A. E., & HENDERSON, P. A., (2003), Explaining the excess of rare species in natural species abundance distributions, *Nature*, 4226933, 714-716.
- MANNING, L. M., PETERSON, C. H., & BISHOP, M. J., (2014), Dominant macrobenthic populations experience sustained impacts from annual disposal of fine sediments on sandy beaches, *Marine Ecology Progress Series*, 508, 1-15.

- MINARD, G., TIKHONOV, G., OVASKAINEN, O., & SAASTAMOINEN, M., (2019), The microbiome of the *Melitaea cinxia* butterfly shows marked variation but is only little explained by the traits of the butterfly or its host plant, *Environmental Microbiology*, 2111, 4253-4269.
- MOMAL, R., ROBIN, S., & AMBROISE, C., (2020), Tree-based inference of species interaction networks from abundance data, *Methods in Ecology and Evolution*, 115, 621-632.
- MOMAL, R., ROBIN, S., & AMBROISE, C., (2021), Accounting for missing actors in interaction network inference from abundance data, *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 705, 1230-1258.
- 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*, 266, 740-752.
- NIKU, J., HUI, F. K. C., TASKINEN, S., & WARTON, D. I., (2019), gllvm : Fast analysis of multivariate abundance data with generalized linear latent variable models in r, *Methods in Ecology and Evolution*, 1012, 2173-2182.
- NORBERG, A., ABREGO, N., BLANCHET, F. G., ADLER, F. R., ANDERSON, B. J., ANTTILA, J., ARAÚJO, M. B., DALLAS, T., DUNSON, D., ELITH, J., FOSTER, S. D., FOX, R., FRANKLIN, J., GODSOE, W., GUISAN, A., O'HARA, B., HILL, N. A., HOLT, R. D., HUI, F. K., ... OVASKAINEN, O., (2019), A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels, *Ecological Monographs*, e01370.
- OVASKAINEN, O., & ABREGO, N., (2020), *Joint Species Distribution Modelling : With Applications in R*, Cambridge University Press.
- OVASKAINEN, O., TIKHONOV, G., DUNSON, D., GRÖTAN, V., ENGEN, S., SÆTHER, B.-E., & ABREGO, N., (2017a), How are species interactions structured in species-rich communities? A new method for analysing time-series data, *Proceedings of the Royal Society B : Biological Sciences*, 2841855, 20170768.
- OVASKAINEN, O., TIKHONOV, G., NORBERG, A., GUILLAUME BLANCHET, F., DUAN, L., DUNSON, D., ROSLIN, T., & ABREGO, N., (2017b), How to make more out of community data? A conceptual framework and its implementation as models and software, *Ecology Letters*, 205, 561-576.
- PARADIS, E., & SCHLIEP, K., (2019), ape 5.0 : an environment for modern phylogenetics and evolutionary analyses in R, *Bioinformatics*, 353, 526-528.
- POGGIATO, G., MÜNKEMÜLLER, T., BYSTROVA, D., ARBEL, J., CLARK, J. S., & THUILLER, W., (2021), On the Interpretations of Joint Modeling in Community Ecology, *Trends in Ecology & Evolution*.
- POLLOCK, L. J., MORRIS, W. K., & VESK, P. A., (2012), The role of functional traits in species distributions revealed through a hierarchical model, *Ecography*, 358, 716-725.
- POLLOCK, L. J., O'CONNOR, L. M., MOKANY, K., ROSAUER, D. F., TALLUTO, M. V., & THUILLER, W., (2020), Protecting Biodiversity (in All Its Complexity) : New Models and Methods, *Trends in Ecology & Evolution*, 3512, 1119-1128.
- POLLOCK, L. J., TINGLEY, R., MORRIS, W. K., GOLDING, N., O'HARA, R. B., PARRIS, K. M., VESK, P. A., & MCCARTHY, M. A., (2014), Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM), *Methods in Ecology and Evolution*, 55, 397-406.

- RIGAL, S., DEVICTOR, V., & DAKOS, V., (2020), A method for classifying and comparing non-linear trajectories of ecological variables, *Ecological Indicators*, 112, 106113.
- RYO, M., ANGELOV, B., MAMMOLA, S., KASS, J. M., BENITO, B. M., & HARTIG, F., (2021), Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models, *Ecography*, 442, 199-205.
- SNELL TAYLOR, S. J., EVANS, B. S., WHITE, E. P., & HURLBERT, A. H., (2018), The prevalence and impact of transient species in ecological communities, *Ecology*, 998, 1825-1835.
- STANICZENKO, P. P., SIVASUBRAMANIAM, P., SUTTLE, K. B., & PEARSON, R. G., (2017), Linking macroecology and community ecology : refining predictions of species distributions using biotic interaction networks, *Ecology Letters*, 206, 693-707.
- STROUD, J. T., BUSH, M. R., LADD, M. C., NOWICKI, R. J., SHANTZ, A. A., & SWEATMAN, J., (2015), Is a community still a community? Reviewing definitions of key terms in community ecology, *Ecology and Evolution*, 521, 4757-4765.
- THIOULOUSE, J., DRAY, S., DUFOUR, A.-B., SIBERCHICOT, A., JOMBART, T., & PAVOINE, S., (2018), *Multivariate Analysis of Ecological Data with ade4*, Springer.
- TOUMI, C., DE CÁCERES, M., GRALL, J., BOYÉ, A., THIÉBAUT, É., MAGUER, M., LE GARREC, V., BROUDIN, C., HOUBIN, C., & GAUTHIER, O., (2023), Long-term coastal macrobenthic Community Trajectory Analysis reveals habitat-dependent stability patterns, *Ecography*, e06489.
- TREDENNICK, A. T., HOOKER, G., ELLNER, S. P., & ADLER, P. B., (2021), A practical guide to selecting models for exploration, inference, and prediction in ecology, *Ecology*, 1026, e03336.
- TYLER, E. H. M., SOMERFIELD, P. J., BERGHE, E. V., BREMNER, J., JACKSON, E., LANGMEAD, O., PALOMARES, M. L. D., & WEBB, T. J., (2012), Extensive gaps and biases in our knowledge of a well-known fauna : implications for integrating biological traits into macroecology, *Global Ecology and Biogeography*, 219, 922-934.
- VESK, P. A., MORRIS, W. K., NEAL, W. C., MOKANY, K., & POLLOCK, L. J., (2021), Transferability of trait-based species distribution models, *Ecography*, 441, 134-147.
- VIOLET, C., (2020), *Caractériser la contribution des facteurs environnementaux et des interactions biotiques à la variabilité des assemblages benthiques* [mém. de mast., Rennes 1].
- WALDOCK, C., STUART-SMITH, R. D., ALBOUY, C., CHEUNG, W. W. L., EDGAR, G. J., MOUILLOT, D., TJIPUTRA, J., & PELLISSIER, L., (2022), A quantitative review of abundance-based species distribution models, *Ecography*, 20221.
- WARTON, D. I., BLANCHET, F. G., O'HARA, R. B., OVASKAINEN, O., TASKINEN, S., WALKER, S. C., & HUI, F. K., (2015), So many variables : joint modeling in community ecology, *Trends in Ecology & Evolution*, 3012, 766-779.
- WIENS, J. J., ACKERLY, D. D., ALLEN, A. P., ANACKER, B. L., BUCKLEY, L. B., CORNELL, H. V., DAMSCHEN, E. I., JONATHAN DAVIES, T., GRYTNES, J.-A., HARRISON, S. P., et al., (2010), Niche conservatism as an emerging principle in ecology and conservation biology, *Ecology letters*, 1310, 1310-1324.
- WILKINSON, D. P., GOLDING, N., GUILLERA-ARROITA, G., TINGLEY, R., & MCCARTHY, M. A., (2019), A comparison of joint species distribution models for presence-absence data, *Methods in Ecology and Evolution*, 102, 198-211.

- WILKINSON, D. P., GOLDING, N., GUILLERA-ARROITA, G., TINGLEY, R., & MCCARTHY, M. A., (2021), Defining and evaluating predictions of joint species distribution models, *Methods in Ecology and Evolution*, *123*, 394-404.
- ZURELL, D., BERGER, U., CABRAL, J. S., JELTSCH, F., MEYNARD, C. N., MÜNKEMÜLLER, T., NEHRBASS, N., PAGEL, J., REINEKING, B., SCHRÖDER, B., & GRIMM, V., (2010), The virtual ecologist approach : simulating data and observers, *Oikos*, *1194*, 622-635.
- ZURELL, D., POLLOCK, L. J., & THUILLER, W., (2018), Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments?, *Ecography*, *4111*, 1812-1819.

Appendix A - Data Sources and Descriptions of the Datasets

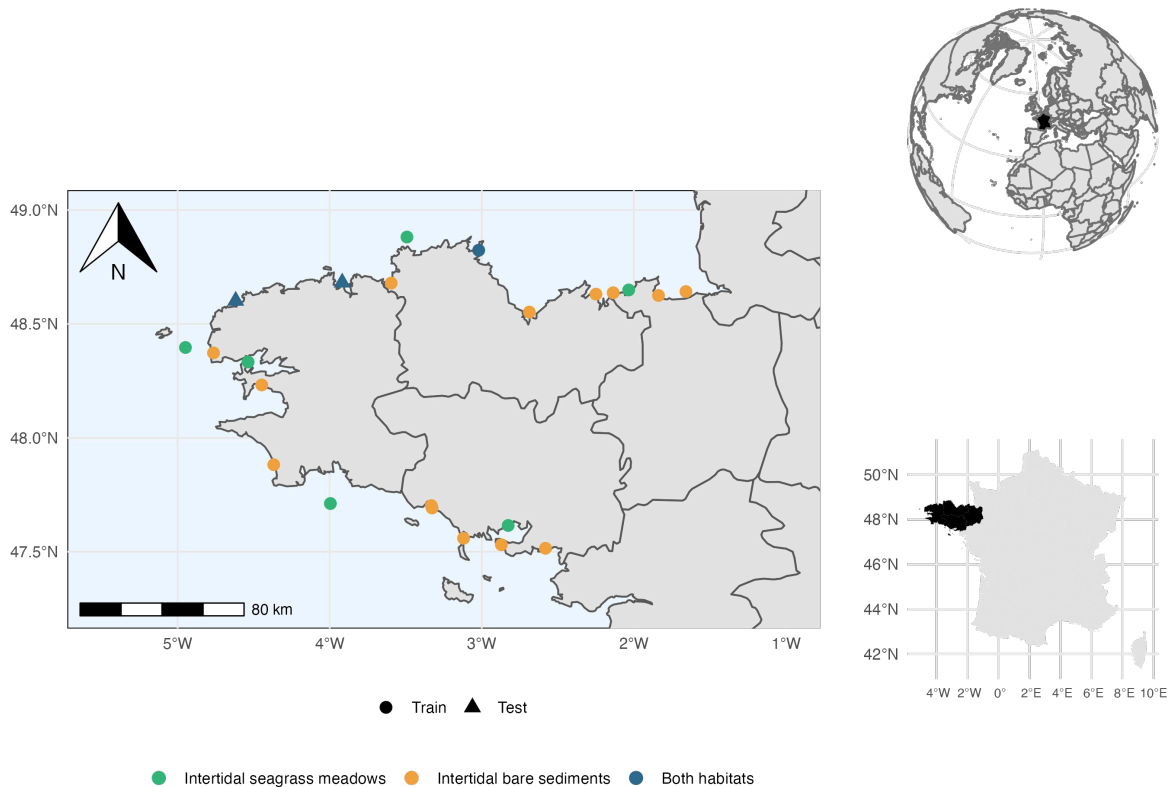


Figure S1 – Map of the sampled sites. Point shapes vary according to their contribution to model training set (circles; used to evaluate model explanatory power) as opposed to the two sites retained for independent model testing (triangles; used to evaluate model predictive power). Point colours vary according to the presence or absence of the two habitats in each site. The two test sites include the two habitats (*i.e.* seagrass and bare sand) and were chosen because they occur in environmental conditions that can be considered average at the scale of the region (thereby limiting extrapolation of the model) but still harbour different communities, representative of the known diversity gradient across the region.

Environmental data acquisition

The models used in this study incorporate a dataset consisting of seven environmental variables related to oceanography, hydrography, and granulometry obtained from BOYÉ (2019). The oceanographic variables include the standard deviation of salinity, surface water temperature, mean velocity of currents, and fetch, which were obtained from the

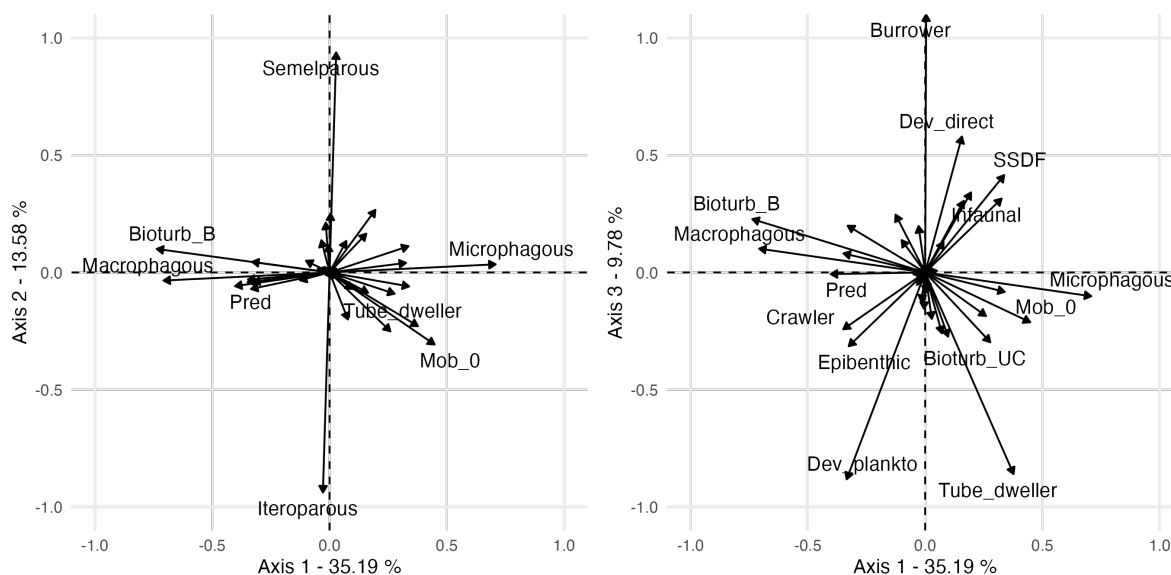


Figure S2 – Fuzzy PCA of the species-by-trait matrix. The first three axes represent 58.55% of the total variance. The first axis distinguishes sessile microphagous species (top positive values) from mobile macrophagous predatory species (bottom negative values). The second axis is a gradient of reproductive strategies (semelparous vs iteroparous). The third axis distinguishes burrowers with direct development from tube-dwellers with planktonic development. For abbreviations and meaning of the trait modalities, see BOYÉ et al. (2019).

PREVIMER database (LECORNU & ROECK, 2009) based on the MARS3D model (LAZURE & DUMAS, 2008). The variables were averaged by extracting daily data for the sampled year at the site coordinates and the eight adjacent cells. The fetch was calculated as the average length of nine radiating fetch segments with a maximum distance of 300km. The granulometry variables were derived from sediment cores that were taken along with associated fauna. The cores were dried, separated into 15 fractions, and the Trask index was calculated as the ratio of the 25th to 75th percentile of the grain distribution. Organic matter mass was estimated through the loss of mass after combustion in an oven.

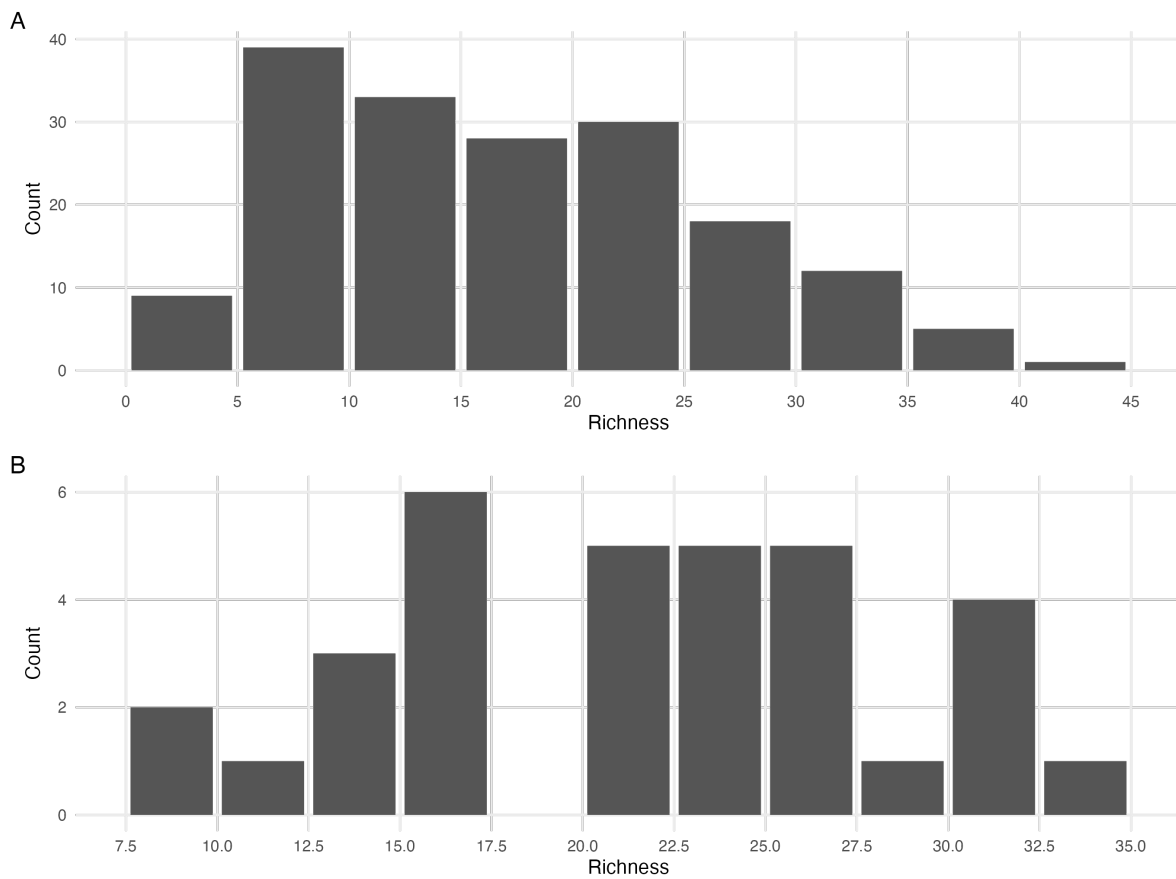


Figure S3 – A. Distribution of the number of species in the samples (site times habitat times year) of the train dataset. B. Distribution of the number of species in the samples (site times habitat times year) of the test dataset.

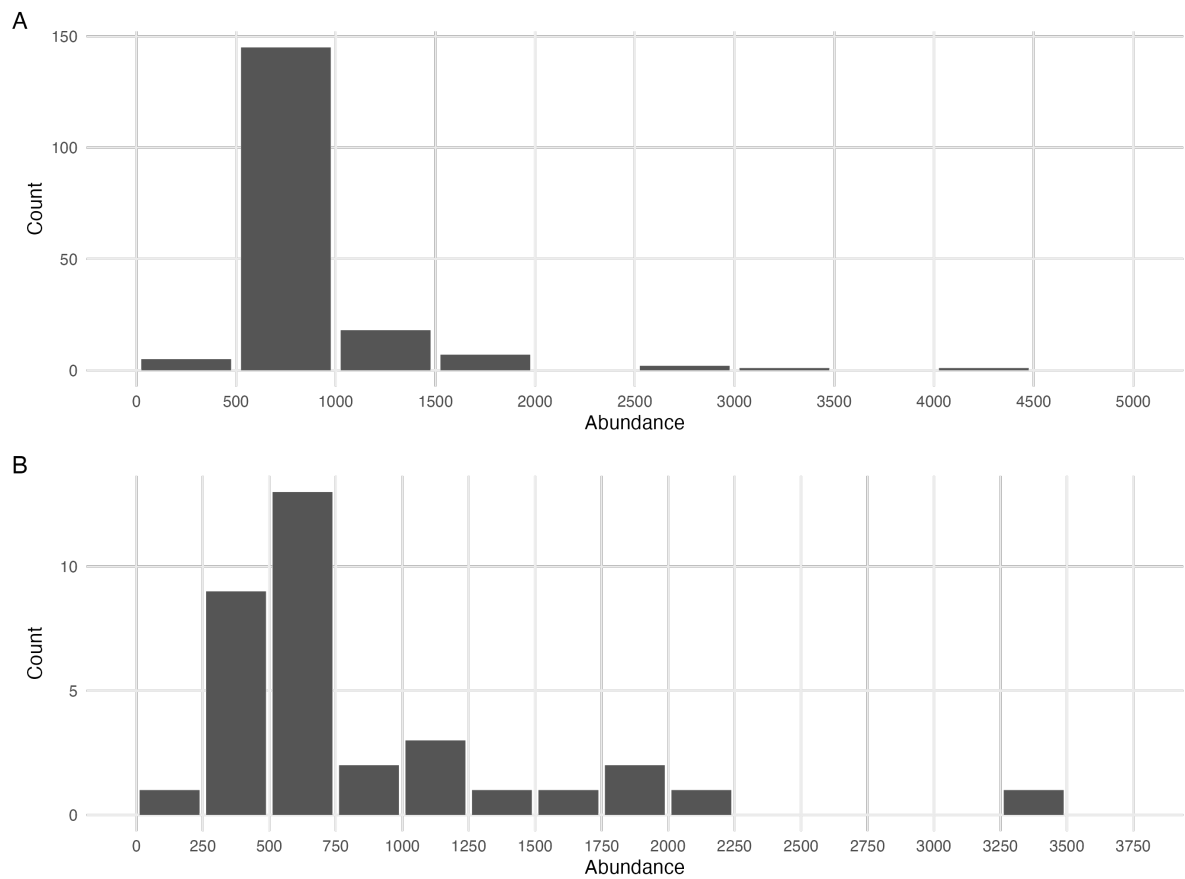


Figure S4 – A. Distribution of the total number of individuals (abundance) in the samples (site times habitat times year) of the train dataset. B. Distribution of the total number of individuals (abundance) in the samples (site times habitat times year) of the test dataset.

Appendix B - Model Convergence

Environmental coefficients

Table S1 – Potential scale reduction factors (PSRF) and effective sample sizes (ESS) for environmental regression parameters (i.e beta coefficients) estimated for the four different models (Bench, Ph, TrPh, WhC) fitted either to abundance or presence-absence data. For further details see Fig. S5 to Fig. S12.

Model	Data Type	Number of coefficients	PSRF (mean \pm sd)	ESS (mean \pm sd)
Benchmark	Abundance	1485	1.18 \pm 0.267	701 \pm 576
Benchmark	Presence/Absence	1485	1.00 \pm 0.002	4967 \pm 417
Phylogeny	Abundance	1485	1.18 \pm 0.204	566 \pm 420
Phylogeny	Presence/Absence	1485	1.00 \pm 0.001	4947 \pm 408
Traits & Phylogeny	Abundance	1485	1.21 \pm 0.317	489 \pm 358
Traits & Phylogeny	Presence/Absence	1485	1.00 \pm 0.008	11459 \pm 2649
Whole Community	Abundance	4170	1.21 \pm 0.287	739 \pm 631
Whole Community	Presence/Absence	4170	1.00 \pm 0.002	4962 \pm 406

Traits coefficients

Table S2 – Potential scale reduction factors (PSRF) and effective sample sizes (ESS) for traits regression parameters (i.e gamma coefficients) estimated for the model including trait information fitted either to abundance or presence-absence data. For further details see Fig. S13 to Fig. S14.

Model	Data type	Number of coefficients	PSRF (mean \pm sd)	ESS (mean \pm sd)
Traits & Phylogeny	Abundance	60	1.08 \pm 0.092	1232 \pm 1209

Model	Data type	Number of coefficients	PSRF (mean \pm sd)	ESS (mean \pm sd)
Traits & Phylogeny	Presence/Absence	60	1.00 \pm 0.001	13227 \pm 1897

Phylogeny coefficients

Table S3 – Potential scale reduction factors (PSRF) and effective sample sizes (ESS) for rho regression parameters (i.e phylogeny coefficient) estimated for the two models including phylogenetic information (TrPh and Ph).

Model	Data type	Number of coefficients	PSRF	ESS
Phylogeny	Abundance	1	1.07	649
Phylogeny	Presence/Absence	1	1.00	9349
Traits & Phylogeny	Abundance	1	1.15	757
Traits & Phylogeny	Presence/Absence	1	1.00	5000

Link between model convergence and species response curves

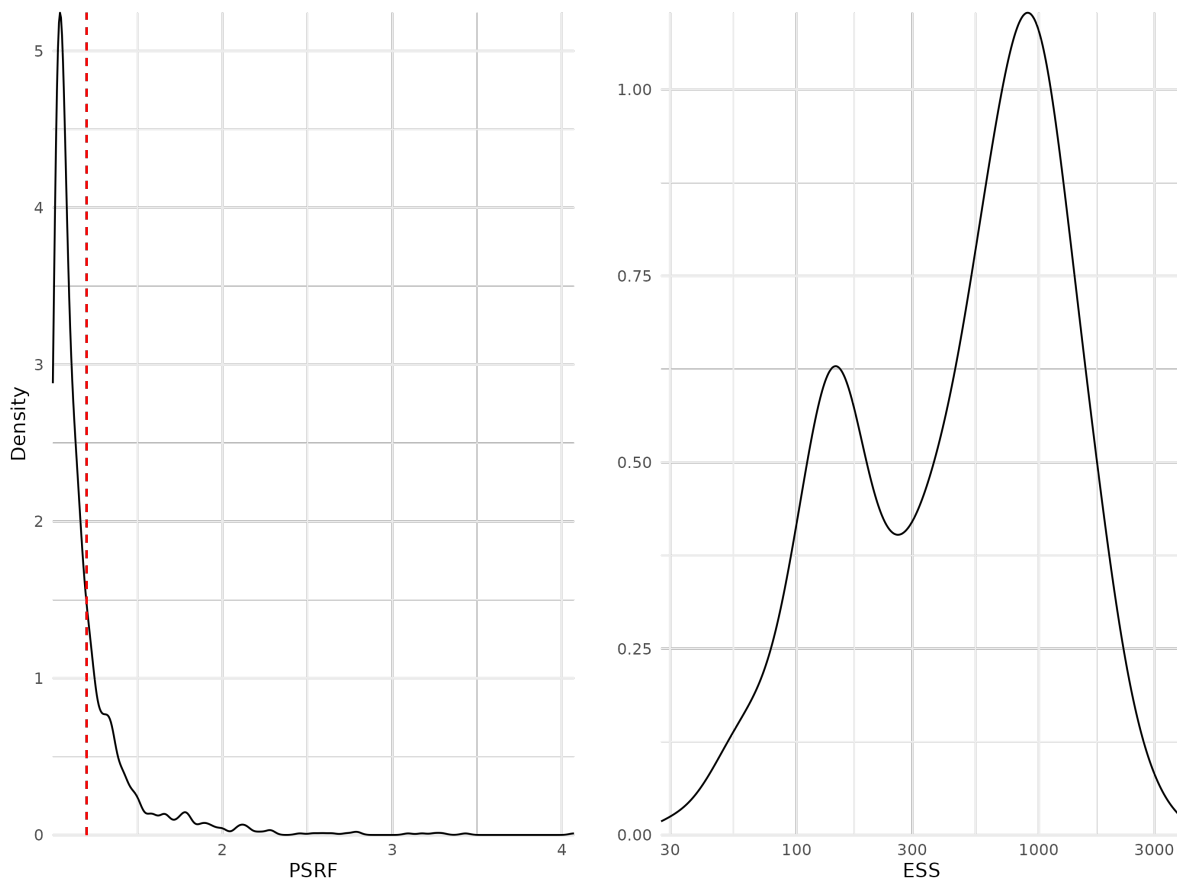


Figure S5 – Density curves of potential scale reduction factors (PSRF see BROOKS et GELMAN (1998); left panel) and effective sample sizes (ESS; right panel) for Beta regression parameters (i.e. environmental coefficients) estimated for the benchmark model fitted with abundance data. For PSRF, values greater than 1.2 (dotted red line) indicate potential convergence issues. ESS estimates the number of independent samples used to estimate each parameter (the more the better).

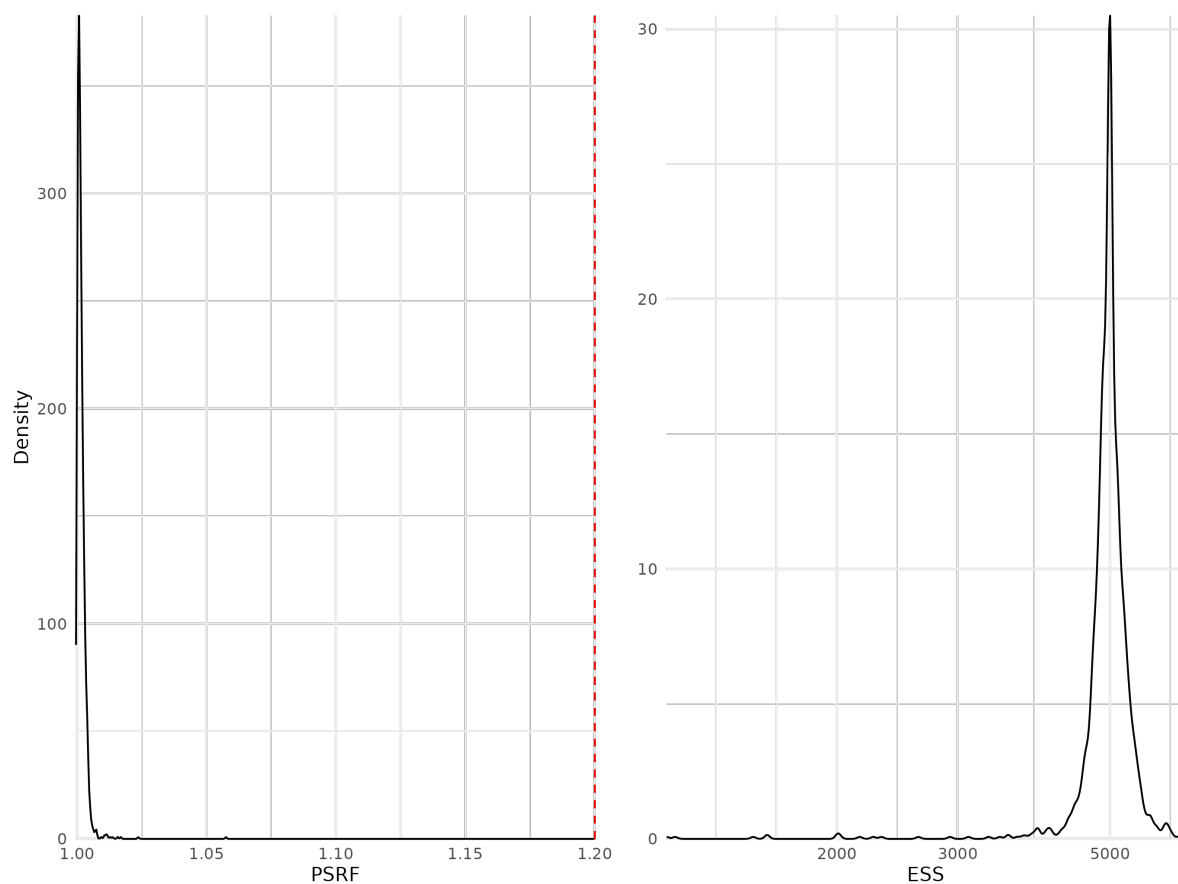


Figure S6 – Density curves of potential scale reduction factors (PSRF see BROOKS et GELMAN (1998); left panel) and effective sample sizes (ESS; right panel) for Beta regression parameters (i.e. environmental coefficients) estimated for the benchmark model fitted with presence/absence data. For further details see Fig. S5.

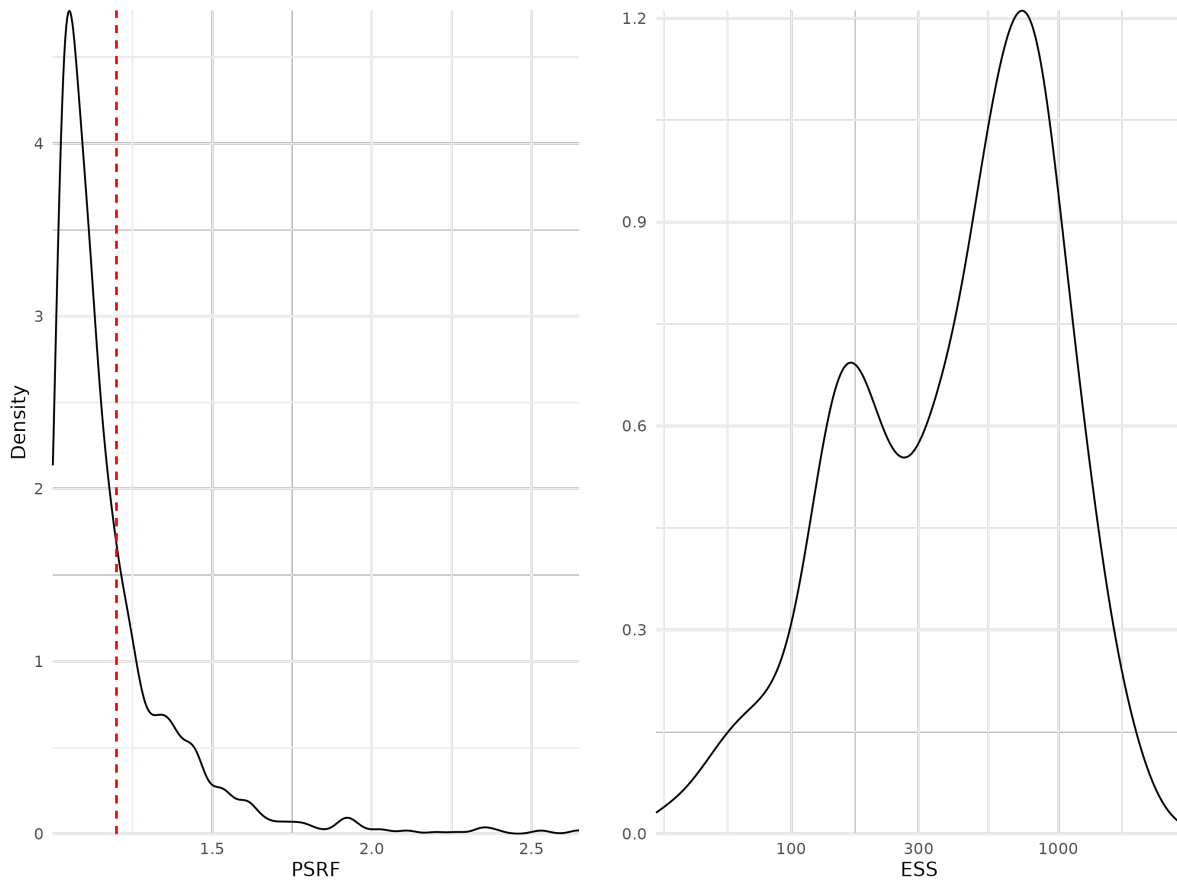


Figure S7 – Density curves of potential scale reduction factors (PSRF see BROOKS et GELMAN (1998); left panel) and effective sample sizes (ESS; right panel) for Beta regression parameters (i.e environmental coefficients) estimated for the phylogeny model fitted with abundance data. For further details see Fig. S5.

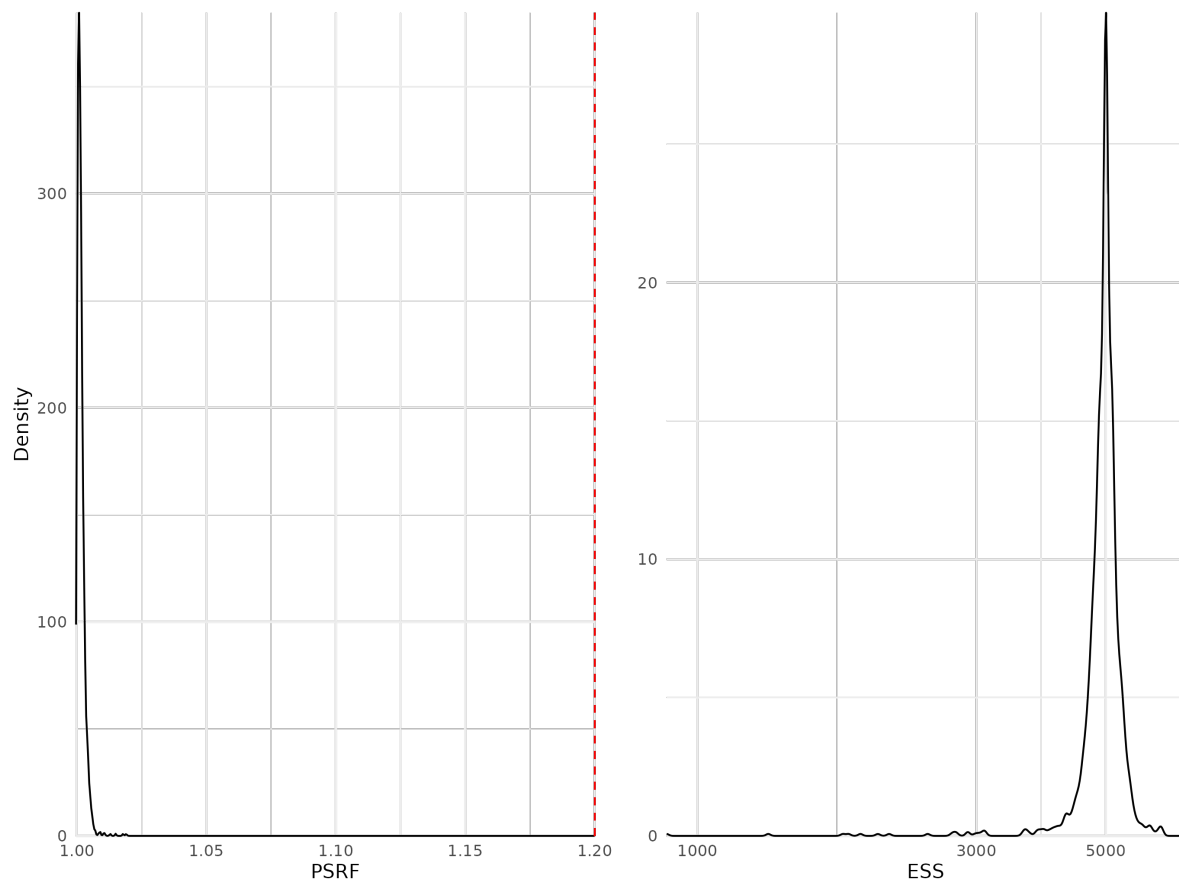


Figure S8 – Density curves of potential scale reduction factors (PSRF see BROOKS et GELMAN (1998); left panel) and effective sample sizes (ESS; right panel) for Beta regression parameters (i.e environmental coefficients) estimated for the phylogeny model fitted with presence/absence data. For further details see Fig. S5.

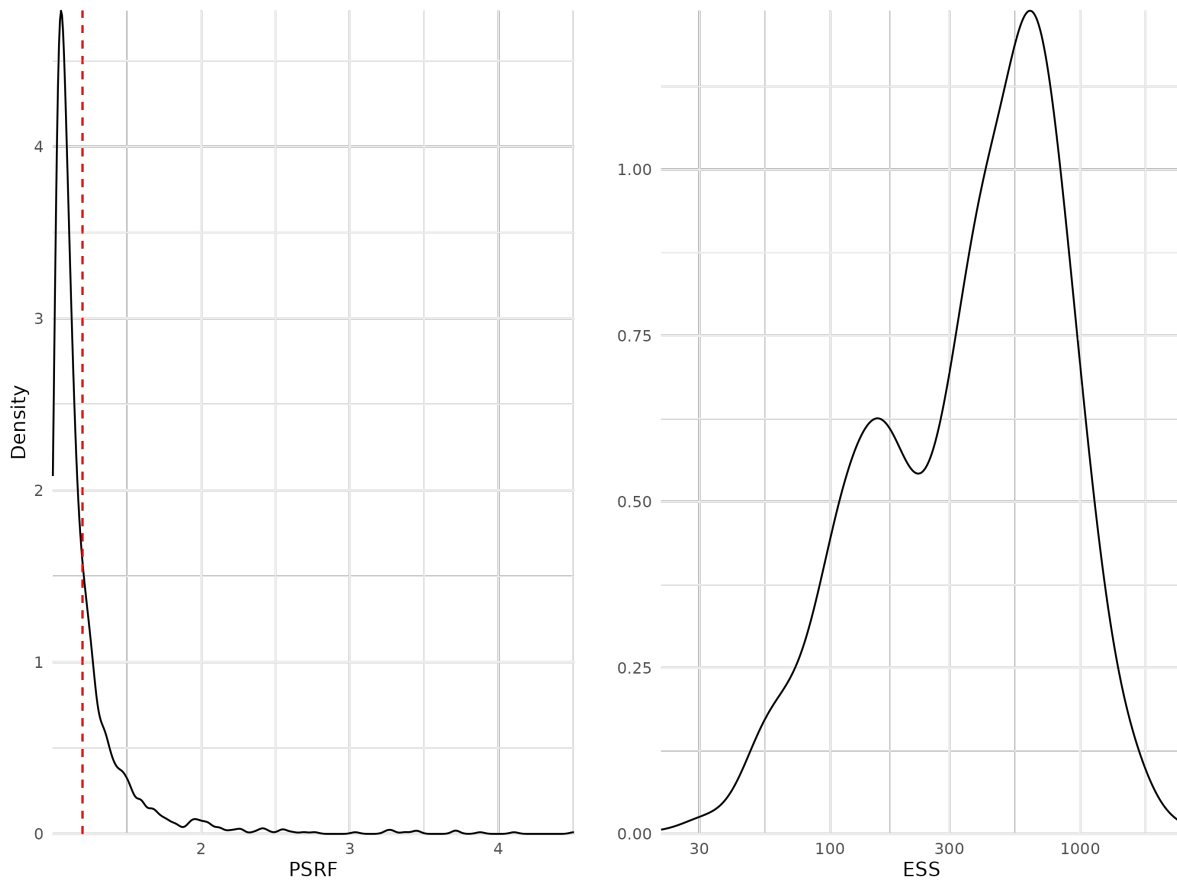


Figure S9 – Density curves of potential scale reduction factors (*PSRF* see *BROOKS et GELMAN (1998)*; left panel) and effective sample sizes (*ESS*; right panel) for Beta regression parameters (*i.e* environmental coefficients) estimated for the traits & phylogeny model fitted with abundance data. For further details see Fig. S5.

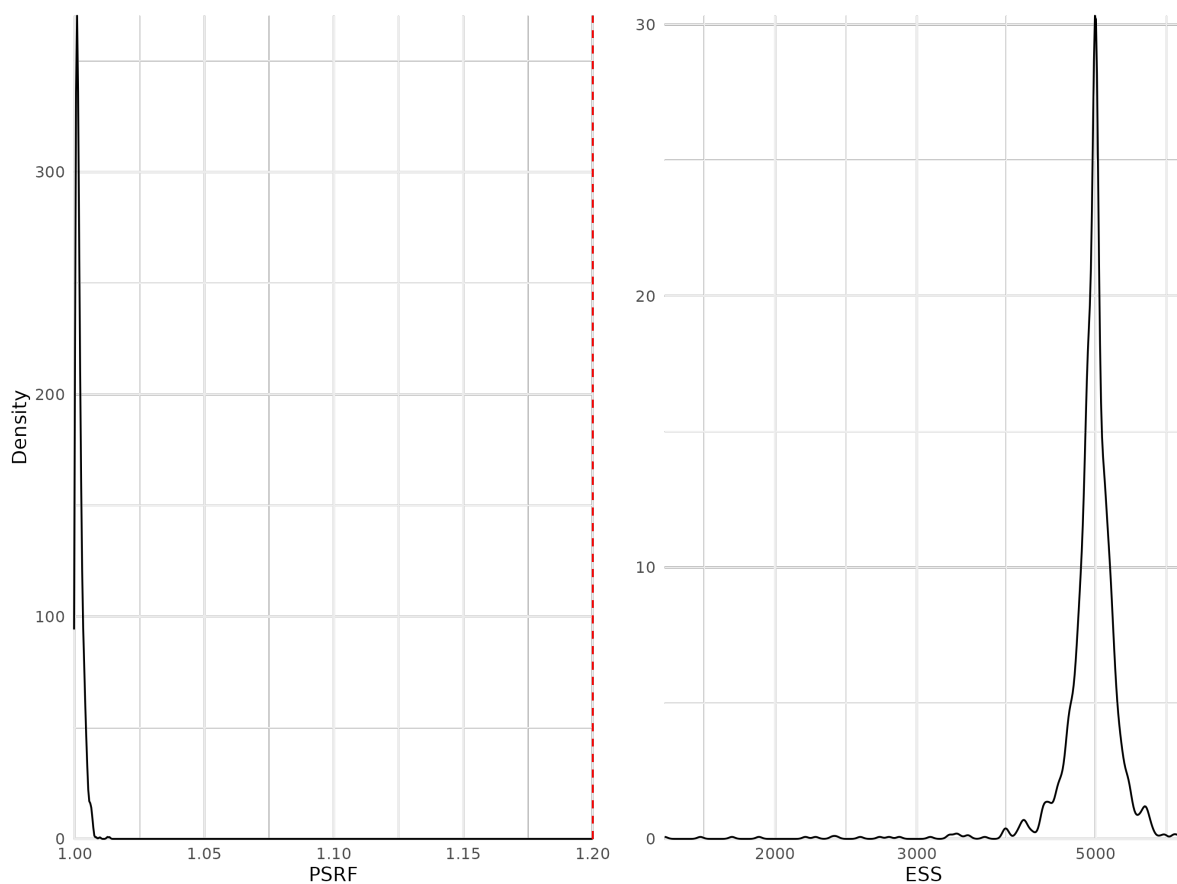


Figure S10 – Density curves of potential scale reduction factors (*PSRF* see *BROOKS et GELMAN (1998)*; left panel) and effective sample sizes (*ESS*; right panel) for Beta regression parameters (i.e. environmental coefficients) estimated for the traits & phylogeny model fitted with presence/absence data. For further details see Fig. S5.

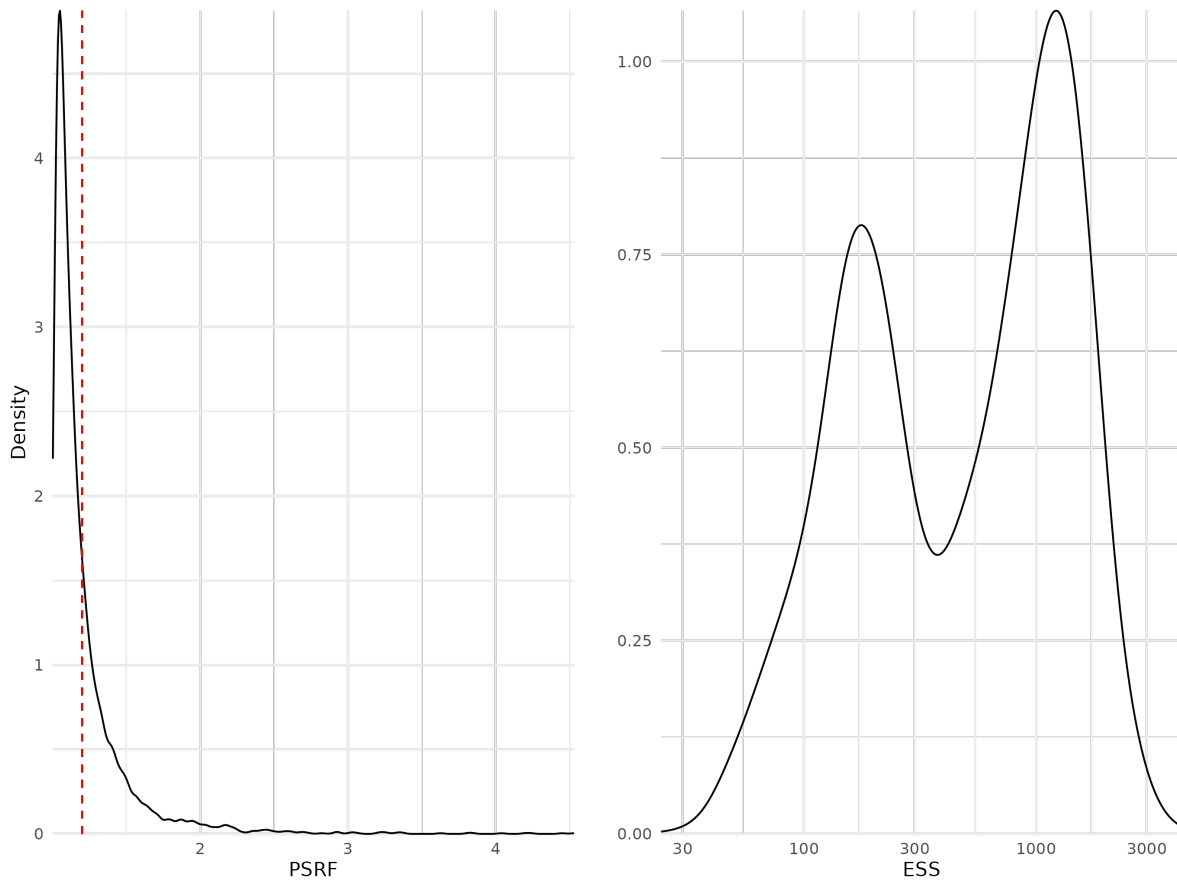


Figure S11 – Density curves of potential scale reduction factors (*PSRF* see *BROOKS et GELMAN (1998)*; left panel) and effective sample sizes (*ESS*; right panel) for Beta regression parameters (*i.e* environmental coefficients) estimated for the whole community model fitted with abundance data. For further details see *Fig. S5*.

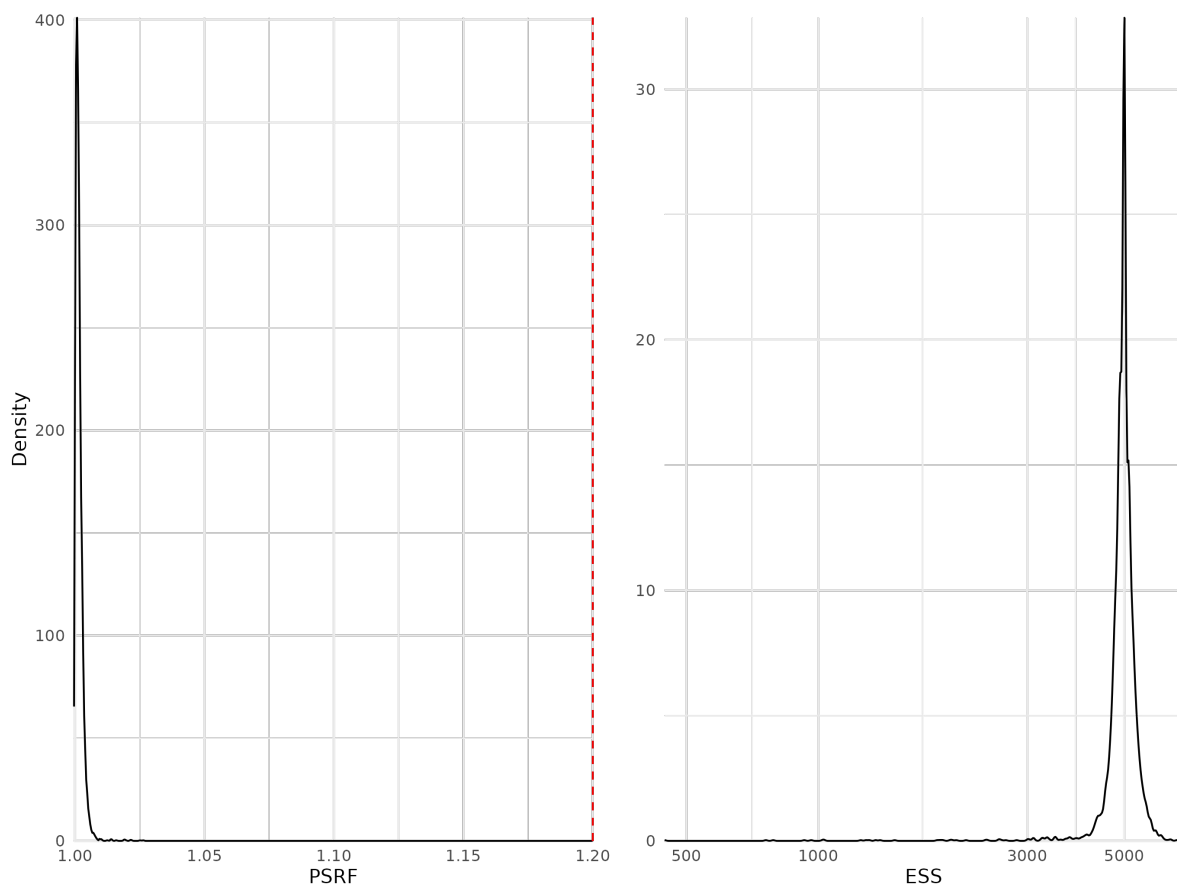


Figure S12 – Density curves of potential scale reduction factors (*PSRF* see *BROOKS et GELMAN (1998)*; left panel) and effective sample sizes (*ESS*; right panel) for Beta regression parameters (i.e. environmental coefficients) estimated for the whole community model fitted with presence/absence data. For further details see Fig. S5.

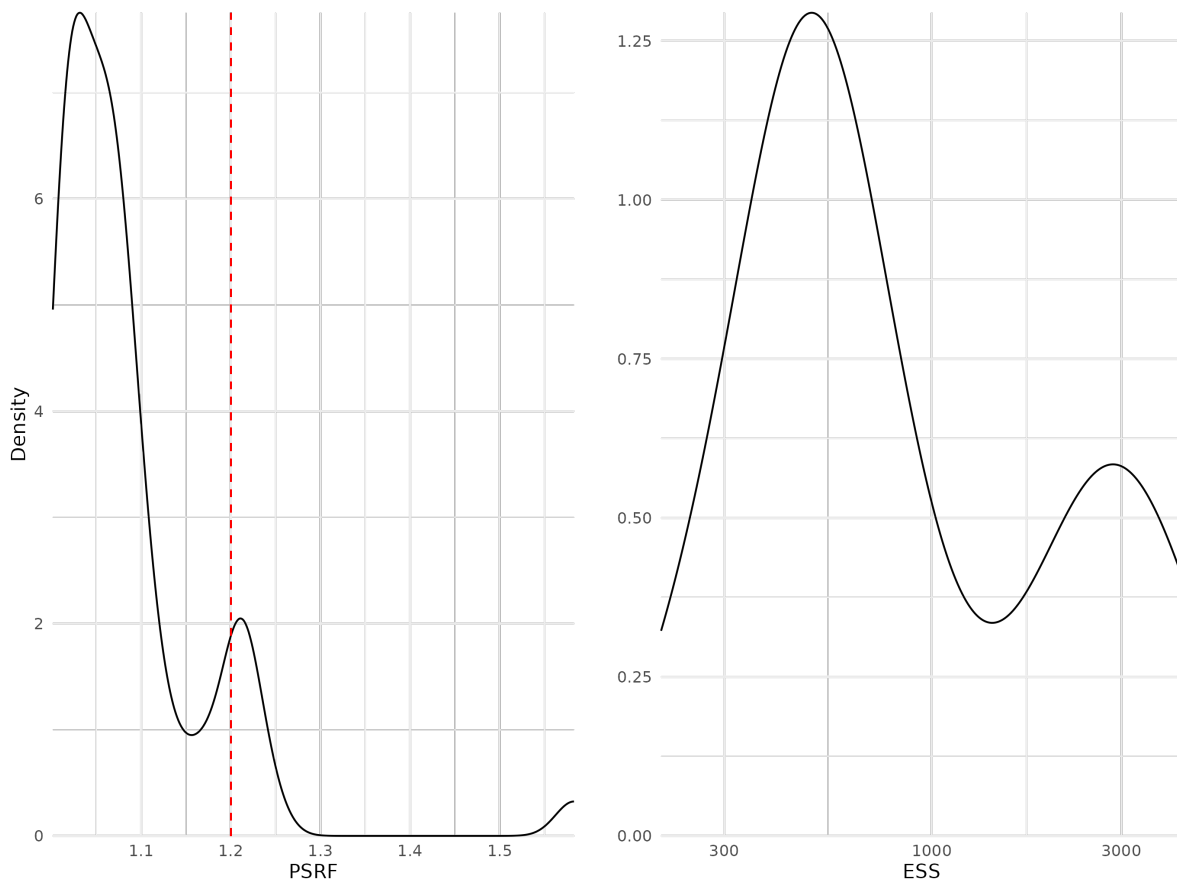


Figure S13 – *Density curves of potential scale reduction factors (PSRF see BROOKS et GELMAN (1998); left panel) and effective sample sizes (ESS; right panel) for Gamma regression parameters (i.e coefficients associated with trait-environment interactions, modeling how species traits influence their niches) estimated for the traits & phylogeny model fitted with abundance data. For further details see Fig. S5.*

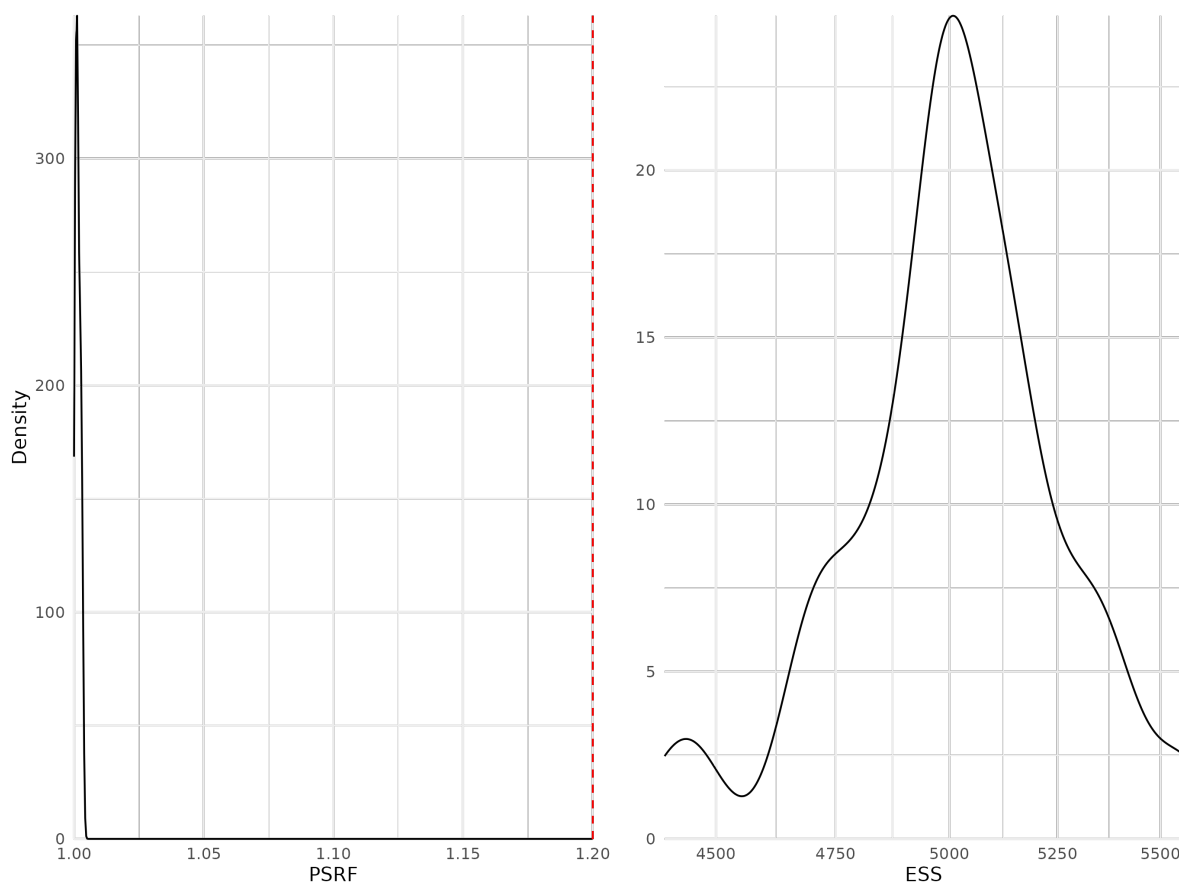


Figure S14 – Density curves of potential scale reduction factors (*PSRF* see *BROOKS et GELMAN (1998)*; left panel) and effective sample sizes (*ESS*; right panel) for Gamma regression parameters (i.e coefficients associated with trait-environment interactions, modeling how species traits influence their niches) estimated for the traits & phylogeny model fitted with presence/absence data. For further details see Fig. S5.

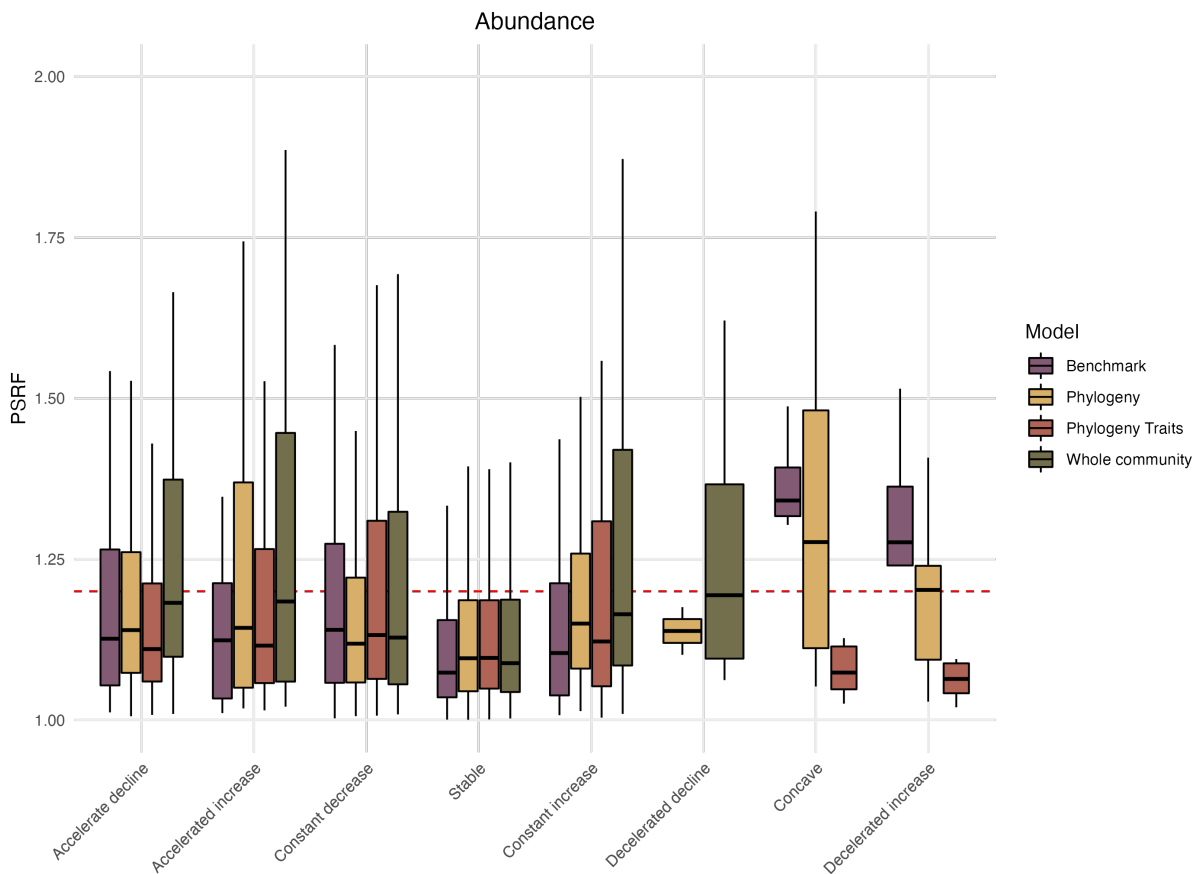


Figure S15 – Distribution of the Potential scale reduction factors (PSRF) as a function of the different shapes of the response curves classified following the methodology proposed by RIGAL et al. (2020) methods (see section “Assessing model performance and interpretability” for more details on the calculation methodology). Results for models fitted with abundance data.

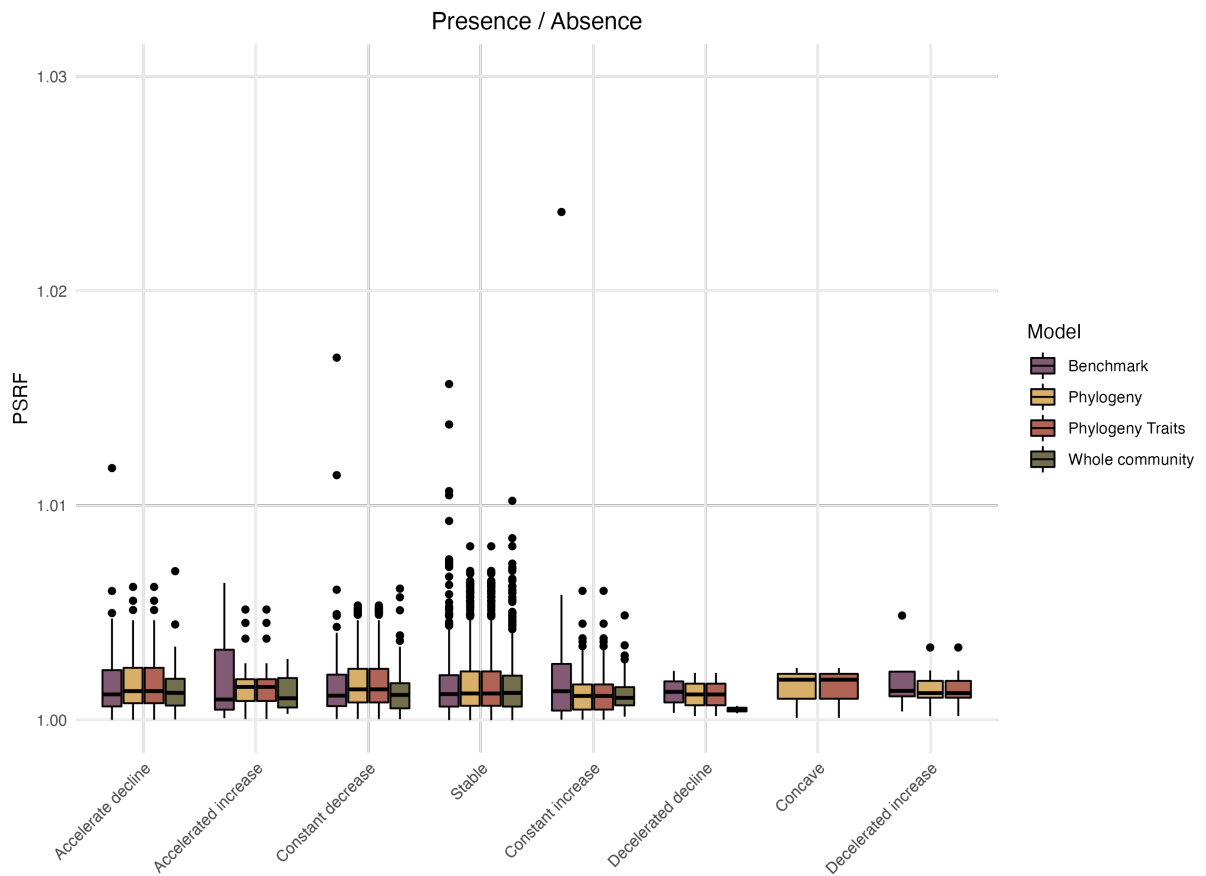


Figure S16 – Distribution of the Potential scale reduction factors (PSRF) as a function of the different shapes of the response curves classified following the methodology proposed by RIGAL *et al.* (2020) methods (see section “Assessing model performance and interpretability” section for more details on the calculation methodology). Results for models fitted with presence/absence data.

Appendix C - Complementary Results

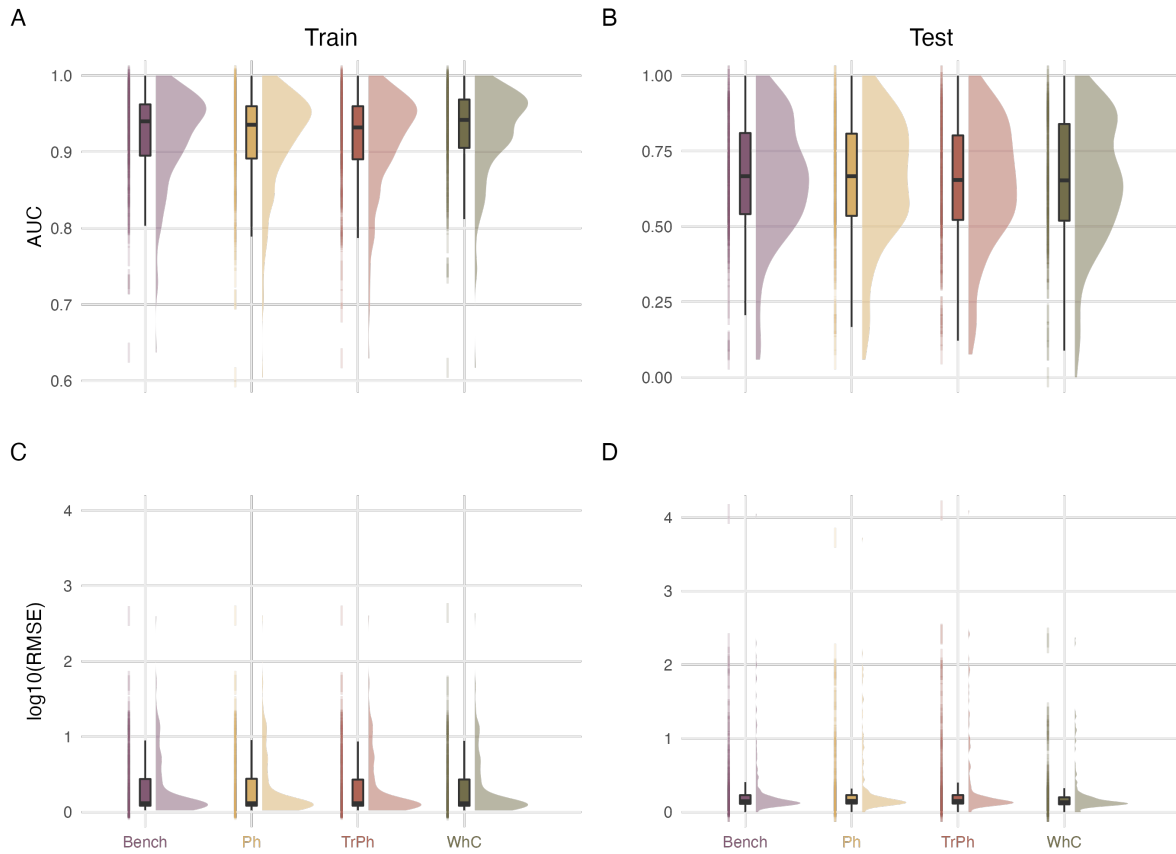


Figure S17 – Comparison of explanatory (left column; Train set) and predictive (right column; Test set) performance capacities of the different model structures fitted on presence/absence (top panels) or abundance (bottom panels) data

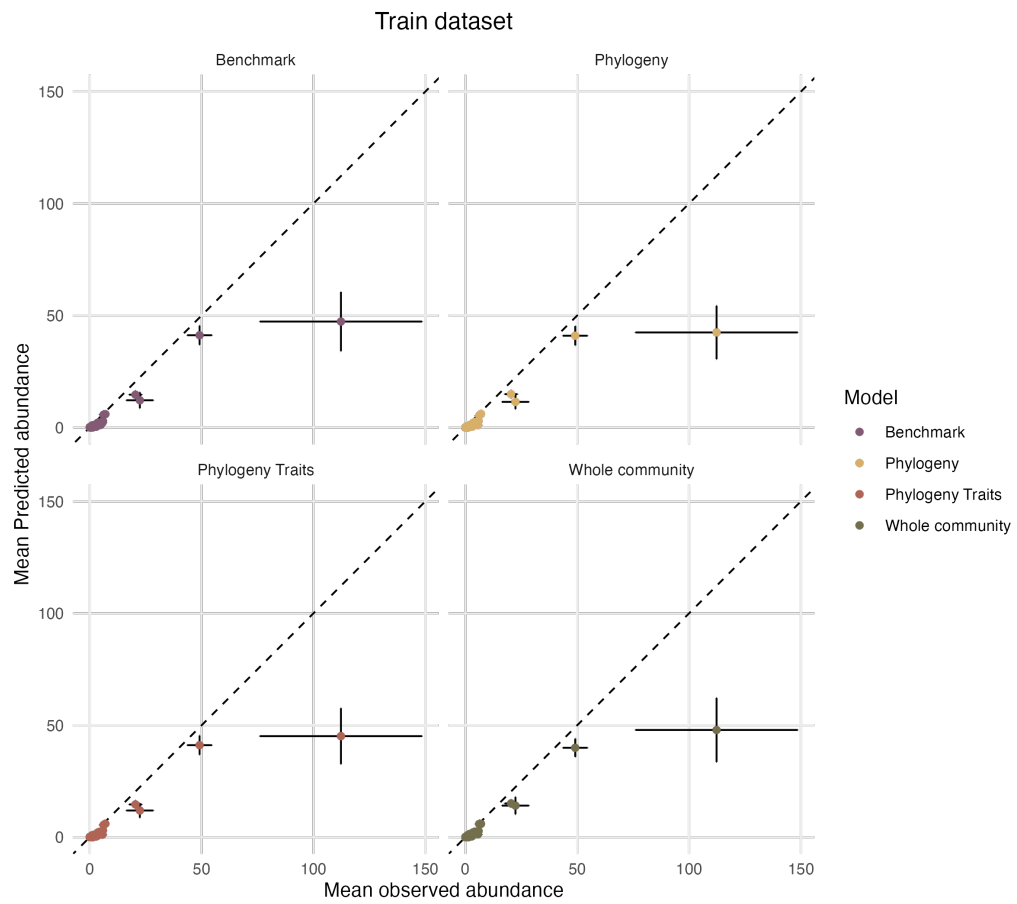


Figure S18 – Mean predicted abundances as a function of mean observed abundances in the training dataset. Each species is represented by a dot, the error bars on each point indicate the standard error around the mean relatively to each axis.

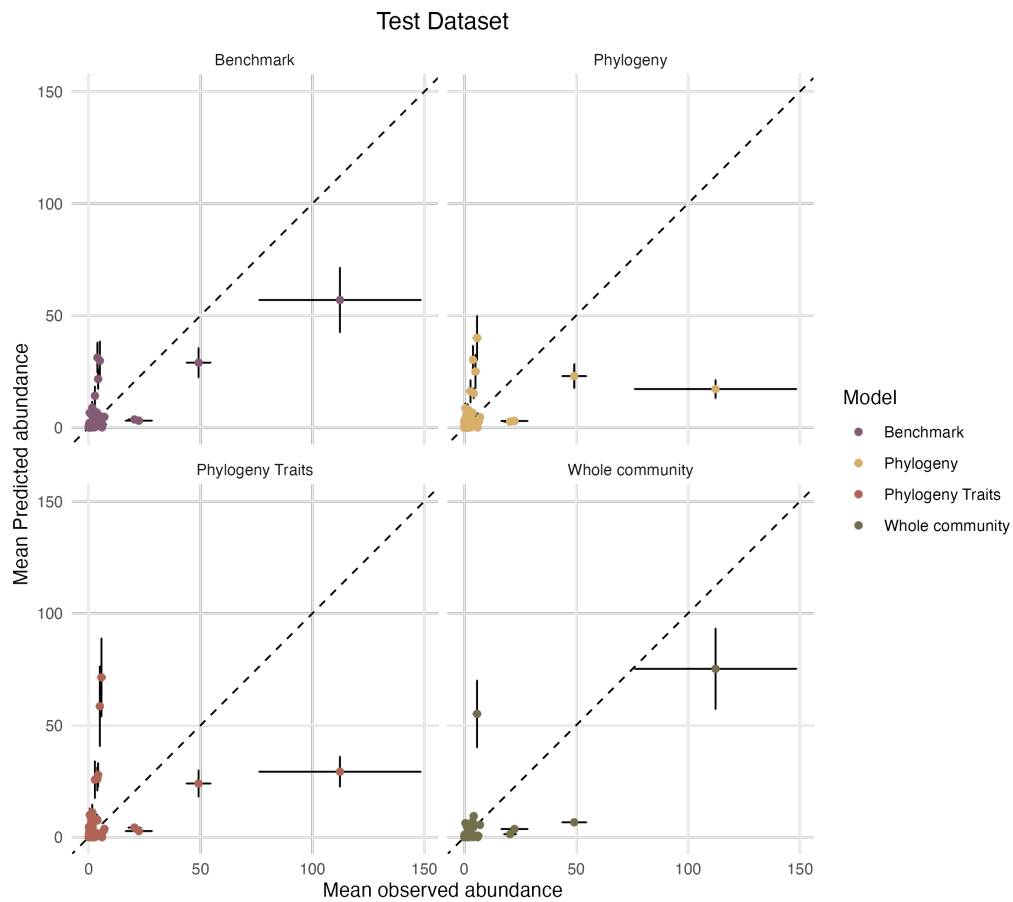


Figure S19 – Mean predicted abundances as a function of mean observed abundances in the training dataset. Each species is represented by a dot, the error bars on each point indicate the standard error around the mean relatively to each axis.

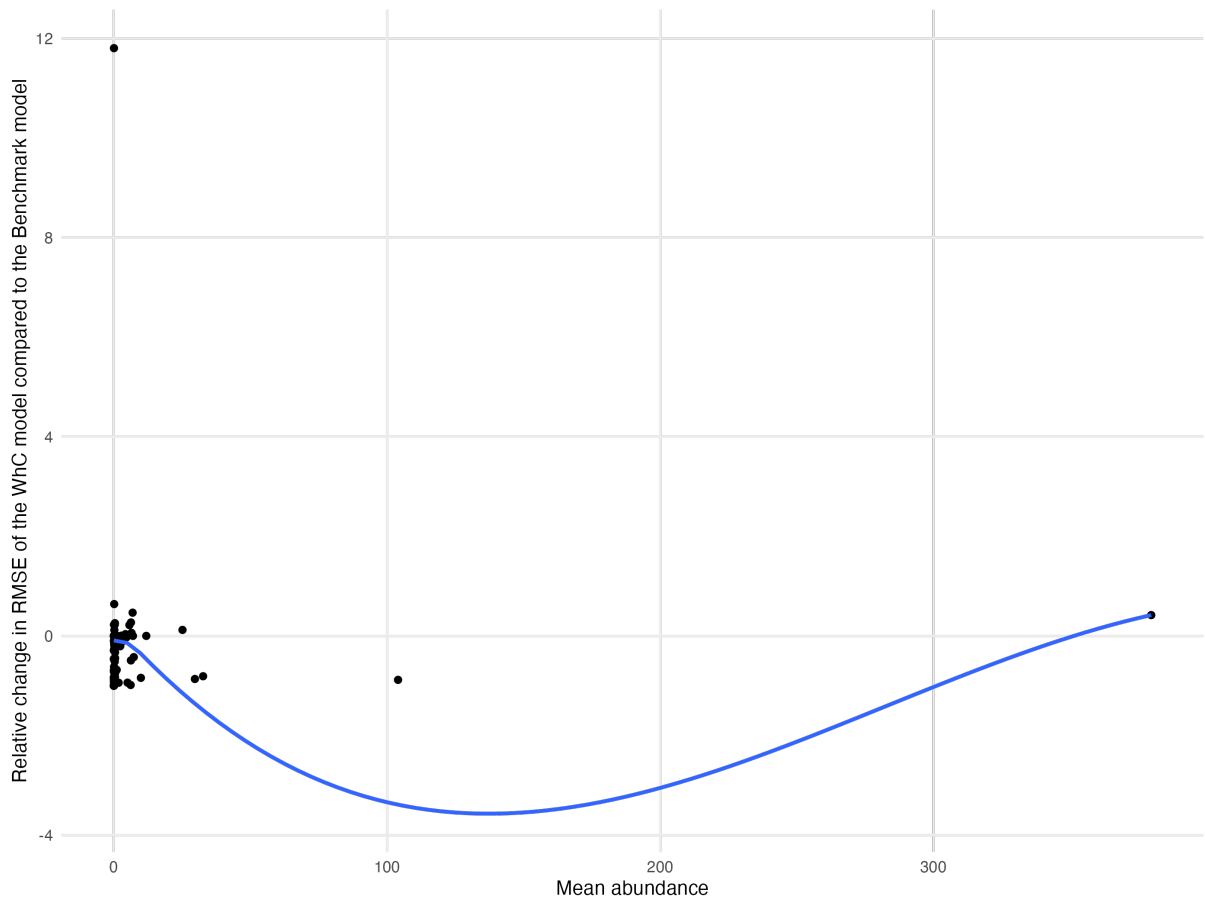


Figure S20 – Relationship between the relative improvement in RMSE for the WhC model compared to Bench model and the mean abundance of species in the training dataset. Each dot represents a species. The blue line represents a fit obtained from a LOESS regression.

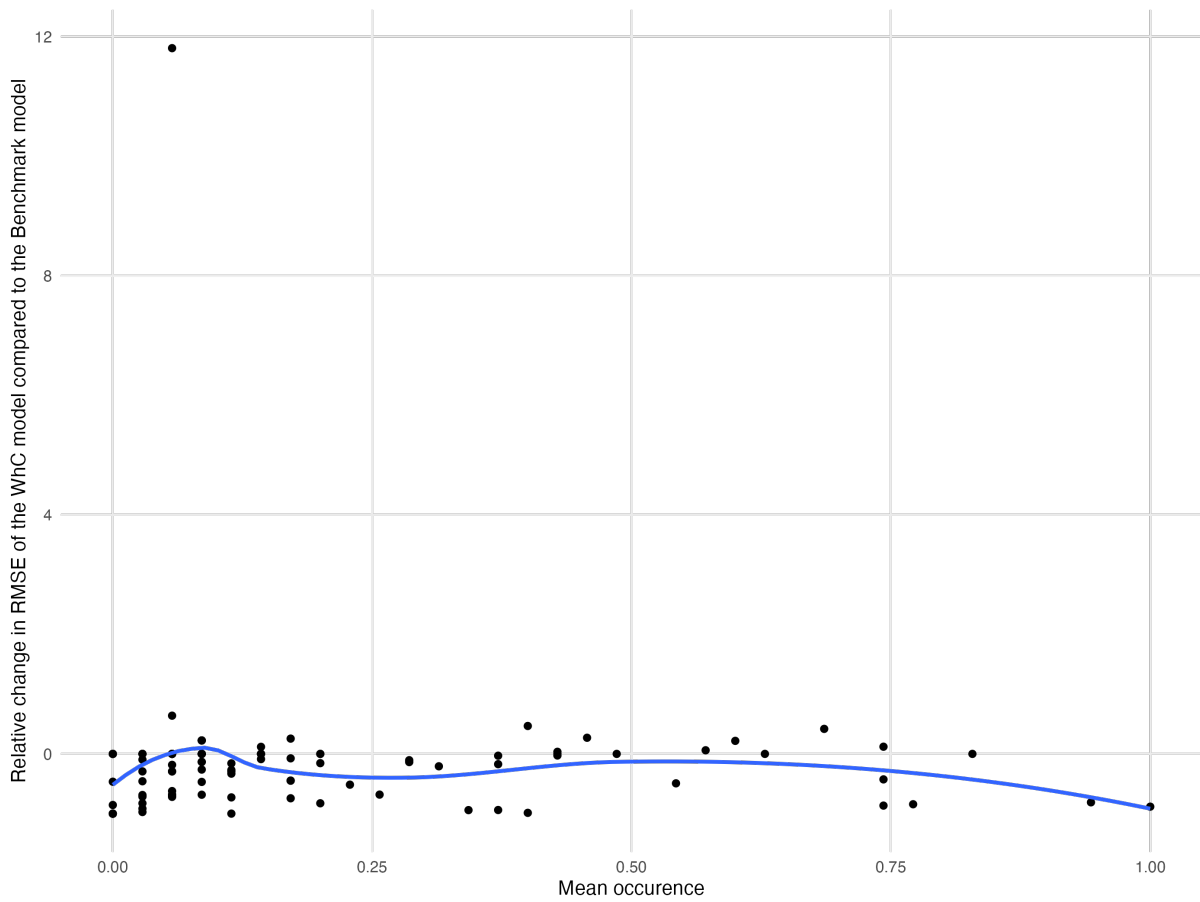


Figure S21 – Relationship between the relative improvement in RMSE for the WhC model compared to Bench model and the mean occurrence of species in the training dataset. Each dot represents a species. The blue line represents a fit obtained from a LOESS regression.

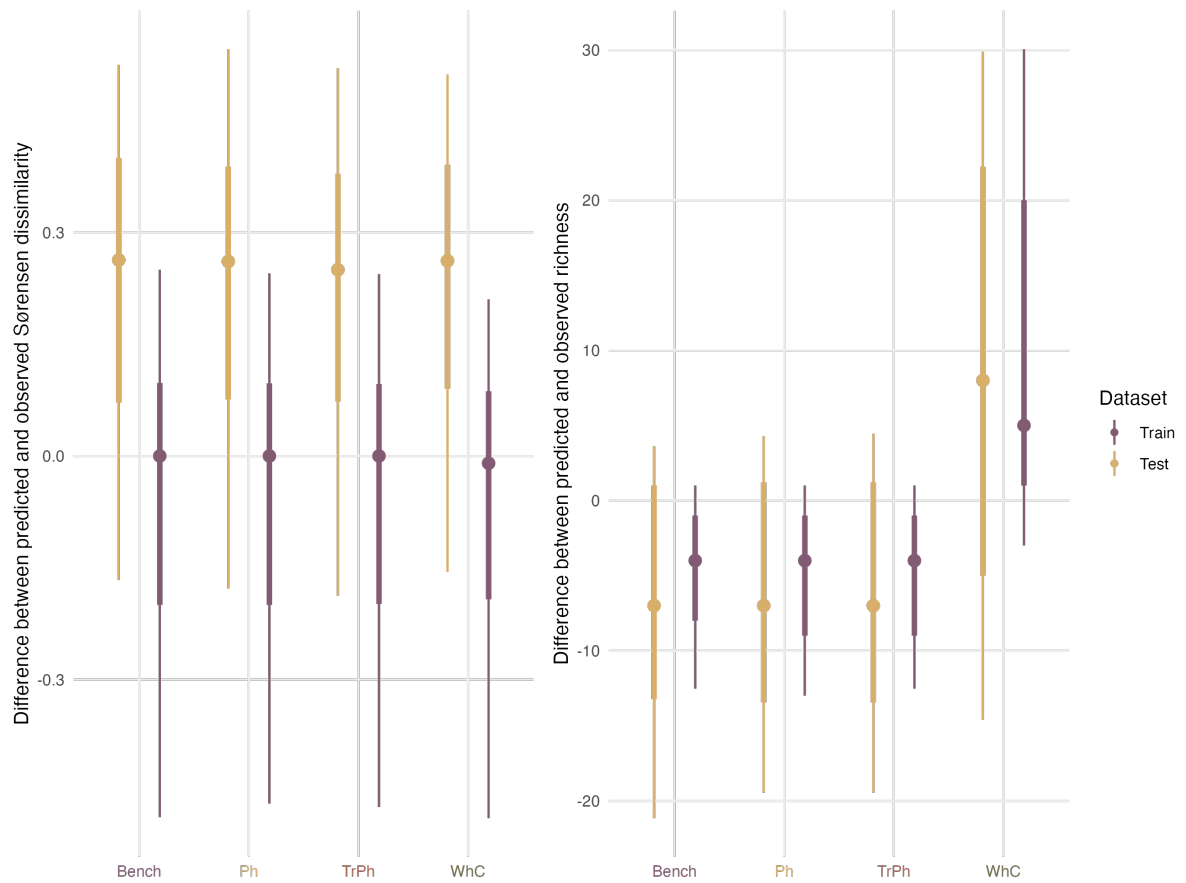


Figure S22 – Comparison of model performances with regards to their ability to predict community structures when fitted with presence/absence data for the train (purple) and test (yellow) dataset. The left column indicates for each model the difference between the pairwise dissimilarities computed on the observed assemblages and those computed on the predicted community. The right column presents the differences in species richness between the observed and predicted assemblages.

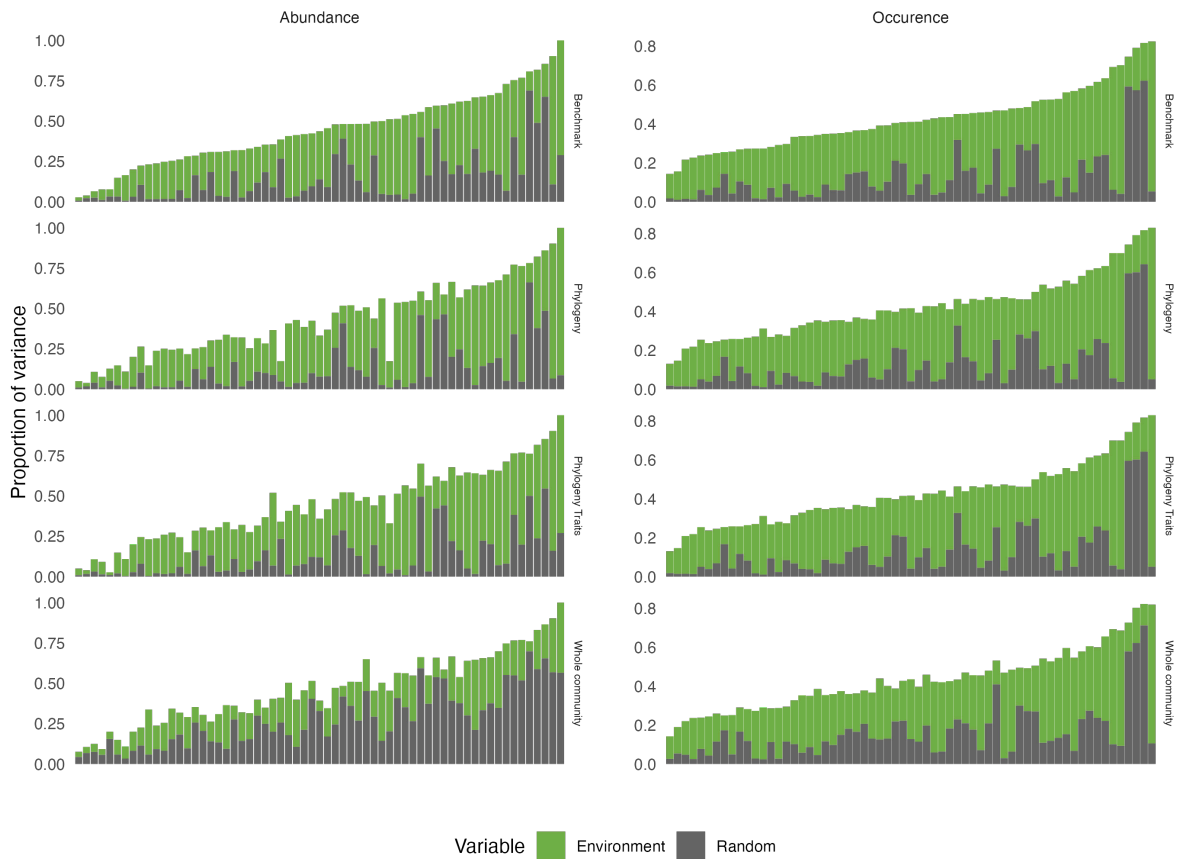


Figure S23 – Comparison across the four alternative model structures of the total amount of variance of each species (along the x-axis) that is explained by (1) the environmental variables (*Environment*) and (2) the three random effects (*Random*). Results are presented for the models fitted with abundance (left) and presence/absence (right) data. Species are ordered by increasing order of total variance explained by the benchmark model.

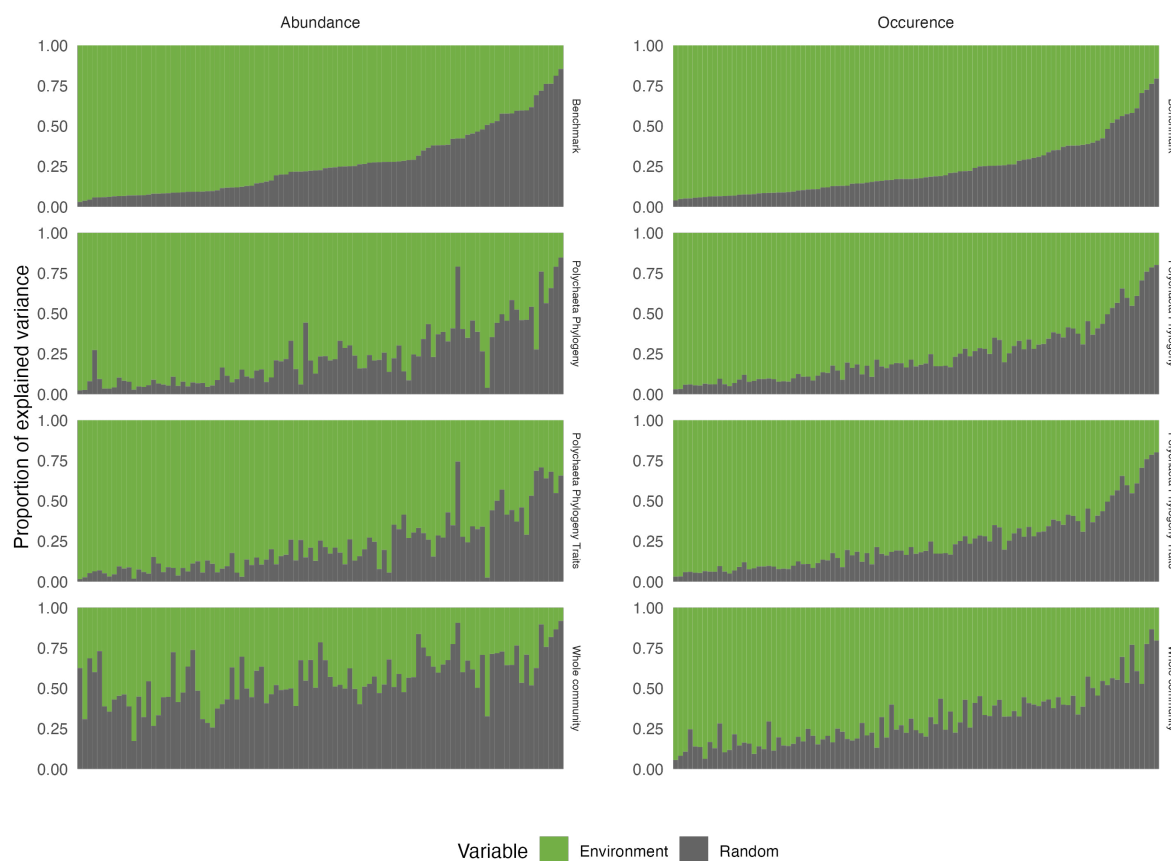


Figure S24 – Comparison across the four alternative model structures of the fraction of variance of each species (along the *x*-axis) that is explained by (1) the environmental variables (*Environment*) and (2) the three random effects (*Random*). Results are presented for the models fitted with abundance (left) and presence/absence (right) data. Species are ordered by decreasing order of variance explained by the environment for the benchmark model.

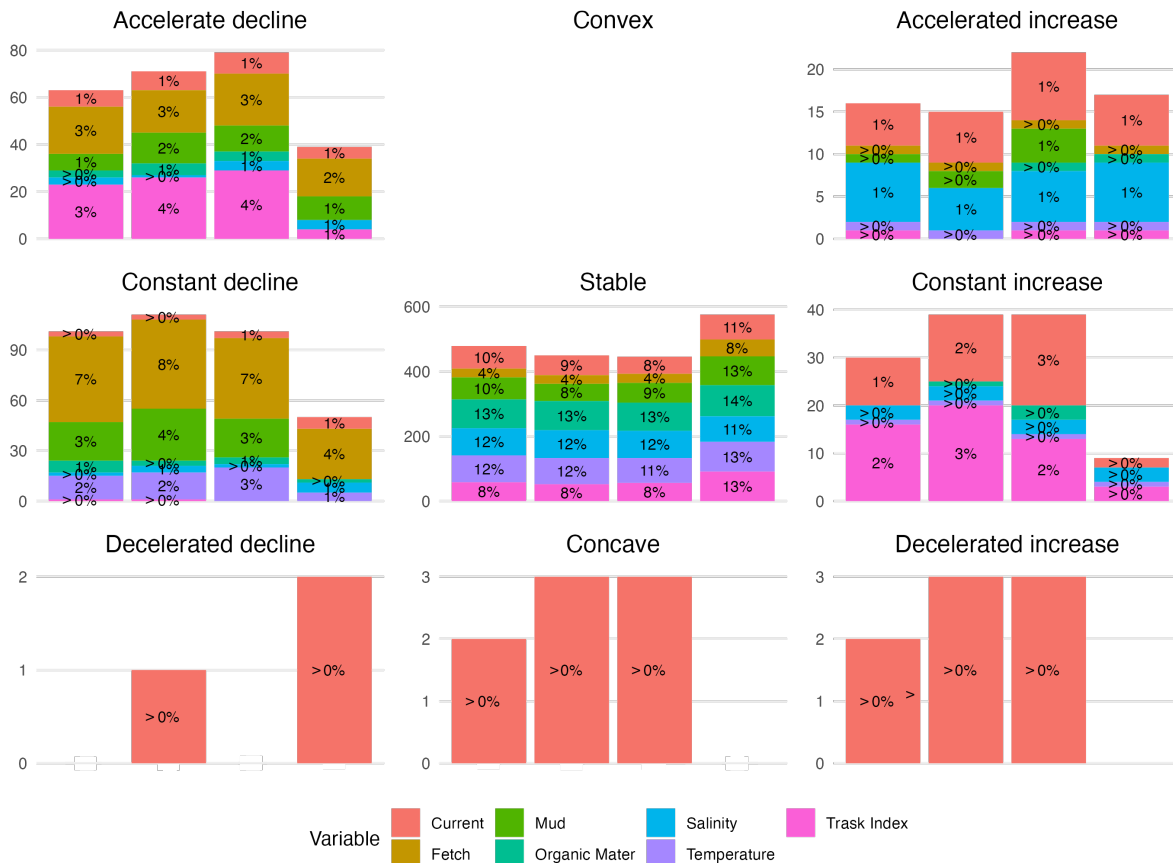


Figure S25 – Same figure as Fig. 4 in the main text for models fitted on abundance data. Number (y-axis) and proportion (indicated above individual bars, rounded to the nearest integer) of response curves (i.e. one for each species-predictor combination) according to the nomenclature (nine shapes highlighted by the black curve in each panel) defined by RIGAL et al. (2020). Results are presented for different model structures : from left to right the Benchmark (Bench), the phylogeny (Ph), the traits & phylogeny (TrPh), and the whole community (WhC) models. Each bar is coloured by the relative contribution of each environmental covariate to this particular shape. For illustrative purposes, note that the scale of variation on the y-axis differs across panels.

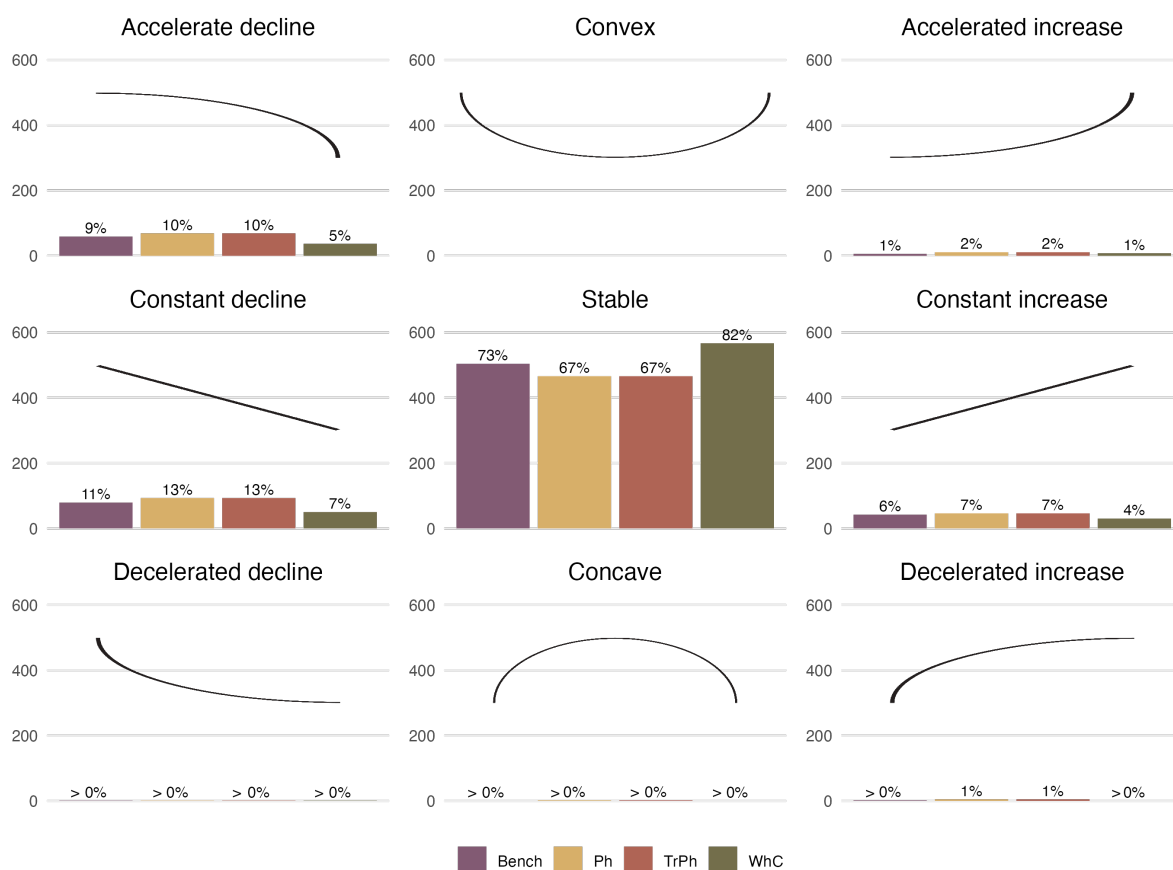


Figure S26 – Same figure as Fig. 4 in the main text but for models fitted to presence/absence data. Number (*y*-axis) and proportion (indicated above individual bars, rounded to the nearest integer) of response curves (*i.e.* one for each species-predictor combination) according to the nomenclature (nine shapes highlighted by the black curve in each panel) defined by RIGAL *et al.* (2020). Results are presented for the different structures : purple for the Benchmark (Bench), yellow for phylogeny (Ph), red for traits & phylogeny (TrPh), and green for whole community (WhC) model.

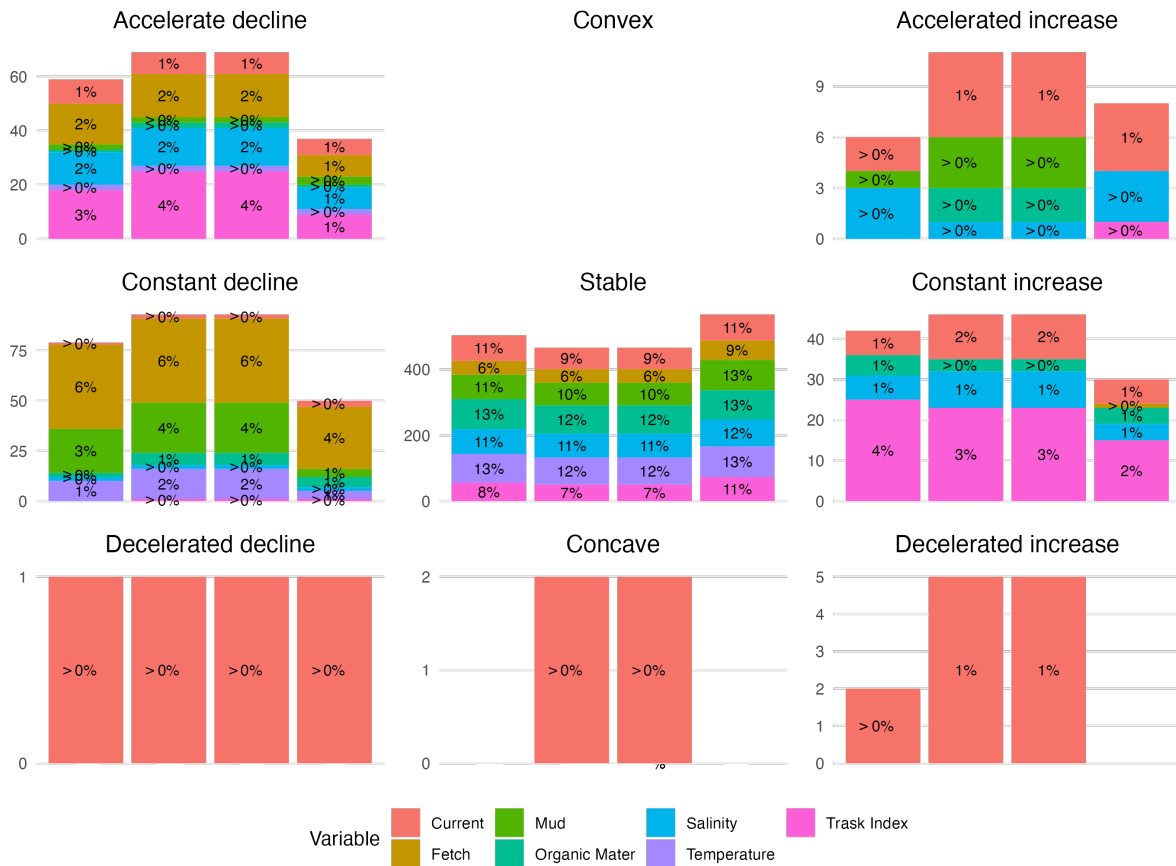


Figure S27 – Same figure as Fig. S25 but for models fitted on presence/absence data. Number (*y*-axis) and proportion (indicated above individual bars, rounded to the nearest integer) of response curves (*i.e.* one for each species-predictor combination) according to the nomenclature (nine shapes highlighted by the black curve in each panel) defined by RIGAL *et al.* (2020) for different presence/absence model structures. Each bar is coloured by the relative contribution of each environmental covariate to this particular shape. For illustrative purposes, note that the scale of variation on the *y*-axis differs across panels.

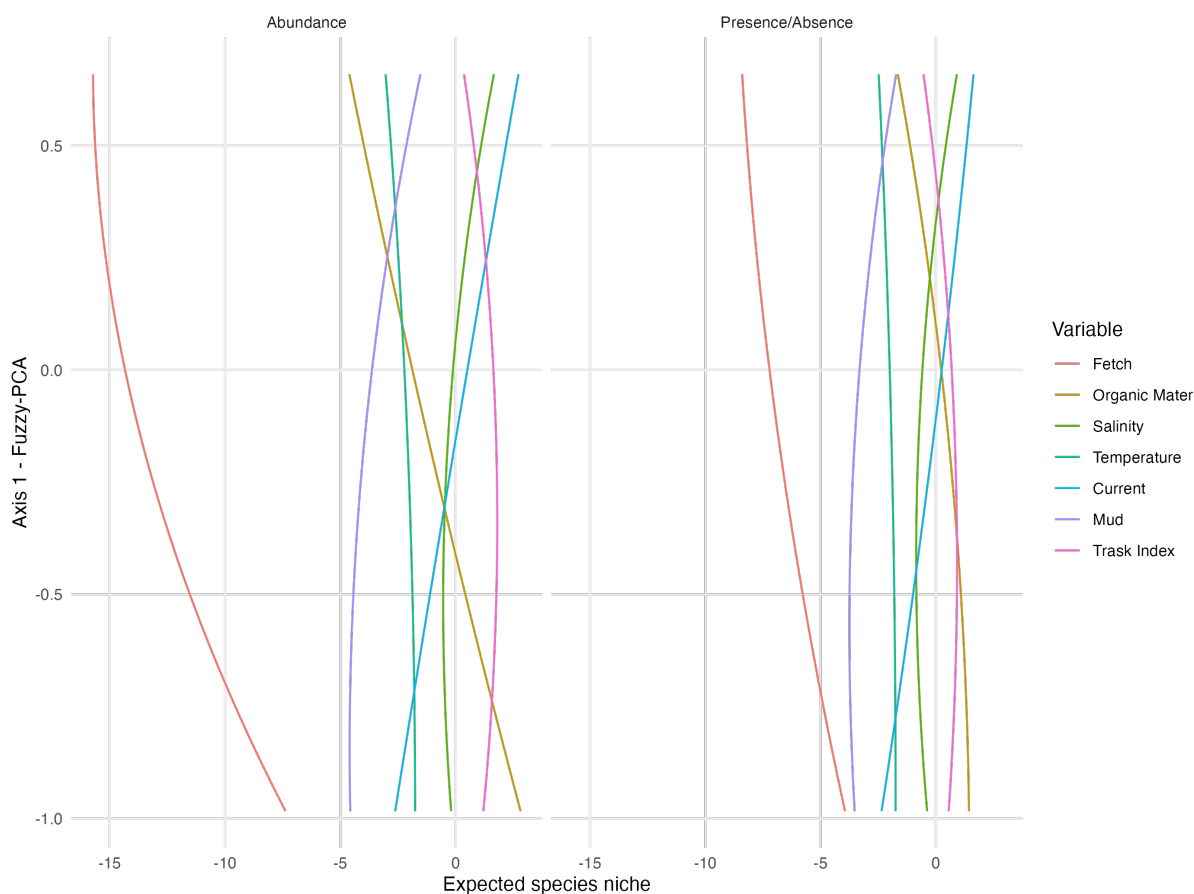


Figure S28 – Relationship between species' position along the first axis of the fuzzy PCA (sessile microphagous-mobile macrophagous gradient) and the different environmental variables used in the models (fitted with abundance data in the left panel, and with presence/absence data in the right panel). Relationships are derived from the regression coefficients estimated for the PhTr model (γ coefficients in HMSC; OVASKAINEN et ABREGO (2020)). The lines are fitted quadratic regressions representing the average response across the different species. As an example of interpretation, the red lines in both graphs indicate that sessile microphagous species are more negatively influenced (lower abundance, low probability for presence) by fetch than macrophagous mobile species.

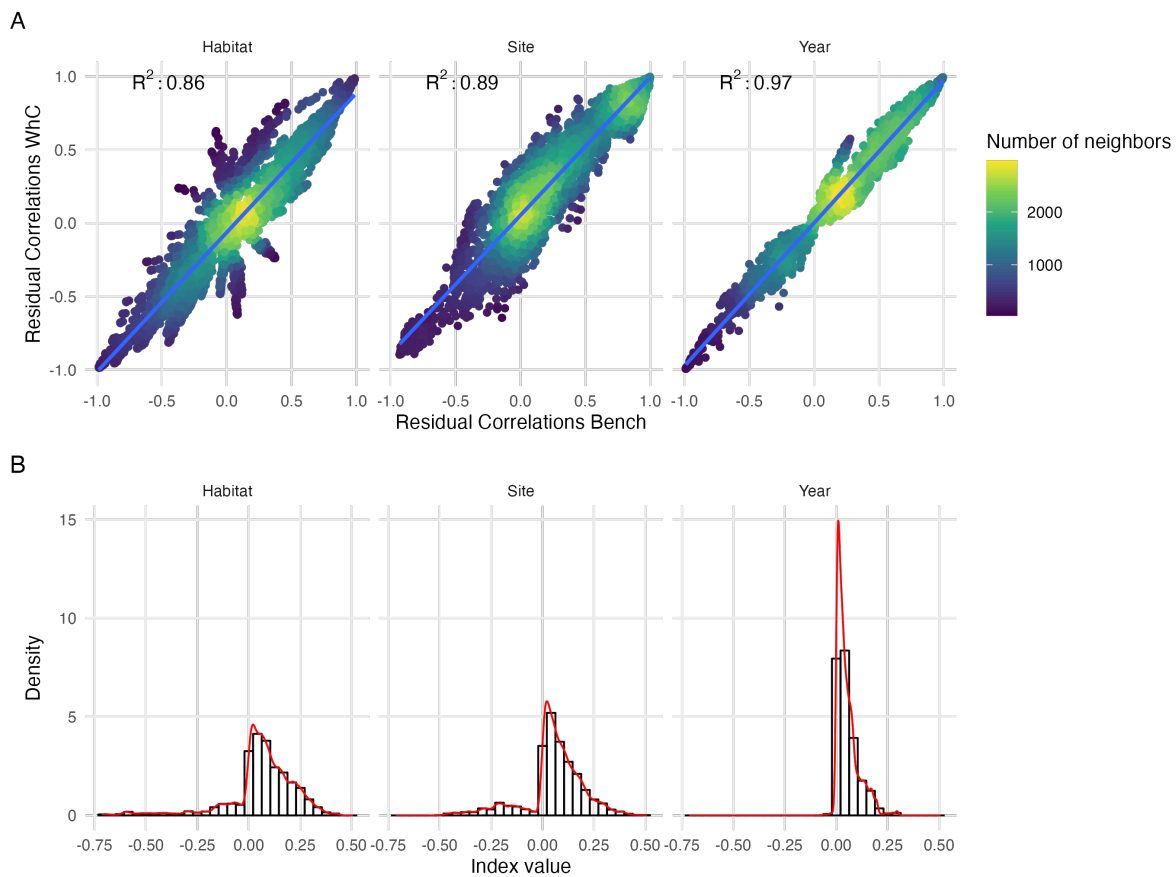


Figure S29 – Same figure as Fig. 5 in the main text but for models fitted on presence/absence data. (A) Comparison of residual correlations associated with the three random effects estimated by the Whole Community Model (y-axis) and the Benchmark model (x-axis). The colour scale highlights the density of points in each scatter plot. (B) Distribution of the index measuring change in sign (sign change left to the zero line, no change to the right) and magnitude (higher departure from the zero line indicate higher difference) between residual correlations estimated by the whole community model and the benchmark model for the three random effects (Habitat, Site, Year).

Bibliographie

- BOYÉ, A., (2019), *Diversité taxinomique et fonctionnelle des habitats benthiques dans l'espace et dans le temps : une perspective régionale et décennale* [thèse de doct.].
- BOYÉ, A., THIÉBAUT, É., GRALL, J., LEGENDRE, P., BROUDIN, C., HOUBIN, C., GARREC, V. L., MAGUER, M., DROUAL, G., & GAUTHIER, O., (2019), Trait-based approach to monitoring marine benthic data along 500 km of coastline, *Diversity and Distributions*, 2512, 1879-1896.
- BROOKS, S. P., & GELMAN, A., (1998), General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, 74, 434-455.
- CRESSWELL, A. K., EDGAR, G. J., STUART-SMITH, R. D., THOMSON, R. J., BARRETT, N. S., & JOHNSON, C. R., (2017), Translating local benthic community structure to national biogenic reef habitat types, *Global Ecology and Biogeography*, 2610, 1112-1125.
- LAZURE, P., & DUMAS, F., (2008), An external–internal mode coupling for a 3D hydrodynamical model for applications at regional scale (MARS), *Advances in Water Resources*, 312, 233-250.
- LECORNU, F., & ROECK, Y.-H. D., (2009), PREVIMER - Coastal observations and forecasts, *La Houille Blanche*, 951, 60-63.
- OVASKAINEN, O., & ABREGO, N., (2020), *Joint Species Distribution Modelling : With Applications in R*, Cambridge University Press.
- RIGAL, S., DEVICTOR, V., & DAKOS, V., (2020), A method for classifying and comparing non-linear trajectories of ecological variables, *Ecological Indicators*, 112, 106113.

From local seafloor imagery to global patterns in benthic habitats : contribution of citizen science to habitat classification across latitudes

Préambule

Le Chapitre 1 a permis de mettre en évidence deux limites majeures de l'application des *jSDM* (*Joint Species Distribution Models*) pour comprendre la structure des communautés et le rôle sous-jacent des habitats. Premièrement, que ce soit sur des données de présence/absence ou bien d'abondance, le framework de *jSDM* testé ici présentait des performances de prédiction de la biodiversité faible, limitant grandement ses capacités d'extrapolation spatiales et/ou temporelles. Deuxièmement, le *jSDM* étudié dans le chapitre précédent nécessitait une importante puissance de calcul pour ajuster aux 99 espèces les plus de 16 000 paramètres nécessaires tout en leur assurant une convergence suffisante des estimations de paramètres. Ainsi, son application sur un jeu de données plus conséquent du *Reef Life Survey*, avec une communauté faunistique de plus de 3 000 taxons, nous a semblé limiter son potentiel pour analyser un tel jeu de données à l'échelle globale. De plus, dans un projet antérieur focalisé sur l'usage des *jSDM* pour l'inférence de réseau de co-occurrence des espèces benthiques (VIOLET, 2020), nous avons fait face à des incertitudes méthodologiques pour caractériser de manière robuste les réseaux d'interactions potentiels. En effet, les réseaux reconstitués avaient été soumis aux jugements d'un panel d'experts benthologues dont les avis discordants n'ont pas permis de valider la véracité des

interactions identifiées. C’est pourquoi pour mieux comprendre comment les habitats biogéniques structurent les communautés associées à l’échelle mondiale, nous avons changé de stratégie et nous avons adopté la méthode “*agréger, puis prédire*” suggérée par FERRIER et GUIBAN (2006). L’objet central de ce chapitre réside dans l’application d’une méthode de groupement peu appliquée en écologie afin d’identifier différents états d’habitats. Le chapitre 3 complètera cette partie d’agrégation des données par une partie prédictive.

Les particularités des données écologiques et notamment celles issues de sciences participatives (hétérogénéité, bruit dans les données, relations non-linéaires entre les variables d’un écosystème, etc.) limitent la pertinence de certaines méthodes de groupement (p. ex. k-means, classification hiérarchique à lien simple complet, ou selon la méthode de Ward) classiquement utilisées en écologie (Fig. 3.1). Nous avons utilisé dans ce chapitre une nouvelle approche de groupement pour (1) distinguer différents états d’habitats biogéniques à partir de données issues des sciences participatives (2) identifier des habitats iconiques, et (3) éventuellement détecter des transitions entre différents types d’habitats. Ce chapitre a également pour but d’étudier la distribution de ces états d’habitats à l’échelle du globe pour valider ou non les groupes iconiques identifiés.

En amont de ce chapitre, j’ai étudié le comportement de plusieurs combinaisons de méthodes de réductions de dimensions, ainsi que de groupement (Fig. 3.1) ce qui a orienté le choix pour ce Chapitre 2 vers une méthodologie qui combine un algorithme de réduction basé sur le principe des graphes de voisinage (*UMAP* ; MCINNES et al. (2020)), et une méthode de groupement basée sur la densité (*HDBSCAN* ; MOULAVI et al. (2014) ; MCINNES et al. (2017)). En appliquant cette méthode à des données estimant le pourcentage de couverture de différents habitats sur la base des photoquadrats obtenus sur plus de 6 500 transects réalisés en plongée par des volontaires du programme *Reef Life Survey*, nous avons identifié 17 états d’habitats. Certains de ces habitats représentent des états d’habitats benthiques iconiques comme les forêts de laminaires ou des herbiers marins, alors que d’autres représentent des états d’habitats considérés comme dégradés, par exemple les états dominés par la présence d’algues gazonnantes.

Ce chapitre de thèse a déjà fait l’objet d’une communication orale “*Predicting reef state and regime shift risk using machine learning and the Reef Life Survey*” au 13th International Temperate Reefs Symposium 2023. Ce chapitre devrait être prochainement proposé pour publication au journal *Global Ecology and Biogeography*.

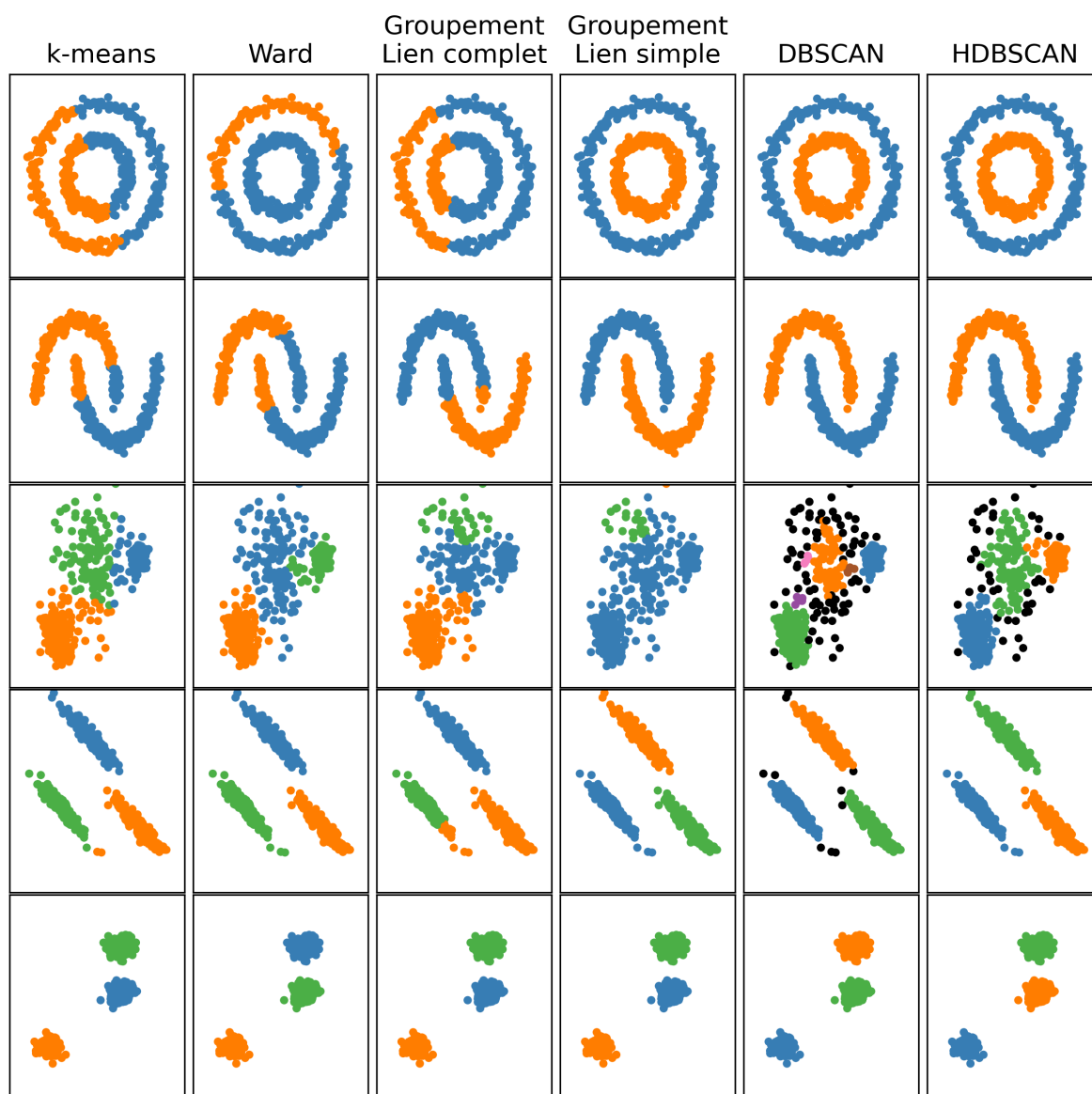


Figure 3.1 – Comparaison de différents algorithmes de groupement utilisés en écologie (en colonne ; à savoir *k*-means, Ward, groupement lien complet ou lien simple, DBSCAN ou HDBSCAN) sur différents jeux de données simulés en deux dimensions (en ligne). Les jeux de données des deux premières lignes contiennent deux groupes, les trois derniers en contiennent trois.

3.1 Abstract

Aim : The aim of this study was to define reef benthic habitat states and explore their spatial and temporal variability at a global scale using an innovative clustering pipeline.

Location : The study uses data on the transects surveyed on shallow (< 20m) reef ecosystems across the globe.

Time period : Transects sampled between 2008 and 2021.

Major taxa studied : Macroalgae, sessile invertebrates, hydrozoans, seagrass, corals.

Methods : Percentage cover was estimated for 24 functional groups of sessile biota and substratum from annotated underwater photoquadrats taken along 6,554 transects by scuba divers contributing to the *Reef Life Survey* dataset. A clustering pipeline combining a dimension-reduction technique, *Uniform Manifold Approximation and Projection (UMAP)*, with *Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)*, was used to identify benthic habitat states. Spatial and temporal variation in habitat distribution was then explored across ecoregions.

Results : The *UMAP-HDBSCAN* pipeline identified 17 distinct clusters representing different benthic habitats and gradients of ecological state. Certain habitat states displayed clear biogeographic patterns, predominantly occurring in temperate regions or tropical waters. Notably, some reefs dominated by turf algae, were ubiquitous regardless of latitudes. Transition zones between temperate and tropical waters emerged as spatial hotspots of habitat state diversity. Temporal patterns revealed, changes in the proportion of certain states showing variations over time, notably an increase in turf algae occurrence.

Main Conclusions : The *UMAP-HDBSCAN* clustering pipeline effectively characterised fine-scale benthic habitat states at a global scale, confirming known broader biogeographic patterns, including the importance of temperate-tropical transition zones as hotspots of habitat state diversity. This fine-scale, yet broadly-scalable habitat classification could be applied as a standardised template for tracking benthic habitat change across space and time at a global scale. The *UMAP-HDBSCAN* pipeline has proven to be a powerful and versatile approach for analysing complex biological datasets and can be applied in various ecological domains.

3.2 Introduction

Benthic habitats contribute to marine coastal ecosystems functioning and the services they provide (BARBIER et al., 2011). More specifically, they contribute to shoreline protec-

tions (BARBIER, 2017), carbon sequestration (FOURQUREAN et al., 2012), support commercial fisheries (BARBIER, 2017) and host diverse species and communities (SUNDAY et al., 2017). As modifiers to abiotic substrates, foundation species, such as kelp, seagrass, and corals engineer biogenic habitats that contribute to specific functions of coastal ecosystems (ELITH & LEATHWICK, 2009). For instance, the tridimensional structure of coral reefs can shelter fish assemblages from predators (HIXON & BEETS, 1993); seaweed or mussel beds can buffer environmental conditions (JURGENS et al., 2022; WHITAKER et al., 2023); and kelp forests are both habitat and food sources for various fish and invertebrate species (EDGAR et al., 2004). Thus, changes in coastal benthic habitats have direct cascading consequences on marine ecosystem structure, functioning and services.

Being hotspots of human activity, coastal ecosystems can be adversely affected by multiple anthropogenic stressors (BOWLER et al., 2020; HALPERN et al., 2019). Global climate change can also lead to fast changes in coastal abiotic conditions (BURROWS et al., 2014; BOWLER et al., 2020). The impact of these multiple stressors on benthic communities and ecosystems are mostly mediated by the response of biogenic habitats like kelp, seagrass, or coral (HARLEY et al., 2006; ROCHA et al., 2015b). For example, in the vicinity of urban areas, eutrophication can induce the replacement of kelp forests by turf algae (FILBEE-DEXTER & WERNBERG, 2018; PESSARRODONA et al., 2021), marine heatwaves can lead to coral bleaching, and their intensification in magnitude and frequency can induce long-term decline in tropical coral reefs (BELLWOOD et al., 2004), and overfishing of herbivorous fish in these reefs can lead to the overgrowth of macroalgae on top of coral reefs (HUGHES et al., 2007). Hence, habitat changes are amongst the greatest symptoms of anthropogenic impacts on shallow marine systems, with large consequences for marine biodiversity (ROCHA et al., 2015a). As anthropogenic stressors on marine ecosystems tend to increase and diversify (HALPERN et al., 2019), habitat changes will likely become more frequent in the future at a global scale (CONVERSI et al., 2015). Yet, both the drivers of habitat change and its consequence on biodiversity remain largely understudied in marine ecosystems (MAZOR et al., 2018).

Detecting and anticipating future habitat changes in benthic ecosystems requires a thorough understanding of the current state and distribution of benthic habitats and characterising of underlying drivers across multiple scales. Currently, detailed knowledge of habitat distribution is mostly local, i.e. at scales ranging from study sites (10 m - 100 m) or bay (100 m - 10 km) up to regions (10 km - 100 km) (e.g. ROBERT et al.

(2015); WICAKSONO et al. (2019)), for instance using remote sensing, acoustic surveys, or rather local monitoring of abiotic conditions and communities (COSTELLO, 2009). At larger scales, habitat distribution maps are largely based on physical, geomorphological and biogeochemical ocean properties (e.g. BROWN et al. (2011); LECOURS et al. (2015); SONNEWALD et al. (2020)). At such scales, habitat maps either disregard biogenic habitats or remain species-specific. Indeed, studies that focus on biogenic habitat distribution tend to consider specific habitat-formers independently (ASSIS et al., 2020; MCKENZIE et al., 2020) and rarely provide information on community composition. Large-scale seafloor habitat maps of either abiotic, or biogenic features also tend to integrate data over large timescales (e.g. decades). Their aim is rather to provide a static picture of the potential distribution of these habitats than to infer changes in habitat states and distribution. Knowledge of benthic habitat changes thus remains highly regional (e.g. CATTANO et al. (2020)). In that context several global studies have collated heterogeneous regional monitoring data to document changes in emblematic habitat-formers, such as seagrass spp. (WAYCOTT et al., 2009; DUNIC et al., 2021), kelp beds (KRUMHANSL et al., 2016; FILBEE-DEXTER & WERNBERG, 2018) or coral reefs (EDDY et al., 2021). Yet, these independent studies on specific habitat-formers are not sufficient to gain a comprehensive understanding of how the seafloor habitat mosaic has changed over the last decade at a global scale and how it is now changing in the face of anthropogenic pressures. Our understanding of current changes in seafloor habitat mosaic is impeded by the lack of large-scale, standardised, data-driven definition and maps of benthic habitat and their potential states.

Identifying changes in benthic habitat states through space and time requires a standardised workflow from data collection through to systematic statistical discrimination between habitat states. In this study, we aim to develop a data-driven pipeline that distinguishes different iconic benthic habitats observed spatially, and apply this to characterise stepwise changes in habitat ecological states through time. Because scientific monitoring programmes are often expensive (EDGAR et al., 2016) and restricted in their spatial and/or temporal coverage (RHODES et al., 2015), participatory science programmes have emerged as valuable means to increase monitoring programme coverage and resolution. In this study, we leverage the benefits of a citizen science program to characterise benthic habitat states at the global scale and overcome the limitations of traditional scientific programs. The *Reef Life Survey (RLS)* relies on standardised diver-based 50-metre-long transects to estimate fish and invertebrate species abundance as well as image-based percentage cover of coastal benthic habitats (EDGAR & STUART-SMITH, 2014a). Estimates of habitat

percentage cover have already proven useful for defining habitat states at a regional scale through the use of unsupervised machine learning techniques (CRESSWELL et al., 2017; PELLETIER et al., 2020). However, the methods proposed in these previous studies come with a number of limitations when upscaling these approaches at a global scale. In particular, the occurrence and abundance of habitat-forming species are expected to show non-linear responses to environmental changes (OKSANEN & MINCHIN, 2002), especially across large environmental gradients. Still, the clustering algorithm used for CRESSWELL et al. (2017) and PELLETIER et al. (2020) are not adapted to take into account the non-linear nature of the dominance patterns between different habitat-forming species.

Hence, we applied a new workflow, combining two algorithms to overcome these challenges : (1) *Uniform Manifold Approximation and Projection (UMAP)* a novel dimension reduction technique preserving complex nonlinear structures and patterns (MCINNES et al., 2020), (2) and the *Hierarchical Density-Based Spatial Clustering of Applications with Noise algorithm (HDBSCAN)* that can identify clusters of varying shapes and sizes while filtering out outlier noise (CAMPELLO et al., 2013 ; MCINNES et al., 2017). While previous ecological studies have successfully applied both *UMAP* for dimension reduction (MILOŠEVIĆ et al., 2022) and *HDBSCAN* for food web classification (OHLSSON & EKLÖF, 2020), our study represents a novel application to coastal marine habitats. We interpret classification results by combining the latest *SHapley Additive exPlanations (SHAP ; LUNDBERG et LEE (2017))* framework with visual inspections of photoquadrats associated with the most representative transects of the different clusters.

Therefore, the aim of this study is to characterise coastal benthic habitat states using a *UMAP-HDBSCAN* pipeline on the *RLS* habitat dataset. Using this pipeline, we identify and classify benthic habitat states at a global scale and characterise their spatial and temporal variability across biogeographical gradients as well as within bioregions.

3.3 Materials & Methods

We used a *UMAP-HDBSCAN* pipeline to cluster the global *RLS* benthic habitat dataset. In the following sections, we sequentially describe : (1) the data used in this study, (2) the clustering pipeline, (3) the interpretation of the identified clusters.

3.3.1 Data

Reef Life Survey photoquadrat dataset

The *RLS* (<http://www.reeflifesurvey.com/>) is a hybrid citizen science/professional researcher program monitoring reef communities around the world using scuba-diving visual census. Details about the survey methods, including protocols, diver training, data quality assurance and data management, are covered by EDGAR et STUART-SMITH (2014b). Here, we used estimates of relative cover of benthic habitats derived from in situ digital photoquadrats : along standardised 50 m transect, 20 photoquadrats, which each approximately covers $0.3 \text{ m} \times 0.3 \text{ m}$, are collected every 2.5 m (EDGAR et al., 2020). Images are then annotated using point counts on the *Squidle+* (<https://squidle.org/>) platform to estimate the percentage covers of about 50 substratum types and functional groups, based on the *CATAMI* benthic imagery classification scheme (ALTHAUS et al. (2015); for further details, see EDGAR et al. (2020)). Based on *RLS* specialists' expertise, these 50 original benthic habitat categories (see Appendix A, Table 1 in Supporting Informations) were grouped into 24 broader categories (Table 3.1) that more consistently capture the range of dominant coastal substratum available along *RLS* transects at the global scale.

We extracted the *RLS* photoquadrat dataset on 24 January 2023. From the original 8,154 transects, we removed partially scored transects. For transects annotated multiple times on *Squidle+* across various research projects, mean percentage cover estimates were considered. After fully curating the dataset, the photoquadrat dataset consisted of 6,554 transects across 2,249 sites over the world. All subsequent analyses were performed at the transect level to consider local-scale variation in the state of benthic habitats.

Table 3.1 – Description of the 24 categories used in this study to overall capture the diversity of habitat types sampled by Reef Life Surveys worldwide. The 50 original *RLS* categories were lumped into these 24 categories that represent ecologically-consistent groups associated with different levels of structural complexity.

Habitat Categories	Description
	<i>Erect algae</i>
Large canopy forming algae	Large overstorey algae forming a canopy, including kelps or large fucoids
Bushy Furoid like	Robust vertical leaf-like shaped brown algae
Other Brown algae	Thick or thin-sheet like vertical algae

Habitat Categories	Description
Red algae	Foliose vertical red algae
Green algae	Thin-sheet like, thick, or ribbon-like, vertical growth algae <i>Erect calcareous algae</i>
Geniculate coralline algae	Red vertical calcified segmented algae
Green calcified algae	Small calcified green algae <i>Encrusting algae</i>
Crustose coralline algae	Red algae forming a small calcified crust over hard substrate
Encrusting algae	Algae forming a leathery crust over a substrate <i>Mat-forming Algae</i>
Filamentous algae	Filamentous algae, epiphyte or rock-attached
Turf algae	Fine and mat-forming filamentous algae growing on hard substrate <i>Plant</i>
Seagrass	Vertical ribbon-like marine plant <i>Sessile invertebrates</i>
Encrusting corals	Stony corals forming a crust over hard substrate
Branching coral	Branching coral forming large colonies
Foliose/Plate corals	Stony corals forming tabular or foliaceous colonies
Massive corals	Stony corals characterised by large, ball- or boulder-shaped colonies with a compact structure
Large-polyp stony corals	Large lobed stony coral, usually free-living
Soft corals and gorgonians	Soft coral or gorgonian in the sub-class Octocorallia
Calcareous hydrocorals and octocorals	Branching or foliaceous coral-like
Other sessile invertebrates	Habitat-forming sessile invertebrates (e.g. sponges, ascidians, bryozoans or molluscs) excluding corals <i>Seabed Materials</i>
Dead coral	Dead attached coral skeleton

Habitat Categories	Description
Bare rocky substrate	Bare rock
Unconsolidated substrate	Gravel, shell, coral rubble
Sand	Sand and fine sediments

3.4 Clustering pipeline

To account for non-linear, high-dimensional and complex nature of the ecological data, we combined a graph theoretical dimension reduction technique and a density-based classification technique, which has successfully identified ecoprovinces using biogeochemical ocean data at a global scale (SONNEWALD et al., 2020). Among the set of methods available, we have chosen the *UMAP* algorithm (MCINNES et al., 2020) and the *HDBSCAN* algorithm (CAMPELLO et al., 2013; MCINNES et al., 2017) for dimension reduction and clustering, respectively.

3.4.1 Dimension reduction - *UMAP*

The *UMAP* algorithm is a non-linear reduction technique (MCINNES et al., 2020). Unlike more traditional methods applied in ecology such as *Principal Component Analysis (PCA)*, *UMAP* preserves both the local structure (preserving the distance between neighbouring points) and the global structure (preserving the distances between the most different points) of the raw dataset (MCINNES et al., 2020). These two key properties have been proven useful for reducing the dimension of complex genomic (DORRITY et al., 2020), or ecological (MILOŠEVIĆ et al., 2022) data prior to clustering. *UMAP* reduces the dimensionality of a dataset by first creating a high-dimensional graph that connects each data point to its k-nearest neighbours. Then, *UMAP* produces a low-dimensional representation of this high-dimensional graph that reflects the original dataset (MCINNES et al., 2020). *UMAP* requires a distance matrix to construct the initial k-nearest-neighbour graph. Here, we applied the Chord transformation to standardise percentage cover data as relative cover per transect before computing euclidean distances between transects (LEGENDRE & GALLAGHER, 2001). In addition to the choice of a suitable distance metric, two *UMAP* hyperparameters can influence dimension reduction. The first one is the num-

ber of neighbours $n_neighbors$ to consider when creating the k-nearest neighbour graph. Low $n_neighbors$ values will allow the embedding to preserve more of the local structure of the original distance matrix and larger ones will preserve more of the global structure (MCINNES et al., 2020). The second parameter is min_dist , which controls the packing density at which *UMAP* is allowed to clump similar points in the reduced dimensional space. A high value of min_dist will tend to preserve the overall topological structure of the data, while a low value allows *UMAP* to clump closely similar points on the embedding. The value of $n_neighbors$ has been tuned in this study, while the value of min_dist has been set to 0.0, since this value allows densification of the low-dimensional representation of the dataset, which is important before using a density-based classification algorithm (VERMEULEN et al., 2021).

3.4.2 Clustering - *HDBSCAN*

After embedding our data into a two-dimensional space, we clustered the generated projections of the data with the unsupervised hierarchical density-based clustering *HDBSCAN* algorithm that can provide both hard (i.e. samples are exclusively assigned to a single cluster) and soft (i.e. samples are assigned probabilities of belonging to the different clusters) clustering solutions. In addition to identifying clusters of various shapes and density from a dendrogram, this algorithm comes with several advantages in ecology both in terms of classification and interpretation: it can exclude noisy observations, which do not get assigned to any clusters, and can also highlight most representative members of each cluster (CAMPELLO et al., 2013; MCINNES et al., 2017).

The *HDBSCAN* clustering algorithm involves a few core steps. First, it computes the core distance for the k-nearest neighbours for all points in the dataset. Then, it computes the extended minimum spanning tree from a weighted graph, where the edges are weighted by the distance between two points while taking into account the density of points around them. Then, *HDBSCAN* builds a hierarchy from the extended minimum spanning tree by cutting it at different levels of density. If the cut results in the creation of clusters smaller than the minimal number of observations set by the user $min_cluster_size$, all points members of these clusters are declared as noise by the algorithm. The algorithm stops when it declares all points as noise and returns to the user a tree-like structure where each node corresponds to a cluster varying in shape and density (CAMPELLO et al., 2013; MCINNES et al., 2017). In this study, we tuned only one parameter for *HDBSCAN* : the

min_cluster_size, controlling for the minimal number of observations required to form a cluster and used the default parameters otherwise.

3.4.3 Evaluation of the clustering output

For this pipeline, we search the best combination of hyperparameters for both *UMAP* (*n_neighbors*) and *HDBSCAN* (*min_cluster_size*) using a complete grid search. We exhaustively explored results sensitivity to the two hyperparameters from 10 to 500 resulting in 241,081 models evaluated. The best combination was found by optimising both the quality of the embedding and the clustering, using two criteria. The *UMAP* embedding was evaluated with the trustworthiness metric (VENNA & KASKI, 2001), ranging from 0 to 1 (the higher the index the more the local structure of the original data is preserved). The quality of the clustering was evaluated with the *DBC*V, which measures both compactness within and separations between clusters (MOULAVI et al., 2014). The *DBC*V index, which ranges between -1 and 1, is appropriate to assess the quality clusters with varying shapes and densities (MOULAVI et al., 2014).

A previous fine-scale analysis by CRESSWELL et al. (2017) on a regional subset of this dataset yielded nine groups of habitats. We expected to find at least that many groups at the global scale and thus restrained our search of the best hyperparameter combinations to the solution yielding at least the same number of cluster than CRESSWELL et al. (2017). Among these solutions, we select the best combination of hyperparameter (*n_neighbors* = 400; *min_cluster_size* = 74) yielding the best performance in terms of both their trustworthiness and *DBC*V scores, while having the maximal number of clusters for a finer granularity.

3.4.4 Interpretation of the clusters

To interpret individual clusters identified with *UMAP-HDBSCAN*, we computed the mean percentage cover of each habitat in each cluster. Then we used the *SHAP* framework to further explore how potential nonlinear interactions between variables may determine clustering outcomes (LUNDBERG & LEE, 2017). Because of the computational cost of applying *SHAP* to our complete pipeline, we used a classification tree (BREIMAN et al., 1984) to approximate the clustering pipeline (i.e. predict label cluster membership based on the raw percentage cover variables) before applying the *SHAP* framework (LUNDBERG &

LEE, 2017). In order to train our classification tree, we used a stratified train-test split to ensure that the relative frequency of each cluster label is preserved in the train and test fold. The training and the test sets contain 80% and 20% of the data, respectively. Then, we used a minimal cost-complexity pruning algorithm to avoid overfitting of our classification tree (BREIMAN et al., 1984) and estimated classification error rates using the F1-score (VAN RIJSBERGEN, 1979). The classification error rates were satisfactory, F1-score of 0.99 and 0.94 on the train and test sets respectively. Based on the *SHAP* values that estimate the influence of each variable to cluster definition, we examined potential interactions between the two most characteristic variables for each cluster by performing a piecewise linear interpolation of the *SHAP* values. Finally, we completed interpretation by extracting the photoquadrats for these transects considered by *HDBSCAN* as the most representative members of their cluster.

3.4.5 Spatio-temporal distribution of benthic habitat states

We first explored the latitudinal distribution of each cluster. We also summarised their occurrence within each of the *Marine Ecoregions of the World (MEOW ; SPALDING et al. (2007))* sampled by the *RLS*. In addition to examining dominant clusters per ecoregion, we also computed the proportion of transects classified as noise, as well as the Gini-Simpson diversity index. We chose this diversity index because it focuses on changes in dominance patterns, more indicative of changes in landscapes and is more robust to low sampling issues than other diversity indices (LANDE et al., 2000).

We identified sites within the *RLS* that had been sampled at least twice. We split our dataset into two equal sets (corresponding to the 2008-2013 and the 2014-2021 periods) based on the median number of transects carried out over the study period in the most surveyed ecoregions (i.e. where more than 30 transects were performed in each period). As a result, only five ecoregions in South-East Australia were analysed. We investigated temporal changes between the two periods by comparing the proportion of habitat identified in each ecoregion, as well as overall when pooling all five ecoregions together.

3.5 Results

Based on extensive exploring of hyperparameter space, both trustworthiness score of 0.98 ± 0.002 (mean \pm sd) and *DBC*V score of 0.46 ± 0.08 (mean \pm sd) for solution

containing at least 9 group suggest that these solutions yield reliable clustering of the *RLS* photoquadrat dataset (Annexe B Fig. S1 in Supporting Informations). Across these best solutions, optimal number of clusters varied between 9 and 184 (22.81 ± 18.37 ; mean \pm sd) while mean number of points classified as noise was $2,207.33 \pm 364.95$ (mean \pm sd). Hereafter, out of these, we chose to focus on the single solution yielding the highest resolution (i.e. the greatest number of clusters; see Fig. 3.2 for a description of the habitat states uncovered), and the smaller number of transects classified as noise (1,464 transects) possible. This solution has a trustworthiness score of 0.98 for *UMAP* and a *DBC*V score of 0.60 for *HDBSCAN*. The number of clusters identified by this set of hyperparameters is 17 (Fig. 3.2).

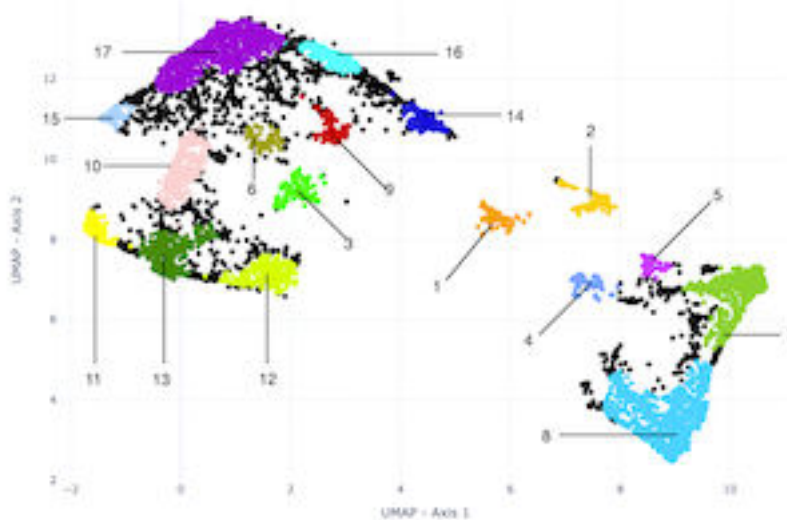


Figure 3.2 – Two-dimensional UMAP embedding of the benthic cover data of the 6,554 RLS transects. Each point corresponds to an RLS transect, coloured according to membership for the selected UMAP-HDBSCAN pipeline. Black dots represent points classified as noise ($n=1464$). The 17 clusters can be interpreted as follows (see Fig. 3.3 and S20-36) : 1. Foliose brown algae ($n=148$) 2. Filamentous algae ($n=208$) 3. Other Sessile invertebrates ($n=185$) 4. Foliose red algae ($n=123$) 5. Seagrass ($n=83$) 6. Soft coral and gorgonians ($n=98$) 7. Bushy fucoids ($n=577$) 8. Canopy forming algae ($n=894$) 9. Unconsolidated substrate ($n=151$) 10. Crustose coralline and turf algae ($n=286$) 11. Green calcified algae ($n=166$) 12. Bare substrates ($n=329$) 13. Crustose coralline algae ($n=409$) 14. Sand ($n=220$) 15. Branching coral ($n=110$) 16. Turf and sand ($n=207$) 17. Turf algae ($n=897$)

The 17 clusters identified can be summarised hereafter according to four broad groups

(Fig. 3.3; Fig. 3.4, see Fig S2-S19 for their distribution on the globe and S20-36 for their interpretation with *SHAP* framework in Supporting Informations) : (1) temperate habitats, (2) subtropical and tropical habitats, (3) broadly-distributed habitats and (4) opportunistic habitats (i.e. habitats with documented ecological dysfunctions - and therefore often habitats under strong anthropogenic influence, characterised by the presence of filamentous algal species or turf).

Transects within temperate regions can be classified according to five major clusters associated with contrasted dominance of sessile invertebrates, foliose red algae, seagrass, bushy foliose algae and canopy-forming algae, as follows : cluster 3 is dominated by at least 30% and on average 42% of sessile invertebrates. Cluster 4 is dominated by at least 40% coverage of foliose red algae. Cluster 5 is dominated by at least 30% and on average 40% seagrass. Cluster 7 is dominated by at least 20% coverage and an average of 56% fucoid bushy algae and an absence of canopy forming algae. Cluster 8 is characterised by a cover of at least 20% and an average of 55% of canopy forming algae with an absence of fucoid bushy algae.

Three clusters correspond to tropical and sub-tropical habitat types. Cluster 6 which is characterised by at least 30% and on average 37% of soft corals and gorgonians. Cluster 11 is composed of 20% coverage and an average of 35% green calcified algae. Finally cluster 15 is composed of at least 35% and on average 55% branching coral. Interestingly, this is the only group of corals identified in the dataset given the four categories of colony-forming corals.

Five clusters correspond to broadly-distributed habitats that can occur across both temperate and tropical latitudes. Cluster 1 is dominated by at least 30% and on average 46% brown foliose algae. Cluster 9 is dominated by the presence of at least 30% and on average 41% unconsolidated substrate. Cluster 12 has at least 30% and on average 42% bare substrate. Cluster 13 is characterised by 40% and on average 51% of crustose coralline algae with an absence of turf algae. Cluster 14 has at least 30% and an average of 53% sand without turf algae.

Finally, four clusters correspond to opportunistic habitats. Cluster 2 is in that respect dominated by at least 30% coverage and an average of 39% filamentous algae. Clusters 10, 11 and 17 are all dominated by turf algae. Cluster 10 is composed of at least 30% and on average 39% of turf algae and at least 20% and on average 28% of crustose coralline

algae. Cluster 16 is characterised by the presence of at least 30% and on average 48% turf algae and a minimum coverage of 20% and on average 26% sand. Cluster 17 is composed of at least 40% and on average 60% turf algae with an absence of crustose coralline.

The clusters identified by the *UMAP-HDBSCAN* pipeline show a marked latitudinal gradient (Fig. 3.4). *Red algae, filamentous algae, fucoïds, canopy-forming algae* and *seagrass* are essentially distributed overall in the temperate zones across latitudes higher than 25° (Fig. 3.4). Conversely, 4 habitat states, namely *soft corals and gorgonians, green calcified algae, sand and turf* and *branching coral* essentially occur in tropical latitudes (lower than 25°) (Fig. 3.4). However, some groups are relatively ubiquitous across all surveyed latitudes such as those associated with transects classified as *noise, bare* and *unconsolidated substrate, brown algae, crustose coralline algae* with and without *turf algae* and *turf algae* (Fig. 3.4).

The spatial distribution of transects sampled by *RLS* volunteers is particularly concentrated in Australia (Fig. 3.5 a). However, other areas such as the Caribbean, the Azores, and French Polynesia have also been extensively surveyed with more than 50 transects (Fig. 3.5 a). Globally, three habitat types dominate in terms of occurrences across all surveyed ecoregions, namely *bare substrate* (n = 20), *turf* (n = 17), and *canopy-forming algae* (n = 11). These three habitat types dominate in 37% of the ecoregions sampled by the *RLS* (Fig. 3.5 b). Two habitat types identified by the *UMAP-HDBSCAN* pipeline, *seagrass* and *red algae*, are not dominant in any of the world's ecoregions. The patterns of dominance of the different clusters also vary along the latitudinal gradient (Fig. 3.5 b), in line with the latitudinal distribution of each cluster (Fig. 3.4). These latitudinal variations of dominance are visible both at a global scale, but also along certain regions. For instance, a decrease in prevalence of sites in the canopy-forming algae cluster accompanies an increase in sites in the turf cluster along the coastline from southern to northern Australia (Fig. 3.5 b).

The proportion of noisy transects is highly heterogeneous across the globe (Fig. 3.5 c). Noisy transects represent 23% of all transects analysed, but are present in some areas more than in others. For example, in the Southern California Bight (western USA), Bight of Sofala/Swamp Coast (Eastern Africa), the Seychelles, and in Three Kings-North Cape (northern New Zealand), at least 60% of transects are classified as noisy (Fig. 3.5 c). While these four ecoregions share in common a low number of transects sampled (Fig. 3.5

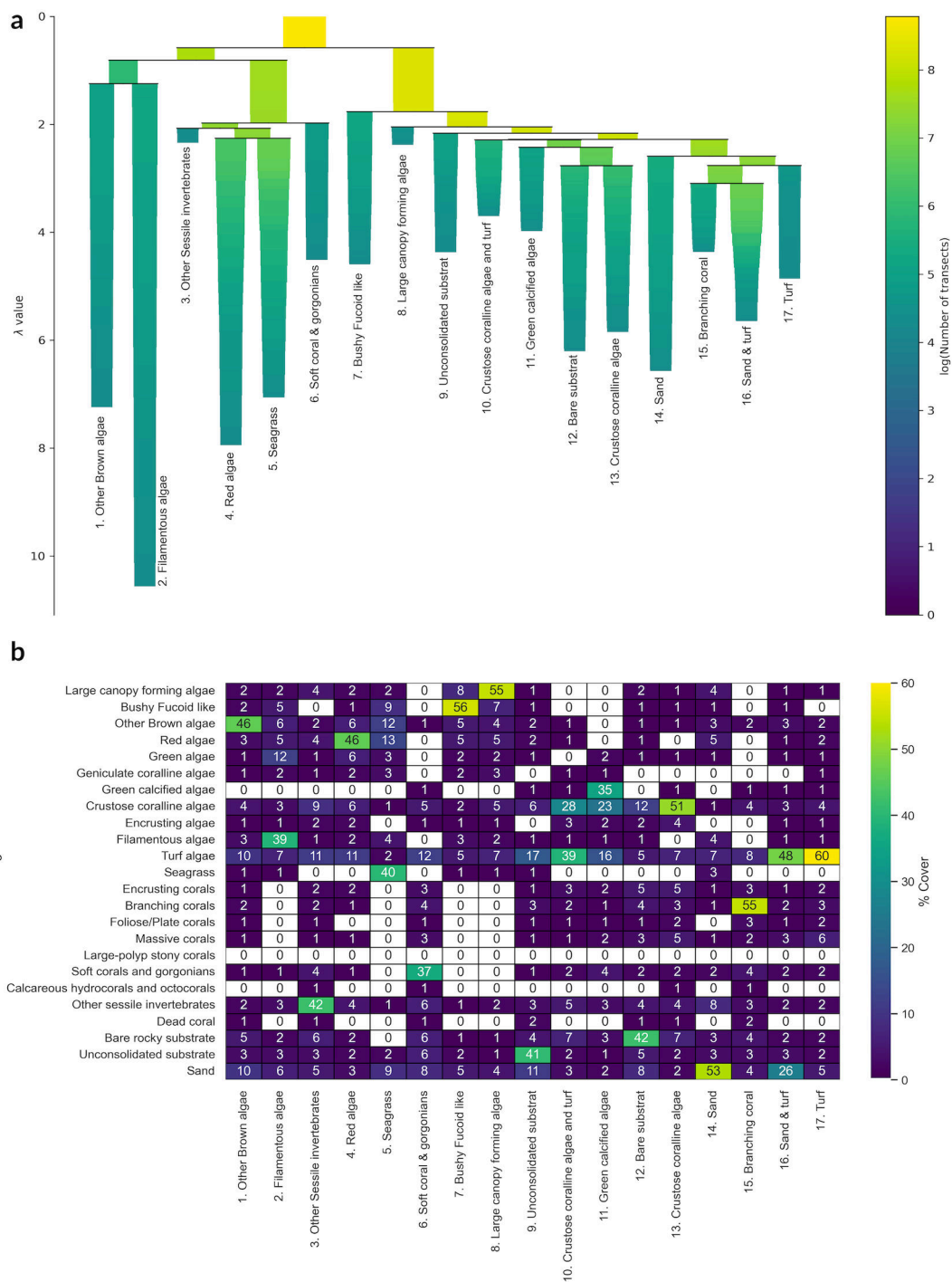


Figure 3.3 – a. HDBSCAN condensed clustering tree of the UMAP 2D embedding b. Heatmap of the mean substrate coverage (rounded to the nearest integer) for each cluster identified by the UMAP-HDBSCAN pipeline.

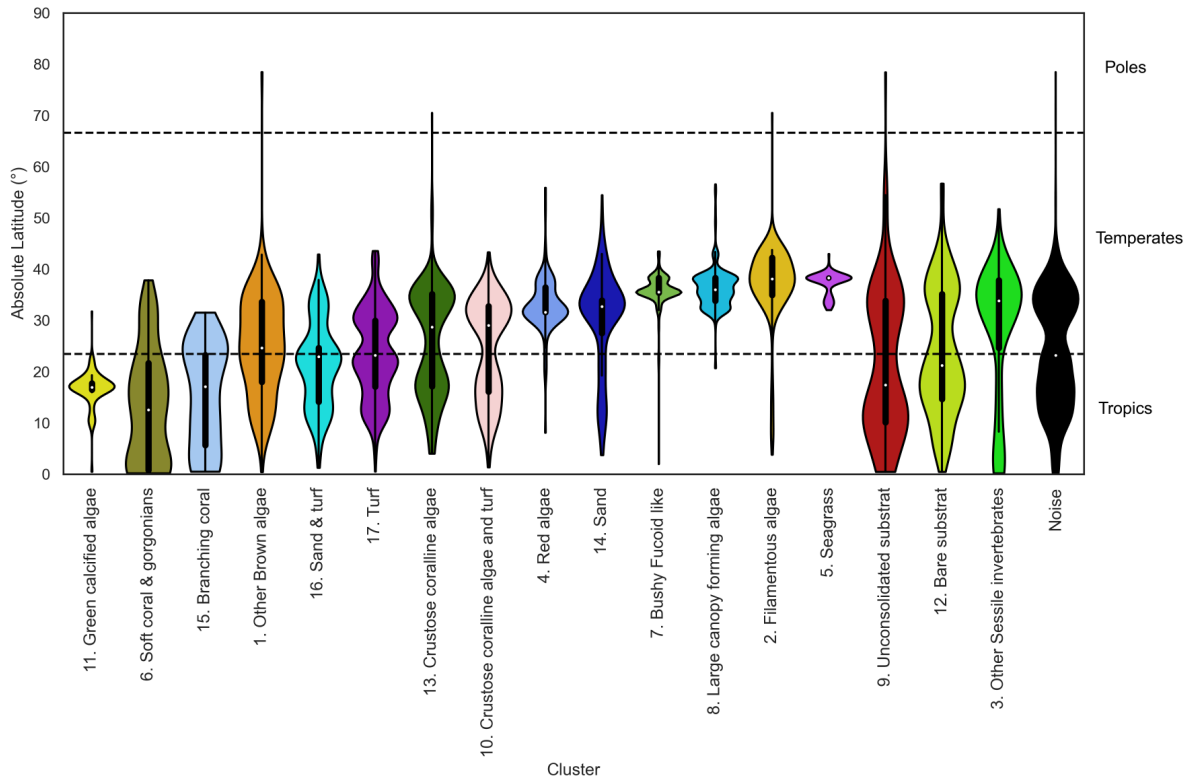


Figure 3.4 – Violin plot of the absolute latitudinal distribution of the different hard cluster solutions.

a), no significant correlation was found between the proportion of transects classified as noisy and the number of transects done in each ecoregion ($\tau_{Kendall} = 0.05$, $p = 0.54$; Fig. S37 Supporting Information). Moreover, 12 ecoregions sampled out of the 83 by the *RLS* had no transects that were classified as noisy (Fig. 3.5 c).

Areas with the highest diversity of habitat types, based on both the number of clusters occurring and on their relative proportions in the ecoregions, are concentrated in eastern and western Australia, as well as in the Caribbean and the Tuamotus (Fig. 3.5 d). Areas with the lowest Gini-Simpson values are the Southern California Bight (western USA) and Bight of Sofala/Swamp (eastern Africa) Coast with a Gini index of 0 (Fig. 3.5 d). It should be noted, nevertheless, that there is a weak correlation between the Gini-Simpson index and the number of transects carried out in the ecoregion ($\tau_{Kendall} = 0.29$, $p < 0.001$; Fig. S38 in Supporting Information).

At a finer spatial scale, it is also possible to identify spatial and temporal transitions

in the occurrence of the different clusters (Fig. 3.6). Along spatial gradients, clusters classified as noise may be a sign of the presence of an ecotone, as in the Cape Howe region in southeastern Australia (Fig. 3.6 b). These transects classified as noise separate an area with transects classified as bare substrate/crustose coralline algae to the north from an area to the south with transects classified as canopy forming algae and red algae (Fig. 3.6 b).

The relative proportion of the different clusters is overall stable overtime when comparing the early and the late 2010s. In this temperate zone, the most predominant cluster remains the canopy forming algae, followed by transects classified as noise (Fig. 3.6 c). Noticeable changes between the two periods include : the proportion of transects classified as sand decreased from 10% to below 8%, while the proportion of transects classified as *turf* increased from 7% to exceed 10% (Fig. 3.6 c & d). Transects classified as *turf* or *crustose coralline and turf* or *sand and turf*, represented approximately 16% of the clusters between 2008-2013 and increased to nearly 19% for the period 2015-2021. The proportion of transects classified as bare substrate declined from approximately 5% in 2008-2013 to only 1% of the transects in 2015-2021.

Chapitre 3 – From local seafloor imagery to global patterns in benthic habitats : contribution of citizen science to habitat classification across latitudes

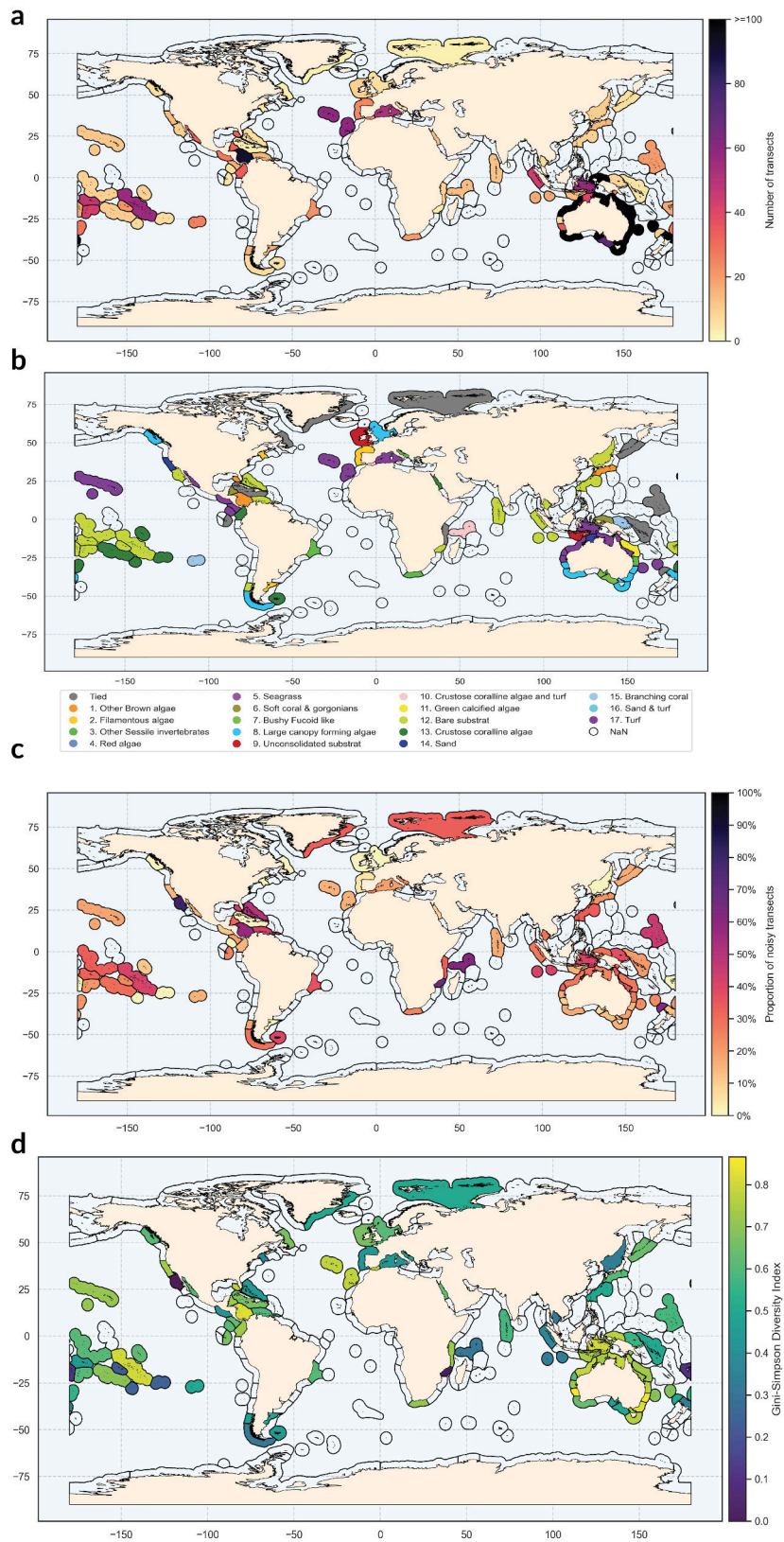


Figure 3.5 – a. Spatial distribution of reef surveys from the Reef Life Survey database used for analyses. b. Map of dominant clusters in each MEOW ecoregion. Dominant clusters were determined as the greatest count of transect labels in each ecoregion. c. Spatial distribution of the proportion of transects classified as noise in each ecoregion. d. Gini-Simpson diversity index calculated by the occurrence of clusters in each ecoregion of the world

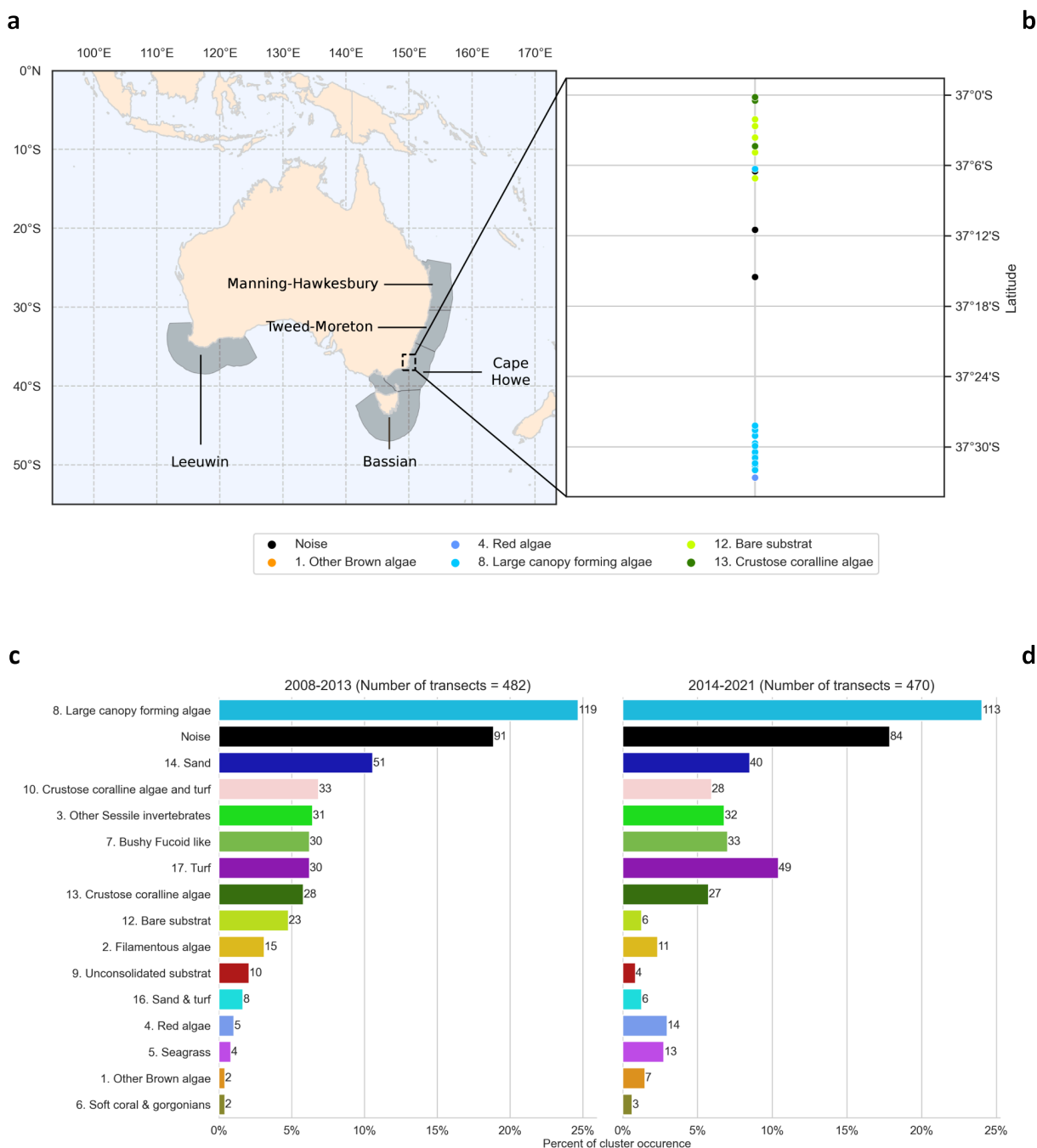


Figure 3.6 – Well-sampled ecoregions in Australia, with high number of transects and temporal replications between 2008 and 2021. *b.* Distribution of 44 sites surveyed between 2008 and 2021 in the Cape Howe region; Colour coding indicates cluster identity. Dots are jittered along the y-axis. *c.* Number and proportion of transects in the different clusters for the period 2008-2013 in the four ecoregions highlighted in *a.* *d.* The same analysis for the period 2014-2021.

3.6 Discussion

The *UMAP-HDBSCAN* clustering pipeline identified 17 distinct clusters within all the *RLS* transects performed globally across a range of coastal temperate and tropical regions. Within these groups, we found different biogenic habitats whose distribution patterns match with current biogeographic knowledge of benthic ecosystems : for example, *bushy furoid-like algae*, and *canopy-forming algae* predominantly occur in temperate waters (ASSIS et al., 2020 ; JAYATHILAKE & COSTELLO, 2020), while *soft corals and gorgonians*, and *branching coral* are more frequent in tropical waters (JONES et al., 2019 ; WIRABUANA et al., 2019). Our analysis also highlights habitat types that occur across the globe, including (1) different granulometric facies like *sand*, *unconsolidated substrate*, and *bare substrate*, as well as (2) different habitat types dominated by low-profile algae , such as *crustose coralline algae* or *turf algae*. The latter are known to occur across the globe and can dominate benthic substrates in diverse conditions (CONNELL et al., 2014 ; LIU et al., 2018).

In addition, this classification also distinguishes between different ecological states of these habitats (hereafter refers to as “habitat state”), including known alternative successional stages, or different degradation states of these habitats (Fig. 3.7). For example, the clusters *crustose coralline algae*, *crustose coralline algae and turf* and *turf* provide an interesting template to describe the habitat transitions described in CORNWALL et al. (2023), which suggests that a shift from *crustose coralline algae* to *turf* domination reduces reef carbonate production. Similarly, the clusters *branching coral*, *turf and sand* and *turf* can be used to describe and quantify in a standardised manner the transitions between corals and turf dominated habitats that are occurring more frequently due to anthropogenic pressures (JOUFFRAY et al., 2015). Fig. 3.6 b also illustrates the occurrence along the southeastern Australian coastline of alternative ecological states on temperate reefs, where dense macroalgal canopies dominated by *Ecklonia radiata* (here *large canopy-forming algae*), can shift to extensive barrens (here *bare substrate* or *crustose coralline algae*) following destructive grazing by the long-spined sea urchin *Centrostephanus rodgersii* (LING et al., 2009). Thus, our approach can classify reef cover data collected across the globe with the *RLS* protocol into an ecologically sound template to explore common reef habitat transitions under anthropogenic pressures (DONOVAN et al., 2018).

Certain of the habitat states identified here on the global *RLS* dataset match well with the habitat states previously identified by CRESSWELL et al. (2017), who applied another

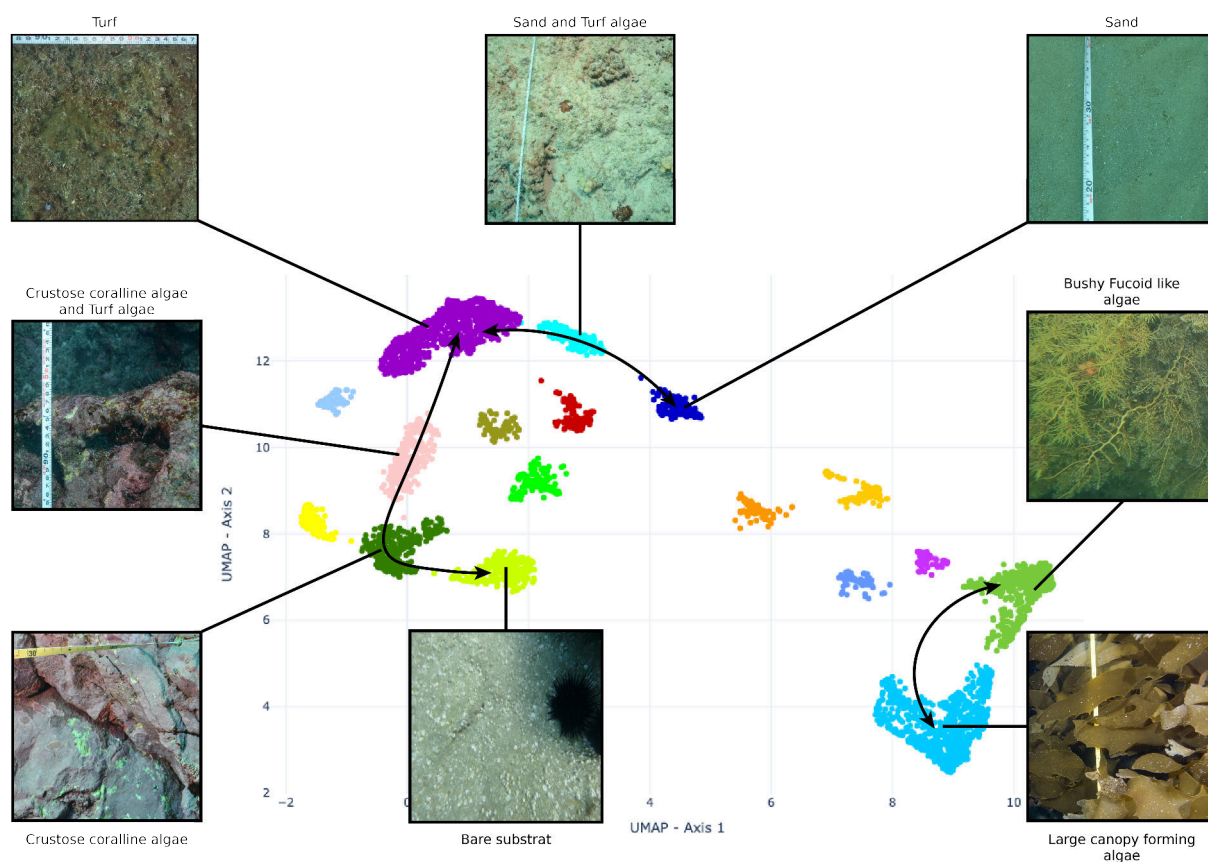


Figure 3.7 – Two-dimensional UMAP embedding of the 5,090 clustered RLS transects. Each picture is representative of its cluster. The arrows indicate potential transitions as identified in JOUFFRAY et al. (2015) and CORNWALL et al. (2023).

clustering approach to an Australian subset of the *RLS* dataset. In particular, some of our groups (i.e. *canopy forming algae*, *turf algae*, *filamentous algae* and *branching coral*), match well with four out of the nine habitat states identified by CRESSWELL et al. (2017) (i.e. “Canopy algae”, “Turf”, “Epiphytic filamentous algae-caulerpa” and “Coral”; see Table 1 in CRESSWELL et al. (2017) for a detailed description of their habitat states and Fig. 3.3 b in this paper for comparison). Furthermore, we identified more finely-resolved habitat states here relative to the classification proposed by CRESSWELL et al. (2017). The clusters *crustose coralline algae* and *bare substrate* identified here are amalgamated into a single “Barren” cluster in CRESSWELL et al. (2017), while the clusters *red algae* and *other brown algae* further detail what CRESSWELL et al. (2017) identified as one single “Foliose algae” group. Thus, our large-scale spatial approach overall confirms the habitat types

defined by CRESSWELL et al. (2017) while to some extent providing a more nuanced distinction between similar habitats. Nevertheless, our analysis only managed to capture one type of coral reef, contrary to what we might have expected, but there may be several reasons for this result. The proportion of coral cover within coral reefs exhibits significant variability, as indicated by (DE'ATH et al., 2012), although multiple sub-categories of coral have been merged together to circumvent this issue (see Appendix A Table S1). Additionally, the morphological diversity of these reefs is extensive, with variations in surface areas (ZAWADA et al., 2019). Such variability might elucidate why a singular group of corals, *Branching coral* is found as a group, since some species of *Acropora spp.* are able to establish colonies with expansive surface areas. Hence, our data-driven approach yields a finely resolved classification that comprises typical benthic habitats (e.g. seagrass meadows, coral reefs, kelp forests) that are common across all major seafloor habitat classification scheme (e.g. *European Nature Information System* ; BAJJOUK et al. (2015)).

Our global classification of the *RLS* data highlights hotspots of diversity in terms of benthic habitats and habitat states. Four ecoregions in particular, the Eastern (Manning-Hawkesbury ecoregion) and Western Australia (Houtman ecoregion), the Caribbean, and the Tuamotus Archipelago, showed a high diversity of habitat types (considering both richness and evenness). The high diversity of habitat types we report in the Caribbean and in the Tuamotus Archipelago potentially contribute to explain the high marine diversity of reef fishes reported in these tropical coral reefs, despite their small area at a global scale (COWMAN et al., 2017). In the transition zones between temperate and tropical waters, such as the Manning-Hawkesbury or the Houtman ecoregions in eastern and western Australia respectively, the high diversity of benthic habitat types we observe could be explained by a high diversity of foundation species. Indeed, high biodiversity is typical of transitional environmental conditions where ecological niches, which are overall disjointed, overlap (FERRO & MORRONE, 2014). This phenomenon is well known for multiple taxon such as birds (ALTAMIRANO et al., 2020), plants (LEMESSA et al., 2023) or reef fish (PINHEIRO et al., 2018) and also seems to apply to a certain extent to biogenic habitats like coralline red algae (SISSINI et al., 2022). Such subtropical or warm temperate zones are also identified as regions where both mobile fauna (VERGÉS et al., 2014) or sessile habitat-forming species assemblages (MARZLOFF et al., 2018) are likely to undergo tropicalisation, which implies that native temperate assemblages can co-occur with warmer species assemblages undergoing poleward climate-driven range shifts. Our finely resolved classification could be modelled against environmental predictors in future work to understand and predict

the state of benthic habitats under current and future conditions (e.g. BELANGER et al. (2012)).

Beyond exploring spatial patterns of benthic biodiversity, our classification of the *RLS* dataset offers a new perspective to explore temporal changes in benthic habitat states. Within the most frequently sampled ecoregions (essentially located in southeastern Australia), we quantified temporal changes in the occurrence of certain habitat types between the period 2008-2013 and 2014-2021. The proportion of transects classified as *large canopy-forming algae* shows an increase in the Manning-Hawkesbury ecoregion (Fig. S39 in Supplementary Informations) while it decreased in the Cape Howe ecoregion (Fig. S40 in Supplementary Informations). These regional differences can be explained by local changes in the environment (KRUMHANSL et al., 2016), including poleward expansion of urchin barrens (LING et al., 2018). We also observe an increase in transects classified as *turf algae* among the five resampled ecoregions (e.g. +67% increase between the two periods), in line with findings of an increase in *turf algae* due to both global climate change and local anthropogenic impacts (FILBEE-DEXTER & WERNBERG, 2018). Note, however, that temporal analysis of changes was restricted to only 5 southeastern Australian data-rich ecoregions out of a total of 232 worldwide. This was due to the distribution of data availability, which is heavily focussed within the region that *RLS* originated from (southern Australia). Moreover, marginal changes in the proportion of certain habitat states, such as *red algae* or *seagrass*, may also be due to the random positioning of transects within a site through time (as opposed to fixed). This is beneficial for greater site replication and for application by citizen scientists, but adds an extra source of noise to observations within sites through time.

Overall benthic habitat changes may reflect a range of processes, including ecological ones such as temporal variability in the cover of habitat-forming species (e.g. WERNBERG et al. (2016)) in relation to climate-driven environmental changes (i.e. tropicalisation of tropical-temperate transition zones (HORTA E COSTA et al., 2014), marine heatwaves (WERNBERG et al., 2016)) or to trends in human stressors (i.e. nutrients and organic pollution runoffs, impacts from coastal human populations; HALPERN et al. (2019)), as well as methodological ones, such as variability in transect location or in sampling effort through time (e.g. STUBLE et al. (2021)). Identifying the processes driving the observed habitat transitions could help better characterise the impact of anthropogenic activities on benthic habitats (see for example DONOVAN et al. (2018) for a similar approach at a finer

spatial scale). Our classification could thus provide an interesting template to further explore changes in benthic habitats across the world (EDGAR et al., 2023).

Nonetheless, not all expected transitions between habitats, or alternative ecological states, come out in the different clusters. Some transitory states may be classified as noise if they are too scarcely observed in the dataset to constitute a cluster of their own. Understanding the drivers behind the transects classified as noise can reveal valuable information about the factors influencing habitat variability and the ecological processes driving shifts between different states. This includes deciphering the reasons for a noise classification, such as variations in environmental conditions, biotic interactions, or anthropogenic disturbances. By investigating these aspects, researchers can gain crucial insights into the dynamics and transitions occurring between habitat states and alternative ecological states.

The *UMAP-HDBSCAN* clustering pipeline presented in this study demonstrates remarkable robustness and versatility, leveraging global data to identify fine-scale patterns within coastal temperate and tropical ecosystems. Because of its hierarchical structure, this pipeline aligns well with established classification standards and facilitates a first data-driven description of global patterns in habitat states, which constitutes a valuable database to explore the influence of local and global drivers of benthic habitat states. Additionally, the pipeline’s capability to handle non-linear data and accommodate noise underscores its adaptability to various ecological contexts and data sources (e.g. citizen science).

Bibliographie

- ALTAMIRANO, T. A., de ZWAAN, D. R., IBARRA, J., WILSON, S., & MARTIN, K., (2020), Treeline ecotones shape the distribution of avian species richness and functional diversity in south temperate mountains, *Scientific Reports*, *101*, 18428.
- ALTHAUS, F., HILL, N., FERRARI, R., EDWARDS, L., PRZESLAWSKI, R., SCHÖNBERG, C. H. L., STUART-SMITH, R., BARRETT, N., EDGAR, G., COLQUHOUN, J., TRAN, M., JORDAN, A., REES, T., & GOWLETT-HOLMES, K., (2015), A Standardised Vocabulary for Identifying Benthic Biota and Substrata from Underwater Imagery : The CATAMI Classification Scheme, *PLOS ONE*, *1010*, 1-18.
- ASSIS, J., FRAGKOPOULOU, E., FRADE, D., NEIVA, J., OLIVEIRA, A., ABECASIS, D., FAUGERON, S., & SERRÃO, E. A., (2020), A fine-tuned global distribution dataset of marine forests, *Scientific Data*, *71*, 119.

- BAJJOUK, T., GUILLAUMONT, B., MICHEZ, N., THOUIN, B., CROGUENNEC, C., POPULUS, J., LOUVEL-GLASER, J., GAUDILLAT, V., CHEVALIER, C., TOUROLLE, J., & HAMON, D., (2015), *Classification EUNIS, Système d'information européen sur la nature : Traduction française des habitats benthiques des Régions Atlantique et Méditerranée. Vol. 2. Habitats subtidaux & complexes d'habitats* (Report), FRANCE.
- BARBIER, E. B., (2017), Marine ecosystem services, *Current Biology*, 2711, R507-R510.
- BARBIER, E. B., HACKER, S. D., KENNEDY, C., KOCH, E. W., STIER, A. C., & SILLIMAN, B. R., (2011), The value of estuarine and coastal ecosystem services, *Ecological Monographs*, 812, 169-193.
- BELANGER, C. L., JABLONSKI, D., ROY, K., BERKE, S. K., KRUG, A. Z., & VALENTINE, J. W., (2012), Global environmental predictors of benthic marine biogeographic structure, *Proceedings of the National Academy of Sciences*, 10935, 14046-14051.
- BELLWOOD, D. R., HUGHES, T. P., FOLKE, C., & NYSTRÖM, M., (2004), Confronting the coral reef crisis, *Nature*, 4296994, 827-833.
- BOWLER, D. E., BJORKMAN, A. D., DORNELAS, M., MYERS-SMITH, I. H., NAVARRO, L. M., NIAMIR, A., SUPP, S. R., WALDOCK, C., WINTER, M., VELLEND, M., BLOWES, S. A., BÖHNING-GAESE, K., BRUELHEIDE, H., ELAHI, R., ANTÃO, L. H., HINES, J., ISBELL, F., JONES, H. P., MAGURRAN, A. E., ... BATES, A. E., (2020), Mapping human pressures on biodiversity across the planet uncovers anthropogenic threat complexes, *People and Nature*, 22, 380-394.
- BREIMAN, L., FRIEDMAN, J., STONE, C. J., & R.A., O., (1984), *Classification and Regression Trees*, Chapman ; Hall/CRC.
- BROWN, C. J., SMITH, S. J., LAWTON, P., & ANDERSON, J. T., (2011), Benthic habitat mapping : A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques, *Estuarine, Coastal and Shelf Science*, 923, 502-520.
- BURROWS, M. T., SCHOEMAN, D. S., RICHARDSON, A. J., MOLINOS, J. G., HOFFMANN, A., BUCKLEY, L. B., MOORE, P. J., BROWN, C. J., BRUNO, J. F., DUARTE, C. M., HALPERN, B. S., HOEGH-GULDBERG, O., KAPPEL, C. V., KIESSLING, W., O'CONNOR, M. I., PANDOLFI, J. M., PARMESAN, C., SYDEMAN, W. J., FERRIER, S., ... POLOCZANSKA, E. S., (2014), Geographical limits to species-range shifts are suggested by climate velocity, *Nature*, 5077493, 492-495.
- CAMPELO, R. J. G. B., MOULAVI, D., & SANDER, J., (2013), Density-Based Clustering Based on Hierarchical Density Estimates, In J. PEI, V. S. TSENG, L. CAO, H. MOTODA & G. XU (Éd.), *Advances in Knowledge Discovery and Data Mining* (p. 160-172), Springer Berlin Heidelberg.
- CATTANO, C., AGOSTINI, S., HARVEY, B. P., WADA, S., QUATTROCCHI, F., TURCO, G., INABA, K., HALL-SPENCER, J. M., & MILAZZO, M., (2020), Changes in fish communities due to benthic habitat shifts under ocean acidification conditions, *Science of The Total Environment*, 725, 138501.
- CONNELL, S., FOSTER, M., & AIROLDI, L., (2014), What are algal turfs? Towards a better description of turfs, *Marine Ecology Progress Series*, 495, 299-307.
- CONVERSI, A., DAKOS, V., GÅRDMARK, A., LING, S., FOLKE, C., MUMBY, P. J., GREENE, C., EDWARDS, M., BLENCKNER, T., CASINI, M., PERSHING, A., & MÖLLMANN, C., (2015), A holistic view of marine regime shifts, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130279.

Chapitre 3 – *From local seafloor imagery to global patterns in benthic habitats : contribution of citizen science to habitat classification across latitudes*

- CORNWALL, C. E., CARLOT, J., BRANSON, O., COURTNEY, T. A., HARVEY, B. P., PERRY, C. T., ANDERSSON, A. J., DIAZ-PULIDO, G., JOHNSON, M. D., KENNEDY, E., KRIEGER, E. C., MALLELA, J., MCCOY, S. J., NUGUES, M. M., QUINTER, E., ROSS, C. L., RYAN, E., SADERNE, V., & COMEAU, S., (2023), Crustose coralline algae can contribute more than corals to coral reef carbonate production, *Communications Earth & Environment*, *41*, 105.
- COSTELLO, M. J., (2009), Distinguishing marine habitat classification concepts for ecological data management, *Marine Ecology Progress Series*, *397*, 253-268.
- COWMAN, P. F., PARRAVICINI, V., KULBICKI, M., & FLOETER, S. R., (2017), The biogeography of tropical reef fishes : endemism and provinciality through time, *Biological Reviews*, *924*, 2112-2130.
- CRESSWELL, A. K., EDGAR, G. J., STUART-SMITH, R. D., THOMSON, R. J., BARRETT, N. S., & JOHNSON, C. R., (2017), Translating local benthic community structure to national biogenic reef habitat types, *Global Ecology and Biogeography*, *2610*, 1112-1125.
- DE'ATH, G., FABRICIUS, K. E., SWEATMAN, H., & PUOTINEN, M., (2012), The 27-year decline of coral cover on the Great Barrier Reef and its causes, *Proceedings of the National Academy of Sciences*, *10944*, 17995-17999.
- DONOVAN, M. K., FRIEDLANDER, A. M., LECKY, J., JOUFFRAY, J.-B., WILLIAMS, G. J., WEDDING, L. M., CROWDER, L. B., ERICKSON, A. L., GRAHAM, N. A. J., GOVE, J. M., KAPPEL, C. V., KARR, K., KITTINGER, J. N., NORSTRÖM, A. V., NYSTRÖM, M., OLESON, K. L. L., STAMOULIS, K. A., WHITE, C., WILLIAMS, I. D., & SELKOE, K. A., (2018), Combining fish and benthic communities into multiple regimes reveals complex reef dynamics, *Scientific Reports*, *81*, 16943.
- DORRITY, M. W., SAUNDERS, L. M., QUEITSCH, C., FIELDS, S., & TRAPNELL, C., (2020), Dimensionality reduction by UMAP to visualize physical and genetic interactions, *Nature Communications*, *111*, 1537.
- DUNIC, J. C., BROWN, C. J., CONNOLLY, R. M., TURSCHWELL, M. P., & CÔTÉ, I. M., (2021), Long-term declines and recovery of meadow area across the world's seagrass bioregions, *Global Change Biology*, *2717*, 4096-4109.
- EDDY, T. D., LAM, V. W., REYDONDEAU, G., CISNEROS-MONTEMAYOR, A. M., GREER, K., PALOMARES, M. L. D., BRUNO, J. F., OTA, Y., & CHEUNG, W. W., (2021), Global decline in capacity of coral reefs to provide ecosystem services, *One Earth*, *49*, 1278-1285.
- EDGAR, G. J., BARRETT, N. S., & MORTON, A. J., (2004), Biases associated with the use of underwater visual census techniques to quantify the density and size-structure of fish populations, *Journal of Experimental Marine Biology and Ecology*, *3082*, 269-290.
- EDGAR, G. J., BATES, A. E., BIRD, T. J., JONES, A. H., KININMONTH, S., STUART-SMITH, R. D., & WEBB, T. J., (2016), New Approaches to Marine Conservation Through the Scaling Up of Ecological Data, *Annual Review of Marine Science*, *81*, 435-461.
- EDGAR, G. J., COOPER, A., BAKER, S. C., BARKER, W., BARRETT, N. S., BECERRO, M. A., BATES, A. E., BROCK, D., CECCARELLI, D. M., CLAUSIUS, E., DAVEY, M., DAVIS, T. R., DAY, P. B., GREEN, A., GRIFFITHS, S. R., HICKS, J., HINOJOSA, I. A., JONES, B. K., KININMONTH, S., ... STUART-SMITH, R. D., (2020), Establishing the ecological basis for conservation of shallow marine life using Reef Life Survey, *Biological Conservation*, *252*, 108855.

- EDGAR, G. J., & STUART-SMITH, R. D., (2014a), Systematic global assessment of reef fish communities by the Reef Life Survey program, *Scientific Data*, 11, 140007.
- EDGAR, G. J., & STUART-SMITH, R. D., (2014b), Systematic global assessment of reef fish communities by the Reef Life Survey program, *Scientific Data*, 11.
- EDGAR, G. J., STUART-SMITH, R. D., HEATHER, F. J., BARRETT, N. S., TURAK, E., SWEATMAN, H., EMSLIE, M. J., BROCK, D. J., HICKS, J., FRENCH, B., BAKER, S. C., HOWE, S. A., JORDAN, A., KNOTT, N. A., MOONEY, P., COOPER, A. T., OH, E. S., SOLER, G. A., MELLIN, C., ... BATES, A. E., (2023), Continent-wide declines in shallow reef life over a decade of ocean warming, *Nature*, 6157954, 858-865.
- ELITH, J., & LEATHWICK, J. R., (2009), Species Distribution Models : Ecological Explanation and Prediction Across Space and Time, *Annual Review of Ecology, Evolution, and Systematics*, 401, 677-697.
- FERRIER, S., & GUISAN, A., (2006), Spatial modelling of biodiversity at the community level, *Journal of Applied Ecology*, 433, 393-404.
- FERRO, I., & MORRONE, J. J., (2014), Biogeographical transition zones : a search for conceptual synthesis, *Biological Journal of the Linnean Society*, 1131, 1-12.
- FILBEE-DEXTER, K., & WERNBERG, T., (2018), Rise of Turfs : A New Battlefield for Globally Declining Kelp Forests, *BioScience*, 682, 64-76.
- FOURQUREAN, J. W., DUARTE, C. M., KENNEDY, H., MARBÀ, N., HOLMER, M., MATEO, M. A., APOSTOLAKI, E. T., KENDRICK, G. A., KRAUSE-JENSEN, D., MCGLATHERY, K. J., & SERRANO, O., (2012), Seagrass ecosystems as a globally significant carbon stock, *Nature Geoscience*, 57, 505-509.
- HALPERN, B. S., FRAZIER, M., AFFLERBACH, J., LOWNDES, J. S., MICHELI, F., O'HARA, C., SCARBOROUGH, C., & SELKOE, K. A., (2019), Recent pace of change in human impact on the world's ocean, *Scientific Reports*, 91, 11609.
- HARLEY, C. D. G., RANDALL HUGHES, A., HULTGREN, K. M., MINER, B. G., SORTE, C. J. B., THORNBER, C. S., RODRIGUEZ, L. F., TOMANEK, L., & WILLIAMS, S. L., (2006), The impacts of climate change in coastal marine systems, *Ecology Letters*, 92, 228-241.
- HIXON, M. A., & BEETS, J. P., (1993), Predation, Prey Refuges, and the Structure of Coral-Reef Fish Assemblages, *Ecological Monographs*, 631, 77-101.
- HORTA E COSTA, B., ASSIS, J., FRANCO, G., ERZINI, K., HENRIQUES, M., J., G. E., & CASELLE, J. E., (2014), Tropicalization of fish assemblages in temperate biogeographic transition zones, *Marine Ecology Progress Series*, 504, 241-252.
- HUGHES, T. P., BELLWOOD, D. R., FOLKE, C. S., MCCOOK, L. J., & PANDOLFI, J. M., (2007), No-take areas, herbivory and coral reef resilience, *Trends in Ecology & Evolution*, 221, 1-3.
- JAYATHILAKE, D. R., & COSTELLO, M. J., (2020), A modelled global distribution of the kelp biome, *Biological Conservation*, 252, 108815.
- JONES, L. A., MANNION, P. D., FARNSWORTH, A., VALDES, P. J., KELLAND, S.-J., & ALLISON, P. A., (2019), Coupling of palaeontological and neontological reef coral data improves forecasts of biodiversity responses under global climatic change, *Royal Society Open Science*, 64, 182111.

Chapitre 3 – *From local seafloor imagery to global patterns in benthic habitats : contribution of citizen science to habitat classification across latitudes*

- JOUFFRAY, J.-B., NYSTRÖM, M., NORSTRÖM, A. V., WILLIAMS, I. D., WEDDING, L. M., KITTINGER, J. N., & WILLIAMS, G. J., (2015), Identifying multiple coral reef regimes and their drivers across the Hawaiian archipelago, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130268.
- JURGENS, L. J., ASHLOCK, L. W., & GAYLORD, B., (2022), Facilitation alters climate change risk on rocky shores, *Ecology*, 1032, e03596.
- KRUMHANSL, K. A., OKAMOTO, D. K., RASSWEILER, A., NOVAK, M., BOLTON, J. J., CAVANAUGH, K. C., CONNELL, S. D., JOHNSON, C. R., KONAR, B., LING, S. D., MICHELI, F., NORDERHAUG, K. M., PÉREZ-MATUS, A., SOUSA-PINTO, I., REED, D. C., SALOMON, A. K., SHEARS, N. T., WERNBERG, T., ANDERSON, R. J., ... BYRNES, J. E. K., (2016), Global patterns of kelp forest change over the past half-century, *Proceedings of the National Academy of Sciences*, 11348, 13785-13790.
- LANDE, R., DEVRIES, P. J., & WALLA, T. R., (2000), When species accumulation curves intersect : implications for ranking diversity using small samples, *Oikos*, 893, 601-605.
- LECOURS, V., DEVILLER, R., SCHNEIDER, D. C., LUCIEER, V. L., BROWN, C. J., & EDINGER, E. N., (2015), Spatial scale and geographic context in benthic habitat mapping : review and future directions, *Marine Ecology Progress Series*, 535, 259-284.
- LEGENDRE, P., & GALLAGHER, E. D., (2001), Ecologically meaningful transformations for ordination of species data, *Oecologia*, 1292, 271-280.
- LEMESSA, D., MEWDED, B., & ALEMU, S., (2023), Vegetation ecotones are rich in unique and endemic woody species and can be a focus of community-based conservation areas, *Botany Letters*, 00, 1-11.
- LING, S. D., BARRETT, N. S., & EDGAR, G. J., (2018), Facilitation of Australia's southernmost reef-building coral by sea urchin herbivory, *Coral Reefs*, 374, 1053-1073.
- LING, S. D., JOHNSON, C. R., FRUSHER, S. D., & RIDGWAY, K. R., (2009), Overfishing reduces resilience of kelp beds to climate-driven catastrophic phase shift, *Proceedings of the National Academy of Sciences*, 10652, 22341-22345.
- LIU, L.-C., LIN, S.-M., CARAGNANO, A., & PAYRI, C., (2018), Species diversity and molecular phylogeny of non-geniculate coralline algae (Corallinophycidae, Rhodophyta) from Taoyuan algal reefs in northern Taiwan, including *Crustaphytum* gen. nov. and three new species, *Journal of Applied Phycology*, 306, 3455-3469.
- LUNDBERG, S. M., & LEE, S.-I., (2017), A Unified Approach to Interpreting Model Predictions, In I. GUYON, U. V. LUXBURG, S. BENGIO, H. WALLACH, R. FERGUS, S. VISHWANATHAN & R. GARNETT (Éd.), *Advances in Neural Information Processing Systems 30* (p. 4765-4774), Curran Associates, Inc.
- MARZLOFF, M. P., OLIVER, E. C. J., BARRETT, N. S., HOLBROOK, N. J., JAMES, L., WOTHERSPOON, S. J., & JOHNSON, C. R., (2018), Differential vulnerability to climate change yields novel deep-reef communities, *Nature Climate Change*, 810, 873-878.
- MAZOR, T., DOROPOULOS, C., SCHWARZMUELLER, F., GLADISH, D. W., KUMARAN, N., MERKEL, K., DI MARCO, M., & GAGIC, V., (2018), Global mismatch of policy and research on drivers of biodiversity loss, *Nature Ecology & Evolution*, 27, 1071-1074.

- MCINNES, L., HEALY, J., & ASTELS, S., (2017), hdbscan : Hierarchical density based clustering, *The Journal of Open Source Software*, 211.
- MCINNES, L., HEALY, J., & MELVILLE, J., (2020), UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction, *arXiv*.
- MCKENZIE, L. J., NORDLUND, L. M., JONES, B. L., CULLEN-UNSWORTH, L. C., ROELFSEMA, C., & UNSWORTH, R. K. F., (2020), The global distribution of seagrass meadows, *Environmental Research Letters*, 157, 074041.
- MILOŠEVIĆ, D., MEDEIROS, A. S., STOJKOVIĆ PIPERAC, M., ČVIJANOVIĆ, D., SOININEN, J., MILOSAVLJEVIĆ, A., & PREDIĆ, B., (2022), The application of Uniform Manifold Approximation and Projection (UMAP) for unconstrained ordination and classification of biological indicators in aquatic ecology, *Science of The Total Environment*, 815, 152365.
- MOULAVI, D., JASKOWIAK, P. A., CAMPELLO, R. J. G. B., ZIMEK, A., & SANDER, J., (2014), Density-Based Clustering Validation, *Proceedings of the 2014 SIAM International Conference on Data Mining (SDM)*, 839-847.
- OHLSSON, M., & EKLÖF, A., (2020), Spatial resolution and location impact group structure in a marine food web, *Ecology Letters*, 2310, 1451-1459.
- OKSANEN, J., & MINCHIN, P. R., (2002), Continuum theory revisited : what shape are species responses along ecological gradients?, *Ecological Modelling*, 1572, 119-129.
- PELLETIER, D., SELMAOUI-FOLCHER, N., BOCKEL, T., & SCHOHN, T., (2020), A regionally scalable habitat typology for assessing benthic habitats and fish communities : Application to New Caledonia reefs and lagoons, *Ecology and Evolution*, 1014, 7021-7049.
- PESSARRODONA, A., FILBEE-DEXTER, K., ALCOVERRO, T., BOADA, J., FEEHAN, C. J., FREDRIKSEN, S., GRACE, S. P., NAKAMURA, Y., NARVAEZ, C. A., NORDERHAUG, K. M., & WERNBERG, T., (2021), Homogenization and miniaturization of habitat structure in temperate marine forests, *Global Change Biology*, 2720, 5262-5275.
- PINHEIRO, H. T., ROCHA, L. A., MACIEIRA, R. M., CARVALHO-FILHO, A., ANDERSON, A. B., BENDER, M. G., DI DARIO, F., FERREIRA, C. E. L., FIGUEIREDO-FILHO, J., FRANCINI-FILHO, R., GASPARINI, J. L., JOYEUX, J.-C., LUIZ, O. J., MINCARONE, M. M., MOURA, R. L., NUNES, J. d. A. C. C., QUIMBAYO, J. P., ROSA, R. S., SAMPAIO, C. L. S., ... FLOETER, S. R., (2018), South-western Atlantic reef fishes : Zoogeographical patterns and ecological drivers reveal a secondary biodiversity centre in the Atlantic Ocean, *Diversity and Distributions*, 247, 951-965.
- RHODES, C. J., HENRYS, P., SIRIWARDENA, G. M., WHITTINGHAM, M. J., & NORTON, L. R., (2015), The relative value of field survey and remote sensing for biodiversity assessment, *Methods in Ecology and Evolution*, 67, 772-781.
- ROBERT, K., JONES, D. O., TYLER, P. A., VAN ROOIJ, D., & HUVENNE, V. A., (2015), Finding the hotspots within a biodiversity hotspot : fine-scale biological predictions within a submarine canyon using high-resolution acoustic mapping techniques, *Marine Ecology*, 364, 1256-1276.
- ROCHA, J., YLETYINEN, J., BIGGS, R., BLENCKNER, T., & PETERSON, G., (2015a), Marine regime shifts : drivers and impacts on ecosystems services, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130273.

Chapitre 3 – *From local seafloor imagery to global patterns in benthic habitats : contribution of citizen science to habitat classification across latitudes*

- ROCHA, J. C., PETERSON, G. D., & BIGGS, R., (2015b), Regime Shifts in the Anthropocene : Drivers, Risks, and Resilience, *PLOS ONE*, 108, 1-16.
- SISSINI, M. N., KOERICH, G., de BARROS-BARRETO, M. B., COUTINHO, L. M., GOMES, F. P., OLIVEIRA, W., COSTA, I. O., de CASTRO NUNES, J., HENRIQUES, M. C., VIEIRA-PINTO, T., TORRANO-SILVA, B. N., OLIVEIRA, M. C., LE GALL, L., & HORTA, P. A., (2022), Diversity, distribution, and environmental drivers of coralline red algae : the major reef builders in the Southwestern Atlantic, *Coral Reefs*, 413, 711-725.
- SONNEWALD, M., DUTKIEWICZ, S., HILL, C., & FORGET, G., (2020), Elucidating ecological complexity : Unsupervised learning determines global marine eco-provinces, *Science Advances*, 622.
- SPALDING, M. D., FOX, H. E., ALLEN, G. R., DAVIDSON, N., FERDAÑA, Z. A., FINLAYSON, M., HALPERN, B. S., JORGE, M. A., LOMBANA, A., LOURIE, S. A., MARTIN, K. D., MCMANUS, E., MOLNAR, J., RECCHIA, C. A., & ROBERTSON, J., (2007), Marine Ecoregions of the World : A Bioregionalization of Coastal and Shelf Areas, *BioScience*, 577, 573-583.
- STUBLE, K. L., BEWICK, S., FISHER, M., FORISTER, M. L., HARRISON, S. P., SHAPIRO, A. M., LATIMER, A. M., & FOX, L. R., (2021), The promise and the perils of resurveying to understand global change impacts, *Ecological Monographs*, 912, e01435.
- SUNDAY, J. M., FABRICIUS, K. E., KROEKER, K. J., ANDERSON, K. M., BROWN, N. E., BARRY, J. P., CONNELL, S. D., DUPONT, S., GAYLORD, B., HALL-SPENCER, J. M., KLINGER, T., MILAZZO, M., MUNDAY, P. L., RUSSELL, B. D., SANFORD, E., THIYAGARAJAN, V., VAUGHAN, M. L. H., WIDDICOMBE, S., & HARLEY, C. D. G., (2017), Ocean acidification can mediate biodiversity shifts by changing biogenic habitat, *Nature Climate Change*, 71, 81-85.
- VAN RIJSBERGEN, C. J., (1979), *Information Retrieval* (2nd), Butterworths.
- VENNA, J., & KASKI, S., (2001), Neighborhood Preservation in Nonlinear Projection Methods : An Experimental Study, In G. DORFFNER, H. BISCHOF & K. HORNIK (Éd.), *Artificial Neural Networks — ICANN 2001* (p. 485-491), Springer Berlin Heidelberg.
- VERGÉS, A., TOMAS, F., CEBRIAN, E., BALLESTEROS, E., KIZILKAYA, Z., DENDRINOS, P., KARAMANLIDIS, A. A., SPIEGEL, D., & SALA, E., (2014), Tropical rabbitfish and the deforestation of a warming temperate sea, *Journal of Ecology*, 1026, 1518-1527.
- VERMEULEN, M., SMITH, K., EREMIN, K., RAYNER, G., & WALTON, M., (2021), Application of Uniform Manifold Approximation and Projection (UMAP) in spectral imaging of artworks, *Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy*, 252, 119547.
- VIOLET, C., (2020), *Caractériser la contribution des facteurs environnementaux et des interactions biotiques à la variabilité des assemblages benthiques* [mém. de mast., Rennes 1].
- WAYCOTT, M., DUARTE, C. M., CARRUTHERS, T. J. B., ORTH, R. J., DENNISON, W. C., OLYARNIK, S., CALLADINE, A., FOURQUIREAN, J. W., HECK, K. L., HUGHES, A. R., KENDRICK, G. A., KENWORTHY, W. J., SHORT, F. T., & WILLIAMS, S. L., (2009), Accelerating loss of seagrasses across the globe threatens coastal ecosystems, *Proceedings of the National Academy of Sciences*, 10630, 12377-12381.
- WERNBERG, T., BENNETT, S., BABCOCK, R. C., de BETTIGNIES, T., CURE, K., DEPCZYNSKI, M., DUFOIS, F., FROMONT, J., FULTON, C. J., HOVEY, R. K., HARVEY, E. S., HOLMES, T. H., KENDRICK, G. A., RADFORD, B., SANTANA-GARCON, J., SAUNDERS, B. J., SMALE, D. A.,

- THOMSEN, M. S., TUCKETT, C. A., ... WILSON, S., (2016), Climate-driven regime shift of a temperate marine ecosystem, *Science*, *353*6295, 169-172.
- WHITAKER, S. G., AMBROSE, R. F., ANDERSON, L. M., FALES, R. J., SMITH, J. R., SUTTON, S., & MILLER, R. J., (2023), Ecological restoration using intertidal foundation species : Considerations and potential for rockweed restoration, *Ecosphere*, *14*3, e4411.
- WICAKSONO, P., ARYAGUNA, P. A., & LAZUARDI, W., (2019), Benthic Habitat Mapping Model and Cross Validation Using Machine-Learning Classification Algorithms, *Remote Sensing*, *11*11.
- WIRABUANA, A., LITAAY, M., MOKA, W., & PRIOSAMBODO, D., (2019), Distribution of Soft Coral Octocorallia (Alcyonacea) in coastal waters of Gonda, Polewali-Mandar, West Sulawesi, *IOP Conference Series : Earth and Environmental Science*, *279*1, 012053.
- ZAWADA, K. J. A., DORNELAS, M., & MADIN, J. S., (2019), Quantifying coral morphology, *Coral Reefs*, *38*6, 1281-1292.

Appendix A - RLS Classification schema

Table S1 – Correspondence of the functional groups of substrates and the habitat groups used in this study.

Photoquadrat Original Categories	Final Classification
Ahermatypic corals	Sessile invertebrates
Ascidians	Sessile invertebrates
Ascidians (stalked)	Sessile invertebrates
Ascidians (unstaked)	Sessile invertebrates
Bare Rock	Bare substrate
Barnacles	Sessile invertebrates
Black & Octocorals	Soft corals and gorgonians
Bottlebrush Acropora corals	Branching corals
Bottlebrush Acropora corals_Bleached	Branching corals
Branching Acropora	Branching corals
Branching Acropora_Bleached	Branching corals
Branching corals	Branching corals
Branching corals_Bleached	Branching corals
Branching Pocillopora	Branching corals
Branching Pocillopora_Bleached	Branching corals
Bryozoa	Sessile invertebrates
Bryozoan (hard)	Sessile invertebrates
Bryozoan (soft)	Sessile invertebrates
Caulerpa	Green algae
Cnidaria	Sessile invertebrates
Cobble	Unconsolidated substrate
Colonial Anemones, Zoanthids and Corallimorphs	Sessile invertebrates
Columnar corals	Massive corals
Columnar corals_Bleached	Massive corals
Coral rubble	Unconsolidated substrate
Coral rubble with turf/encrusting algae	Turf
Corymbose Acropora corals	Branching corals
Corymbose Acropora corals_Bleached	Branching corals
Crustose coralline algae	Crustose coralline algae

Photoquadrat Original Categories	Final Classification
Dead coral	Bare substrate
Desmarestia and Himantothallus	Canopy forming algae
Digitate corals	Massive corals
Digitate corals_Bleached	Massive corals
Durvillaea	Canopy forming algae
Ecklonia radiata	Canopy forming algae
Encrusting corals	Encrusting corals
Encrusting corals_Bleached	Encrusting corals
Encrusting leathery algae	Encrusting leathery algae
Filamentous algae_epiphyte	Filamentous algae
Filamentous brown algae_epiphyte	Filamentous algae
Filamentous green algae_epiphyte	Filamentous algae
Filamentous red algae_epiphyte	Filamentous algae
Filamentous rock-attached algae	Filamentous algae
Foliose/Plate corals	Foliose/Plate corals
Foliose/Plate corals_Bleached	Foliose/Plate corals
Geniculate coralline algae	Geniculate coralline algae
Green calcified algae (Halimeda)	Green calcified algae (Halimeda)
Heliopora coerulea (blue coral)	Calcareous hydrocorals and octocorals
Hydrocoral	Calcareous hydrocorals and octocorals
Hydrocoral_Bleached	Calcareous hydrocorals and octocorals
Hydroids	Sessile invertebrates
Large brown laminarian kelps	Canopy forming algae
Large-polyp stony corals (free-living)	Large-polyp stony corals (free-living)
Large-polyp stony corals (free-living)_Bleached	Large-polyp stony corals (free-living)
Macroalgae	Canopy forming algae
Macroalgae_canopy forming	Canopy forming algae
Macrocystis	Canopy forming algae

Chapitre 3 – *From local seafloor imagery to global patterns in benthic habitats : contribution of citizen science to habitat classification across latitudes*

Photoquadrat Original Categories	Final Classification
Massive corals	Massive corals
Massive corals_Bleached	Massive corals
Medium foliose brown algae	Brown algae
Medium foliose green algae	Green algae
Medium foliose red algae	Red algae
Molluscs	Sessile invertebrates
Organ-pipe coral (Tubipora)	Calcareous hydrocorals and octocorals
Other fucoids	Canopy forming algae
Pebbles/gravel/shell	Unconsolidated substrate
Phyllospora	Canopy forming algae
Polychaete	Sessile invertebrates
Sand	Unconsolidated substrate
Seagrass (Halophila)	Seagrass
Seagrass (straplike)	Seagrass
Seagrasses	Seagrass
Sessile bivalves	Sessile invertebrates
Sessile gastropods	Sessile invertebrates
Slime (not trapping sediment)	Bare substrate
Small <2cm foliose algal cover (not trapping sediment)	Turf
Soft corals and gorgonians	Soft corals and gorgonians
Solitary Anemones	Sessile invertebrates
Sponges	Sessile invertebrates
Sponges (encrusting)	Sessile invertebrates
Sponges (erect)	Sessile invertebrates
Sponges (hollow)	Sessile invertebrates
Sponges (massive)	Sessile invertebrates
Stony corals	Encrusting corals
Stony corals_Bleached	Encrusting corals
Sub-massive corals	Massive corals
Sub-massive corals_Bleached	Massive corals
Substrate	Bare substrate

Photoquadrat Original Categories	Final Classification
Tabular Acropora corals	Foliose/Plate corals
Tabular Acropora corals_Bleached	Foliose/Plate corals
Turfing algae (<2 cm high algal/sediment mat on rock)	Turf
Worms	Sessile invertebrates

Appendix B - Clustering pipeline results

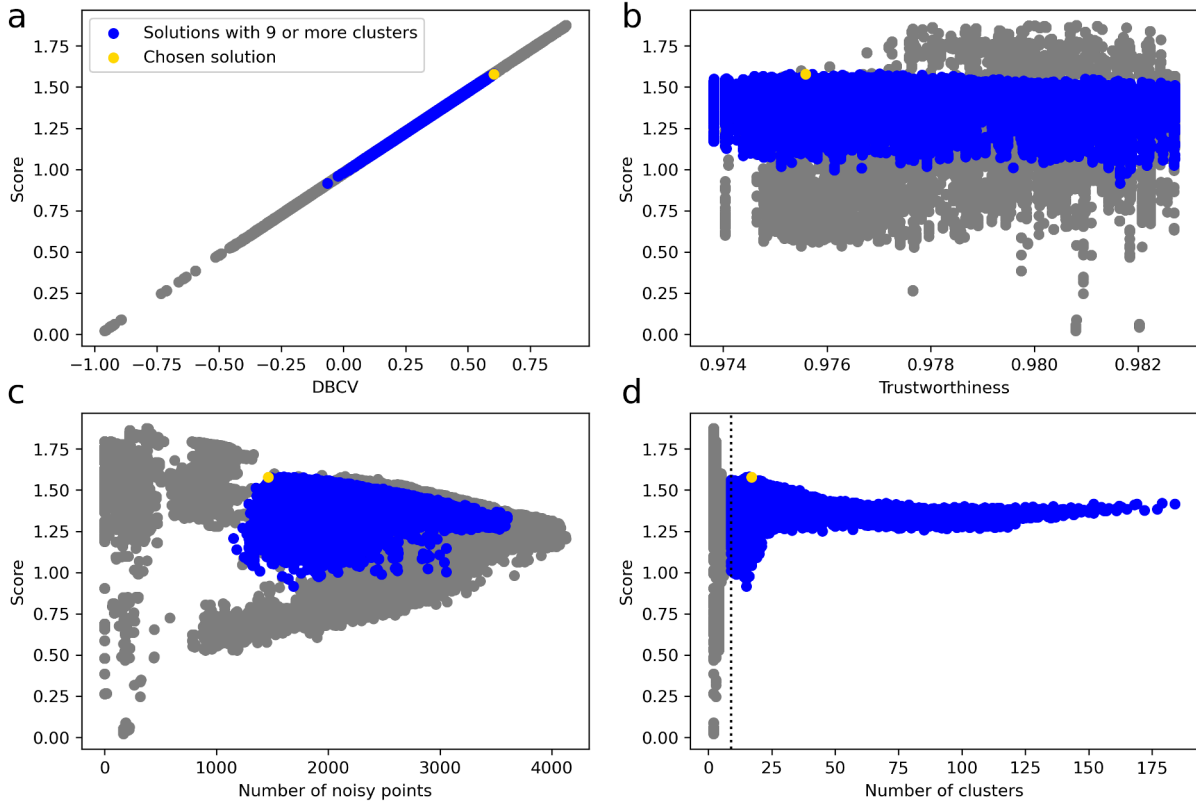


Figure S1 – Plot of the evolution of the score (i.e. sum of the DBCV value and the Trustworthiness) according to the different parameters measured (a. DBCV ; b. Trustworthiness ; c. Number of noisy points ; d. Number of clusters (the dotted line represents the number of clusters uncovered by CRESSWELL et al. (2017))). Each of the 241,100 dots represent the evaluation of a unique combination of hyperparameter values for UMAP (n_neighbors) and HDBSCAN (min_cluster_size). Blue dots represent the best solutions with the higher DBCV and Trustworthiness having at least 9 groups, while the yellow dot represents the best solution found according to our criteria. All other grey dots represent the other evaluated solutions

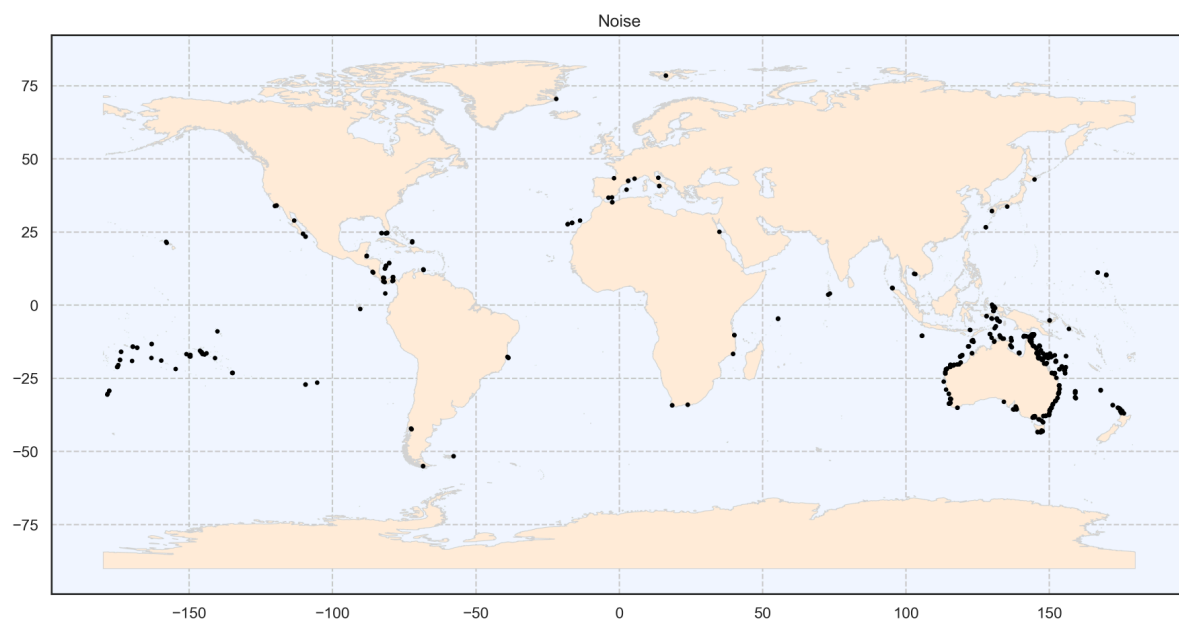


Figure S2 – *Spatial distribution of the cluster noise at the global scale. Each point represents a transect.*

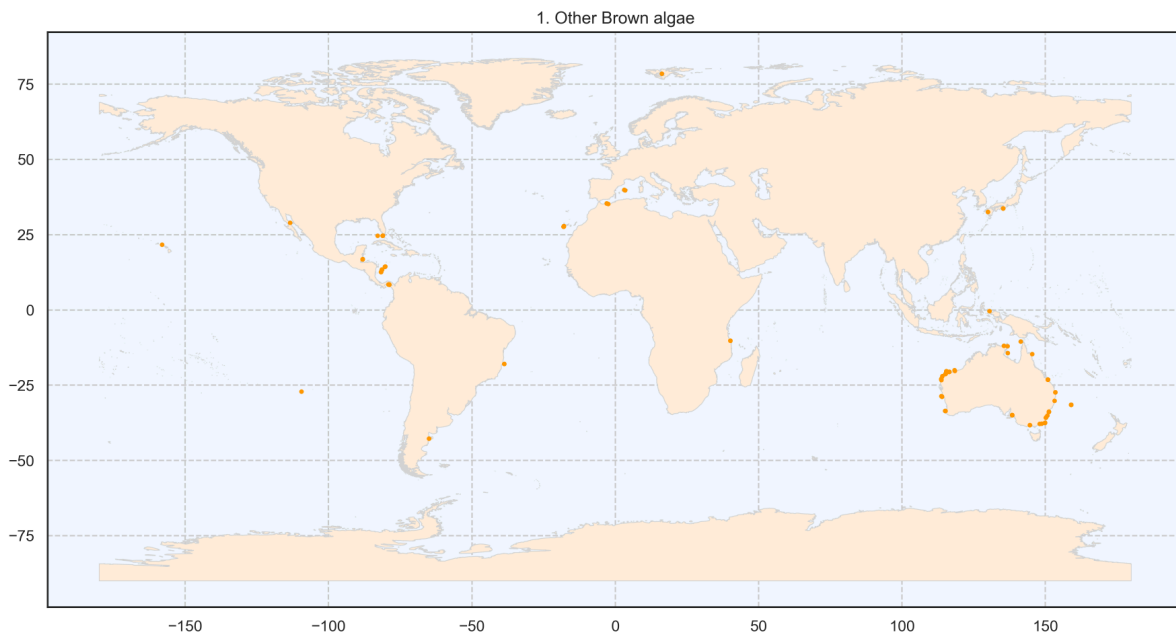


Figure S3 – *Spatial distribution of the cluster foliose brown algae at the global scale. Each point represents a transect.*

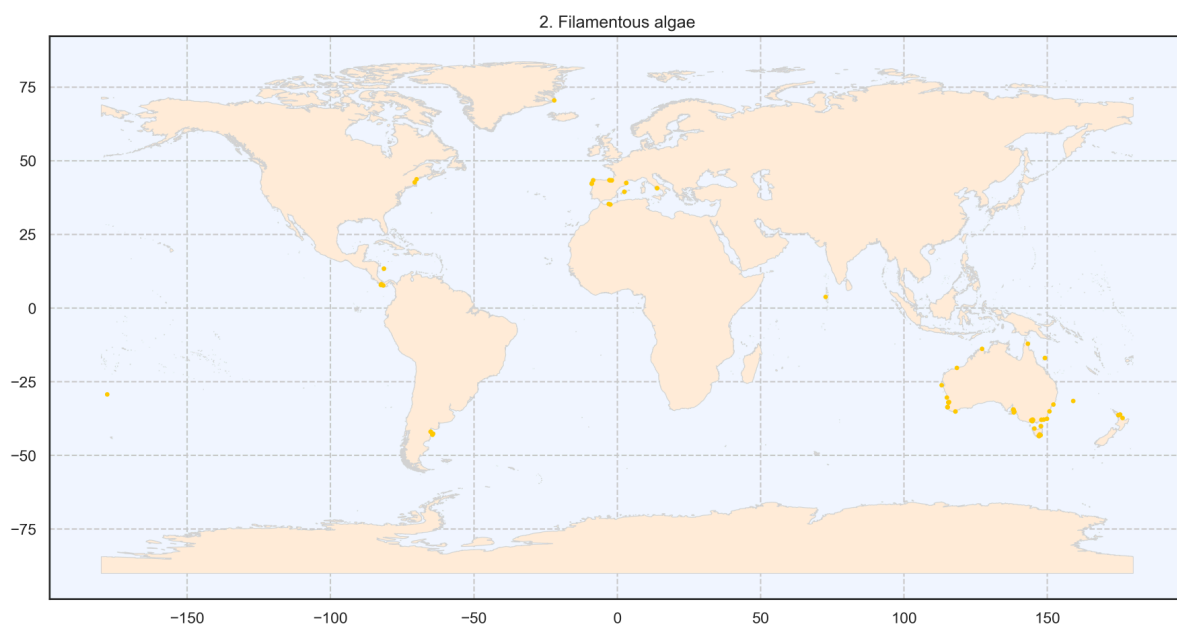


Figure S4 – *Spatial distribution of the cluster filamentous algae at the global scale. Each point represents a transect.*

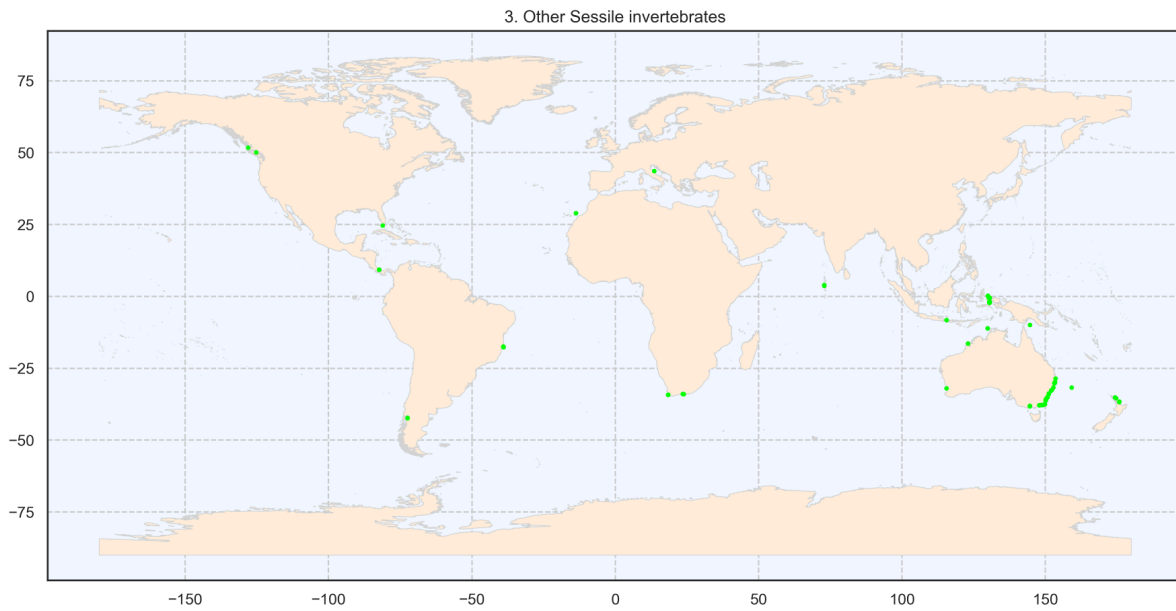


Figure S5 – *Spatial distribution of the cluster other Sessile invertebrates at the global scale. Each point represents a transect.*

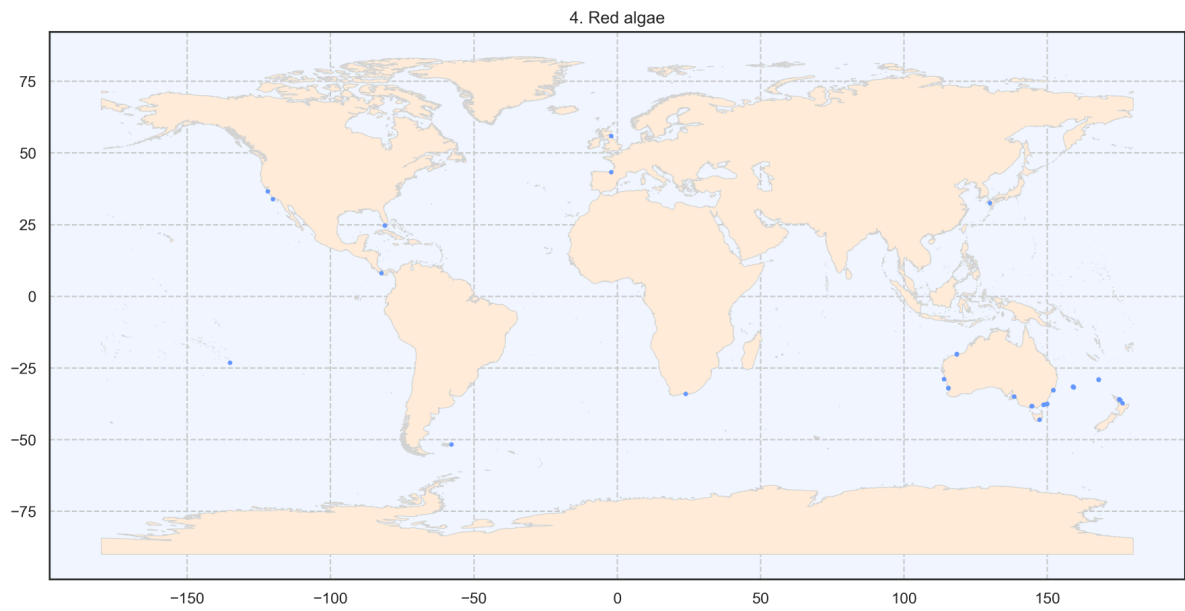


Figure S6 – *Spatial distribution of the cluster red algae at the global scale. Each point represents a transect.*

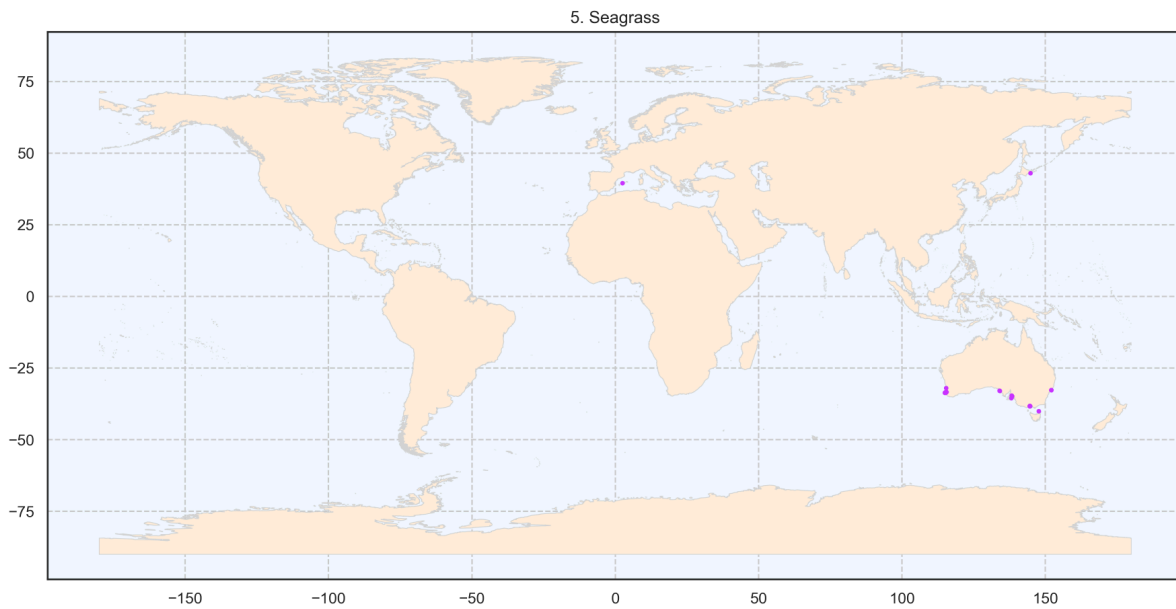


Figure S7 – *Spatial distribution of the cluster seagrass at the global scale. Each point represents a transect.*

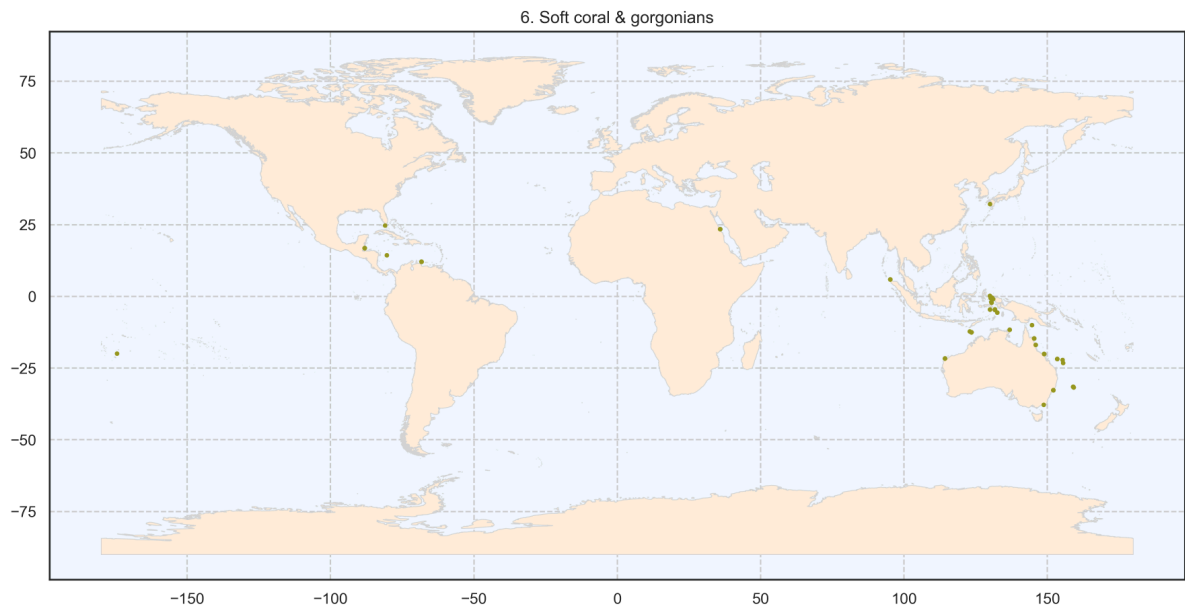


Figure S8 – *Spatial distribution of the cluster soft coral and gorgonians at the global scale. Each point represents a transect.*

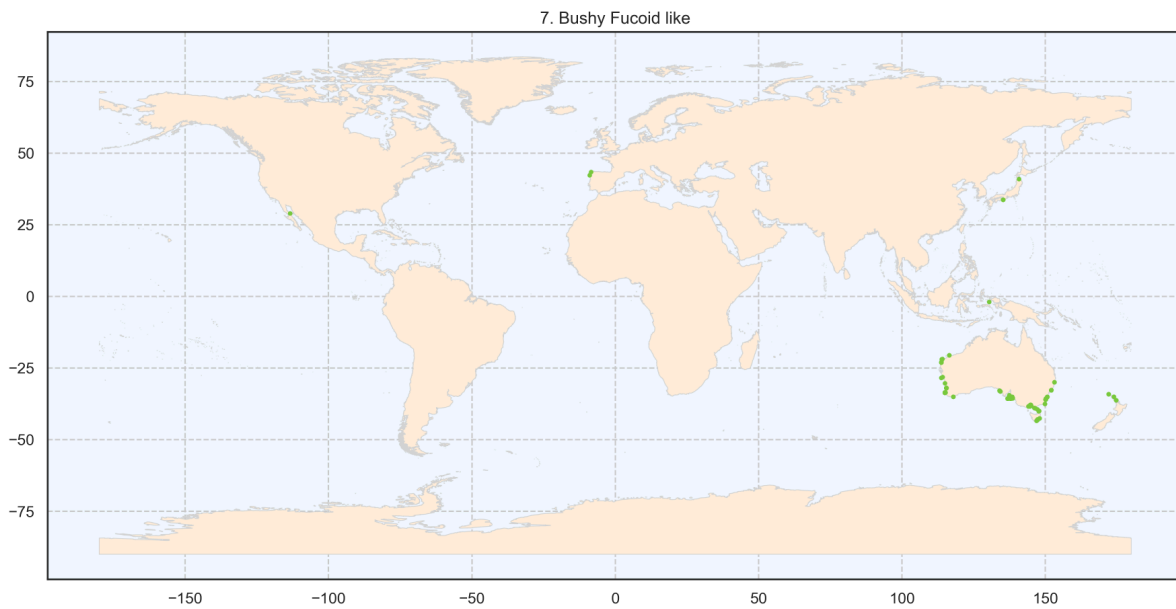


Figure S9 – *Spatial distribution of the cluster bushy furoid-like algae at the global scale. Each point represents a transect.*

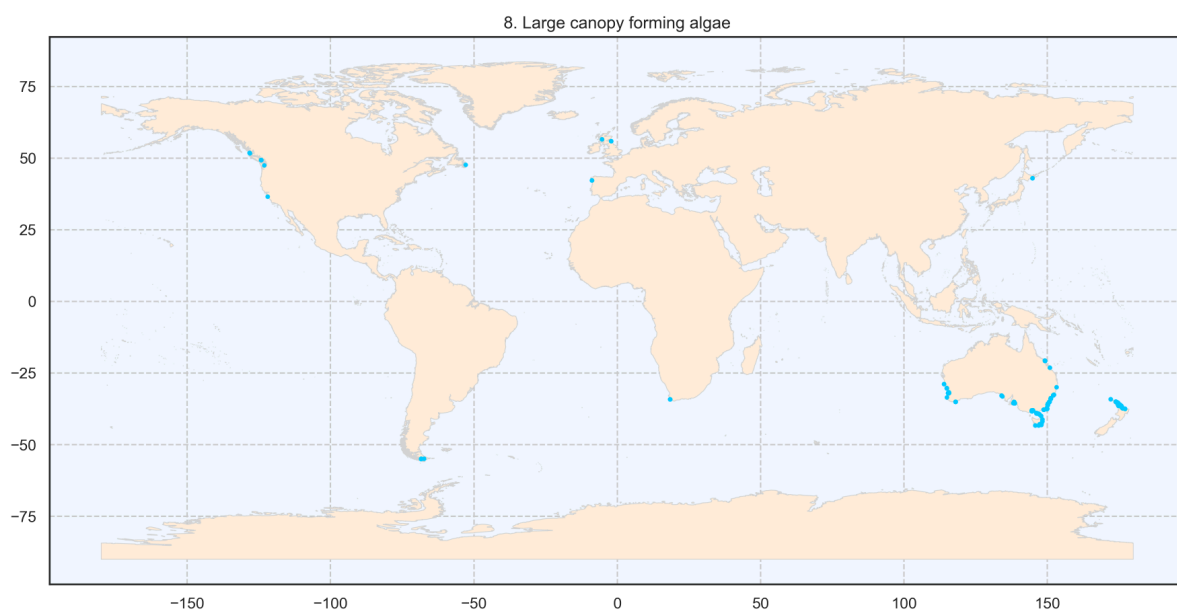


Figure S10 – *Spatial distribution of the cluster large canopy forming algae at the global scale. Each point represents a transect.*

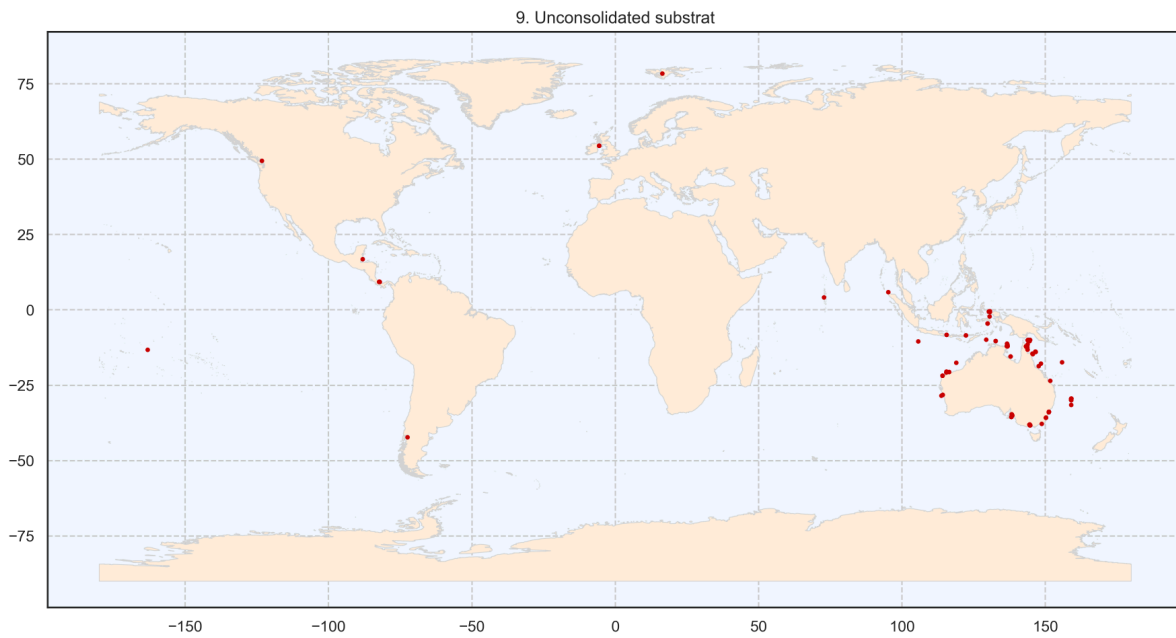


Figure S11 – *Spatial distribution of the cluster unconsolidated substrate at the global scale. Each point represents a transect.*

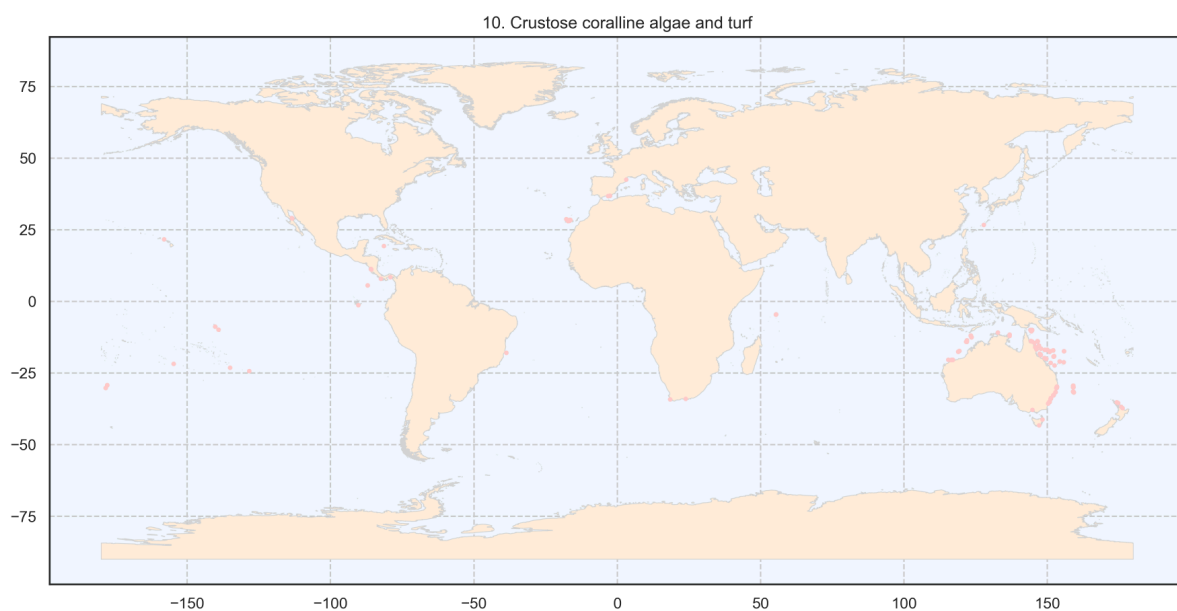


Figure S12 – *Spatial distribution of the cluster crustose coralline algae and turf at the global scale. Each point represents a transect.*

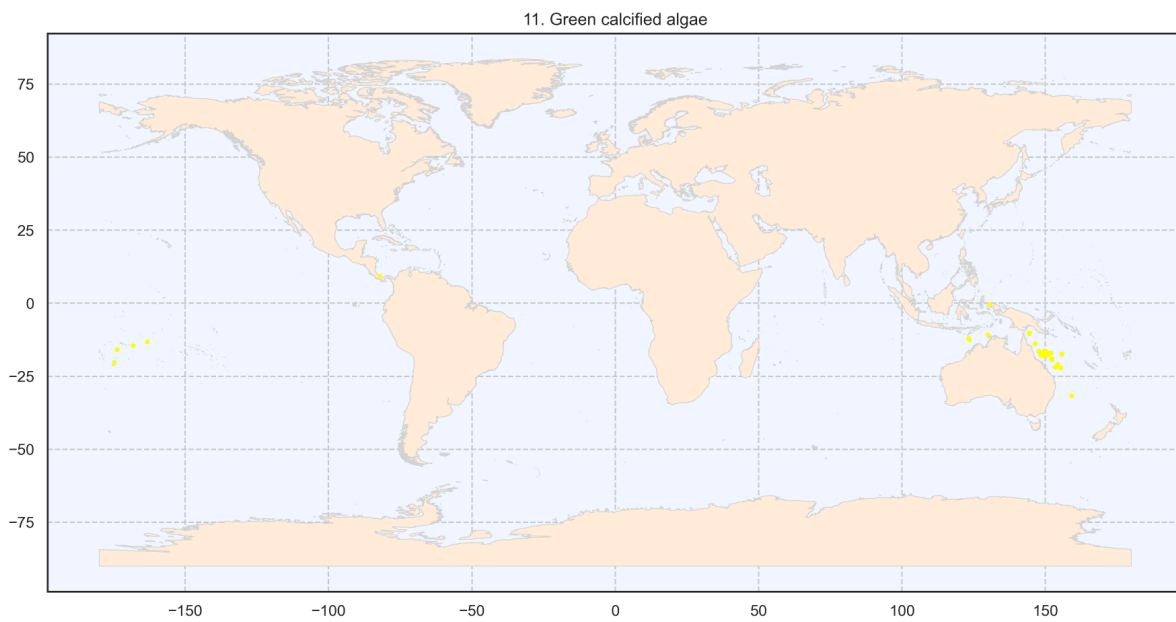


Figure S13 – *Spatial distribution of the cluster green calcified algae at the global scale. Each point represents a transect.*

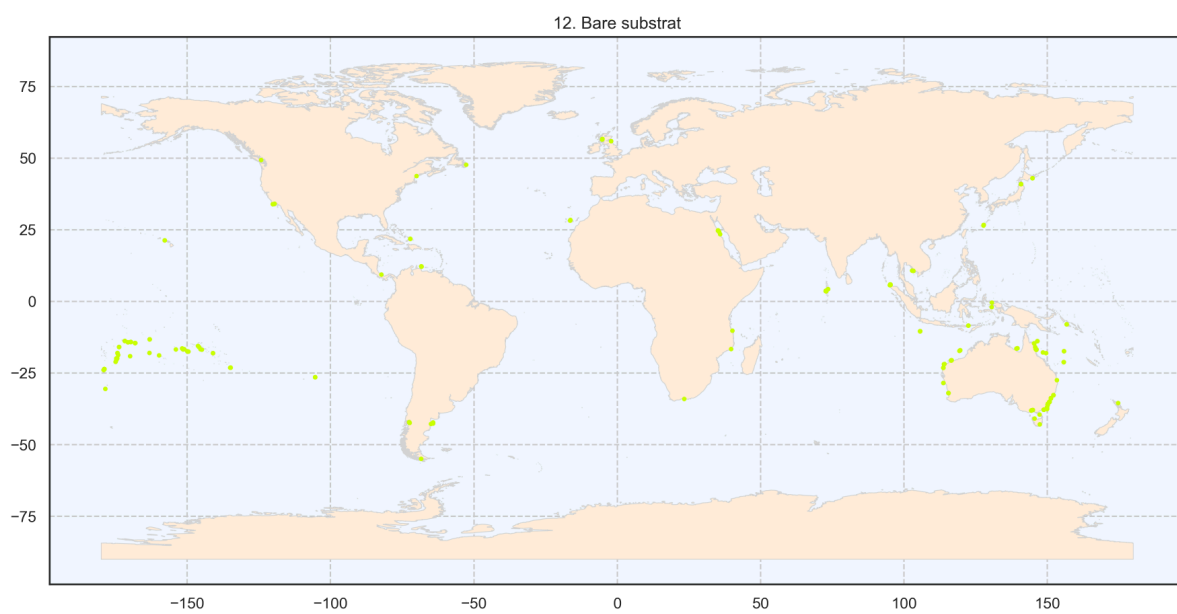


Figure S14 – *Spatial distribution of the cluster bare substrate at the global scale. Each point represents a transect.*

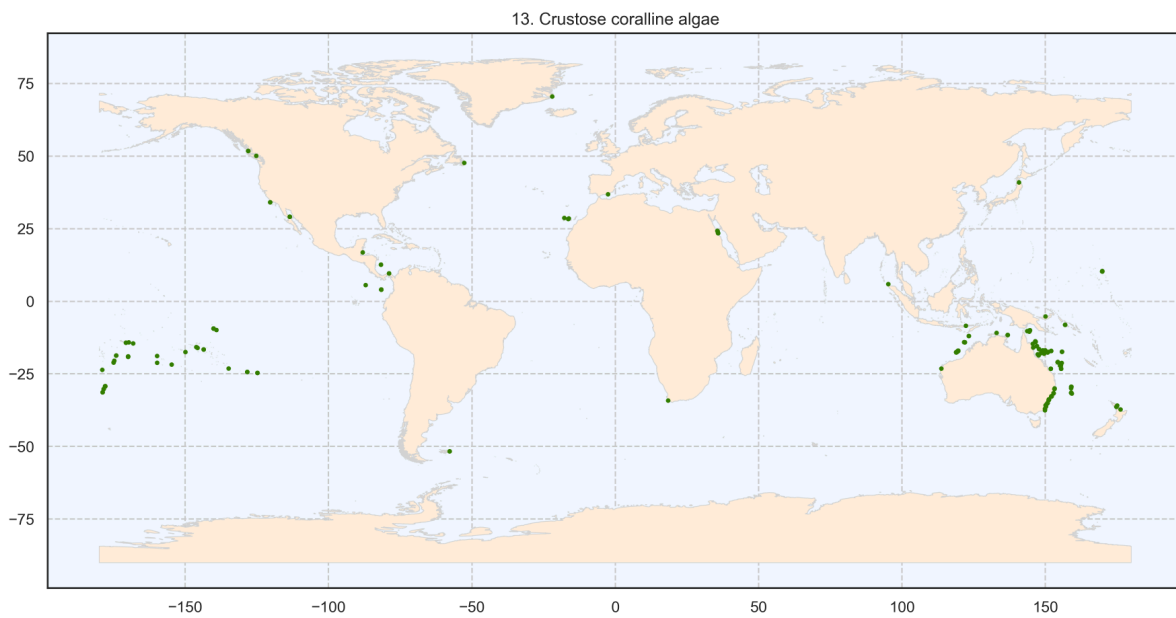


Figure S15 – *Spatial distribution of the cluster crustose coralline algae at the global scale. Each point represents a transect.*

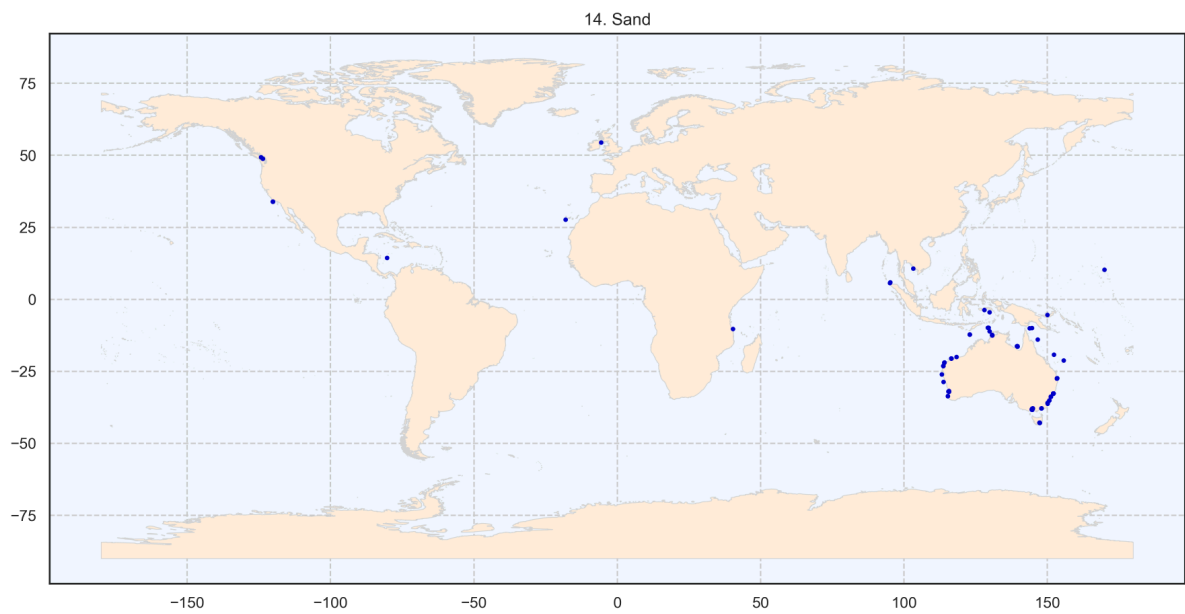


Figure S16 – *Spatial distribution of the cluster sand at the global scale. Each point represents a transect.*

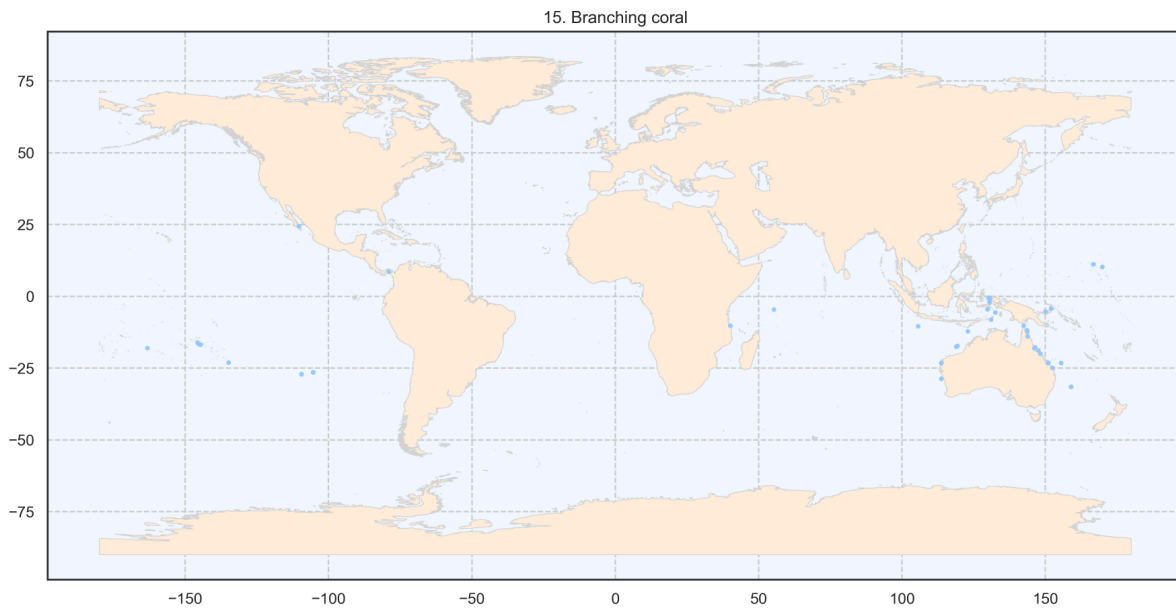


Figure S17 – *Spatial distribution of the cluster branching coral at the global scale. Each point represents a transect.*

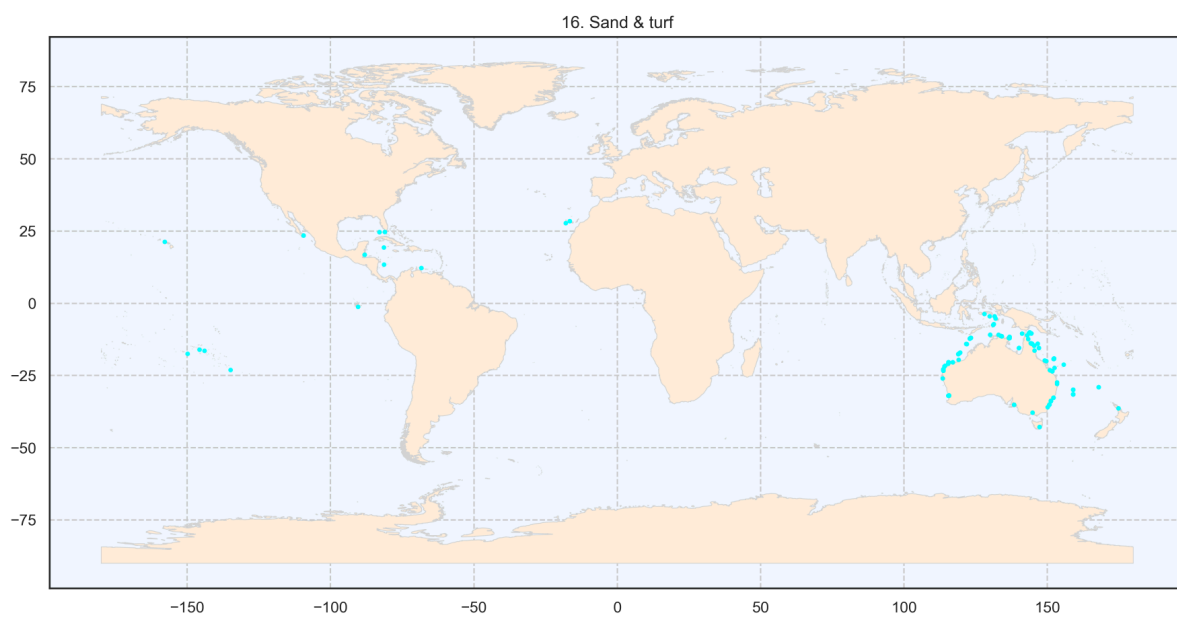


Figure S18 – *Spatial distribution of the cluster sand and turf algae at the global scale. Each point represents a transect.*

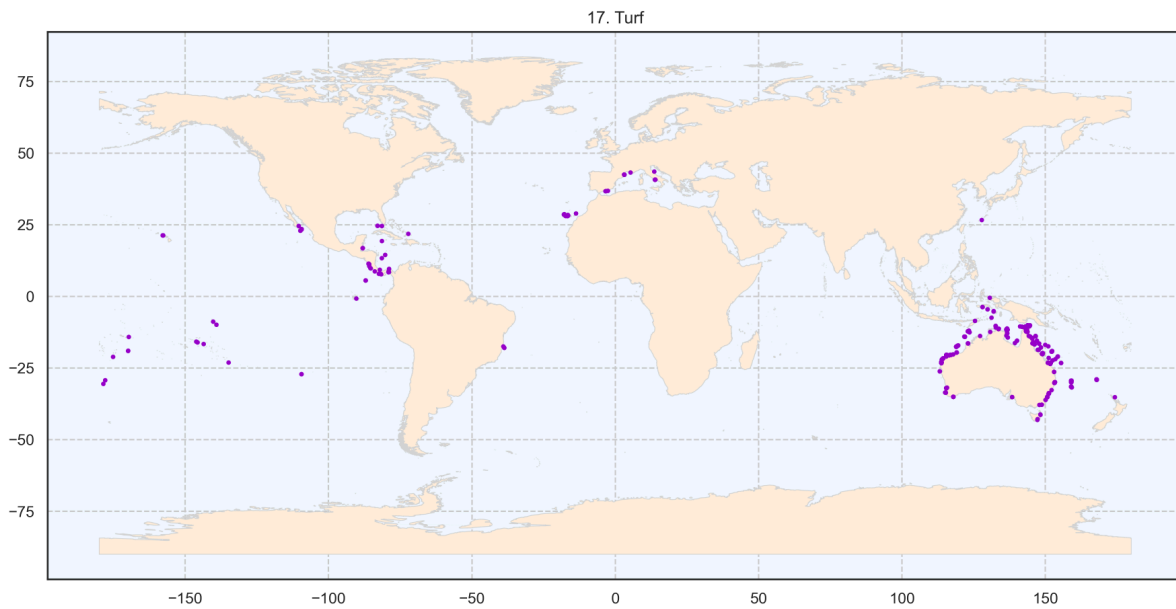


Figure S19 – *Spatial distribution of the cluster turf algae at the global scale. Each point represents a transect.*

Appendix C - Interpretation of the uncovered clusters

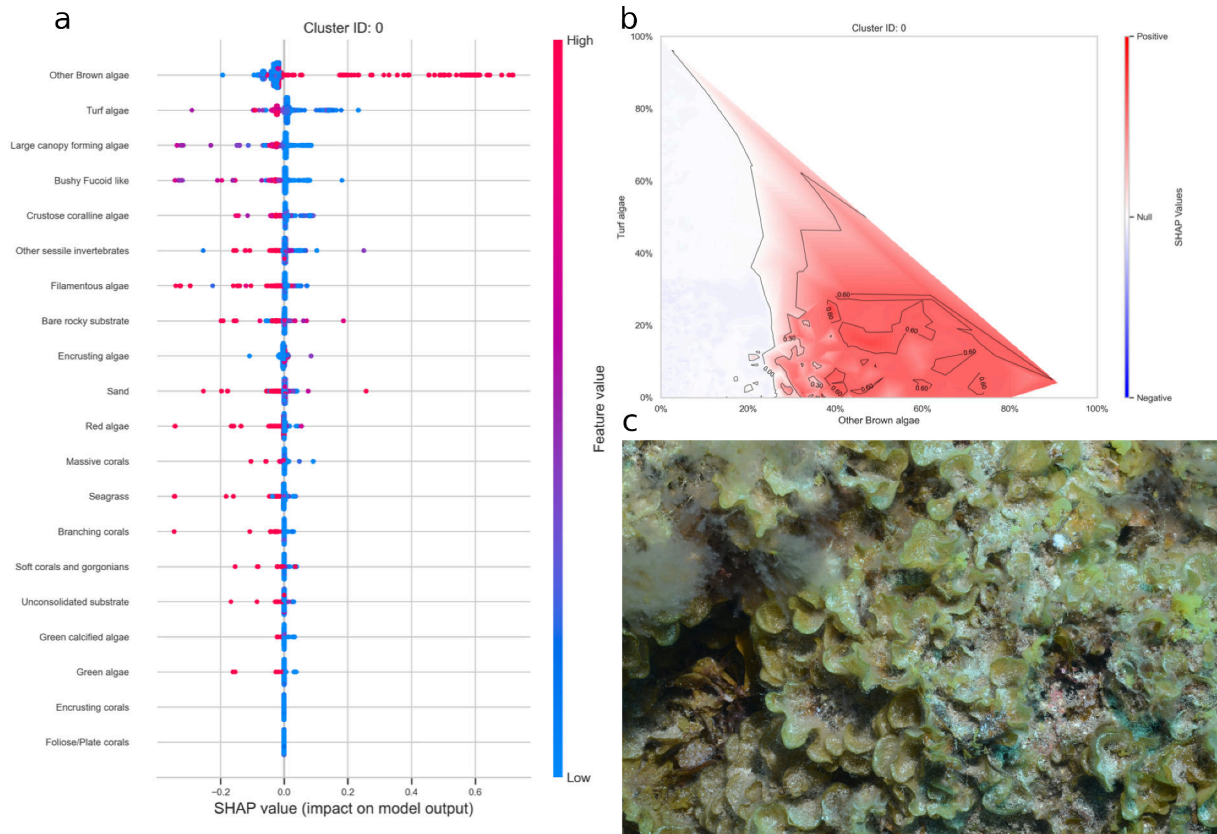


Figure S20 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the *y*-axis indicates its relative importance for the considered cluster. The position of each point along the *x*-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster brown algae. *c.* Example of photo-quadrate for one transect of the cluster brown algae categorised by HDBSCAN as exemplary.

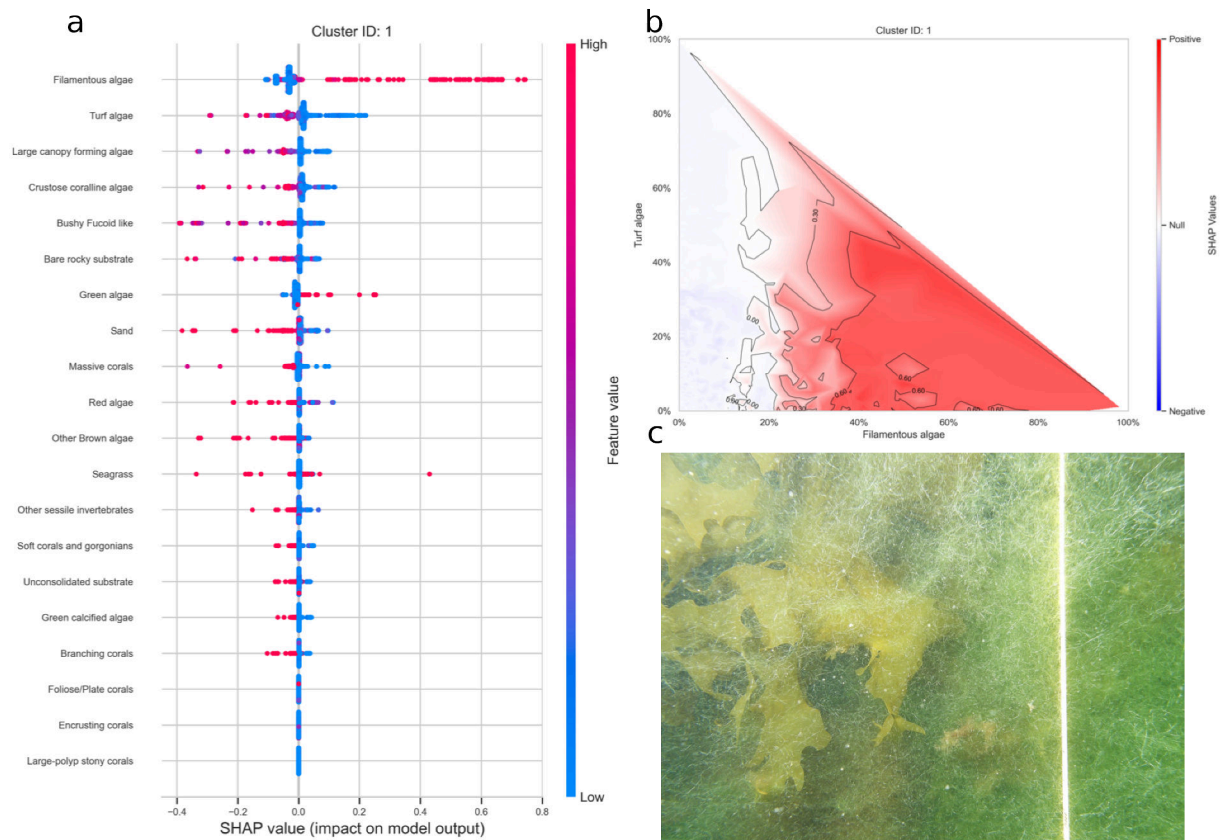


Figure S21 – a. SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. b. Linear interpolation of the SHAP values for the two most influential variables for the cluster filamentous algae. c. Example of photoquadrat for one transect of the cluster filamentous algae categorised by HDBSCAN as exemplary.

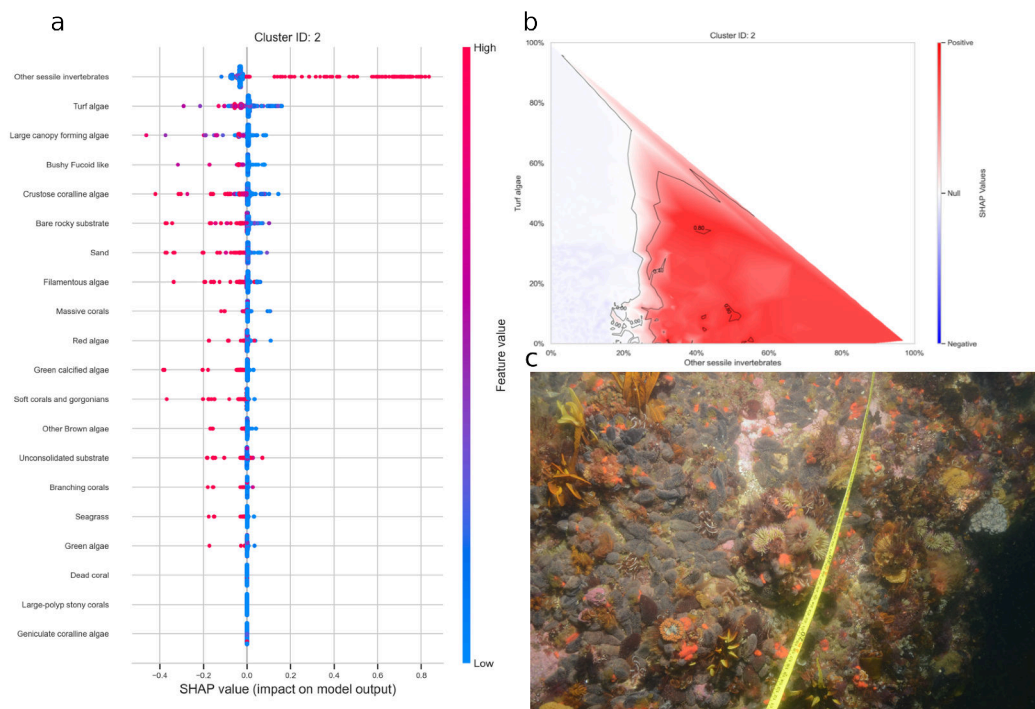


Figure S22 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster sessile invertebrates. *c.* Example of photoquadrat for one transect of the cluster sessile invertebrates categorised by HDBSCAN as exemplary.

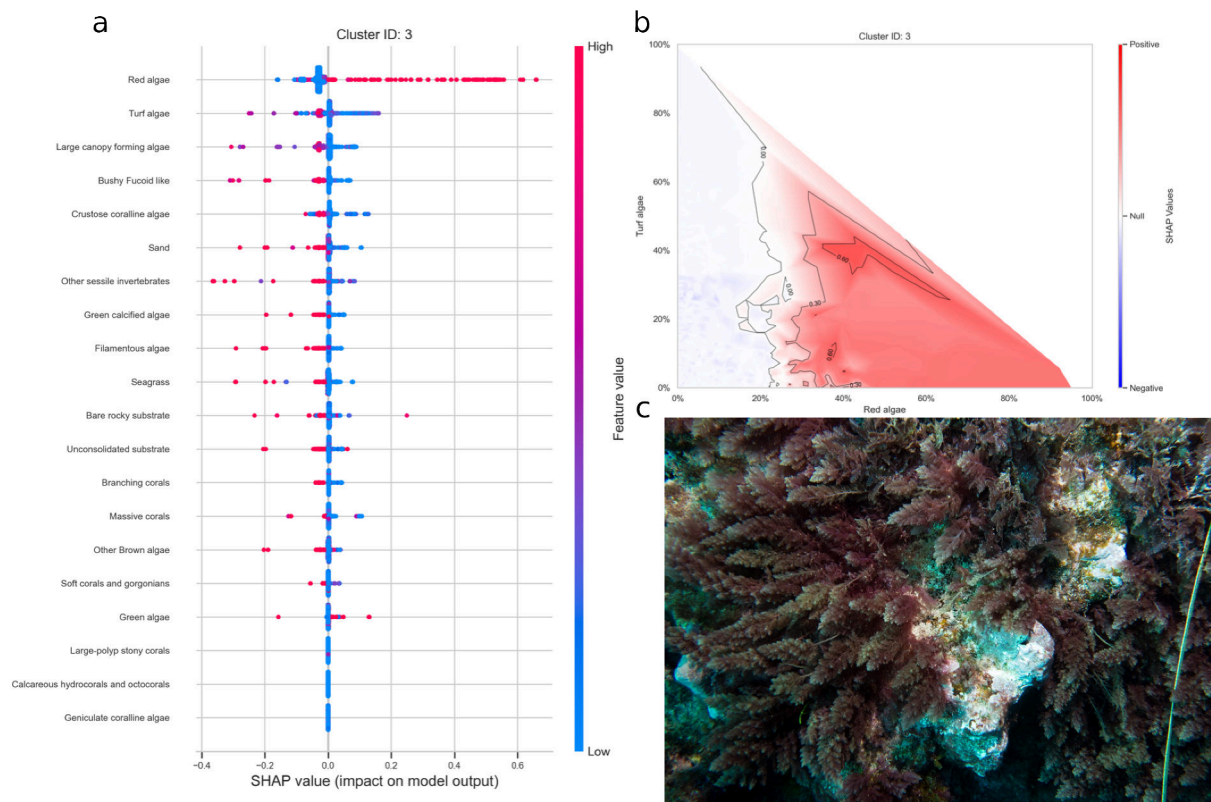


Figure S23 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster red algae. *c.* Example of photoquadrat for one transect of the cluster red algae categorised by HDBSCAN as exemplary.

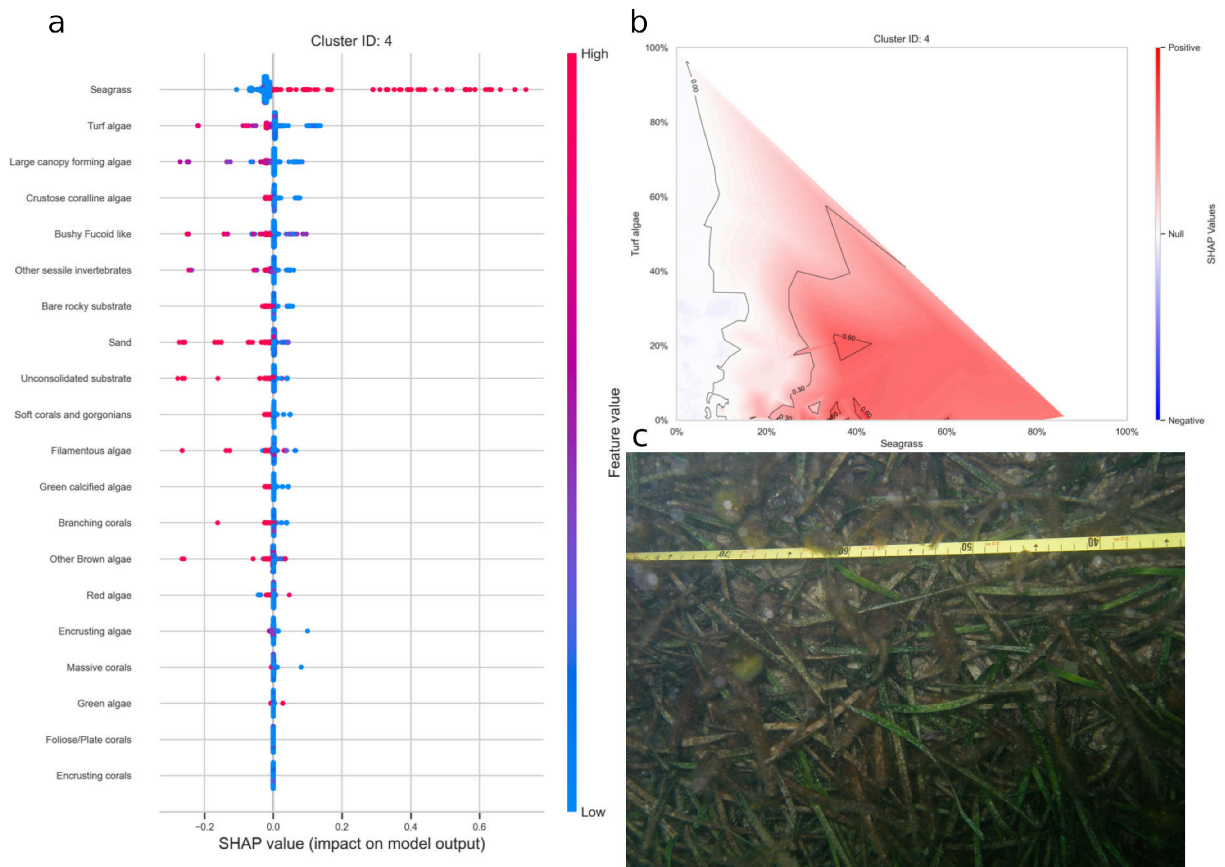


Figure S24 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster seagrass. *c.* Example of photoquadrat for one transect of the cluster seagrass categorised by HDBSCAN as exemplary.

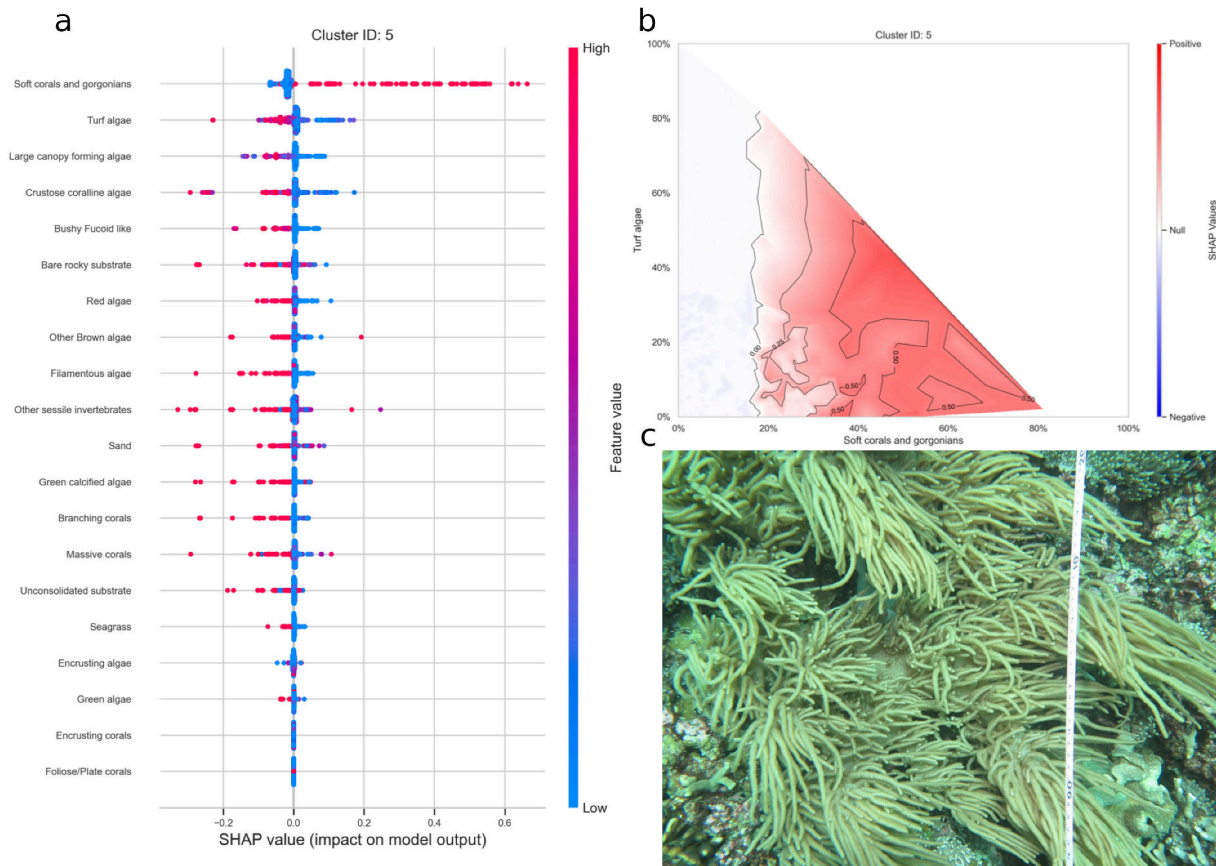


Figure S25 – a. SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. b. Linear interpolation of the SHAP values for the two most influential variables for the cluster soft coral and gorgonians. c. Example of photoquadrat for one transect of the cluster soft coral and gorgonians categorised by HDBSCAN as exemplary.

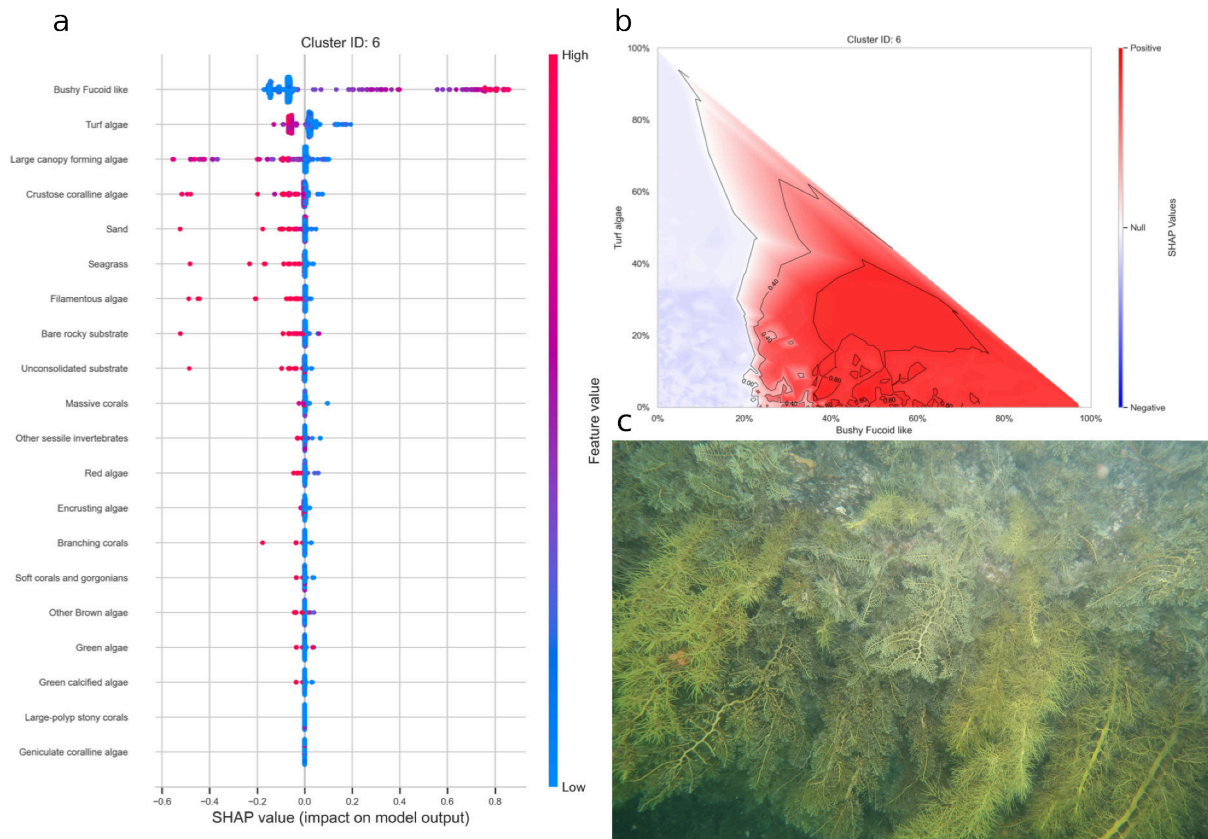


Figure S26 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster bushy fucoid-like. *c.* Example of photoquadrat for one transect of the cluster bushy fucoid-like categorised by HDBSCAN as exemplary.

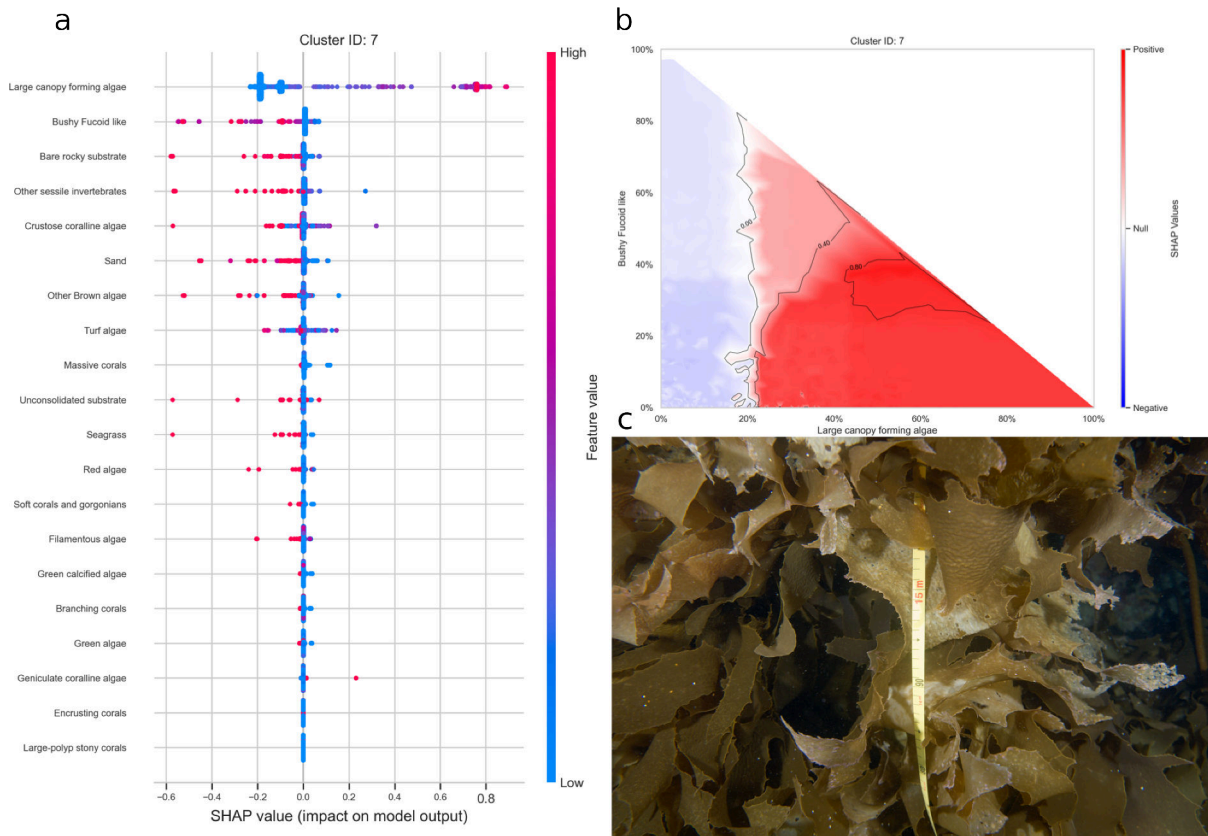


Figure S27 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster large canopy forming algae. *c.* Example of photoquadrat for one transect of the cluster large canopy forming algae categorised by HDBSCAN as exemplary.

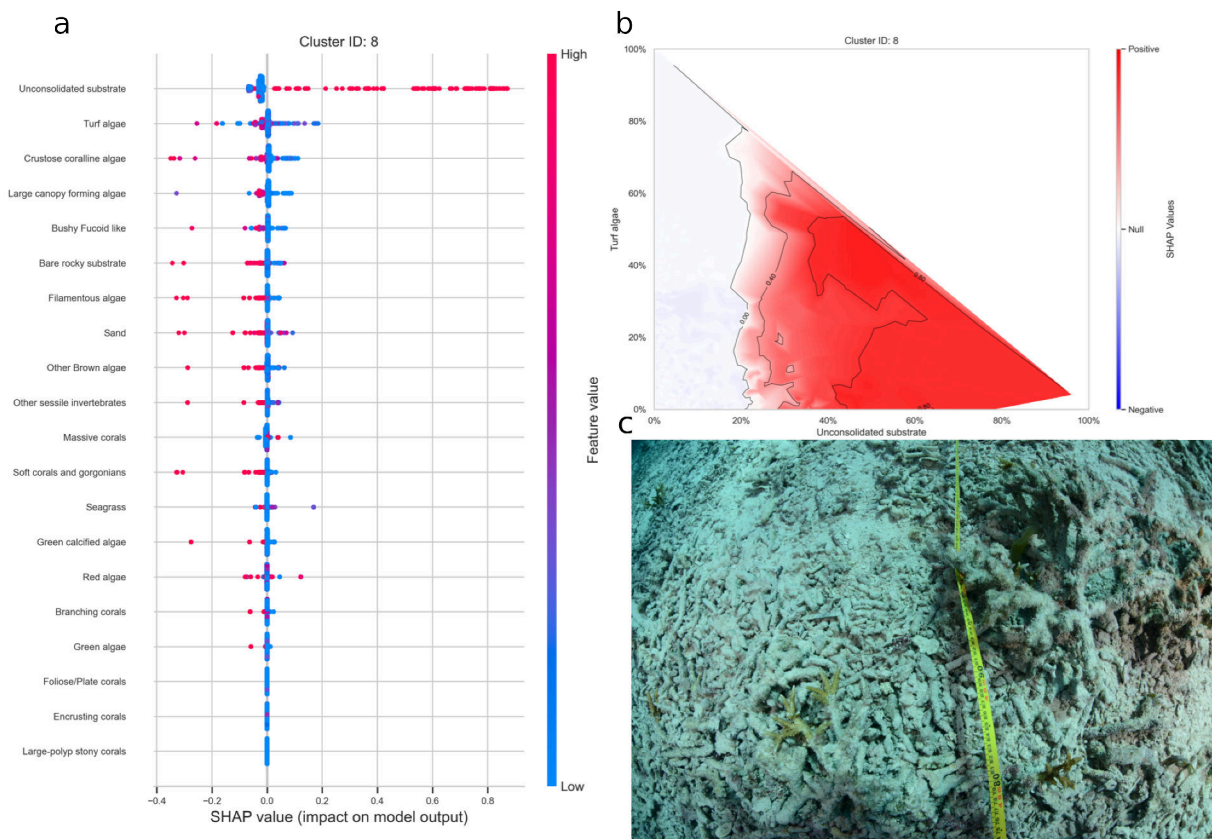


Figure S28 – a. SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. b. Linear interpolation of the SHAP values for the two most influential variables for the cluster unconsolidated substrat. c. Example of photoquadrat for one transect of the cluster unconsolidated substrat categorised by HDBS-CAN as exemplary.

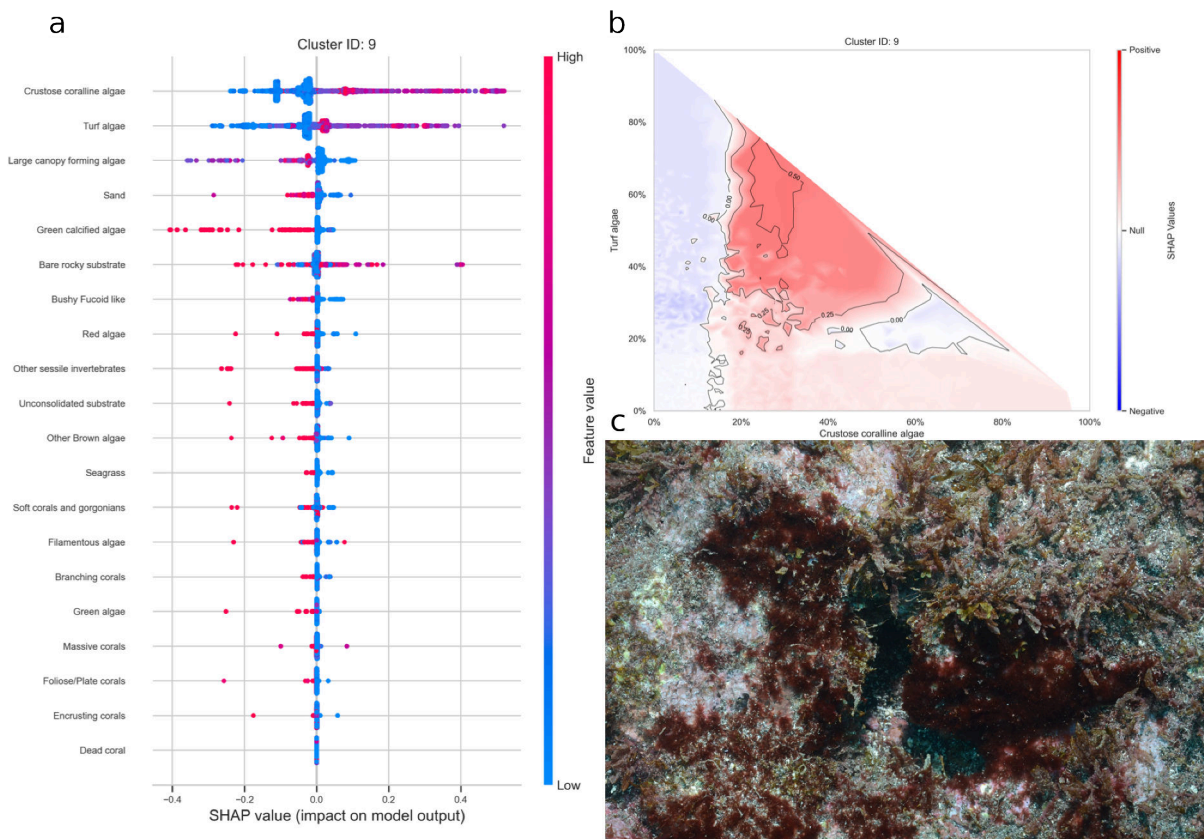


Figure S29 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster crustose coralline algae and turf. *c.* Example of photoquadrat for one transect of the cluster crustose coralline algae and turf categorised by HDBSCAN as exemplary.

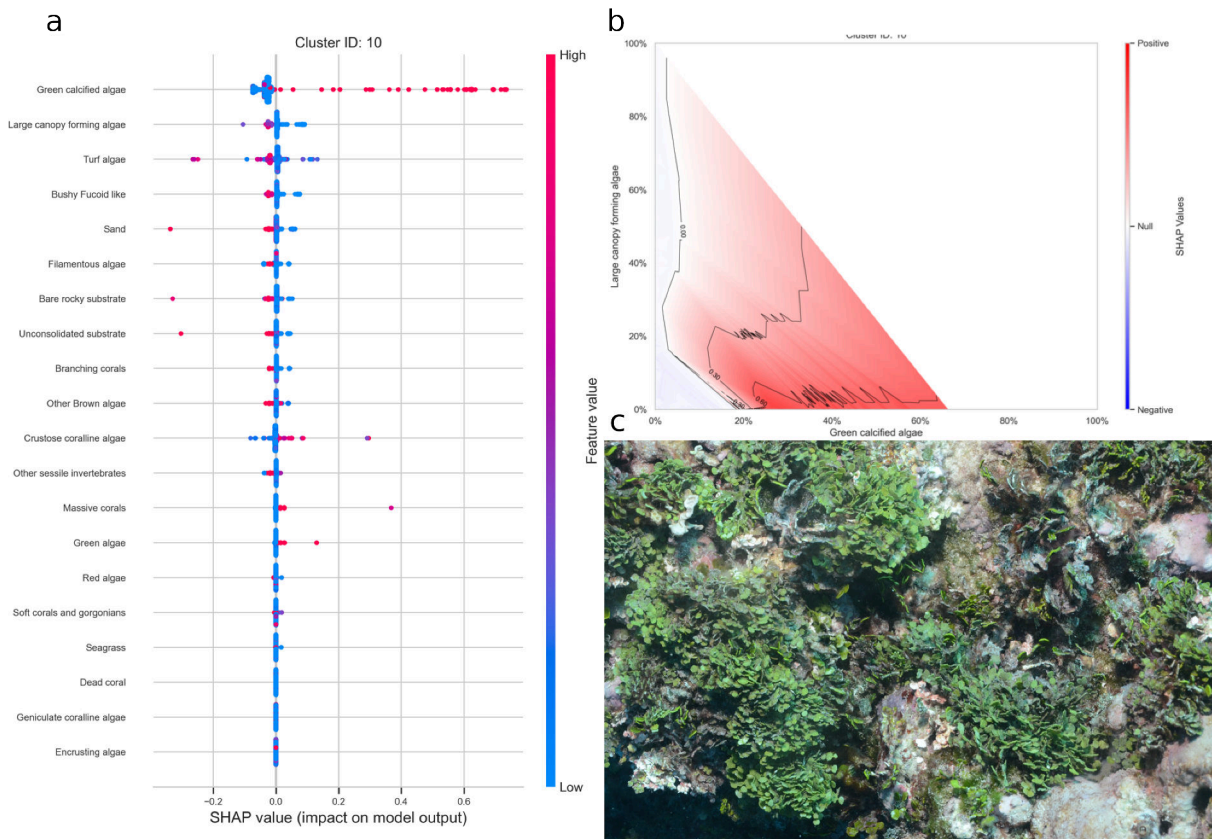


Figure S30 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster green calcified algae. *c.* Example of photoquadrat for one transect of the cluster green calcified algae categorised by HDBSCAN as exemplary.

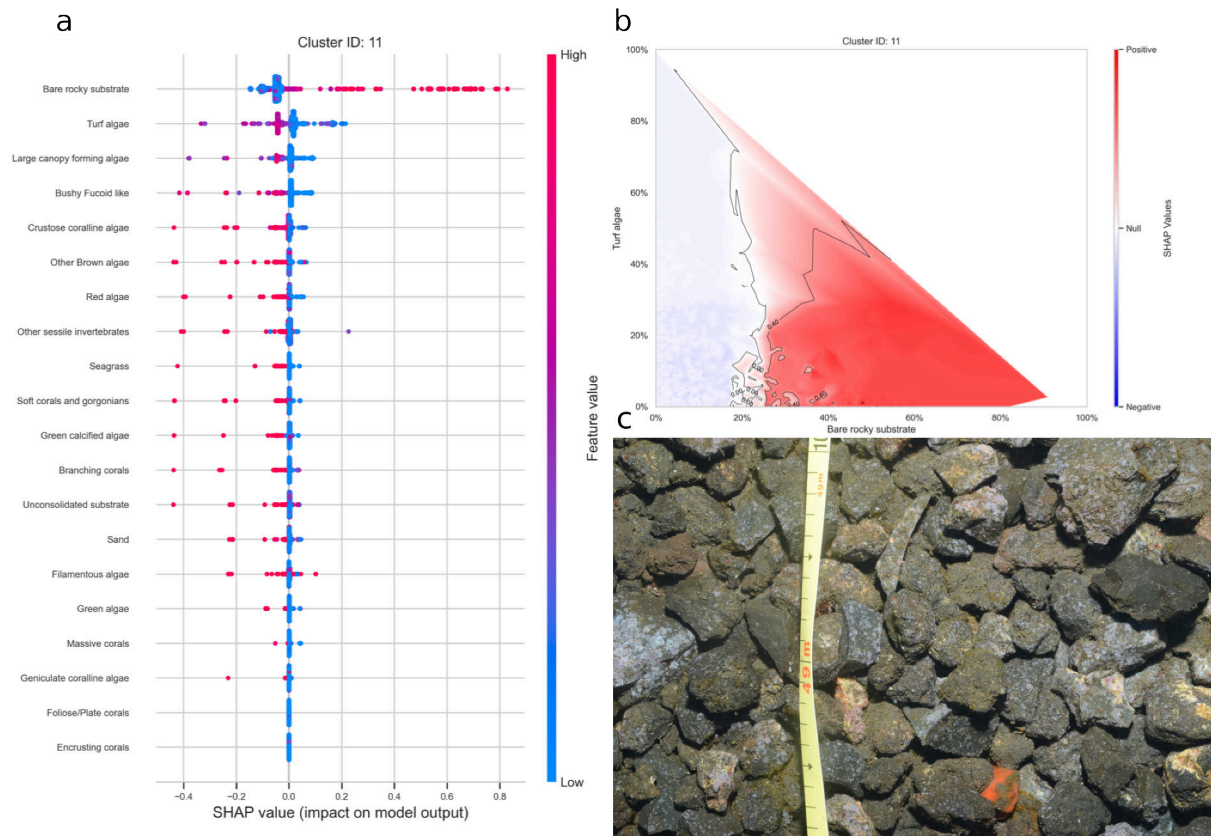


Figure S31 – a. SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. b. Linear interpolation of the SHAP values for the two most influential variables for the cluster bare substrate. c. Example of photo-quadrate for one transect of the cluster bare substrate categorised by HDBSCAN as exemplary.

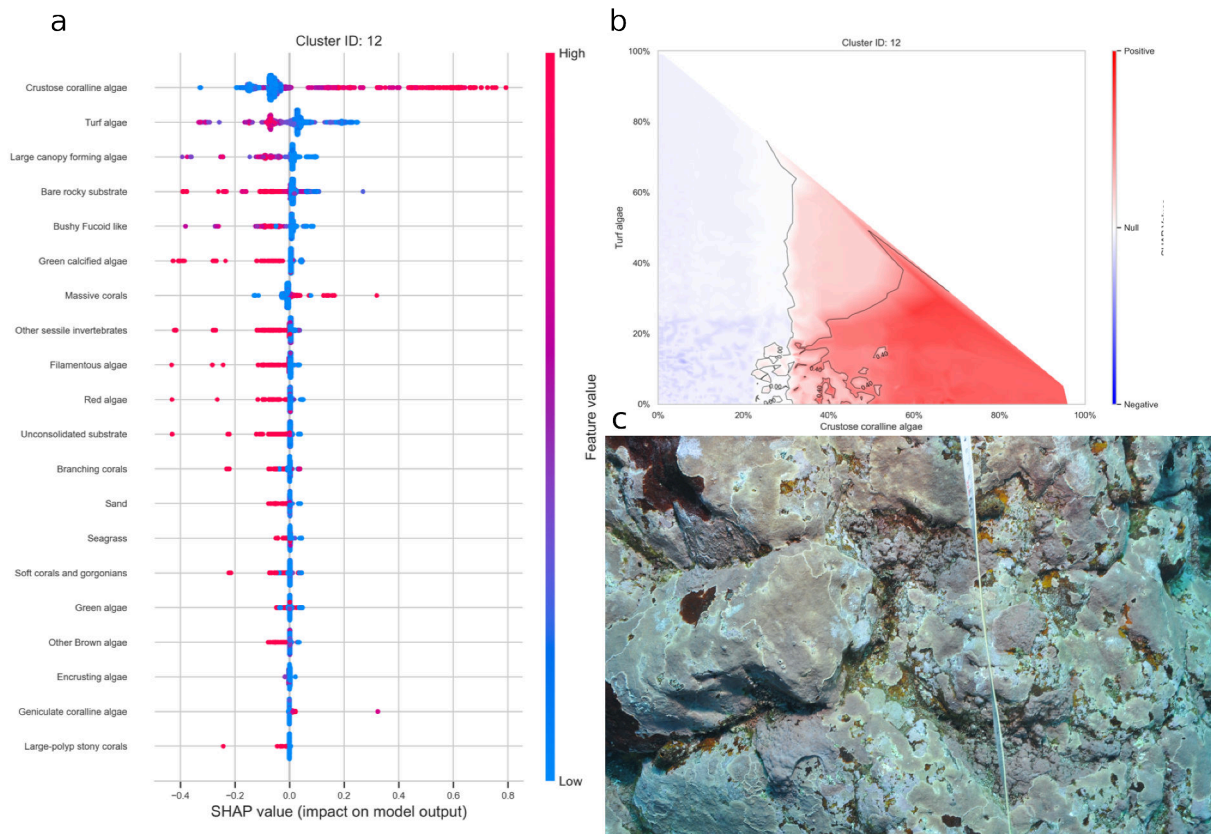


Figure S32 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster crustose coralline algae. *c.* Example of photoquadrat for one transect of the cluster crustose coralline algae categorised by HDBSCAN as exemplary.

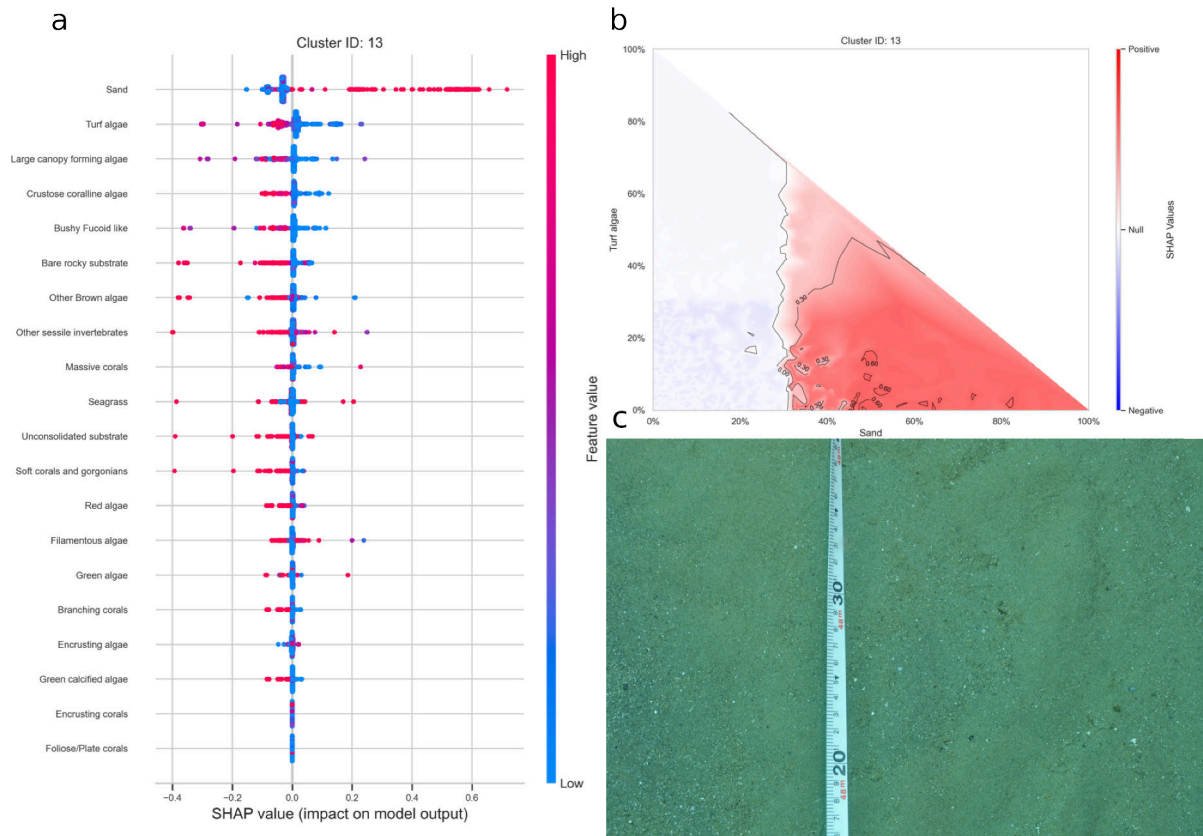


Figure S33 – a. SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. b. Linear interpolation of the SHAP values for the two most influential variables for the cluster sand. c. Example of photoquadrat for one transect of the cluster sand categorised by HDBSCAN as exemplary.

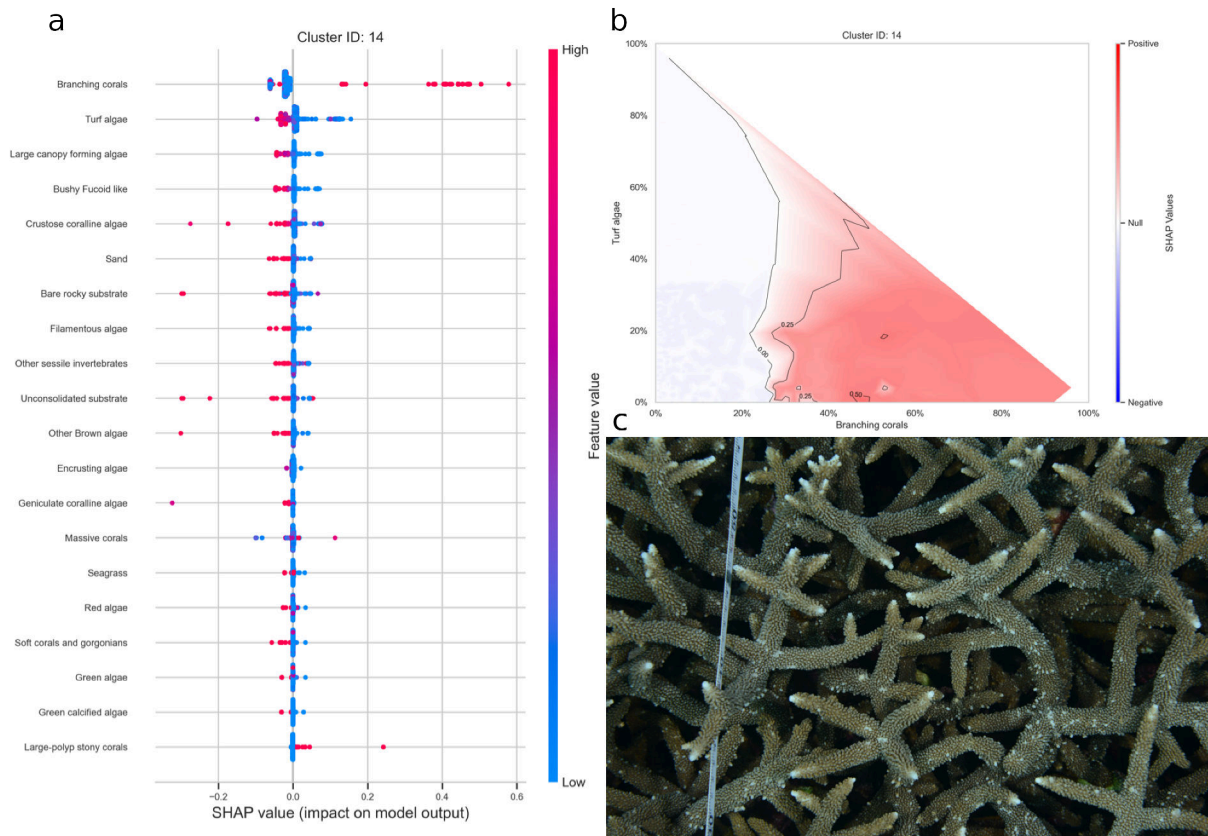


Figure S34 – *a.* SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. *b.* Linear interpolation of the SHAP values for the two most influential variables for the cluster branching coral. *c.* Example of photoquadrat for one transect of the cluster branching coral categorised by HDBSCAN as exemplary.

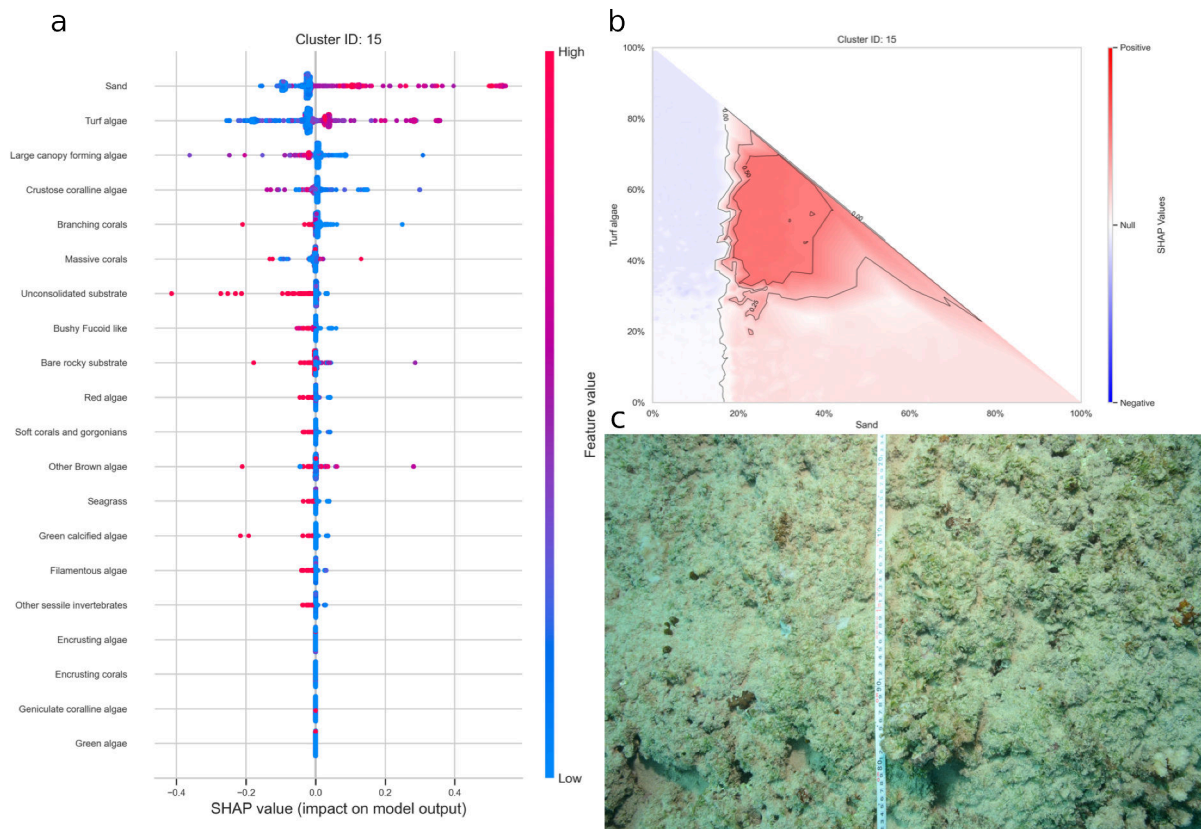


Figure S35 – a. SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. b. Linear interpolation of the SHAP values for the two most influential variables for the cluster sand and turf algae. c. Example of photoquadrat for one transect of the cluster sand and turf algae categorised by HDBSCAN as exemplary.

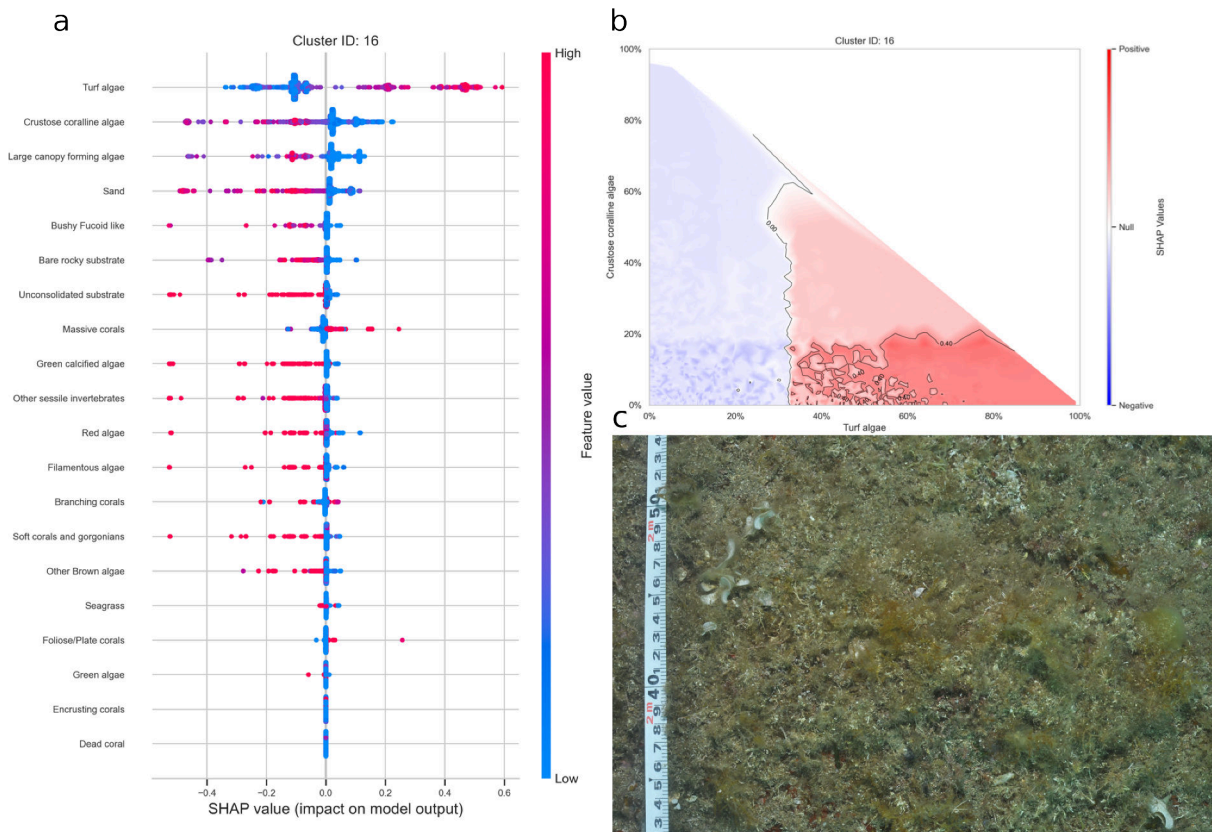


Figure S36 – a. SHAP summary plot showing the impact of each habitat substrate on the classification. The position of each habitat group on the y-axis indicates its relative importance for the considered cluster. The position of each point along the x-axis indicates if the observation is associated with a lower or higher affinity with the cluster and the colour of the point indicates if the value of the cover of this habitat is rather high or low. b. Linear interpolation of the SHAP values for the two most influential variables for the cluster turf algae. c. Example of photoquadrat for one transect of the cluster turf algae categorised by HDBSCAN as exemplary.

Appendix D - Complementary results

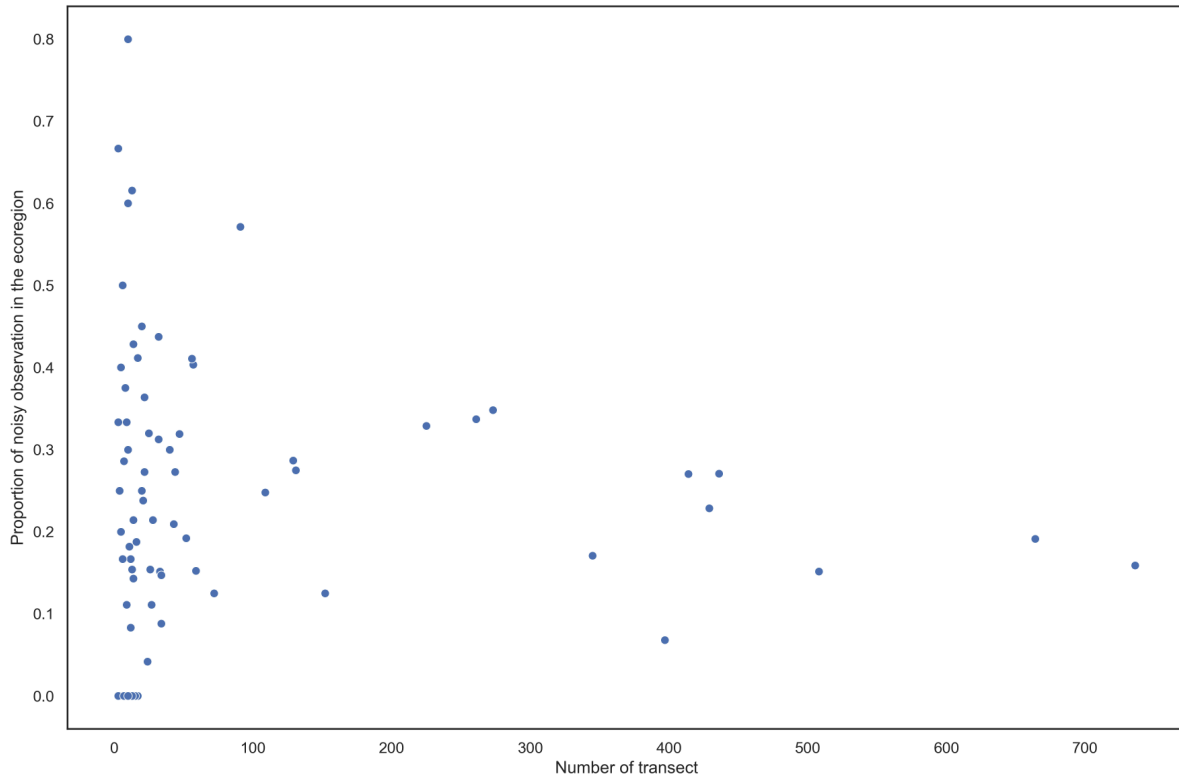


Figure S37 – *Plot of the proportion of transect classified as noisy by the UMAP-HDBSCAN pipeline as a function of the \log_{10} (number of transects) sampled within an ecoregion*

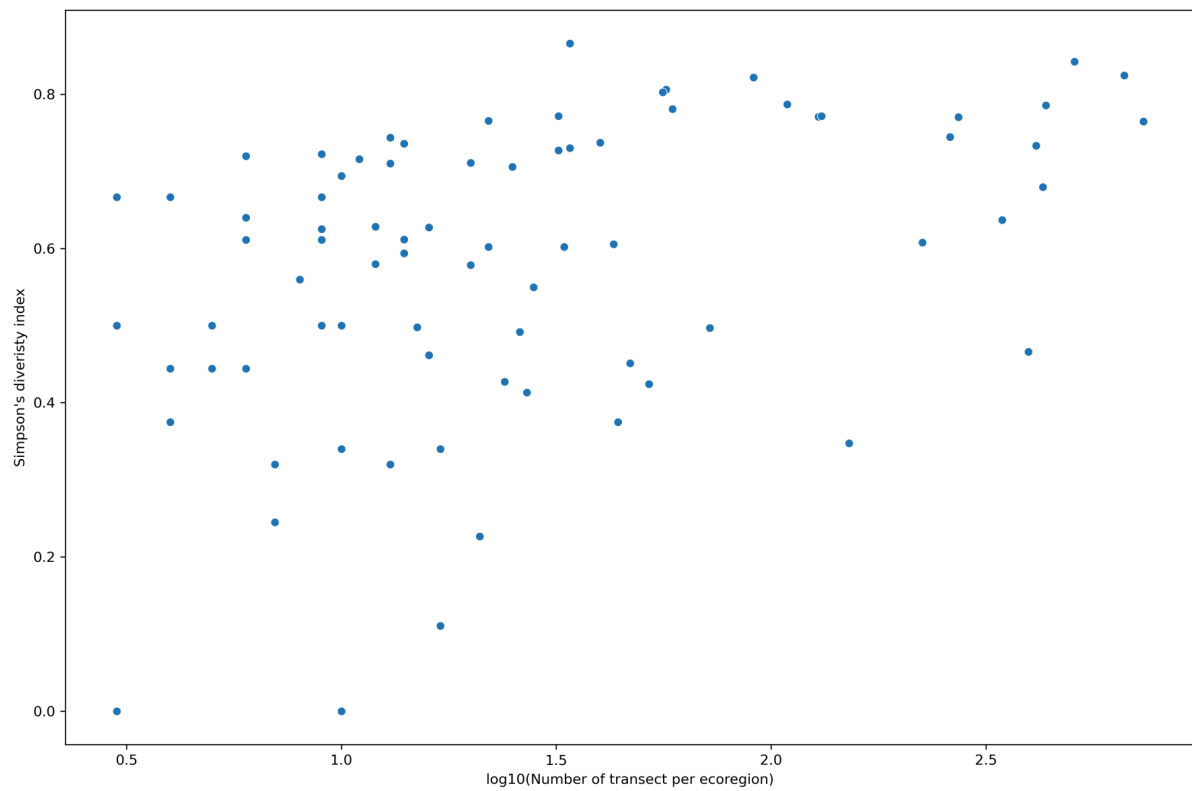


Figure S38 – Plot of the Gini-Simpson diversity index as a function of the \log_{10} (number of transects) sampled within an ecoregion

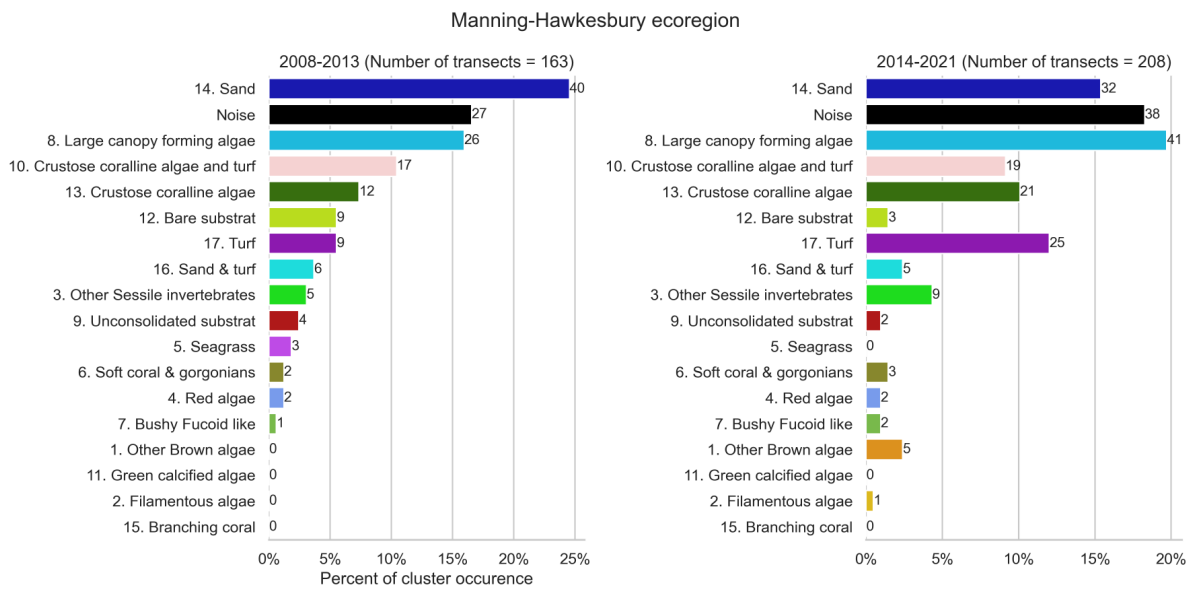


Figure S39 – Bar graph of the proportion of different clusters for the period 2008-2013 and 2015-2021 for sites that were sampled at least once in each of the two periods for the Manning-Hawkesbury ecoregion.

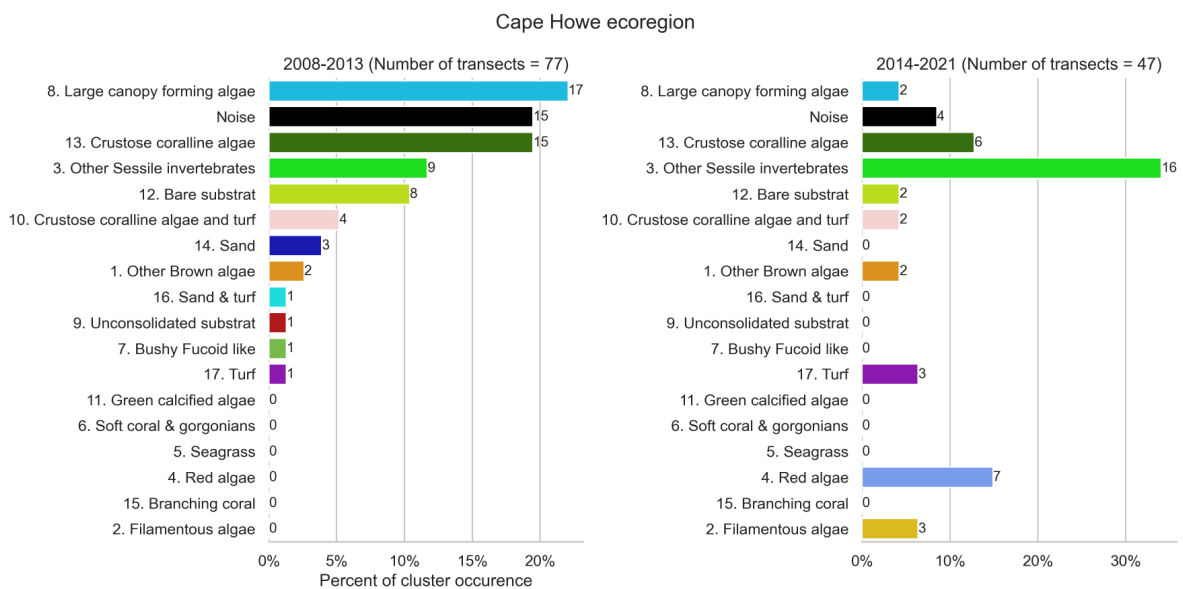


Figure S40 – Bar graph of the proportion of different clusters for the period 2008-2013 and 2015-2021 for sites that were sampled at least once in each of the two periods for the Cape Howe ecoregion.

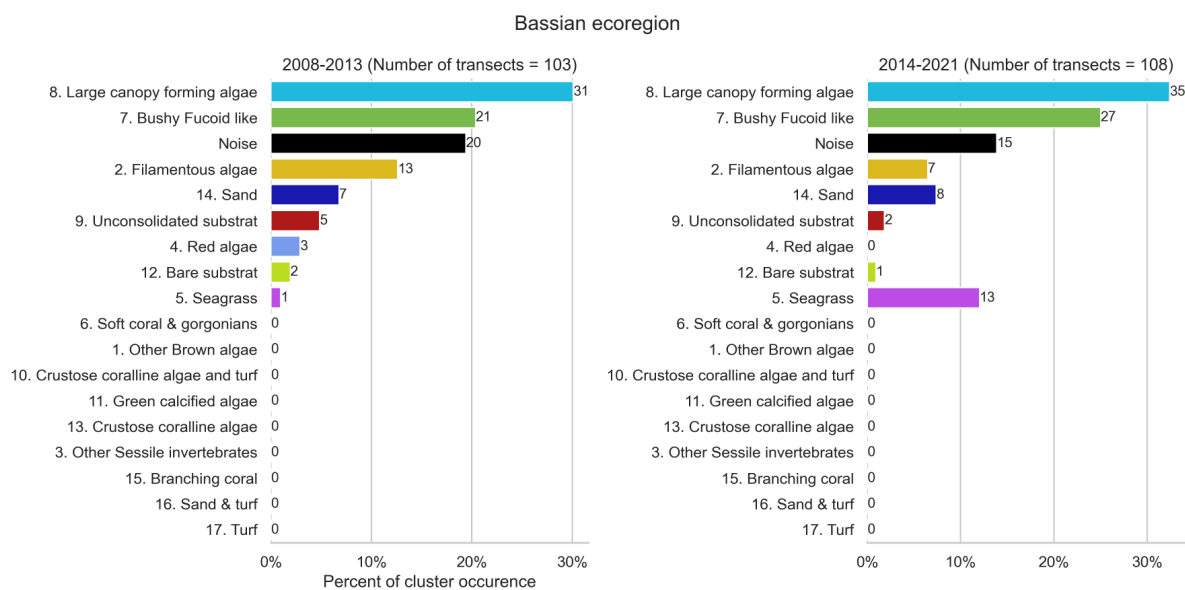


Figure S41 – Bar graph of the proportion of different clusters for the period 2008-2013 and 2015-2021 for sites that were sampled at least once in each of the two periods for the Bassian ecoregion.

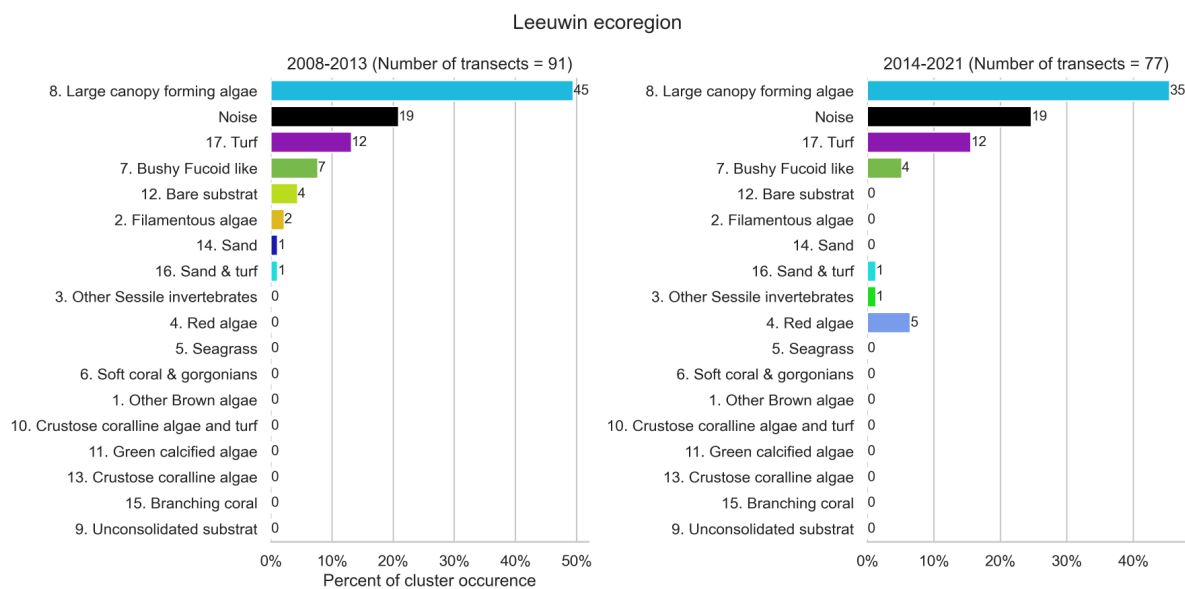


Figure S42 – Bar graph of the proportion of different clusters for the period 2008-2013 and 2015-2021 for sites that were sampled at least once in each of the two periods for the Leeuwin ecoregion.

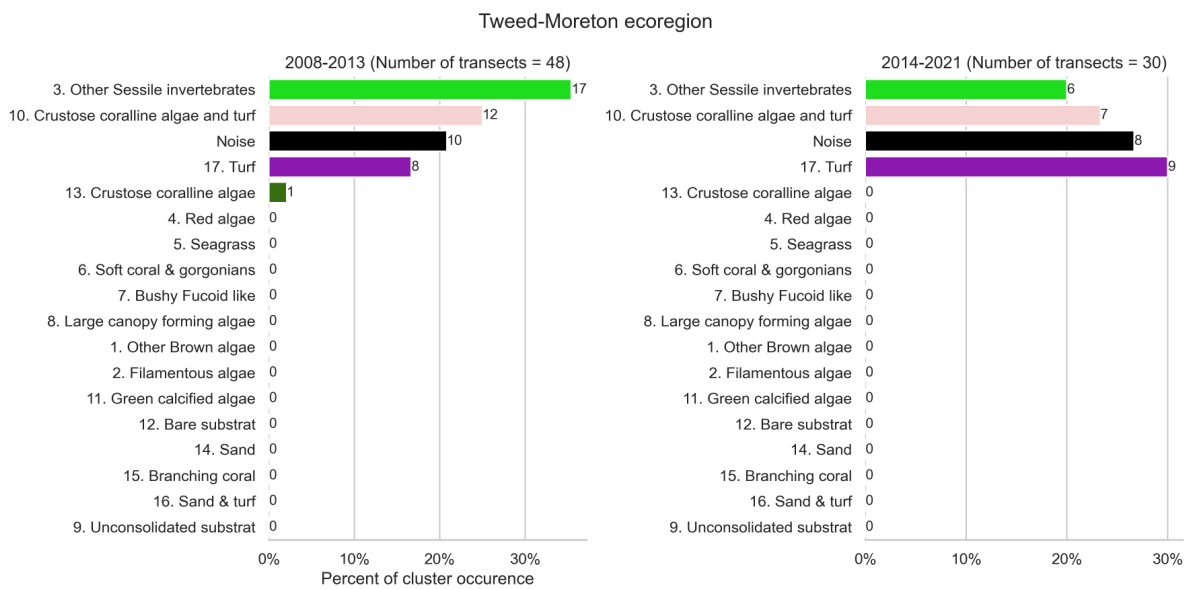


Figure S43 – Bar graph of the proportion of different clusters for the period 2008-2013 and 2015-2021 for sites that were sampled at least once in each of the two periods for the Tweed-Moreton ecoregion.

Bibliographie

- BOYÉ, A., (2019), *Diversité taxinomique et fonctionnelle des habitats benthiques dans l'espace et dans le temps : une perspective régionale et décennale* [thèse de doct.].
- BOYÉ, A., THIÉBAUT, É., GRALL, J., LEGENDRE, P., BROUDIN, C., HOUBIN, C., GARREC, V. L., MAGUER, M., DROUAL, G., & GAUTHIER, O., (2019), Trait-based approach to monitoring marine benthic data along 500 km of coastline, *Diversity and Distributions*, 2512, 1879-1896.
- BROOKS, S. P., & GELMAN, A., (1998), General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, 74, 434-455.
- CRESSWELL, A. K., EDGAR, G. J., STUART-SMITH, R. D., THOMSON, R. J., BARRETT, N. S., & JOHNSON, C. R., (2017), Translating local benthic community structure to national biogenic reef habitat types, *Global Ecology and Biogeography*, 2610, 1112-1125.
- LAZURE, P., & DUMAS, F., (2008), An external–internal mode coupling for a 3D hydrodynamical model for applications at regional scale (MARS), *Advances in Water Resources*, 312, 233-250.
- LECORNU, F., & ROECK, Y.-H. D., (2009), PREVIMER - Coastal observations and forecasts, *La Houille Blanche*, 951, 60-63.
- OVASKAINEN, O., & ABREGO, N., (2020), *Joint Species Distribution Modelling : With Applications in R*, Cambridge University Press.
- RIGAL, S., DEVICTOR, V., & DAKOS, V., (2020), A method for classifying and comparing non-linear trajectories of ecological variables, *Ecological Indicators*, 112, 106113.

Patterns of Co-occurrence and Spatial Richness in Marine Habitats : Insights from Large-Scale Modelling

Préambule

Grâce au Chapitre 2, nous avons identifié différents états d'habitats biogéniques à l'échelle mondiale en appliquant une nouvelle méthode d'agrégation que nous avons appliquée à des données d'un programme de sciences participatives. Maintenant que les états d'habitats ont été définis, ce chapitre vise à modéliser leur distribution à l'aide d'une approche spatiale semblable à celle utilisée dans le cadre des *SDM*. Nous avons choisi de recentrer ces analyses du Chapitre 3 sur la région australo-pacifique, cœur géographique historique du programme *Reef Life Survey* (GRAHAM & STUART-SMITH, 2009). L'objectif premier est de comprendre les facteurs environnementaux et anthropiques influençant leurs occurrences. En caractérisant les niches environnementales, un second objectif est de déterminer les habitats partageant des niches similaires donc susceptibles de coexister le long des côtes australiennes. Dans ce contexte, ce chapitre vise à étudier la distribution de 14 états d'habitats sur les côtes australiennes.

Nous avons utilisé un *SDM* de type *Random Forest* multivarié pour caractériser les niches environnementales des états d'habitats en considérant comme prédicteurs des données environnementales (*Bio-Oracle*; ASSIS et al. (2018)), et des données de pressions anthropiques (*Ocean Health Index*; HALPERN et al. (2019)). Nos résultats indiquent que le modèle a un bon pouvoir explicatif, mais un pouvoir prédictif faible, laissant penser que la transférabilité de notre modèle dans le temps et l'espace est restreinte. Nos résultats

indiquent que les facteurs les plus importants régissant la distribution de ces différents états d'habitats sont la température de surface de l'eau, l'intensité de la pression de pêche et les anomalies dans la fréquence des vagues de chaleur marine. Nous montrons également dans ce chapitre que tous les états d'habitat identifiés ne pouvaient co-exister dans le même transect ce qui suggère que certaines combinaisons d'états d'habitats sont invraisemblables (ou jamais observées à ce jour). Ce travail a permis également de mettre en avant que les zones de transitions tempérées/tropicales sont susceptibles d'abriter un plus grand nombre d'états d'habitats. Cela peut impliquer que ces zones sont plus exposées aux changements de régime, puisque la disparition de l'état dominant actuel pourrait laisser place au développement d'un état d'habitat différent, notamment dans un contexte de tropicalisation des zones tempérées chaudes des côtes est (VERGÉS et al., 2014; VERGÉS et al., 2019) et ouest (WERNBERG et al., 2016) australiennes. Ces changements sont d'autant plus importants à surveiller puisqu'ils pourraient avoir un impact important sur la composition et les fonctions des communautés associées et par conséquent modifier le fonctionnement même des écosystèmes.

4.1 Abstract

Benthic reef habitats serve as foundational structures in coastal ecosystems, facilitating essential ecological processes and significantly amplifying biodiversity. However they are increasingly subject to regime shifts that can be detrimental for marine ecosystem functioning and biodiversity. In this study we aim to investigate the distribution and ecological drivers of reef benthic habitat states across Australia, in order to understand and anticipate their potential/risk of ecological regime shifts in the face of changing environments and anthropogenic impacts. Using random forest models applied to occurrences of reef habitat states stemming from a previous analyses of the *Reef Life Survey* transect data, we quantified the relative contribution of anthropogenic and environmental drivers in explaining their spatial distribution. Specifically, we focused on 14 reef states including turf, kelp or coral reefs. Models showed a good explanatory power with AUC (mean = 0.87; sd = 0.07). For most habitats, we found that SST, Fishing intensity and Marine Heatwaves were the most important driver of their distribution while anthropogenic factors only had a marginal influence. However, models presented a poor predictive power suggesting limited transferability in time or space. On the calibration areas, we found that several habitats (up to eight) can occur under similar environmental conditions, sugges-

ting that the likelihood of state changes (including regime shifts) is spatially variable. Overall, this approach makes it possible to highlight areas and environmental conditions where the occurrence of alternative stable states is more likely.

4.2 Introduction

Regime shifts are persistent changes in the structure and/or the function of ecosystems occurring more or less suddenly (SCHEFFER et al., 2001; SCHEFFER & CARPENTER, 2003). Ecological regime shifts have been identified in many different ecosystems : forest to savannahs (STAVER et al., 2011), tundra to boreal forests (SCHEFFER et al., 2012), coral reefs to macroalgae dominated reefs (MCMANUS & POLSENBERG, 2004), or fisheries collapse (GÅRDMARK et al., 2015). Areas affected by regime shifts tend usually display profound and lasting changes in the ecosystem services they provide to human populations (ROCHA et al., 2015a), with economic impacts through the disappearance of certain activities (CRÉPIN et al., 2012), and societal and cultural impacts through the loss of cultural and traditional knowledge (MUSTONEN & SHADRIN, 2021). Thus, one of the common features of all these regime shifts is that they drive the ecosystem under consideration towards a new state, considered from a human perspective to be less desirable (SCHEFFER, 2009).

Multiple factors can trigger a change in ecological regime towards a new ecosystem states (ROCHA et al., 2015b). These factors include climatic variations caused by global climate change (ROCHA et al., 2015b), increase in the number of extreme climatic events (WERNBERG et al., 2016), introduction of invasive species (KOTTA et al., 2018), increase in anthropogenic pressure (HALPERN et al., 2019), and loss of biogenic habitats (AIROLDI et al., 2008).

Biogenic habitats, formed by engineers' species, play a crucial role in the functioning of marine coastal ecosystems and in supporting components of biodiversity (VOZZO et al., 2023; ALBERTSON et al., 2022). For instance, mangroves serve as a nursery for coral-reef fish (NAGELKERKEN et al., 2002), coral act as a refuge from predators for many fish and invertebrates species (ALMANY, 2004), marine seagrasses enhance water quality by for instance increasing water pH (RICART et al., 2021), and their presence increases infauna diversity (GONZÁLEZ-ORTIZ et al., 2016; BOYÉ et al., 2017). Because biogenic habitats are also vulnerable to human pressures and natural disturbances (AIROLDI & BECK, 2007; BUTT et al., 2022; WERNBERG et al., 2023), they play a central role in mediating the effects

of current environmental changes on marine coastal ecosystems and biodiversity (SUNDAY et al., 2017; BULLERI et al., 2018).

Over the last few decades, regime shifts involving persistent changes in biogenic benthic habitats have been observed worldwide across temperate and tropical regions. Examples include shifts from macroalgal forests to either sea urchin barrens (LING et al., 2015) or turf (FILBEE-DEXTER & WERNBERG, 2018), from crustose coralline algae to turf (CORNWALL et al., 2023), and also from coral-dominated reefs to macroalgae (MCMANUS & POLSENBERG, 2004), or from seagrass to bare sediments (MAXWELL et al., 2017). These regime shifts share in common that as the biogenic habitats degrade or change, novel ecological dynamics take root to push and maintain the ecosystems into new states (NYSTRÖM et al., 2012). As these new states are often undesirable from a management perspective while also displaying hysteresis making it difficult to return to the initial state (NYSTRÖM et al., 2012), substantial efforts have been devoted to identifying, predicting and preventing regime shifts before they occur (BIGGS et al., 2009). Detection of early warning signals of regime shifts has been intensively investigated by analysing temporal ecosystem dynamics when long-term time series are available (SCHEFFER et al., 2009), or spatial patterns emerging from ecological dynamics (KÉFI et al., 2014; NIJP et al., 2019; WARD et al., 2018). Although promising theoretically based on numerical or mesocosm experiments, these data-greedy methods remain of limited practical interest in real world ecosystems across large spatial scales, as (i) they are often computed retrospectively after the shift has occurred (e.g. LITZOW et MUETER (2014)) and (ii) to date, no single metrics amongst the alternative candidates (e.g. increasing variance, skewness, etc...) has proven reliable across all case studies (HASTINGS & WYSHAM, 2010). An alternative that has gained increasing interests is to identify the environmental thresholds beyond which ecosystems shift to new states (KELLY et al., 2015). However, this approach has also proved challenging (TURNER et al., 2020; HILLEBRAND et al., 2020). Process-based simulation models can for instance contribute to estimating ecological thresholds and informing effective management strategies (MARZLOFF et al., 2013; MARZLOFF et al., 2016) but they remain restricted to theoretical cases or data-rich ecosystems. Overall, different strategies may be necessary to anticipate regime shifts across ecosystems, depending on data availability and ecosystem properties (e.g. non-linear responses, lags in response to effects of stressors; LITZOW et al. (2016); TURNER et al. (2020)).

In the case of coral reefs, defining and modelling the occurrence of alternative reef

states have improved our understanding and ability to predict the impact of anthropogenic pressures on reef states (JOUFFRAY et al., 2015; JOUFFRAY et al., 2019; DONOVAN et al., 2018). After identifying potential reef states (e.g. DONOVAN et al. (2018); Chapitre 2), modelling their spatial distribution can indeed help assess the relative importance of different environmental drivers (including climate-driven or local anthropogenic pressures), evaluate the extent to which drivers have synergistic or antagonistic effects on ecosystem states, and also characterise potential non-linear ecosystem responses across environmental gradients (JOUFFRAY et al., 2015; JOUFFRAY et al., 2019). All these elements are key to guide anticipatory strategies and define management actions (LITZOW & HUNSICKER, 2016; TURNER et al., 2020). An unexplored avenue concerns the potential for distribution models to identify areas where multiple ecosystem states are possible and hence, where supposedly changes in ecosystem states are more likely. To date however, such modelling strategy has only been led on coral reefs at a local scale (JOUFFRAY et al., 2019) so its transposability to other ecosystems across larger areas remains untested.

In this context, Chapter 2 defined habitat states using the *Reef Life Survey (RLS)* database, which provides a solid basis for a modelling exercise on a regional scale. Building on the dataset of reef states occurrences around Australia consolidated in Chapter 2, this study aims to explain and predict their distribution using SDM approaches, with the specific goals of (1) assessing the influence of biophysical factors and anthropogenic pressures on the occurrence of the different reef habitat states, (2) determining the transferability of these SDMs in order to quantify the predictability of reef habitat states at large scales, (3) identify reef states that share similar environmental niches and can potentially coexist, and (4) identify areas that may host several states, as these areas are the most likely to undergo state changes.

4.3 Materials & Methods

4.3.1 Datasets

Biotic dataset

In this study, we used the dataset produced in Chapter 2 from the *RLS* citizen science program where 50-m diver transects were classified into different habitat states based on percentage cover estimates of major benthic habitat groups. As a reminder of Chapter

2, we used a clustering pipeline composed of two distinct algorithm *UMAP* (MCINNES et al., 2020) for dimension reduction and *HDBSCAN* (MOULAVI et al., 2014; MCINNES et al., 2017) to classify the *RLS* benthic habitat relative cover (estimated through underwater dives; for further details see EDGAR et STUART-SMITH (2014)) for 6,554 transects into 17 distinct states. These 17 reef states include iconic biogenic habitats such as kelp forests, encrusting red algae, corals, and seagrass beds, as well as alternative degradation states of these habitats (i.e. branching coral vs. turf algae). Out of the 6,554 transects classified, 5,122 were sampled in Australian waters, which are the focus of this study. Not that this classification methodology excludes certain observations considered as noisy. Thus, from this dataset, we removed all transects classified as “noisy” (935 transects) while also removing transects for which at least one covariate had missing value due to the immediate proximity of the coastline (871 transects). Thus, the following analyses are based on 3,316 transects. As our goal in this study is to understand the distribution of biogenic habitats, we merged three out of the 17 initial habitat states (i.e. “Unconsolidated substrat”, “Bare substrat” and “Sand”) into a single group that corresponds to non-living substrate (“Substrate” hereafter). In addition, since turf algae can trap sediments and produce chemical substances preventing the recruitment of canopy-forming species (BUREK et al., 2018), we decided to merge the initial group “Sand and Turf algae” with the group “Turf algae”. Overall, 14 reef states are modelled in this study (See appendix A, Table S1 for their description and Fig. S1 for their distribution across the dataset).

Predictors dataset

The predictors used for model fitting were extracted from two sources : *Bio-Oracle* (ASSIS et al., 2018) for biophysical predictors and *Ocean Health Index* (HALPERN et al., 2019) for anthropogenic pressures. *Bio-Oracle* provides aggregated data for the period 2000-2014 at a resolution of 0.083° (9.28 km at the equator). From this database, we extracted the mean values of four biophysical predictors known to be important for coastal ecosystems (Table 4.1).

The Ocean Health Index database provides yearly estimates for 14 predictors for the period 2003-2013 at a resolution of 0.0083° (928 m at the equator). We extracted yearly data for the eight predictors known to impact coastal ecosystems and computed the mean for all predictors. We also computed the standard deviation for the predictors linked to fishing activity (i.e. Artisanal fishing, Fishing High and Low Bycatch, Fishing Non-

Destructive High and low Bycatch) . We then performed a principal component analysis on all predictors related to fisheries so as to synthesise the core information related to fishing impacts. The first two axes explained 44.3% and 21.3% of the variability, respectively. The first axis mostly describes variation in the intensity of destructive fishing activity for marine benthic habitats, whereas the second axis is related to non-destructive fishing activity (See Fig. S2 appendix A). In addition, we ensured that the selected predictors were not too highly correlated with each other (all absolute values of Pearson’s correlations < 0.7 ; Fig. S3 in appendix A).

Table 4.1 – *Environmental variables extracted from Bio-Oracle and anthropogenic variables extracted from Ocean Health Index. See ASSIS et al. (2018) for a detailed description of the Bio-Oracle variables and their acquisition method and Table S1 to S3 in HALPERN et al. (2019) for detailed description and acquisition method of anthropogenic variables*

Data source	Variable	Metric	
<i>Bio-oracle</i>	Current velocity	Mean	
	Diffuse attenuation		
	Primary productivity		
	Sea Surface Temperature		
	Artisanal Fishing		
	Fishing High Bycatch		
<i>Ocean Health Index</i>	Fishing Low Bycatch	PCA axis on Mean and Standard deviation	
	Fishing Non Destructive High Bycatch		
	Fishing Non Destructive Low Bycatch		
	Direct Human Disturbance ¹		
	Nutrient Pollution		Mean
	Marine Heatwaves Anomaly ²		

Since the *Bio-Oracle* and *Ocean Health Index* datasets have different resolutions and

1. Coastal density population within a 10 km radius of the coastline. See Table S1 and S2 in HALPERN et al. (2019) supplementary information for more details.

2. Marine Heatwaves Anomaly is calculated by subtracting the count of extreme sea surface temperature weeks during a five-year period from the count of extreme sea surface temperature weeks during a baseline five-year period (1985-1989). See Table S1 and S2 in HALPERN et al. (2019) supplementary information for more details.

projections, all predictors were reprojected to a common projection (EPSG:4326) and a common resolution of 0.0083° , which is the highest resolution offered by *Ocean Health Index*. Hence, bio-oracle predictors were downscaled to match the *Ocean Health Indices* resolution.

4.3.2 Statistical analyses

Model fitting and evaluation

We modelled habitat states using a multivariate Random Forest model (BREIMAN, 2001) fitted using the Python library Scikit-Learn (PEDREGOSA et al., 2011). The Random Forest model was trained on 163,800 combinations of hyperparameters “*max_depth*”, “*max_features*”, and “*n_estimators*” (see Table S2 in appendix B). For each combination of hyperparameters, we binarised the probabilities obtained from the Random Forest model according to reef state-specific thresholds that maximized the True Skill Statistic (MaxTSS; ALLOUCHE et al. (2006)), and only retained combination of hyperparameters having the highest MaxTSS on the explanatory folds (see below for the train/test split of the dataset).

To address the proximity of some transects and the resulting spatial autocorrelation, we used a 3-fold cross-validation procedure with spatial blocks for model training. Using the *blockCV* R package (VALAVI et al., 2019), the autocorrelation range of the presences and absences between the different habitats states was estimated as 524 km^2 , which was then used to split our study area into 37 blocks of that size (equal area across blocks) to ensure that the spatial correlation between any two blocks was negligible. Finally, each block and all the transects it contains were randomly assigned into three folds. This whole procedure ensured that spatially close transects were not used to both train and validate the model (See Fig S5. in appendix A). Using this method, we allocated the 3,316 transects into three folds that were used for training (and assessing explanatory power) as well as for estimating predictive power (training with two folds to predict the remaining one, in a classical cross-validation scheme). The distribution of different habitat states within the different training and test folds is presented in Fig S6 in appendix A.

The performance of explanatory power and predictive power were evaluated using a set of complementary metrics including the maximum of the True Skill Statistic (Max TSS; ALLOUCHE et al. (2006)), the F1-score, the Area Under the Precision-Recall Curve (AUPRC; FLACH et KULL (2015)) and the AUC (FAWCETT, 2006). Because model predictive

power was poor, tThe best hyperparameter combination based on explanatory power was then used to fit the model on the whole dataset.

Model predictions

The most important variables in explaining the distribution of each reef habitat state were identified using the *SHAP* (*SHapley Additive exPlanations*) framework (LUNDBERG & LEE, 2017). is a framework rooted in cooperative game theory (SHAPLEY, 1953) that quantifies the contribution of each predictor to model predictions (LUNDBERG & LEE, 2017). *SHAP* values were then averaged across states to identify the three most important variables across all states. Then, for each habitat state, we modelled their dependence against these three most influential variables using Partial Dependence Plots (i.e. marginal response curves ; MOLNAR (2022)) across the observed range of our predictors.

Then, we investigated whether certain environmental conditions could be suitable for multiple habitats. To this end, we used the threshold that maximised the TSS for each habitat state during the hypertuning of the parameters of the Random Forest model (Fig. S7), to binarise occurrence probabilities into presences and absences. We then stacked these binary predictions across habitats to compute the number of states predicted as present in each pixel/set of environmental conditions. As performed for individual reef states, we explored the response of this number of predicted states against the three most influential variables, identified above, using Partial Dependence Plots.

From the stacked predictions, we also identified the habitat states that can co-occur in the same transect. We calculated the relative frequency with which each habitat state was associated with zero, one, two or more of the other habitat states. In order to study the spatial distribution of the number of predicted habitat states, we further calculated the minimal, median and maximum number of habitat states per spatial block. That scale was chosen since spatial blocks were built to be uncorrelated (and therefore independent) spatial units. Using a linear regression, we then investigated whether these spatial summary statistics varied depending on the coordinates of the spatial blocks (using the centroid of the block), and the number of transects within each spatial block. Quadratic effects were included for the latitude and longitude of the spatial blocks in order to account for potential non-linear effects.

4.4 Results

4.4.1 Model Performance & Predictors importance

The models with the selected hyperparameters showed good explanatory power (measured on the fold on which the model was trained) with an average Max TSS across all habitat states of 0.78 ± 0.03 (mean \pm standard deviation ; Fig. S6 in appendix B). AUC values ranged from 0.89 to 0.99 (Fig. S8-21 in appendix B), while AUPRC values ranged from 0.48 to 0.71 (Fig. S22-35). The best explanatory power according to maxTSS was obtained for “Soft coral and gorgonians” (TSS = 0.94), while the worst was obtained for “Turf” (TSS = 0.49). The predictive power (calculated on the fold not used for model training) was significantly worse, with a mean Max TSS of 0.13 ± 0.04 (mean \pm standard deviation).

Our results show that habitat states usually respond to a set of weak predictors (i.e. most predictors contributed similarly to the explanatory power, although with some slight differences ; Fig. 4.1). The three most important variables were Sea Surface Temperature, Fishing PC1 and Marine Heatwave anomalies (see Fig. S36-50 in appendix C). Although these three variables were the most important considering all habitats, the occurrence of certain habitats was best predicted by other covariates. For example, Primary Productivity appeared as a relevant predictor of Crustose coralline algae distribution, whereas Current Velocity was important for both Filamentous algae (Fig. S42 in appendix C), and Seagrass (Fig. S47. in appendix C). Overall, Direct Human Disturbance had a low explanatory power for all habitat states.

4.4.2 Habitat states patterns

Overall, stacking predictions showed that most conditions are suitable to multiple habitat states, with a richness (i.e. number of predicted reef states for a given set of environmental conditions) of 3.43 ± 1.50 (mean \pm standard deviation) ranging from one to eight potential habitat states occurring per transect. The richness varied differently along environmental gradients (Fig. 4.2 a) : for instance, large variations around an average richness value of 4 were observed for Sea Surface Temperature whereas a global (non-linear) decrease was observed when fishing pressure or Marine heat-wave anomalies increased.

Changes in habitat state richness along environmental gradients depends on the res-

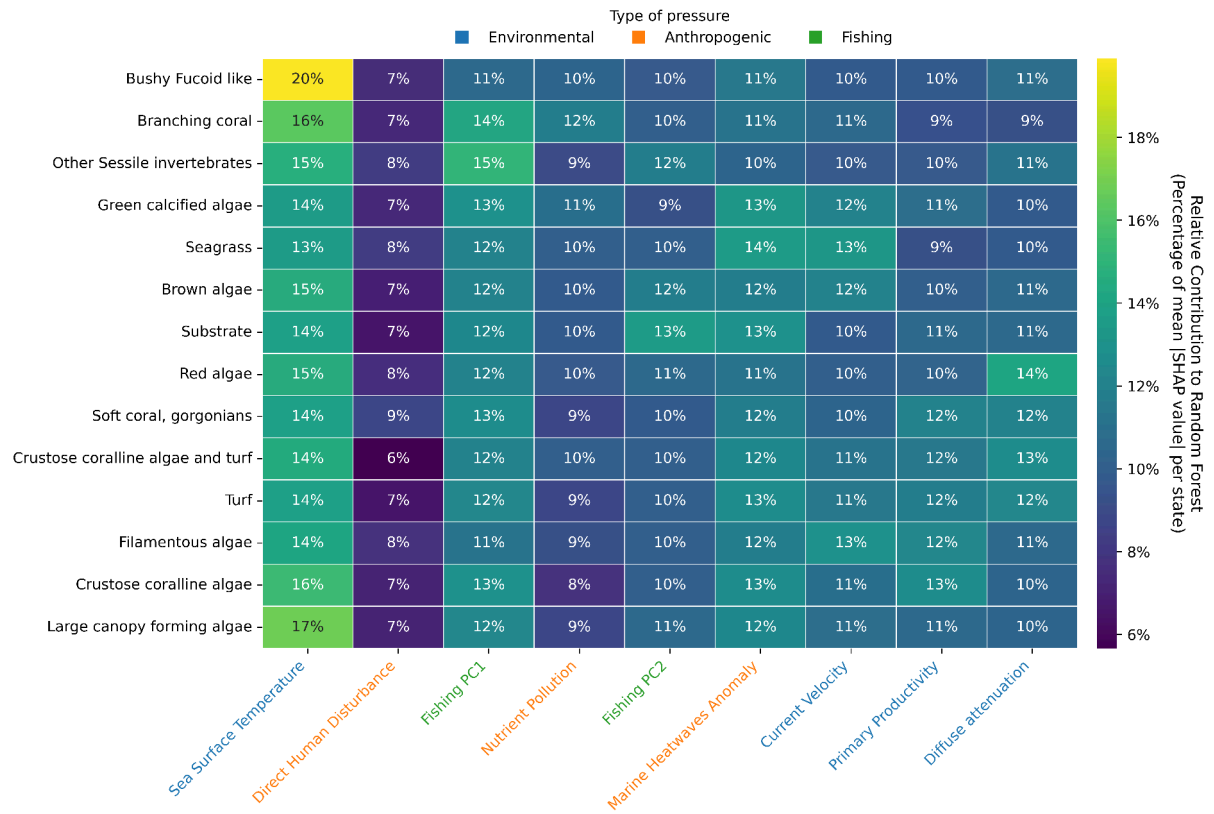


Figure 4.1 – Heatmap of the relative contribution of variables to the Random Forest explanatory power. The variables are colour coded according to the type of predictor (blue : abiotic conditions, green : fishing pressure, orange : other anthropogenic pressures). The relative contribution of each variable for a given state is expressed relative to the sum of all variable’s contributions for that state.

ponse of the 14 reef habitat states along these gradients (Fig. 4.2 b). Indeed, the relative stability of habitat state richness along the mean temperature gradient does not reflect a stability in the response of the different states but rather that some states are replacing others at certain temperature values. For instance, when Sea Surface Temperature exceeds 20°C, the model predicts a strong decrease in the probability of occurrence of Canopy Forming algae and Bushy Furoid states but an increase in the probability of states dominated by Crustose coralline algae with and without Turf and by Green calcified algae. Above that same 20°C threshold, the probability of occurrence of the Turf state is maximal. With regards to fishing pressure, the probability of occurrence of Large canopy forming algae increases with fishing pressure. For Marine Heatwaves, while the probability of occurrence Green calcified algae decreases as the frequency of Marine Heatwaves

increases, the opposite is observed for Bushy Furoid like states.

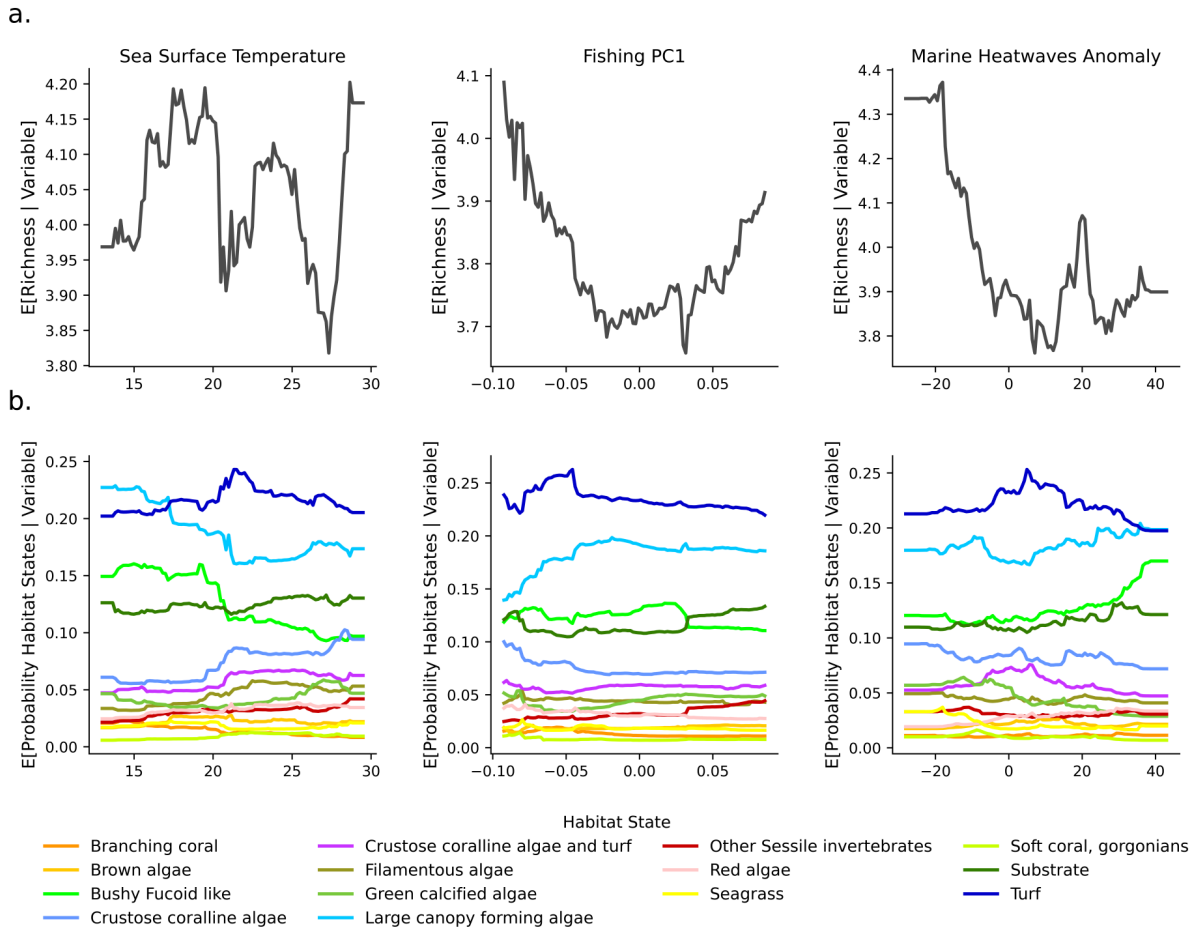


Figure 4.2 – Partial Dependence Plots of the a. Habitat states richness b. probability of each of the 14 habitat states against the three most important variables identified previously for these habitat states, namely : Sea Surface Temperature, Fishing PC1 (Fishing pressure increases as Fishing PC1 increases) and Marine Heatwaves Anomaly. For all plots, the X-axis is expressed in the original unit of the variables. For graphical representation, we only plot here the [0% ; 90%] values of Fishing PC1, excluding the few positive extreme values.

Among the 3,316 transects analysed, 222 (~6.5%) are predicted to be only suitable to a single habitat state while 71% exhibit favourable conditions to accommodate two to four different habitat states (Fig. 4.3 a). Out of the 16,384 possible combinations of occurrence and co-occurrence of the 14 habitat states, we predict only 302 (~2%) combinations of co-occurring habitat states. Among these combinations, 12 comprised one habitat state (note that Brown algae as well as Soft coral and gorgonians were never predicted to occur alone), while a maximum of 112 combinations were observed in transects predicted to be

suitable for three different habitat states.

Interestingly, some reef habitat states (e.g. Turf or Large canopy forming algae) are more likely to occur in conditions suitable for only a few habitat states (one or two habitat states) while others (e.g. Filamentous algae, Red algae, Brown algae and Other Sessile invertebrates) are more likely to occur in conditions supposedly favourable to a large number of habitat states (Fig. 4.3 b).

Further, our results indicate a preferential association between certain habitat states (Fig. 4.3 c). For instance, some iconic reef habitat states such as Branching coral are mainly predicted in areas that are also favourable to Turf (19% of the time) and Large Canopy Forming algae (14%). Other reef states such as Bushy Furoid like are mainly associated with Large canopy forming algae (15%), Turf (12%), Substrate (11%), and Filamentous algae (11%).

4.4.3 Spatial distribution

At the Australian geographical scale, different spatial patterns of richness (i.e. number of predicted reef states) were observed depending on the region (Fig. 4.4). Most spatial blocks around Australia have transects where only one unique habitat state can occur (Fig. 4.4 a). The majority of spatial blocks have a median richness of 3 (Fig. 4.4 b). One has a richness of 6, but the number of transects in this spatial block is relatively low, with only seven transects sampled. On the east coast of Australia, in the Tweed-Moreton and Manning-Hawkesbury ecoregions, two ecoregions at the transition between temperate and tropical zones, the max richness was 8 whereas the median value of the adjacent temperate or tropical zones was lower. Over 73% of spatial blocks can accommodate up to seven or eight different reef habitat states (Fig. 4.4 c). Only the maximum of predicted reef states showed a significant relationship, albeit weak (linear coefficient of 0.007), with the number of transects carried out ($p < 0.01$).

Chapitre 4 – Patterns of Co-occurrence and Spatial Richness in Marine Habitats : Insights from Large-Scale Modelling

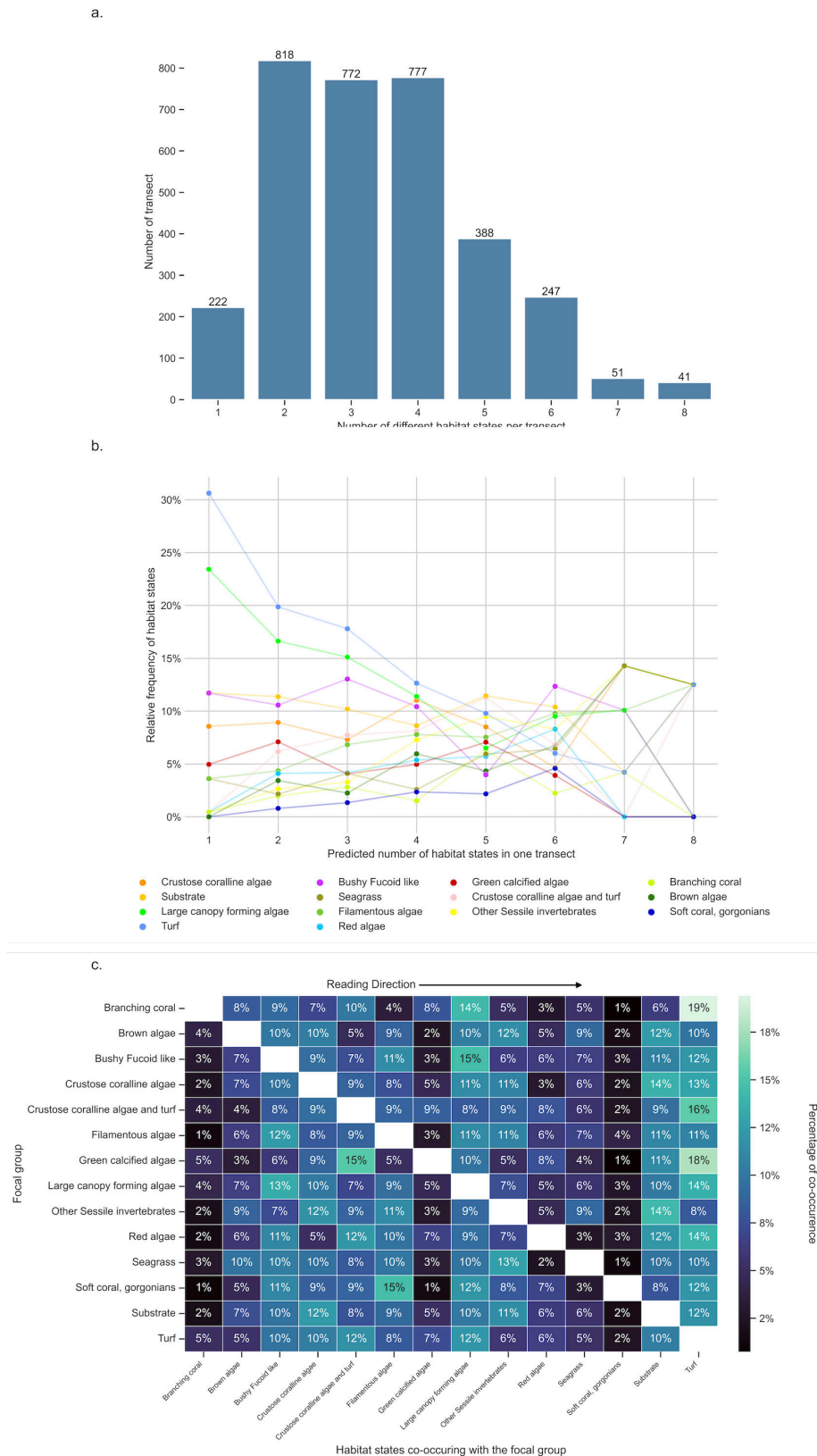


Figure 4.3 – a. Distribution of the number of possible habitat states per transects, the numbers above the bar represent the number of transects predicted to have this number of habitat states. b. Relative frequency of habitat states as a function of predicted number habitat states in one transect. c. Heatmap of the frequency of co-occurrence of each habitat state. Each line of the heatmap has been standardised by the total number of co-occurrences with the focal habitat state.

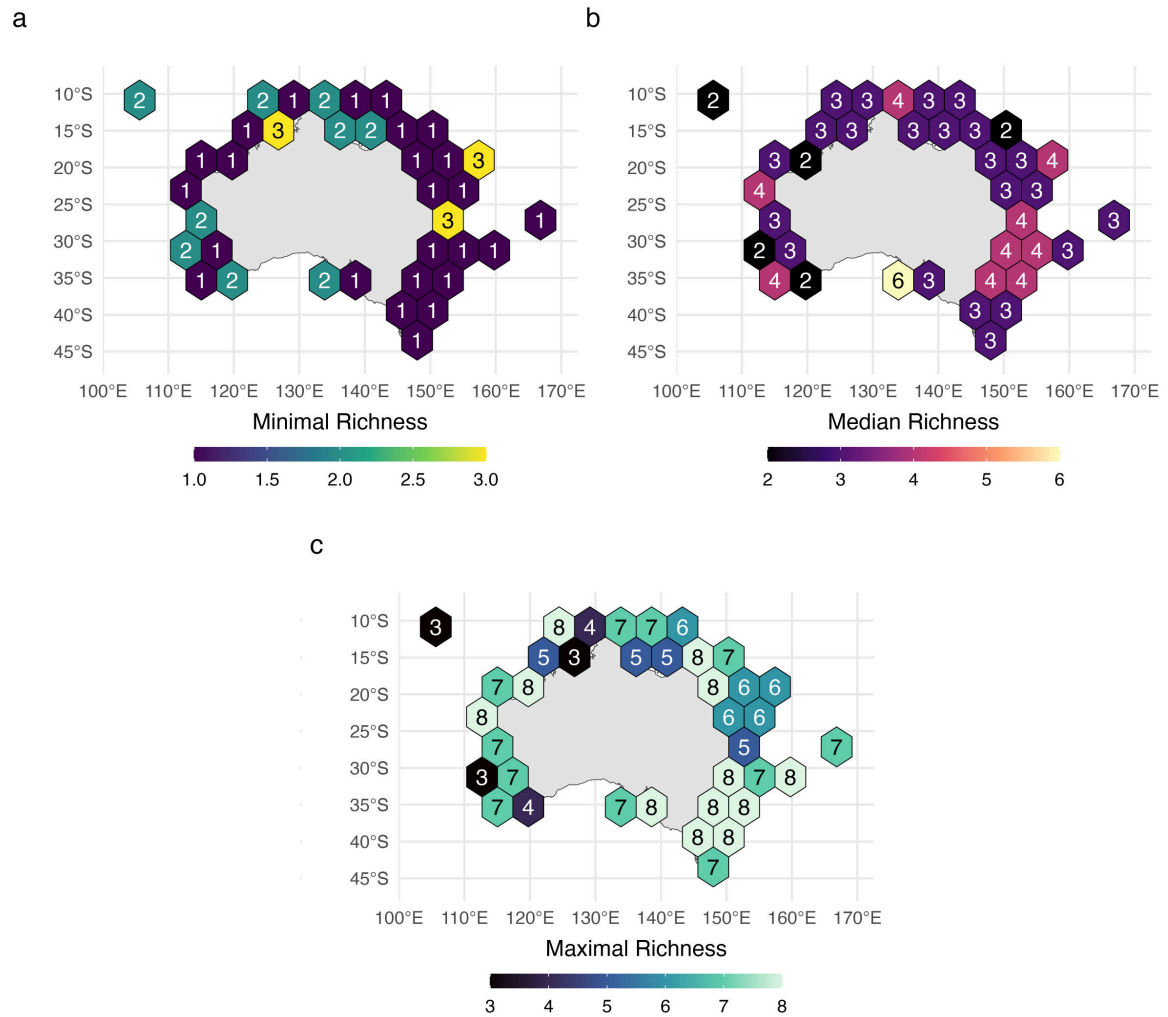


Figure 4.4 – a. Minimal, b. median and c. maximal richness (i.e. number of predicted reef states) per spatial block

4.5 Discussion

In this study, we built upon Chapter 2 to explain and predict the spatial distribution of multiple reef benthic habitats. Overall, the random forest multivariate modelling approach we have used showed substantial explanatory power but a low predictive power, suggesting poor transferability. Despite evidence for a good explanatory power, all predictors presented a similar contribution to the variance explained precluding us to clearly identify the drivers of the spatial distribution of most habitats. Partial dependence plots highlighted some meaningful relationships and showed a differential response of habitats along environmental gradients that may ultimately trigger the emergence of alternative stable states. Importantly, some environmental conditions appeared to be suitable for several habitats suggesting that the current existence of an habitat in a given area may be due to historical contingencies and/or biotic settings.

4.5.1 Explanatory and predictive power of the model

Our model demonstrates some explanatory power, especially when it comes to elucidating the distribution of the rarest habitat states in the dataset such as Branching coral or Soft coral and gorgonians (See Fig. S1 in appendix A and Fig. S6 in appendix B). This outcome suggests that the model is able to decipher the reason why habitat states are absent and gives us insight about the complex relationships between environmental factors and habitat states. This strongly contrasts with the model's predictive power, which is quite low overall. Several factors may contribute to the low transferability of the models. First, the reef habitat states defined in Chapter 2 are derived from cover data of about 50 substratum types defined accordingly to the CATAMI benthic imagery classification scheme (ALTHAUS et al. (2015); for further details, see EDGAR et al. (2020)), merged into 24 functional group. Hence, the initial reef state categorisation is based on cover data of morpho-types that does not take into account the identity of the underlying species. As a result, some reef states may have large environmental niches in our model while they are in fact driven by species with much narrower niches and located in different areas. For example, the diversity of Crustose coralline algae is estimated at several hundred different species (DEAN et al., 2015; TWIST et al., 2019), which can be found in both temperate and tropical zones (SISSINI et al., 2022). The same goes for Turf, which is a rather loosely defined functional group (CONNELL et al., 2014). Turf can include nearly a hundred different species, sometimes with a lack of consistency in its definition across space

(CONNELL et al., 2014). Hence, while Turf can be found in both tropical and temperate zones, the underlying species and their ecology may greatly differ across different areas (FILBEE-DEXTER & WERNBERG, 2018). Furthermore, it's empirically observed that benthic assemblages exhibit more reliable and quantifiable spatial patterns compared to fish assemblages. This distinction has considerable implications when it comes to delineating appropriate management zones (SANDIN et al., 2022).

Furthermore, discrepancies in predictor resolutions and misalignment with habitat states pose additional hurdles (POTTER et al., 2013). Our environmental dataset operates at best at a resolution of 1 km², while the photo quadrats from *RLS* transects, cover an approximate area of 20 m². This spatial disparity hampers the models to make accurate predictions (CONNOR et al., 2018; MOUDRÝ et al., 2023), the estimation of environment-species relationships may be flattened (MEYNARD et al., 2023), resulting in an incorrect estimation of environmental niches. Variability of sampling effort per spatial block and differences in habitat states prevalence could be another potential explanatory factor for the model's limited predictive power. Variations in data collection intensity across different spatial blocks or habitats could introduce bias and affect the MaxTSS threshold and thus model prediction performances (SOMODI et al., 2017; LEROY et al., 2018). However, it is unlikely that our predictive performances were particularly affected by the choice of metrics, since for the others the predictive power was also very low. Finally, prevalence issues in sampling data hampered our predictive power in two ways. Firstly, The most prevalent groups are more prevalent by orders of magnitude. Turf, Canopy forming algae, Furoid like algae represent respectively 24%, 17% and 13%, comparatively to Brown algae, Branching coral and Soft cora and gorgonians that represent only 2% and 1% each for the last two in the whole dataset (Fig. S1).

In light of these challenges, it is important to recognise that while our model apparently performs well in explaining the distribution of habitat states, its practical utility for making precise predictions is rather limited. Nonetheless, there are strategies for enhancing predictive power, such as improving predictor resolution to be closer to the scale at which transects are sampled, increasing sampling in the less well sampled spatial block, or the use of machine learning techniques to over-sample less prevalent habitats or under-sample more prevalent ones (HE & GARCIA, 2009). This latter strategy has not been applied in this study, as it could be an alternative to the spatial-block approach that we used in this study (GAUL et al., 2022) especially when the data come from a citizen science program

(ROBINSON et al., 2018; ROBINSON et al., 2020). All the avenues presented here could be used to improve the predictive power of this type of model, but further studies are still needed to assess the extent to which each of the approaches presented, such as upscaling methods (e.g. MEYNARD et al. (2023)) or over/under-sampling, can improve the predictive power of *SDMs*.

4.5.2 Influence of biophysical factors and anthropogenic pressures

The most important variables identified for the different habitat states were, in order, Sea Surface Temperature, Fishing PC1 and Marine Heatwaves Anaomaly. Conversely, the variables manifesting the least influence across all habitat states were anthropogenic pressures, specifically Nutrient Pollution and Direct Human Disturbance. Our findings align with those of JOUFFRAY et al. (2019), yet it is crucial to acknowledge a limitation that both our study and theirs share. Certain anthropogenic pressures, such as Nutrient Pollution, can be highly localized, posing challenges for regional models to accurately capture. Consistent with the observations of JOUFFRAY et al. (2019), our model underscores the critical role of Sea Surface Temperature in shaping habitat states. This aligns with previous literature emphasizing temperature's significance in predicting benthic diversity patterns, as evidenced by BELANGER et al. (2012).

Response curves provide insights into the relationships between environmental factors and certain habitat states. Overall, the response curves we have generated exhibit a commendable degree of consistency with the existing literature. For instance, we confirm that Filamentous algae tend to thrive in areas less exposed to marine currents (QUINTANO et al., 2015), while Branching coral are more prevalent in warm temperate waters (HIGUCHI et al., 2015). These patterns provide valuable ecological insights and corroborate our understanding of the response of these habitats along environmental gradients. However, it is important to recognise that not all relationships exhibit such consistency. Notably, the positive link between fishing pressure or marine heatwaves and the prevalence of Large canopy forming algae, is less straightforward and even seems contradictory (WERNBERG et al., 2016). It is crucial to contextualise these findings within the framework of our modelling approach. The SDM model employed in this study is correlative by essence (SHABANI et al., 2016), only capturing statistical associations between environmental variables and habitat states. While these correlations can reveal important ecological patterns, they do not establish causal relationships. Understanding the drivers behind these relationships

requires a nuanced perspective. For example, the observed positive correlation between Large canopy forming algae and commercial fishing activity may be the result of the high fishing commercial value of kelp forests, where approximately \$30,000 of fish products are extracted per hectare and per year (EGER et al., 2023). Similarly, the relationship between marine heatwaves and the probability of occurrence of Large canopy forming algae should be interpreted in the context of ecological dynamics that Australian shores recently experienced. Indeed, major episodes of marine heatwaves occurred in 2011 and 2015 (Hobday et al. 2018) that have led to the decline of kelp, a common large canopy-forming species, in the following years (WERNBERG et al., 2016). This may lead to a spurious correlation in our dataset, since transect may have been sampled in these areas before the kelps disappeared. Furthermore, Large Canopy Forming algae contains more than just temperate macroalgae, some species of tropical fleshy macroalgae such as *Sargassum* spp. have been incorporated as Large canopy forming categories. This kind of fleshy macroalgae can be found growing on the top of coral reef after a regime shift to a macroalgae dominated reef (SMITH et al., 2022), especially if the coral reef has been degraded due to marine heatwaves (DONOVAN et al., 2021).

4.5.3 Co-occurrence patterns

Our exploration of co-occurrence patterns among reef habitat states has yielded intriguing findings shedding light on the ecological dynamics of these underwater ecosystems. It's striking to note that only a mere 2% of the possible combinations among the 14 habitat states were observed in our data. This limited co-occurrence of habitat states within transects suggests a complex interplay of ecological factors governing their distribution. Certain habitat states exhibit weak co-occurrence patterns with others, implying distinct environmental responses or the ability for these states to thrive in inhospitable conditions. For instance, Branching coral and Soft corals and gorgonians rarely co-occur (Fig. 4.3 c.), this can be explain by the small number of corals habitats states in our dataset (see Discussion in Chapter 2). It can also be a manifestation of niche differences between the two different groups, although competition between these two groups may also contribute to this pattern (SAMMARCO et al., 1983).

Another interesting observation concerns Large canopy forming algae, and Turf that are frequently predicted alone. While this pattern could be the result of a bias in our model where the differential prevalence of habitat states combined with the variable number of

transects per spatial block could influence the joint occurrence probabilities predicted by the model in each observational unit, an ecological explanation is also possible. Specifically, we found that other habitat types presenting similar characteristics with wide ecological niches and a high prevalence (e.g. Bushy Furoid-like, Crustose coralline algae; Fig. S1 in appendix A) do not exhibit similar patterns (Fig. 4.3 b), this observation suggests that the influence of the this bias may be less pronounced than initially perceived, indicating the potential involvement of other underlying factors. For Turf it may be due to areas that have already undergone a regime shift (i.e. the original habitat has disappeared and has been replaced by Turf; JOUFFRAY et al. (2015)), regime shift that struck hundreds of kilometres of Australian coastline (FILBEE-DEXTER & WERNBERG, 2018). For Large canopy forming algae, it may be due to a bias in the *RLS* sampling method, which, with its photoquadrat, does not take into account understory habitat, such as Red algae, Brown algae and Crustose coralline algae, thus underestimating the niche of the low-profile understory habitat states.

4.5.4 Richness Spatial Patterns

Our examination of spatial richness patterns within spatial-blocks has unveiled insights into the distribution of habitat states along the Australian coast. This analysis raises important questions about the ecological significance of these patterns and their potential implications for the region’s marine ecosystems. One notable observation is that the majority of spatial-blocks predominantly exhibit a single ecological status. This phenomenon highlights the prevalence of specific habitat types in various regions, which can be either desirable, such as areas characterised by Large canopy forming, Bushy Furoid like, or seagrass (COLEMAN & WERNBERG, 2017; FILBEE-DEXTER & WERNBERG, 2018; JÄNES et al., 2021), or less desirable, such as those dominated by Turf (FILBEE-DEXTER & WERNBERG, 2018). This raises an intriguing question : do these unique habitat states persist in areas that have remained largely unaffected by regime shifts, or do they represent remnants of past ecological transformations? Understanding the historical context and ecological drivers behind these spatial patterns is crucial to unravelling their significance.

Intriguingly, the eastern Australian transition zone stands out with a greater median habitat diversity. This may be linked to the ongoing process of tropicalisation occurring in the region (FIGUEIRA & BOOTH, 2010). As the climate warms, there is increasing evidence of shifts in marine ecosystems, including the expansion of tropical species into temperate

zones (VERGÉS et al., 2014). This phenomenon could potentially lead to regime shifts in the near future (VERGÉS et al., 2019), making the eastern Australian transition zone a focal point for monitoring and research. Furthermore, it's worth noting that a substantial number of spatial blocks exhibit the maximum number of habitat states. This intriguing finding suggests the possibility of ongoing or impending state changes across the entire Australian coast.

4.6 Conclusion

Our study delves into the spatial distribution of multiple reef benthic habitats and highlighted that some areas are more prone to exhibit multiple habitat states based on environmental niche similarities. Yet while our model provides some insights in explaining the distribution of habitat states and its associated ecological drivers, its predictive power appeared limited for several reasons including resolution mismatch between response and predictor variables or the way habitats were clustered. Despite these limits, which are mostly tailored to the datasets used, this approach could prove useful to identify the conditions under which multiple habitats can be found, and therefore the areas where regime shifts are more likely to take place.

Bibliographie

- AIROLDI, L., & BECK, M., (2007), Loss, status and trends for coastal marine habitats of Europe, In R. N. GIBSON, R. J. A. ATKINSON & J. D. M. GORDON (Éd.), *Oceanography and Marine Biology An Annual Review* (1st, T. 45), CRC Press.
- AIROLDI, L., BALATA, D., & BECK, M. W., (2008), The Gray Zone : Relationships between habitat loss and marine diversity and their applications in conservation, *Journal of Experimental Marine Biology and Ecology*, 3661, 8-15.
- ALBERTSON, L. K., SKLAR, L. S., TUMOLO, B. B., CROSS, W. F., COLLINS, S. F., & WOODS, H. A., (2022), The ghosts of ecosystem engineers : Legacy effects of biogenic modifications, *Functional Ecology*.
- ALLOUCHE, O., ASAF, T., & KADMON, R., (2006), Assessing the accuracy of species distribution models : prevalence, kappa and the true skill statistic (TSS), *Journal of Applied Ecology*, 436, 1223-1232.
- ALMANY, G. R., (2004), Does increased habitat complexity reduce predation and competition in coral reef fish assemblages?, *Oikos*, 1062, 275-284.

- ALTHAUS, F., HILL, N., FERRARI, R., EDWARDS, L., PRZESLAWSKI, R., SCHÖNBERG, C. H. L., STUART-SMITH, R., BARRETT, N., EDGAR, G., COLQUHOUN, J., TRAN, M., JORDAN, A., REES, T., & GOWLETT-HOLMES, K., (2015), A Standardised Vocabulary for Identifying Benthic Biota and Substrata from Underwater Imagery : The CATAMI Classification Scheme, *PLOS ONE*, *1010*, 1-18.
- 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*, *273*, 277-284.
- BELANGER, C. L., JABLONSKI, D., ROY, K., BERKE, S. K., KRUG, A. Z., & VALENTINE, J. W., (2012), Global environmental predictors of benthic marine biogeographic structure, *Proceedings of the National Academy of Sciences*, *10935*, 14046-14051.
- BIGGS, R., CARPENTER, S. R., & BROCK, W. A., (2009), Turning back from the brink : Detecting an impending regime shift in time to avert it, *Proceedings of the National Academy of Sciences*, *1063*, 826-831.
- BOYÉ, A., LEGENDRE, P., GRALL, J., & GAUTHIER, O., (2017), Constancy despite variability : Local and regional macrofaunal diversity in intertidal seagrass beds, *Journal of Sea Research*, *130*, 107-122.
- BREIMAN, L., (2001), Random Forests, *Machine Learning*, *451*, 5-32.
- BULLERI, F., ERIKSSON, B. K., QUEIRÓS, A., AIROLDI, L., ARENAS, F., ARVANITIDIS, C., BOUMA, T. J., CROWE, T. P., DAVOULT, D., GUIZIEN, K., IVEŠA, L., JENKINS, S. R., MICHALET, R., OLABARRIA, C., PROCACCINI, G., SERRÃO, E. A., WAHL, M., & BENEDETTI-CECCHI, L., (2018), Harnessing positive species interactions as a tool against climate-driven loss of coastal biodiversity, *PLOS Biology*, *169*, 1-19.
- BUREK, K. E., O'BRIEN, J. M., & SCHEIBLING, R. E., (2018), Wasted effort : recruitment and persistence of kelp on algal turf, *Marine Ecology Progress Series*, *600*, 3-19.
- BUTT, N., HALPERN, B. S., O'HARA, C. C., ALLCOCK, A. L., POLIDORO, B., SHERMAN, S., BYRNE, M., BIRKELAND, C., DWYER, R. G., FRAZIER, M., WOODWORTH, B. K., ARANGO, C. P., KINGSFORD, M. J., UDYAWER, V., HUTCHINGS, P., SCANES, E., MCCLAREN, E. J., MAXWELL, S. M., DIAZ-PULIDO, G., ... KLEIN, C. J., (2022), A trait-based framework for assessing the vulnerability of marine species to human impacts, *Ecosphere*, *132*, e3919.
- COLEMAN, M. A., & WERNBERG, T., (2017), Forgotten underwater forests : The key role of fucoids on Australian temperate reefs, *Ecology and Evolution*, *720*, 8406-8418.
- CONNELL, S., FOSTER, M., & AIROLDI, L., (2014), What are algal turfs? Towards a better description of turfs, *Marine Ecology Progress Series*, *495*, 299-307.
- CONNOR, T., HULL, V., VIÑA, A., SHORTRIDGE, A., TANG, Y., ZHANG, J., WANG, F., & LIU, J., (2018), Effects of grain size and niche breadth on species distribution modeling, *Ecography*, *418*, 1270-1282.
- CORNWALL, C. E., CARLOT, J., BRANSON, O., COURTNEY, T. A., HARVEY, B. P., PERRY, C. T., ANDERSSON, A. J., DIAZ-PULIDO, G., JOHNSON, M. D., KENNEDY, E., KRIEGER, E. C., MALLELA, J., MCCOY, S. J., NUGUES, M. M., QUINTER, E., ROSS, C. L., RYAN, E., SADERNE, V., & COMEAU, S., (2023), Crustose coralline algae can contribute more than corals to coral reef carbonate production, *Communications Earth & Environment*, *41*, 105.

- CRÉPIN, A.-S., BIGGS, R., POLASKY, S., TROELL, M., & DE ZEEUW, A., (2012), Regime shifts and management, *Ecological Economics*, 84, 15-22.
- DEAN, A. J., STENECK, R. S., TAGER, D., & PANDOLFI, J. M., (2015), Distribution, abundance and diversity of crustose coralline algae on the Great Barrier Reef, *Coral Reefs*, 342, 581-594.
- DONOVAN, M. K., BURKEPILE, D. E., KRATOCHWILL, C., SHLESINGER, T., SULLY, S., OLIVER, T. A., HODGSON, G., FREIWALD, J., & van WOESIK, R., (2021), Local conditions magnify coral loss after marine heatwaves, *Science*, 3726545, 977-980.
- DONOVAN, M. K., FRIEDLANDER, A. M., LECKY, J., JOUFFRAY, J.-B., WILLIAMS, G. J., WEDDING, L. M., CROWDER, L. B., ERICKSON, A. L., GRAHAM, N. A. J., GOVE, J. M., KAPPEL, C. V., KARR, K., KITTINGER, J. N., NORSTRÖM, A. V., NYSTRÖM, M., OLESON, K. L. L., STAMOULIS, K. A., WHITE, C., WILLIAMS, I. D., & SELKOE, K. A., (2018), Combining fish and benthic communities into multiple regimes reveals complex reef dynamics, *Scientific Reports*, 81, 16943.
- EDGAR, G. J., COOPER, A., BAKER, S. C., BARKER, W., BARRETT, N. S., BECERRO, M. A., BATES, A. E., BROCK, D., CECCARELLI, D. M., CLAUSIUS, E., DAVEY, M., DAVIS, T. R., DAY, P. B., GREEN, A., GRIFFITHS, S. R., HICKS, J., HINOJOSA, I. A., JONES, B. K., KININMONTH, S., ... STUART-SMITH, R. D., (2020), Establishing the ecological basis for conservation of shallow marine life using Reef Life Survey, *Biological Conservation*, 252, 108855.
- EDGAR, G. J., & STUART-SMITH, R. D., (2014), Systematic global assessment of reef fish communities by the Reef Life Survey program, *Scientific Data*, 11.
- EGER, A. M., MARZINELLI, E. M., BEAS-LUNA, R., BLAIN, C. O., BLAMEY, L. K., BYRNES, J. E. K., CARNELL, P. E., CHOI, C. G., HESSING-LEWIS, M., KIM, K. Y., KUMAGAI, N. H., LORDA, J., MOORE, P., NAKAMURA, Y., PÉREZ-MATUS, A., PONTIER, O., SMALE, D., STEINBERG, P. D., & VERGÉS, A., (2023), The value of ecosystem services in global marine kelp forests, *Nature Communications*, 141, 1894.
- FAWCETT, T., (2006), An introduction to ROC analysis, *Pattern Recognition Letters*, 278, 861-874.
- FIGUEIRA, W. F., & BOOTH, D. J., (2010), Increasing ocean temperatures allow tropical fishes to survive overwinter in temperate waters, *Global Change Biology*, 162, 506-516.
- FILBEE-DEXTER, K., & WERNBERG, T., (2018), Rise of Turfs : A New Battlefront for Globally Declining Kelp Forests, *BioScience*, 682, 64-76.
- FLACH, P., & KULL, M., (2015), Precision-Recall-Gain Curves : PR Analysis Done Right, In C. CORTES, N. LAWRENCE, D. LEE, M. SUGIYAMA & R. GARNETT (Éd.), *Advances in Neural Information Processing Systems* (T. 28), Curran Associates, Inc.
- GÅRDMARK, A., CASINI, M., HUSS, M., van LEEUWEN, A., HJELM, J., PERSSON, L., & de ROOS, A. M., (2015), Regime shifts in exploited marine food webs : detecting mechanisms underlying alternative stable states using size-structured community dynamics theory, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130262.
- GAUL, W., SADYKOVA, D., WHITE, H. J., LEÓN-SÁNCHEZ, L., CAPLAT, P., EMMERSON, M. C., & YEARSLEY, J. M., (2022), Modelling the distribution of rare invertebrates by correcting class imbalance and spatial bias, *Diversity and Distributions*, 2810, 2171-2186.

- GONZÁLEZ-ORTIZ, V., EGEA, L. G., JIMÉNEZ-RAMOS, R., MORENO-MARÍN, F., PÉREZ-LLORÉNS, J. L., BOUMA, T., & BRUN, F., (2016), Submerged vegetation complexity modifies benthic infauna communities : the hidden role of the belowground system, *Marine Ecology*, 373, 543-552.
- GRAHAM, J. E., & STUART-SMITH, R. D., (2009), Ecological effects of marine protected areas on rocky reef communities—a continental-scale analysis, *Marine Ecology Progress Series*, 388, 51-62.
- HALPERN, B. S., FRAZIER, M., AFFLERBACH, J., LOWNDES, J. S., MICHELI, F., O'HARA, C., SCARBOROUGH, C., & SELKOE, K. A., (2019), Recent pace of change in human impact on the world's ocean, *Scientific Reports*, 91, 11609.
- HASTINGS, A., & WYSHAM, D. B., (2010), Regime shifts in ecological systems can occur with no warning, *Ecology Letters*, 134, 464-472.
- HE, H., & GARCIA, E. A., (2009), Learning from Imbalanced Data, *IEEE Transactions on Knowledge and Data Engineering*, 219, 1263-1284.
- HIGUCHI, T., AGOSTINI, S., CASARETO, B. E., SUZUKI, Y., & YUYAMA, I., (2015), The northern limit of corals of the genus *Acropora* in temperate zones is determined by their resilience to cold bleaching, *Scientific Reports*, 51, 18467.
- HILLEBRAND, H., DONOHUE, I., HARPOLE, W. S., HODAPP, D., KUCERA, M., LEWANDOWSKA, A. M., MERDER, J., MONTOYA, J. M., & FREUND, J. A., (2020), Thresholds for ecological responses to global change do not emerge from empirical data, *Nature Ecology & Evolution*, 411, 1502-1509.
- JÄNES, H., CARNELL, P., YOUNG, M., IERODIACONOU, D., JENKINS, G. P., HAMER, P., ZU ERMGASSEN, P. S., GAIR, J. R., & MACREADIE, P. I., (2021), Seagrass valuation from fish abundance, biomass and recreational catch, *Ecological Indicators*, 130, 108097.
- JOUFFRAY, J.-B., NYSTRÖM, M., NORSTRÖM, A. V., WILLIAMS, I. D., WEDDING, L. M., KITTINGER, J. N., & WILLIAMS, G. J., (2015), Identifying multiple coral reef regimes and their drivers across the Hawaiian archipelago, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130268.
- JOUFFRAY, J.-B., WEDDING, L. M., NORSTRÖM, A. V., DONOVAN, M. K., WILLIAMS, G. J., CROWDER, L. B., ERICKSON, A. L., FRIEDLANDER, A. M., GRAHAM, N. A. J., GOVE, J. M., KAPPEL, C. V., KITTINGER, J. N., LECKY, J., OLESON, K. L. L., SELKOE, K. A., WHITE, C., WILLIAMS, I. D., & NYSTRÖM, M., (2019), Parsing human and biophysical drivers of coral reef regimes, *Proceedings of the Royal Society B : Biological Sciences*, 2861896, 20182544.
- KÉFI, S., GUTTAL, V., BROCK, W. A., CARPENTER, S. R., ELLISON, A. M., LIVINA, V. N., SEEKELL, D. A., SCHEFFER, M., van NES, E. H., & DAKOS, V., (2014), Early Warning Signals of Ecological Transitions : Methods for Spatial Patterns, *PLOS ONE*, 93, 1-13.
- KELLY, R. P., ERICKSON, A. L., MEASE, L. A., BATTISTA, W., KITTINGER, J. N., & FUJITA, R., (2015), Embracing thresholds for better environmental management, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130276.
- KOTTA, J., WERNBERG, T., JÄNES, H., KOTTA, I., NURKSE, K., PÄRNOJA, M., & ORAV-KOTTA, H., (2018), Novel crab predator causes marine ecosystem regime shift, *Scientific Reports*, 81, 4956.
- 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*, 459, 1994-2002.
- LING, S. D., SCHEIBLING, R. E., RASSWEILER, A., JOHNSON, C. R., SHEARS, N., CONNELL, S. D., SALOMON, A. K., NORDERHAUG, K. M., PÉREZ-MATUS, A., HERNÁNDEZ, J. C., CLEMENTE, S., BLAMEY, L. K., HEREU, B., BALLESTEROS, E., SALA, E., GARRABOU, J., CEBRIAN, E., ZABALA, M., FUJITA, D., & JOHNSON, L. E., (2015), Global regime shift dynamics of catastrophic sea urchin overgrazing, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130269.
- LITZOW, M. A., HOBDAV, A. J., FRUSHER, S. D., DANN, P., & TUCK, G. N., (2016), Detecting regime shifts in marine systems with limited biological data : An example from southeast Australia, *Progress in Oceanography*, 141, 96-108.
- LITZOW, M. A., & HUNSICKER, M. E., (2016), Early warning signals, nonlinearity, and signs of hysteresis in real ecosystems, *Ecosphere*, 712, e01614.
- LITZOW, M. A., & MUETER, F. J., (2014), Assessing the ecological importance of climate regime shifts : An approach from the North Pacific Ocean, *Progress in Oceanography*, 120, 110-119.
- LUNDBERG, S. M., & LEE, S.-I., (2017), A Unified Approach to Interpreting Model Predictions, In I. GUYON, U. V. LUXBURG, S. BENGIO, H. WALLACH, R. FERGUS, S. VISHWANATHAN & R. GARNETT (Éd.), *Advances in Neural Information Processing Systems 30* (p. 4765-4774), Curran Associates, Inc.
- MARZLOFF, M. P., JOHNSON, C. R., LITTLE, L. R., SOULIÉ, J.-C., LING, S. D., & FRUSHER, S. D., (2013), Sensitivity analysis and pattern-oriented validation of TRITON, a model with alternative community states : Insights on temperate rocky reefs dynamics, *Ecological Modelling*, 258, 16-32.
- MARZLOFF, M. P., LITTLE, L. R., & JOHNSON, C. R., (2016), Building Resilience Against Climate-Driven Shifts in a Temperate Reef System : Staying Away from Context-Dependent Ecological Thresholds, *Ecosystems*, 191, 1-15.
- MAXWELL, P. S., EKLÖF, J. S., van KATWIJK, M. M., O'BRIEN, K. R., de la TORRE-CASTRO, M., BOSTRÖM, C., BOUMA, T. J., KRAUSE-JENSEN, D., UNSWORTH, R. K. F., van TUSSENBROEK, B. I., & van der HEIDE, T., (2017), The fundamental role of ecological feedback mechanisms for the adaptive management of seagrass ecosystems – a review, *Biological Reviews*, 923, 1521-1538.
- MCINNIS, L., HEALY, J., & ASTELS, S., (2017), hdbscan : Hierarchical density based clustering, *The Journal of Open Source Software*, 211.
- MCINNIS, L., HEALY, J., & MELVILLE, J., (2020), UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction, *arXiv*.
- MCMANUS, J. W., & POLSENBERG, J. F., (2004), Coral–algal phase shifts on coral reefs : Ecological and environmental aspects, *Progress in Oceanography*, 602, 263-279.
- MEYNARD, C. N., PIOUS, C., & KAPLAN, D. M., (2023), A theoretical framework for upscaling species distribution models, *Methods in Ecology and Evolution*.
- MOLNAR, C., (2022), *Interpretable machine learning : Explaining the decisions and behaviour of machine learning models* (2nd).
- MOUDRÝ, V., KEIL, P., GÁBOR, L., LECOURS, V., ZARZO-ARIAS, A., BARTÁK, V., MALAVASI, M., ROCCHINI, D., TORRESANI, M., GDULOVÁ, K., GRATTAROLA, F., LEROY, F., MARCHETTO,

- E., THOUVERAI, E., PROŠEK, J., WILD, J., & ŠÍMOVÁ, P., (2023), Scale mismatches between predictor and response variables in species distribution modelling : A review of practices for appropriate grain selection, *Progress in Physical Geography : Earth and Environment*, 473, 467-482.
- MOULAVI, D., JASKOWIAK, P. A., CAMPELLO, R. J. G. B., ZIMEK, A., & SANDER, J., (2014), Density-Based Clustering Validation, *Proceedings of the 2014 SIAM International Conference on Data Mining (SDM)*, 839-847.
- MUSTONEN, T., & SHADRIN, V. I., (2021), The River Alazeya : Shifting Socio-Ecological Systems Connected to a Northeastern Siberian River, *Arctic*, 74, 67-86.
- NAGELKERKEN, I., ROBERTS, C. M., van der VELDE, G., DORENBOSCH, M., van RIEL, M. C., de la MORINÈRE, E. C., & NIENHUIS, P. H., (2002), How important are mangroves and seagrass beds for coral-reef fish? The nursery hypothesis tested on an island scale, *Marine Ecology Progress Series*, 244, 299-305.
- NIJP, J. J., TEMME, A. J., van VOORN, G. A., KOOISTRA, L., HENGEVELD, G. M., SOONS, M. B., TEULING, A. J., & WALLINGA, J., (2019), Spatial early warning signals for impending regime shifts : A practical framework for application in real-world landscapes, *Global Change Biology*, 256, 1905-1921.
- NYSTRÖM, M., NORSTRÖM, A. V., BLECKNER, T., de la TORRE-CASTRO, M., EKLÖF, J. S., FOLKE, C., ÖSTERBLUM, H., STENECK, R. S., THYRESSON, M., & TROELL, M., (2012), Confronting Feedbacks of Degraded Marine Ecosystems, *Ecosystems*, 155, 695-710.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., & DUCHESNAY, E., (2011), Scikit-learn : Machine Learning in Python, *Journal of Machine Learning Research*, 12, 2825-2830.
- POTTER, K. A., ARTHUR WOODS, H., & PINCEBOURDE, S., (2013), Microclimatic challenges in global change biology, *Global Change Biology*, 1910, 2932-2939.
- QUINTANO, E., DÍEZ, I., MUGUERZA, N., SANTOLARIA, A., & GOROSTIAGA, J. M., (2015), Epiphytic flora on *Gelidium corneum* (Rhodophyta : Gelidiales) in relation to wave exposure and depth, *Scientia Marina*, 794, 479-486.
- RICART, A. M., WARD, M., HILL, T. M., SANFORD, E., KROEKER, K. J., TAKESHITA, Y., MEROLLA, S., SHUKLA, P., NINOKAWA, A. T., ELSMORE, K., & GAYLORD, B., (2021), Coast-wide evidence of low pH amelioration by seagrass ecosystems, *Global Change Biology*, 2711, 2580-2591.
- ROBINSON, O. J., RUIZ-GUTIERREZ, V., & FINK, D., (2018), Correcting for bias in distribution modelling for rare species using citizen science data, *Diversity and Distributions*, 244, 460-472.
- ROBINSON, O. J., RUIZ-GUTIERREZ, V., REYNOLDS, M. D., GOLET, G. H., STRIMAS-MACKEY, M., & FINK, D., (2020), Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models, *Diversity and Distributions*, 268, 976-986.
- ROCHA, J., YLETYINEN, J., BIGGS, R., BLECKNER, T., & PETERSON, G., (2015a), Marine regime shifts : drivers and impacts on ecosystems services, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3701659, 20130273.

- ROCHA, J. C., PETERSON, G. D., & BIGGS, R., (2015b), Regime Shifts in the Anthropocene : Drivers, Risks, and Resilience, *PLOS ONE*, 108, 1-16.
- SAMMARCO, P. W., COLL, J. C., LA BARRE, S., & WILLIS, B., (1983), Competitive strategies of soft corals (Coelenterata : Octocorallia) : Allelopathic effects on selected scleractinian corals, *Coral Reefs*, 13, 173-178.
- SANDIN, S. A., ALCANTAR, E., CLARK, R., de LEÓN, R., DILROSUN, F., EDWARDS, C. B., ESTEP, A. J., EYNAUD, Y., FRENCH, B. J., FOX, M. D., GREY, D., HAMILTON, S. L., KRAMP, H., MARHAVER, K. L., MILLER, S. D., ROACH, T. N. F., SEFERINA, G., SILVEIRA, C. B., SMITH, J. E., ... VERMEIJ, M. J. A., (2022), Benthic assemblages are more predictable than fish assemblages at an island scale, *Coral Reefs*, 414, 1031-1043.
- SCHEFFER, M., (2009), *Critical Transitions in Nature and Society*, Princeton University Press.
- SCHEFFER, M., BASCOMPTE, J., BROCK, W. A., BROVKIN, V., CARPENTER, S. R., DAKOS, V., HELD, H., van NES, E. H., RIETKERK, M., & SUGIHARA, G., (2009), Early-warning signals for critical transitions, *Nature*, 4617260, 53-59.
- SCHEFFER, M., & CARPENTER, S. R., (2003), Catastrophic regime shifts in ecosystems : linking theory to observation, *Trends in Ecology & Evolution*, 1812, 648-656.
- SCHEFFER, M., CARPENTER, S., FOLEY, J. A., FOLKE, C., & WALKER, B., (2001), Catastrophic shifts in ecosystems, *Nature*, 4136856, 591-596.
- SCHEFFER, M., HIROTA, M., HOLMGREN, M., NES, E. H. V., & CHAPIN, F. S., (2012), Thresholds for boreal biome transitions, *Proceedings of the National Academy of Sciences*, 10952, 21384-21389.
- SHABANI, F., KUMAR, L., & AHMADI, M., (2016), A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area, *Ecology and Evolution*, 616, 5973-5986.
- SHAPLEY, L. S., (1953), A Value for n-Person Games, In H. W. KUHN & A. W. TUCKER (Éd.), Princeton University Press.
- SISSINI, M. N., KOERICH, G., de BARROS-BARRETO, M. B., COUTINHO, L. M., GOMES, F. P., OLIVEIRA, W., COSTA, I. O., de CASTRO NUNES, J., HENRIQUES, M. C., VIEIRA-PINTO, T., TORRANO-SILVA, B. N., OLIVEIRA, M. C., LE GALL, L., & HORTA, P. A., (2022), Diversity, distribution, and environmental drivers of coralline red algae : the major reef builders in the Southwestern Atlantic, *Coral Reefs*, 413, 711-725.
- SMITH, H. A., PRENZLAU, T., WHITMAN, T., FULTON, S. E., BORGHI, S., LOGAN, M., HERON, S. F., & BOURNE, D. G., (2022), Macroalgal canopies provide corals limited protection from bleaching and impede post-bleaching recovery, *Journal of Experimental Marine Biology and Ecology*, 553, 151762.
- SOMODI, I., LEPESI, N., & BOTTA-DUKÁT, Z., (2017), Prevalence dependence in model goodness measures with special emphasis on true skill statistics, *Ecology and Evolution*, 73, 863-872.
- STAVER, A. C., ARCHIBALD, S., & LEVIN, S., (2011), Tree cover in sub-Saharan Africa : Rainfall and fire constrain forest and savanna as alternative stable states, *Ecology*, 925, 1063-1072.
- SUNDAY, J. M., FABRICIUS, K. E., KROEKER, K. J., ANDERSON, K. M., BROWN, N. E., BARRY, J. P., CONNELL, S. D., DUPONT, S., GAYLORD, B., HALL-SPENCER, J. M., KLINGER, T., MILAZZO, M., MUNDAY, P. L., RUSSELL, B. D., SANFORD, E., THIYAGARAJAN, V., VAUGHAN, M. L. H.,

- WIDDICOMBE, S., & HARLEY, C. D. G., (2017), Ocean acidification can mediate biodiversity shifts by changing biogenic habitat, *Nature Climate Change*, 71, 81-85.
- TURNER, M. G., CALDER, W. J., CUMMING, G. S., HUGHES, T. P., JENTSCH, A., LADEAU, S. L., LENTON, T. M., SHUMAN, B. N., TURETSKY, M. R., RATAJCZAK, Z., WILLIAMS, J. W., WILLIAMS, A. P., & CARPENTER, S. R., (2020), Climate change, ecosystems and abrupt change : science priorities, *Philosophical Transactions of the Royal Society B : Biological Sciences*, 3751794, 20190105.
- TWIST, B. A., NEILL, K. F., BILEWITCH, J., JEONG, S. Y., SUTHERLAND, J. E., & NELSON, W. A., (2019), High diversity of coralline algae in New Zealand revealed : Knowledge gaps and implications for future research, *PLOS ONE*, 1412, 1-21.
- VALAVI, R., ELITH, J., LAHOZ-MONFORT, J. J., & GUILLERA-ARROITA, G., (2019), blockCV : An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models, *Methods in Ecology and Evolution*, 102, 225-232.
- VERGÉS, A., MCCOSKER, E., MAYER-PINTO, M., COLEMAN, M. A., WERNBERG, T., AINSWORTH, T., & STEINBERG, P. D., (2019), Tropicalisation of temperate reefs : Implications for ecosystem functions and management actions, *Functional Ecology*, 336, 1000-1013.
- VERGÉS, A., TOMAS, F., CEBRIAN, E., BALLESTEROS, E., KIZILKAYA, Z., DENDRINOS, P., KARAMANLIDIS, A. A., SPIEGEL, D., & SALA, E., (2014), Tropical rabbitfish and the deforestation of a warming temperate sea, *Journal of Ecology*, 1026, 1518-1527.
- VOZZO, M. L., DOROPOULOS, C., SILLIMAN, B. R., STEVEN, A., REEVES, S. E., ter HOFSTEDÉ, R., van KONINGSVELD, M., van de KOPPEL, J., MCPHERSON, T., RONAN, M., & SAUNDERS, M. I., (2023), To restore coastal marine areas, we need to work across multiple habitats simultaneously, *Proceedings of the National Academy of Sciences*, 12026, e2300546120.
- WARD, D. F. L., WOTHERSPOON, S., MELBOURNE-THOMAS, J., HAAPKYLÄ, J., & JOHNSON, C. R., (2018), Detecting ecological regime shifts from transect data, *Ecological Monographs*, 884, 694-715.
- WERNBERG, T., BENNETT, S., BABCOCK, R. C., de BETTIGNIES, T., CURE, K., DEPCZYNSKI, M., DUFOIS, F., FROMONT, J., FULTON, C. J., HOVEY, R. K., HARVEY, E. S., HOLMES, T. H., KENDRICK, G. A., RADFORD, B., SANTANA-GARCON, J., SAUNDERS, B. J., SMALE, D. A., THOMSEN, M. S., TUCKETT, C. A., ... WILSON, S., (2016), Climate-driven regime shift of a temperate marine ecosystem, *Science*, 3536295, 169-172.
- WERNBERG, T., S. THOMSEN, M., K. BAUM, J., J. BISHOP, M., F. BRUNO, J., A. COLEMAN, M., FILBEE-DEXTER, K., GAGNON, K., HE, Q., MURDIYARSO, D., ROGERS, K., R. SILLIMAN, B., A. SMALE, D., STARKO, S., & A. VANDERKLIFT, M., (2023), Impacts of Climate Change on Marine Foundation Species, *Annual Review of Marine Science*, 161.

Appendix A - Datasets

Table S1 – *Habitat states identified in the Chapter 2 and their description included in this study.*

Habitat States	Description
Branching coral	Branching coral of the Acroporidae family
Brown algale	Brown foliose algae
Bushy Furoid like	Bushy algae with a morphology similar to that of <i>Fucus spp.</i>
Crustose coralline algae	Crustose (nongeniculate) coralline algae forming a thin calcereous cruste
Crustose coralline algae and turf	Crustose (nongeniculate) coralline algae forming a thin calcereous cruste with turf algae (see below)
Filamentous algae	Long filamentous epiphytic algae growing over large algae
Green calcified algae	Green calcified algae of the <i>Halimeda</i> genus with a green calcified thallus
Large canopy forming algae	Large canopy forming algae as species of genus <i>Laminaria</i> , <i>Macrocystis</i> , <i>Sargassum</i> , <i>Durvillaea</i> , <i>Ecklonia</i> and others
Other Sessile invertebrates	Other sessile invertebrates like mussel beds, oyster reefs
Red algae	Red foliose algae
Seagrass	Marine angiosperms of various families as Cymodoceaceae, Posidoniaceae, Potamogetonaceae, Ruppiaceae, Zannichelliaceae and Zosteraceae
Soft coral and gorgonians	Soft coral and gorgonians
Substrate	Substrate including bare substrate, sand and unconsolidate substrate
Turf	Small algae (2cm < in hight) forming dense and tangled mats or carpets on marine substrates

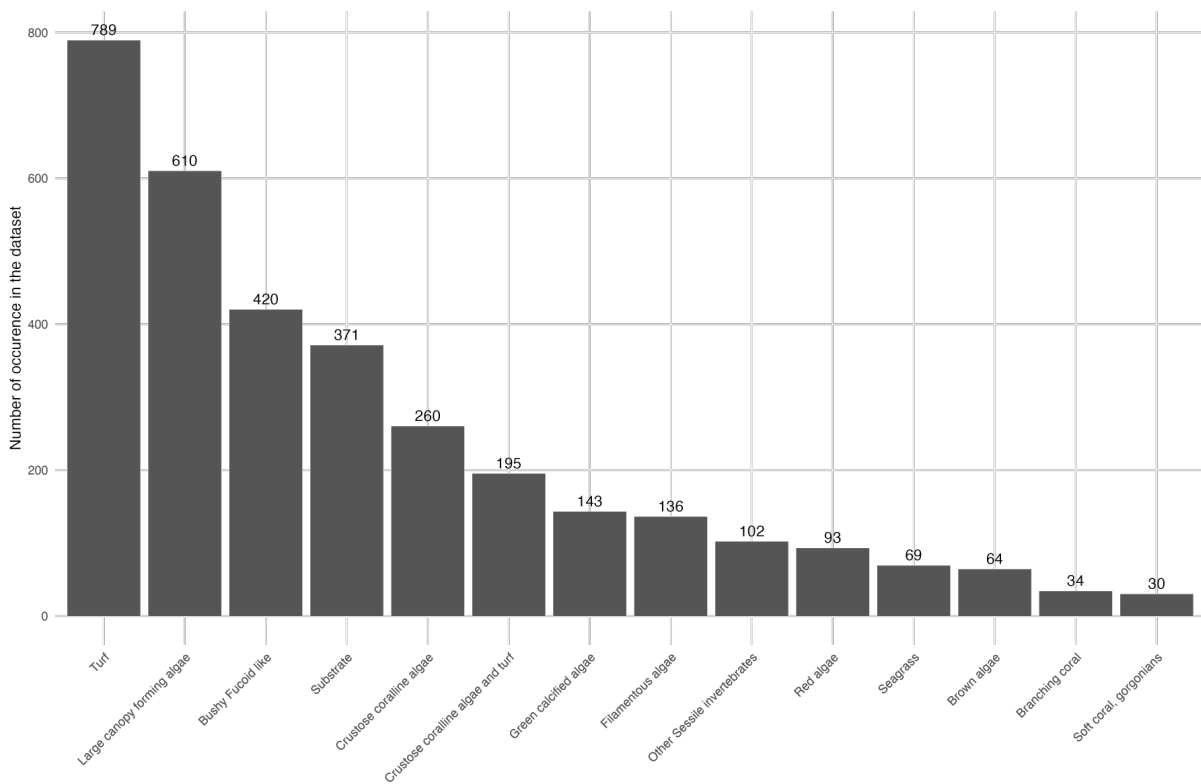


Figure S1 – Occurrence of the different habitat states in the dataset

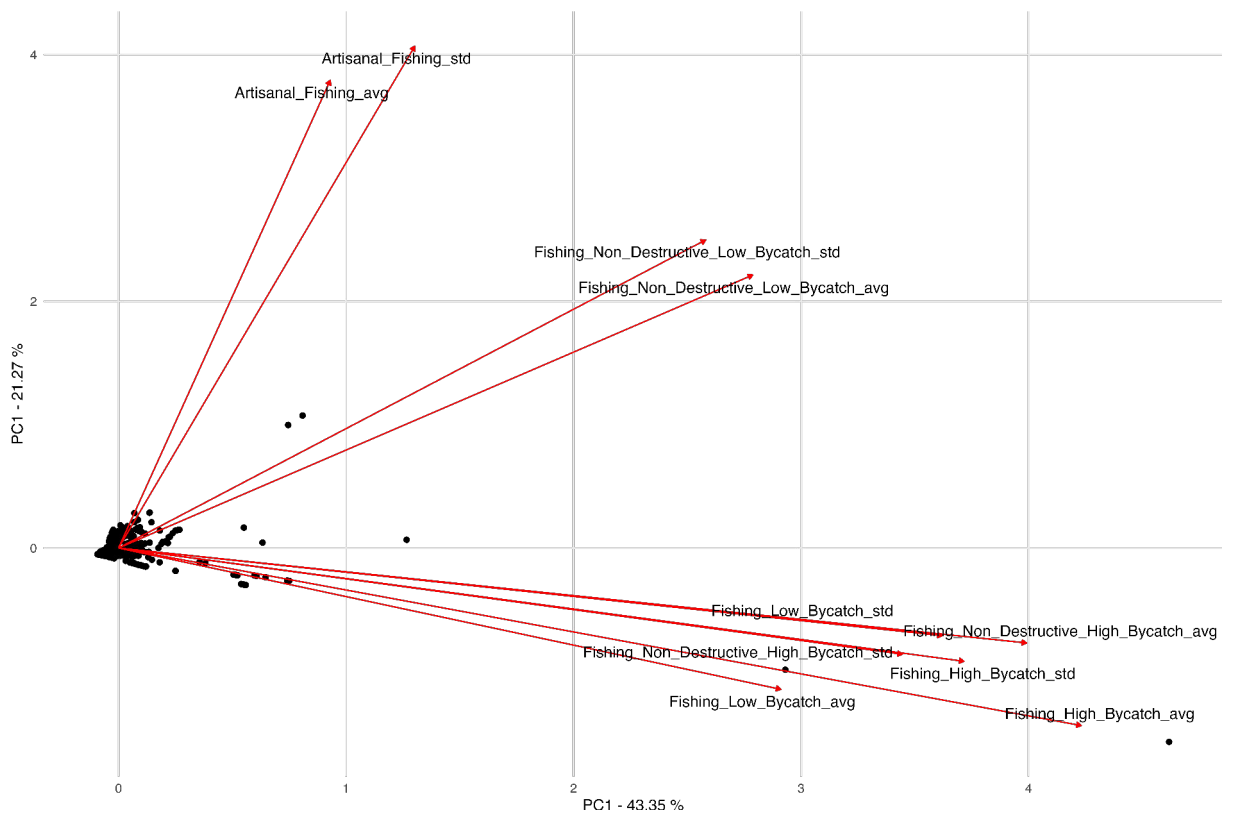


Figure S2 – *PCA biplots of the fishing data.*

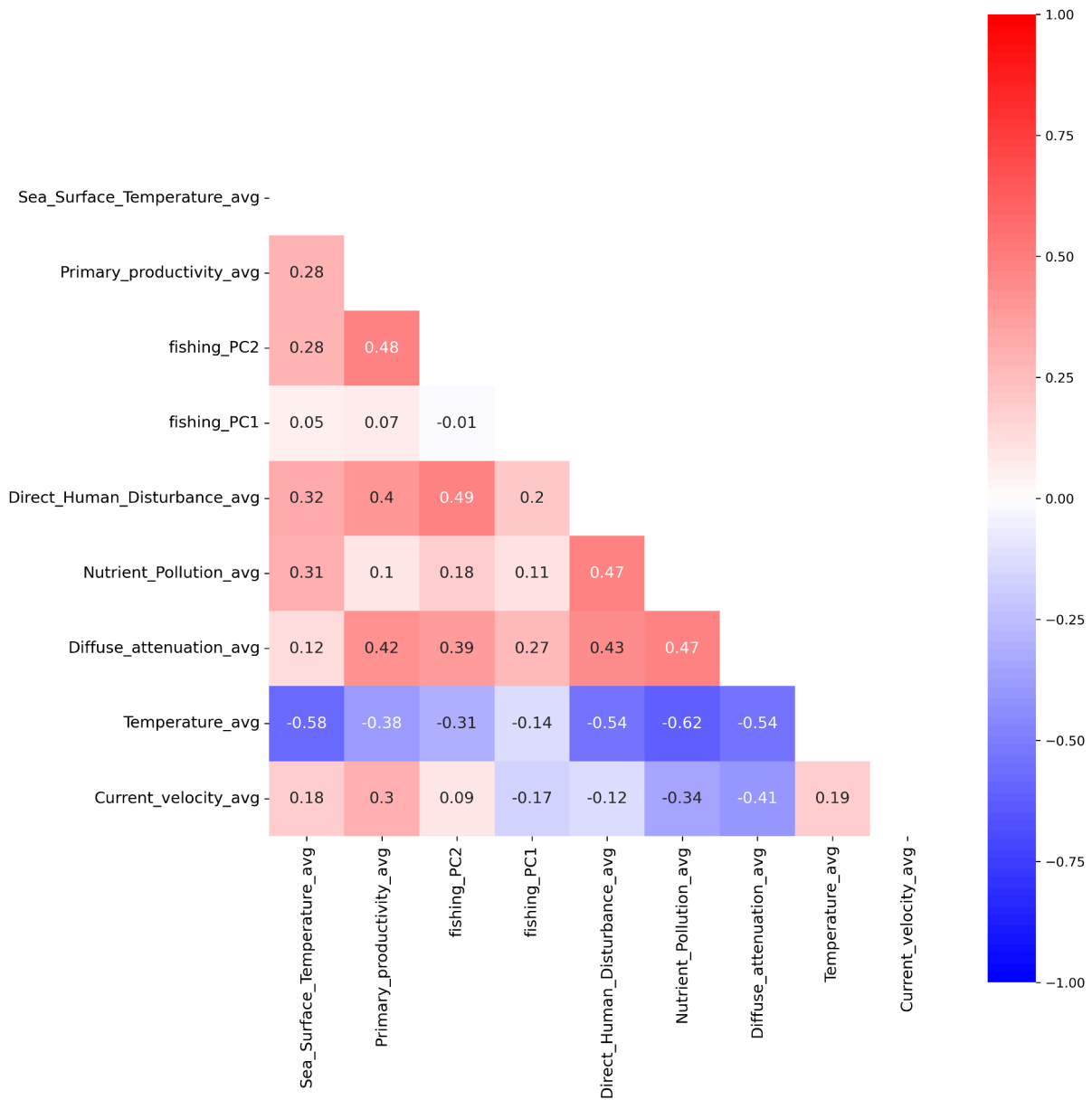


Figure S3 – Pearson's correlation matrix of selected predictors.

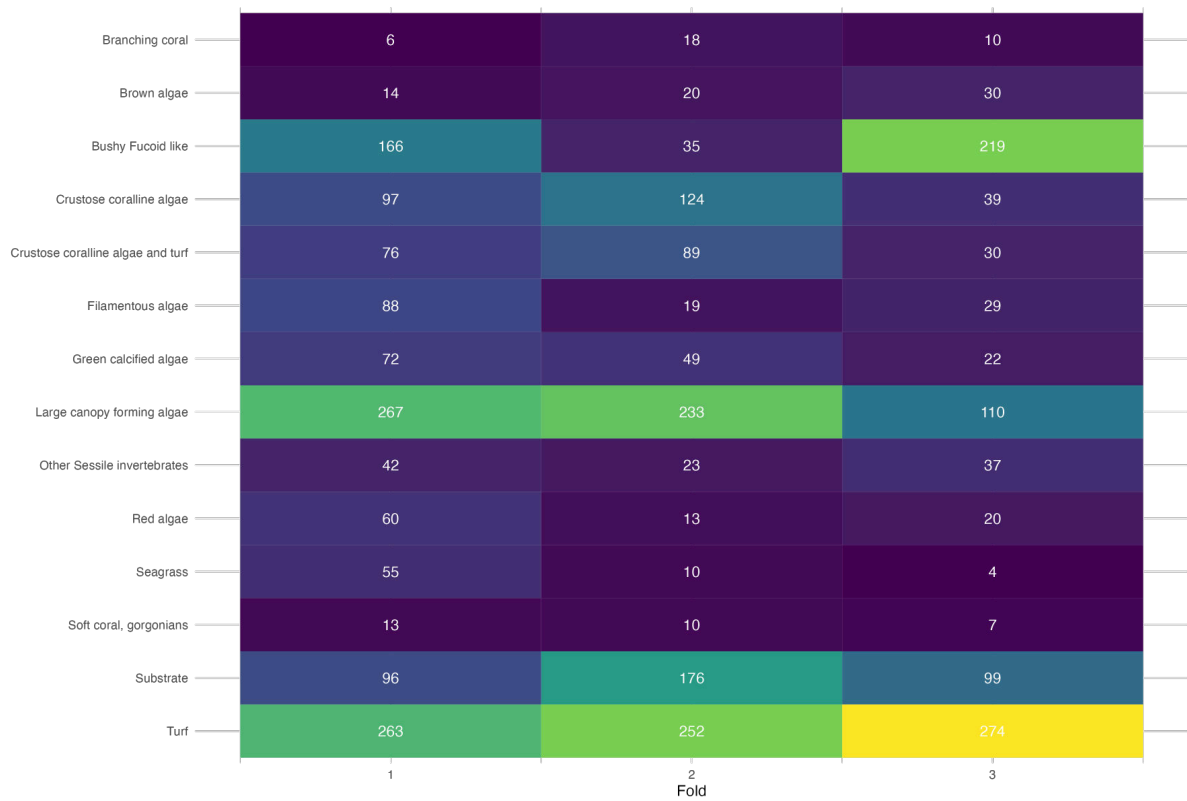


Figure S5 – *Number of occurrences of each habitat state per training and test fold. The colour is based on the proportion represented by each habitat state within the considered fold.*

Appendix B - Model fitting & Evaluation

Model Fit

Hyperparameter tuning

Table S2 – Description of the hyperparameters tuned and their values according to the scikit-learn documentation.

Hyperparameter		
name	Description	Values
<i>n_estimators</i>	The number of trees in the forest.	{100, 200, 300, ..., 3000}
<i>max_features</i>	The number of predictors to consider when looking for the best split.	{2, 3, 4, ..., 8}
<i>max_depth</i>	The maximum depth of each tree. If ∞ , then nodes are expanded until all leaves are pure or contain less than 2 samples	{5, 10, 15, ..., 100, ∞ }

Confusion Matrix

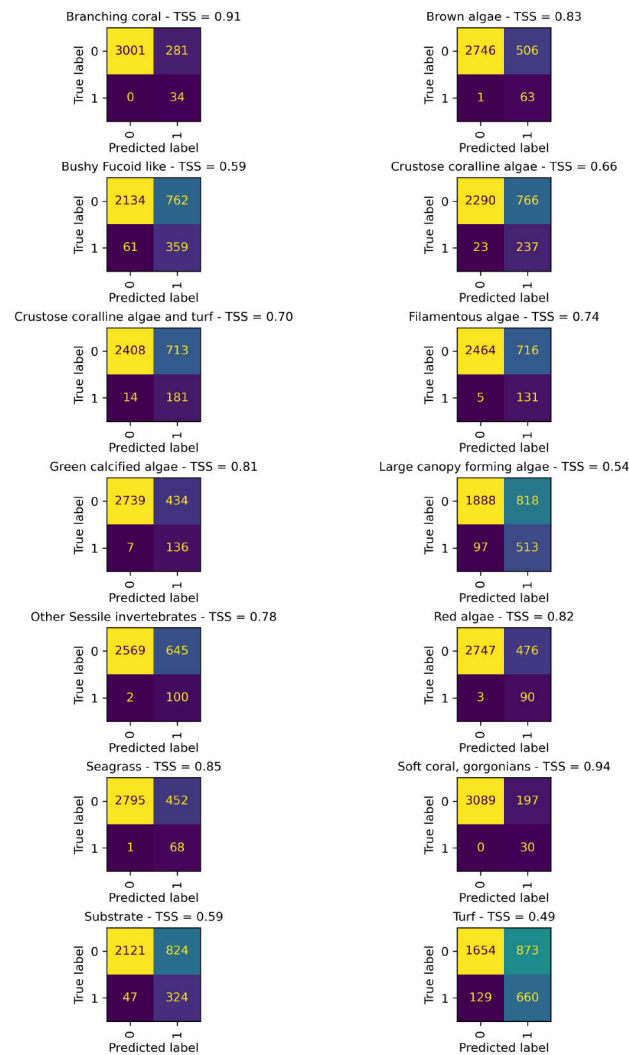


Figure S6 – Confusion Matrix of the explanatory power.

Probability distribution of habitat states

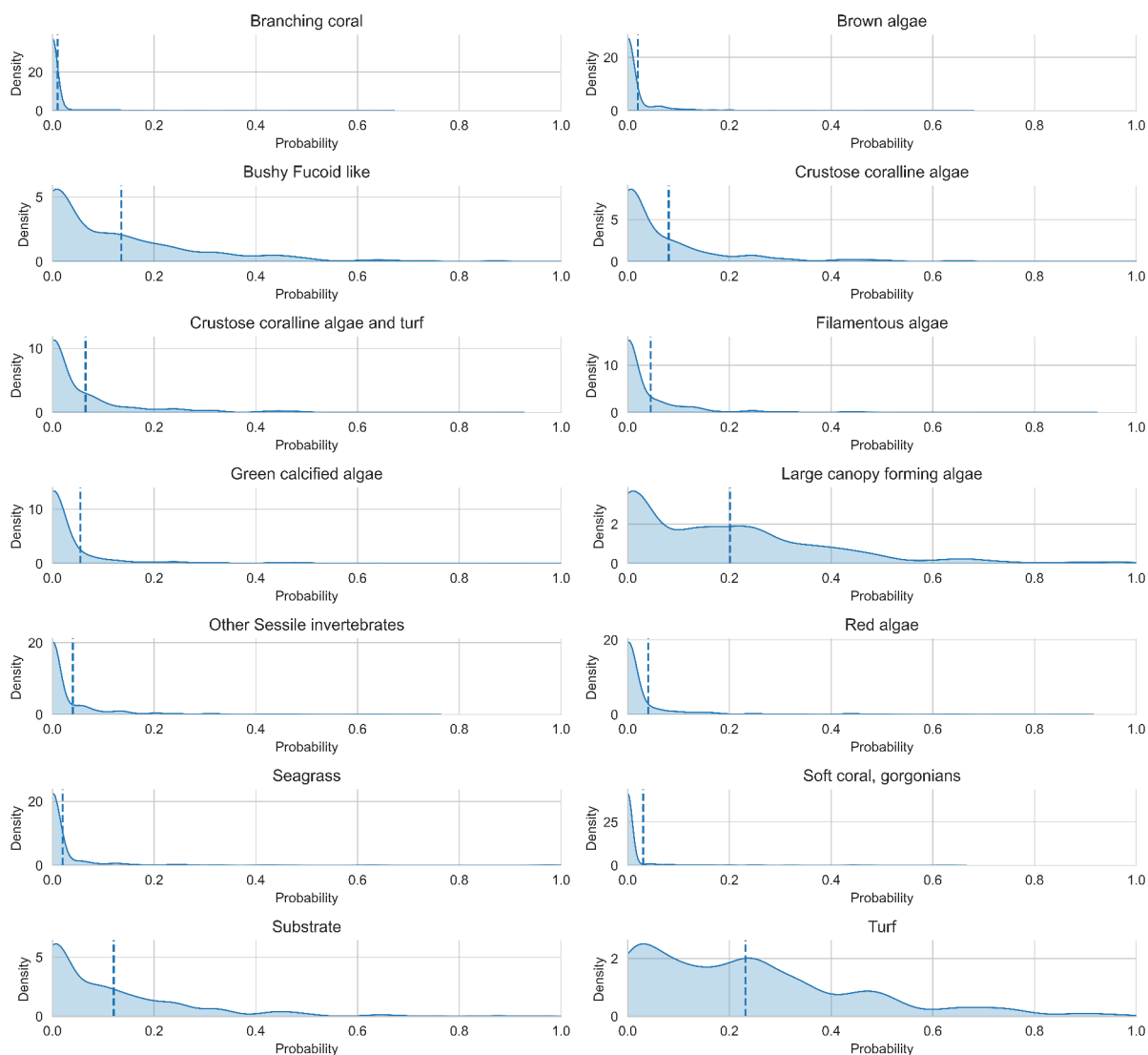


Figure S7 – *Distribution of occurrence probabilities for each habitat state. The dotted line represents the threshold optimised by MaxTSS to binarise the predictions.*

AUC Curves

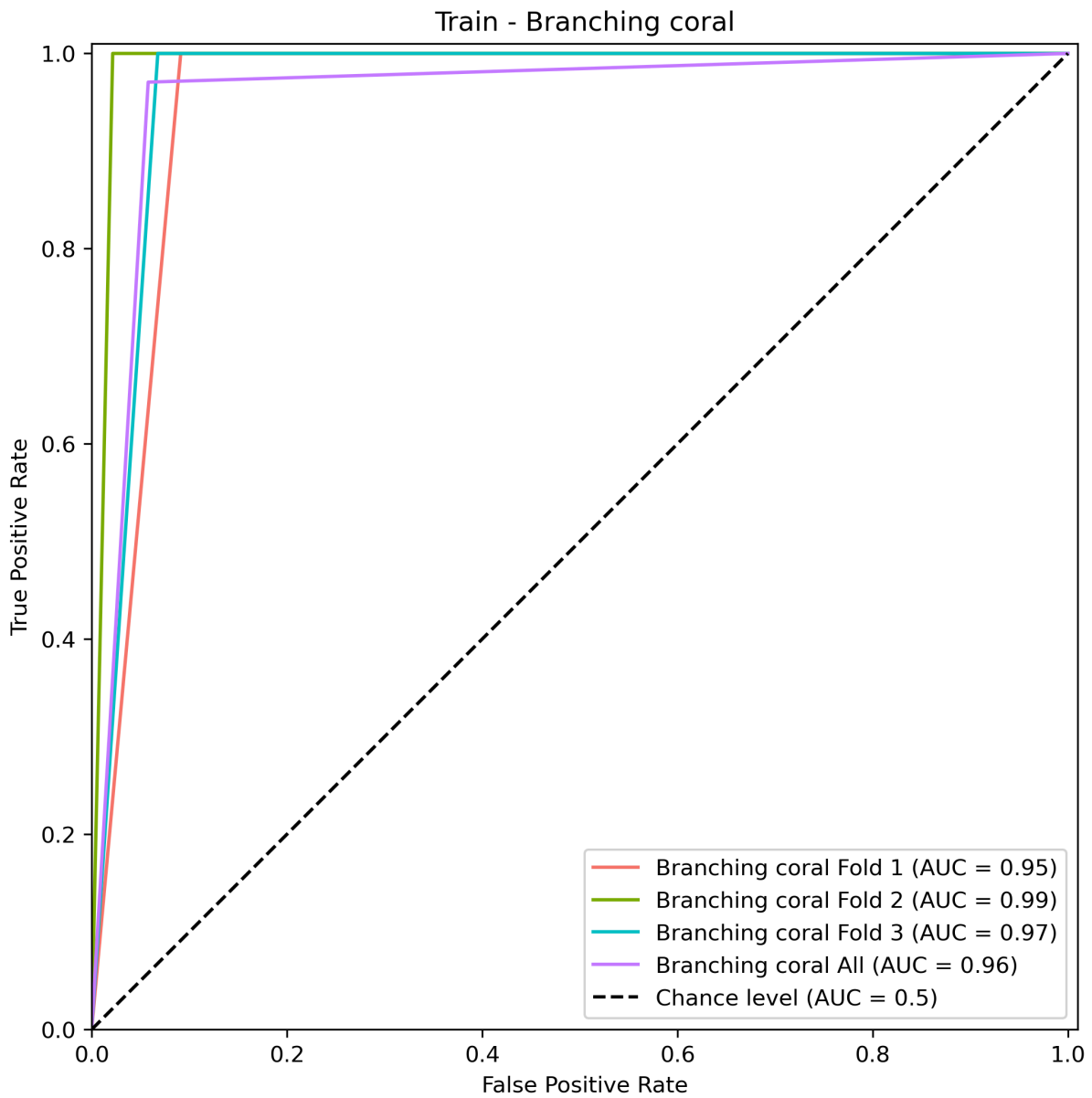


Figure S8 – AUC curves of the explanatory power for the Branching coral.

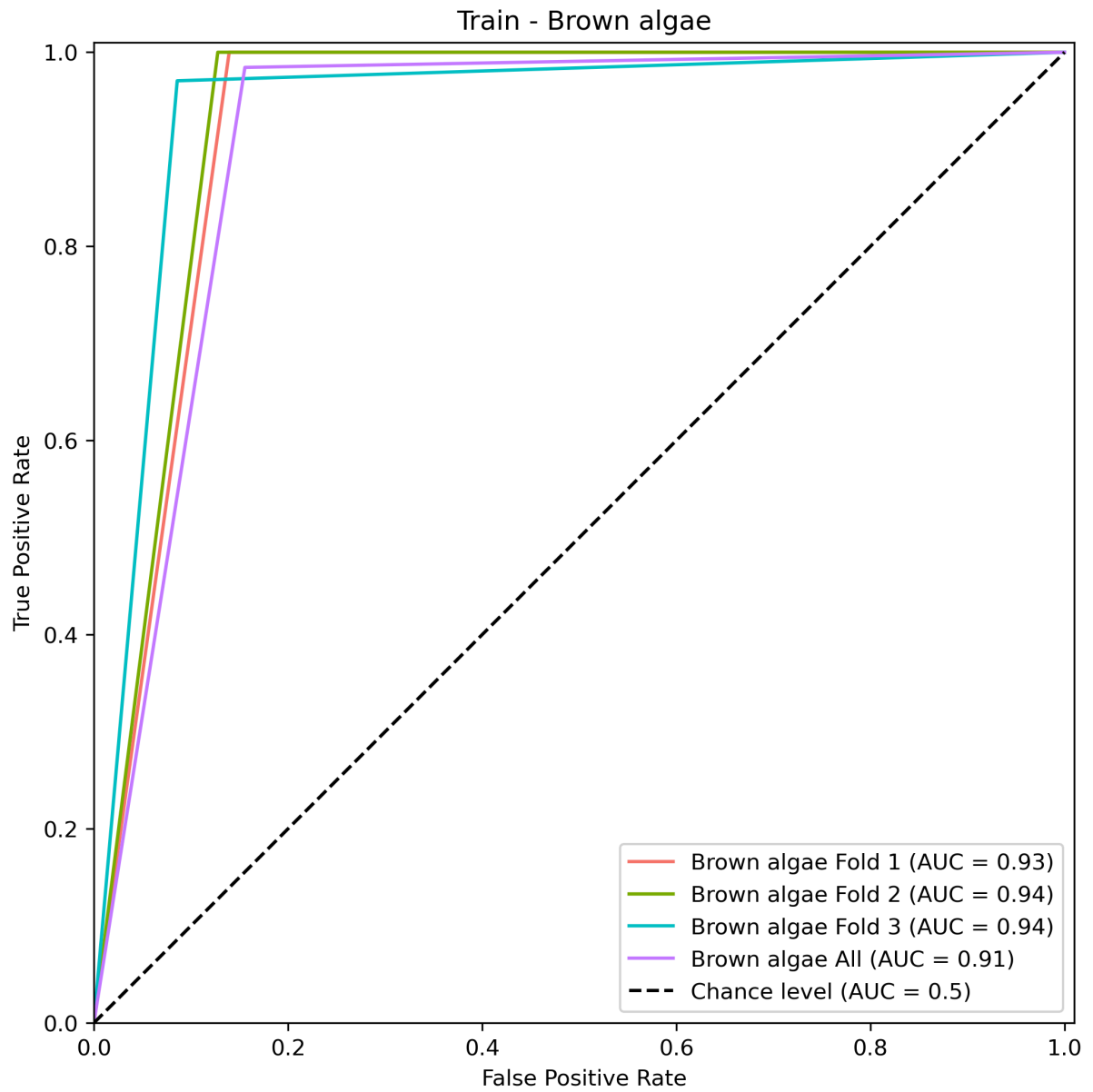


Figure S9 – *AUC curves of the explanatory power for the Brown algae.*

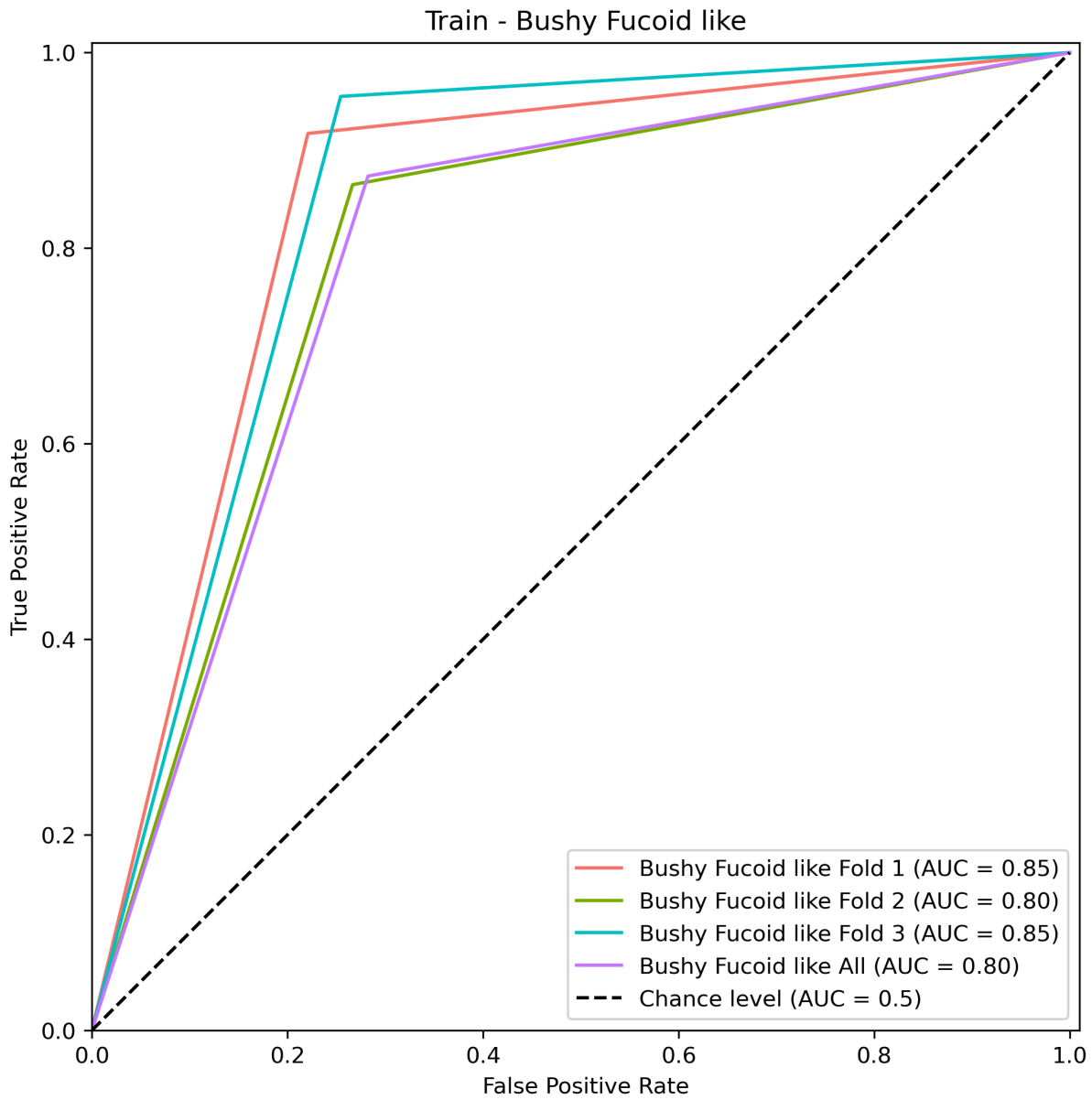


Figure S10 – AUC curves of the explanatory power for the Bushy Furoid like.

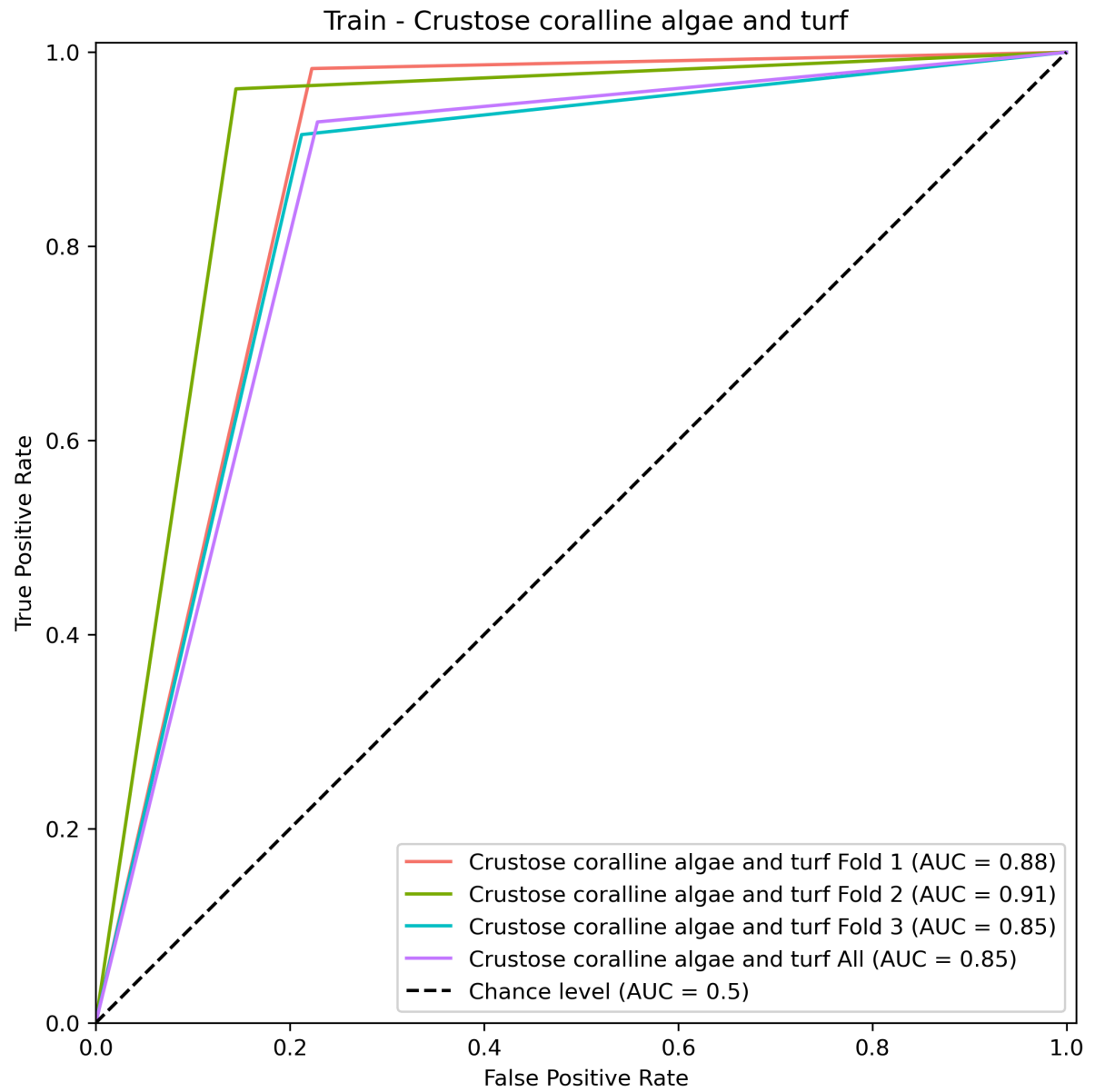


Figure S11 – *AUC curves of the explanatory power for the Crustose coralline algae and turf.*

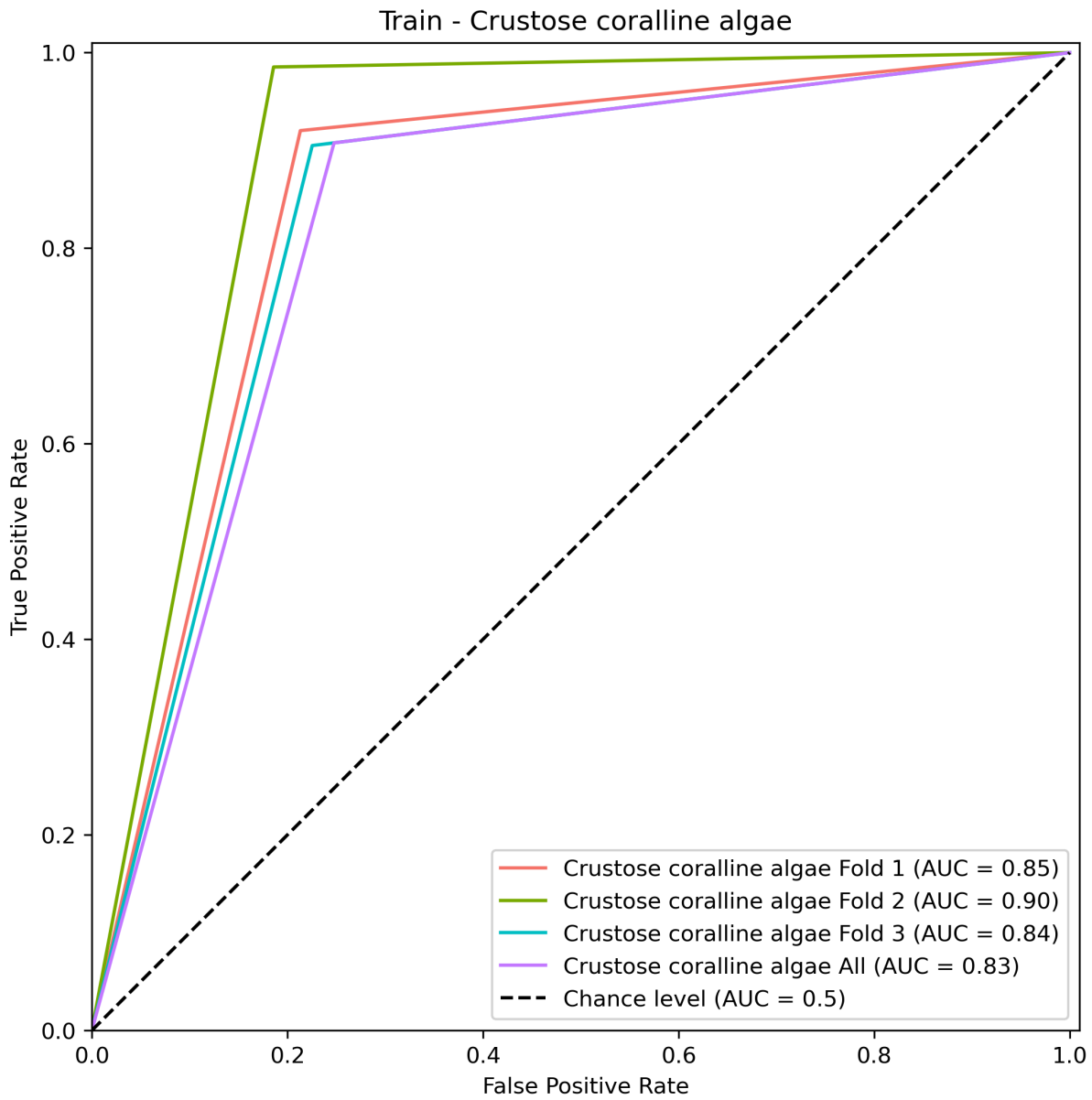


Figure S12 – AUC curves of the explanatory power for the Crustose coralline algae.

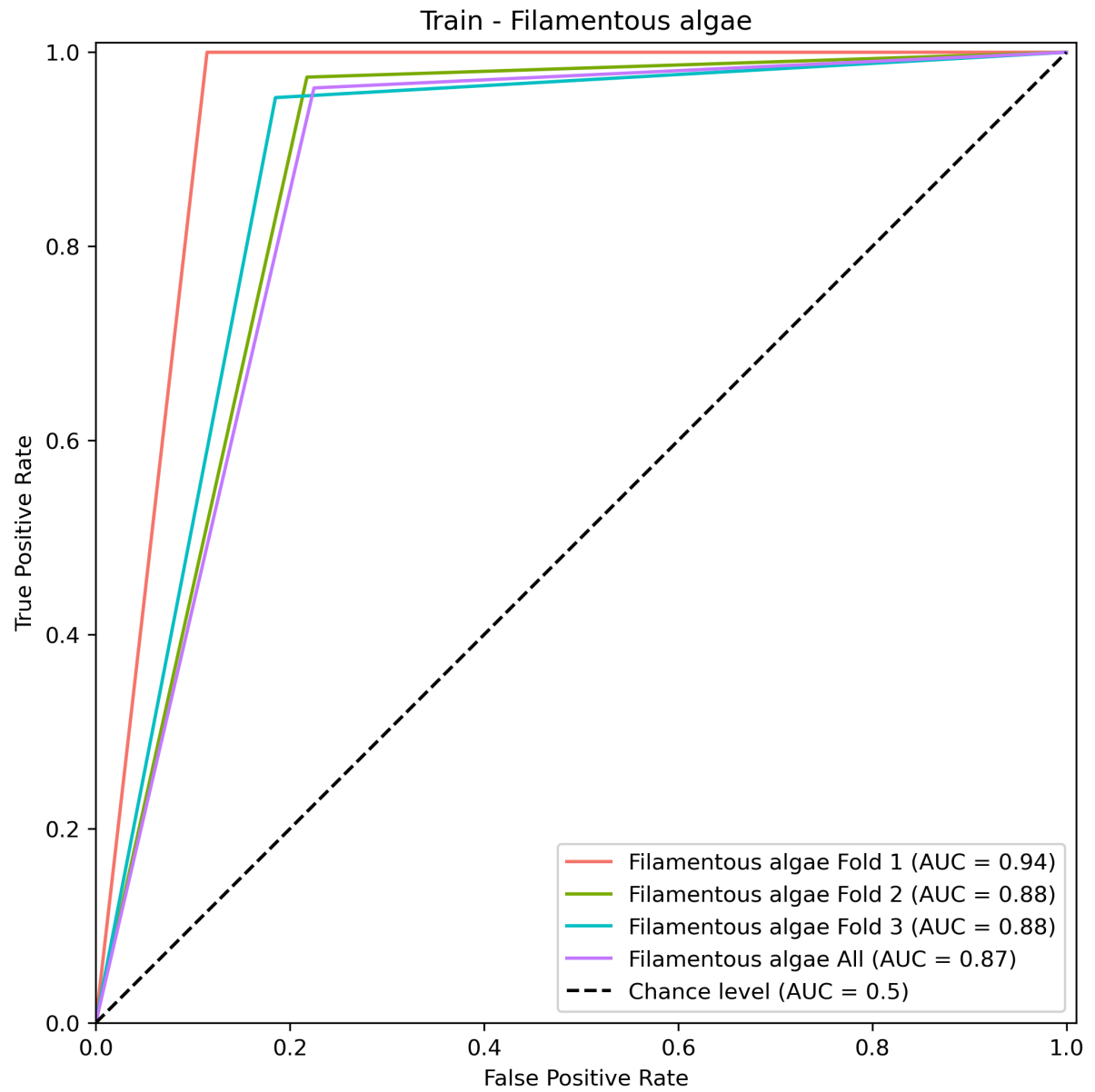


Figure S13 – *AUC curves of the explanatory power for the Filamentous algae.*

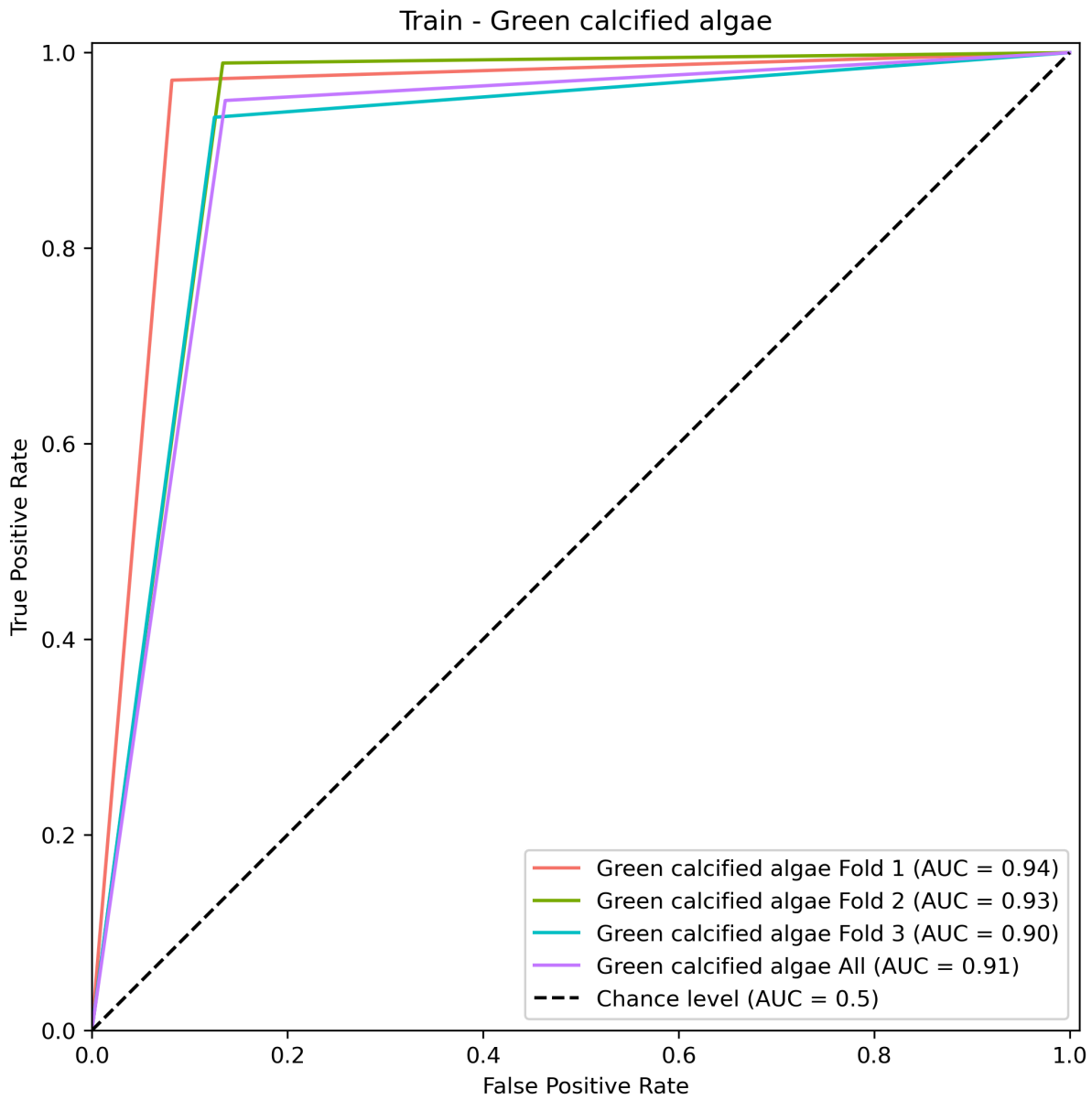


Figure S14 – AUC curves of the explanatory power for the Green Calcified algae.

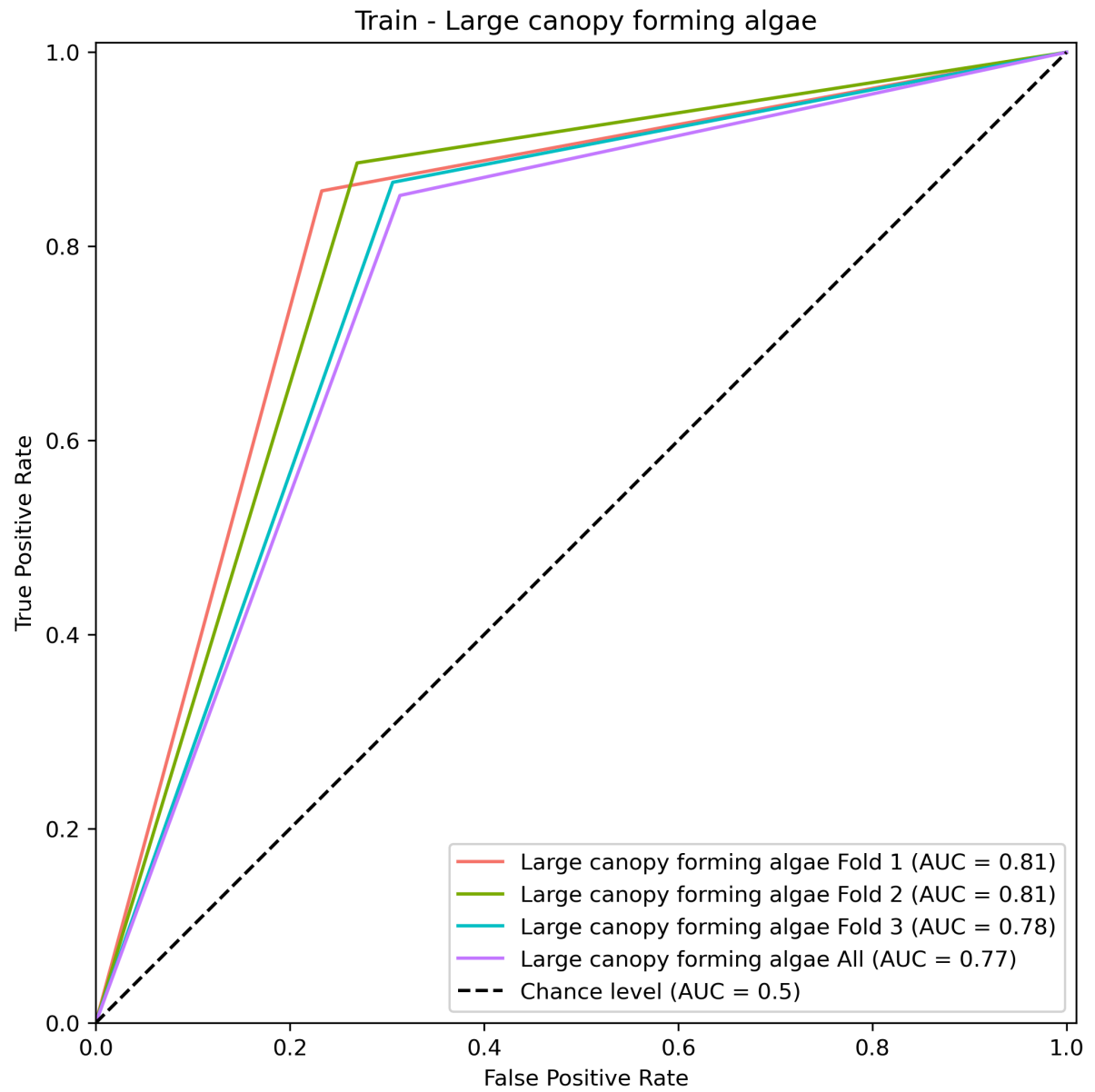


Figure S15 – *AUC curves of the explanatory power for the Large canopy forming algae.*

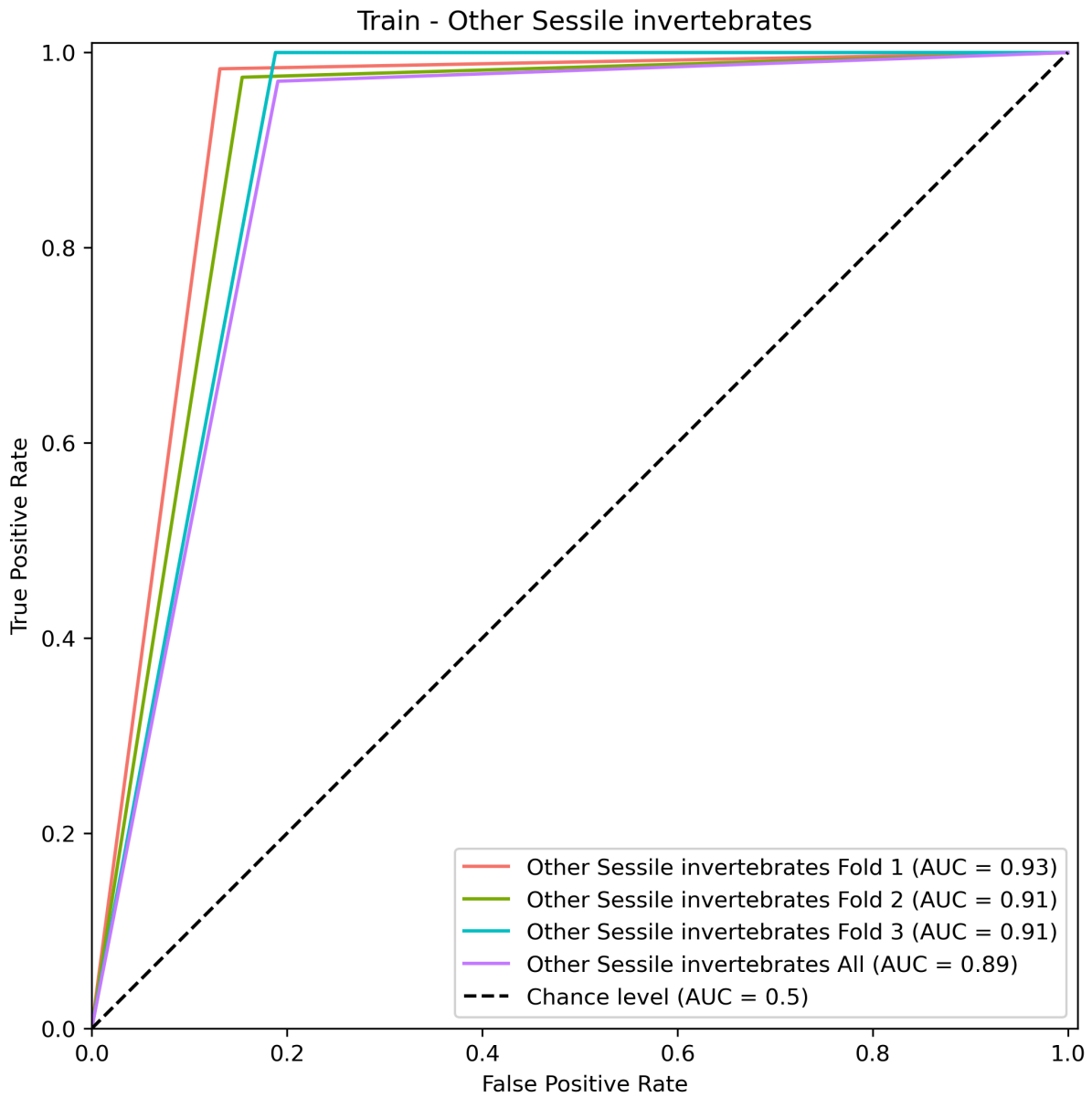


Figure S16 – AUC curves of the explanatory power for the *Other Sessile invertebrates*.

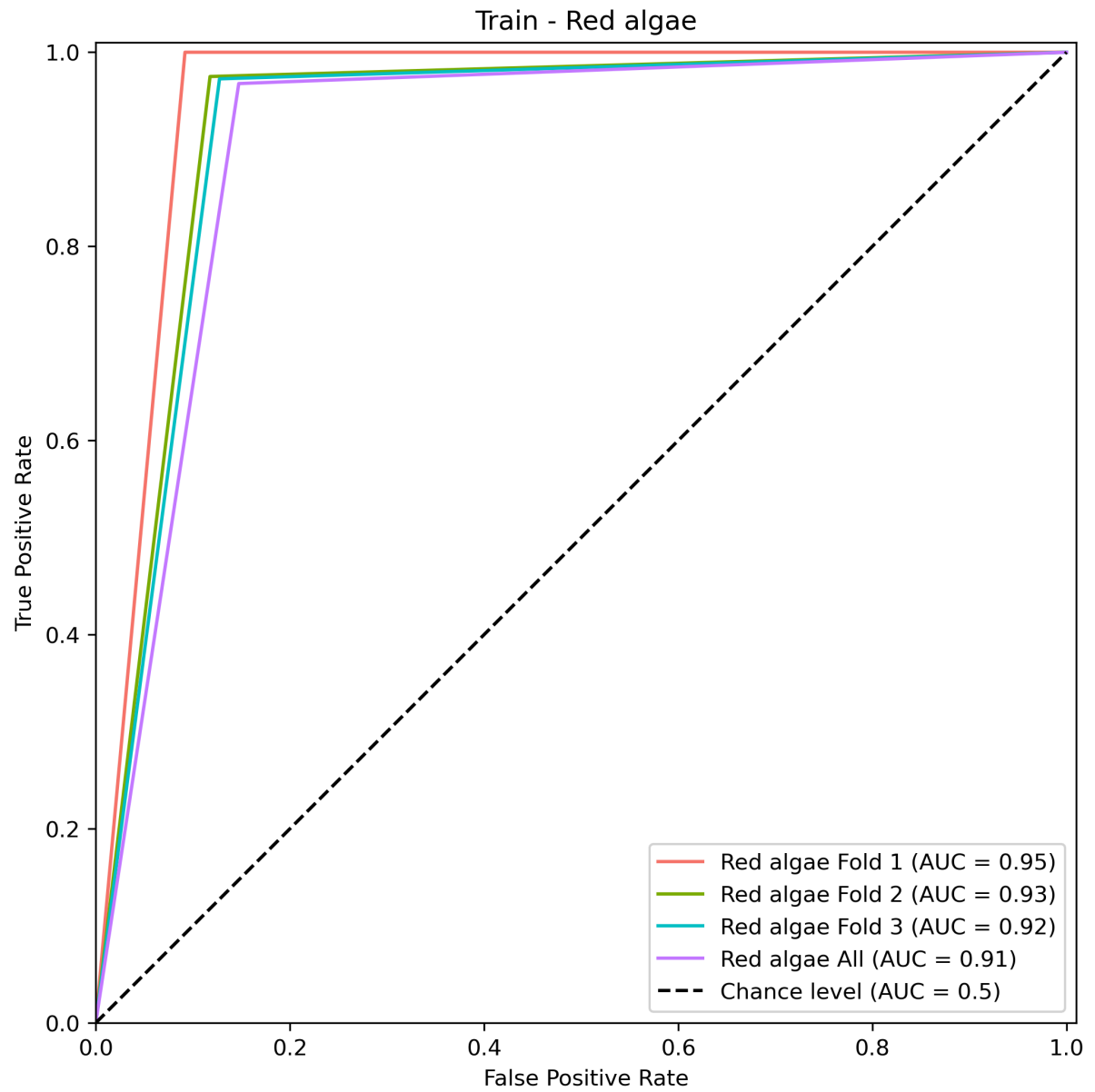


Figure S17 – *AUC curves of the explanatory power for the Red algae.*

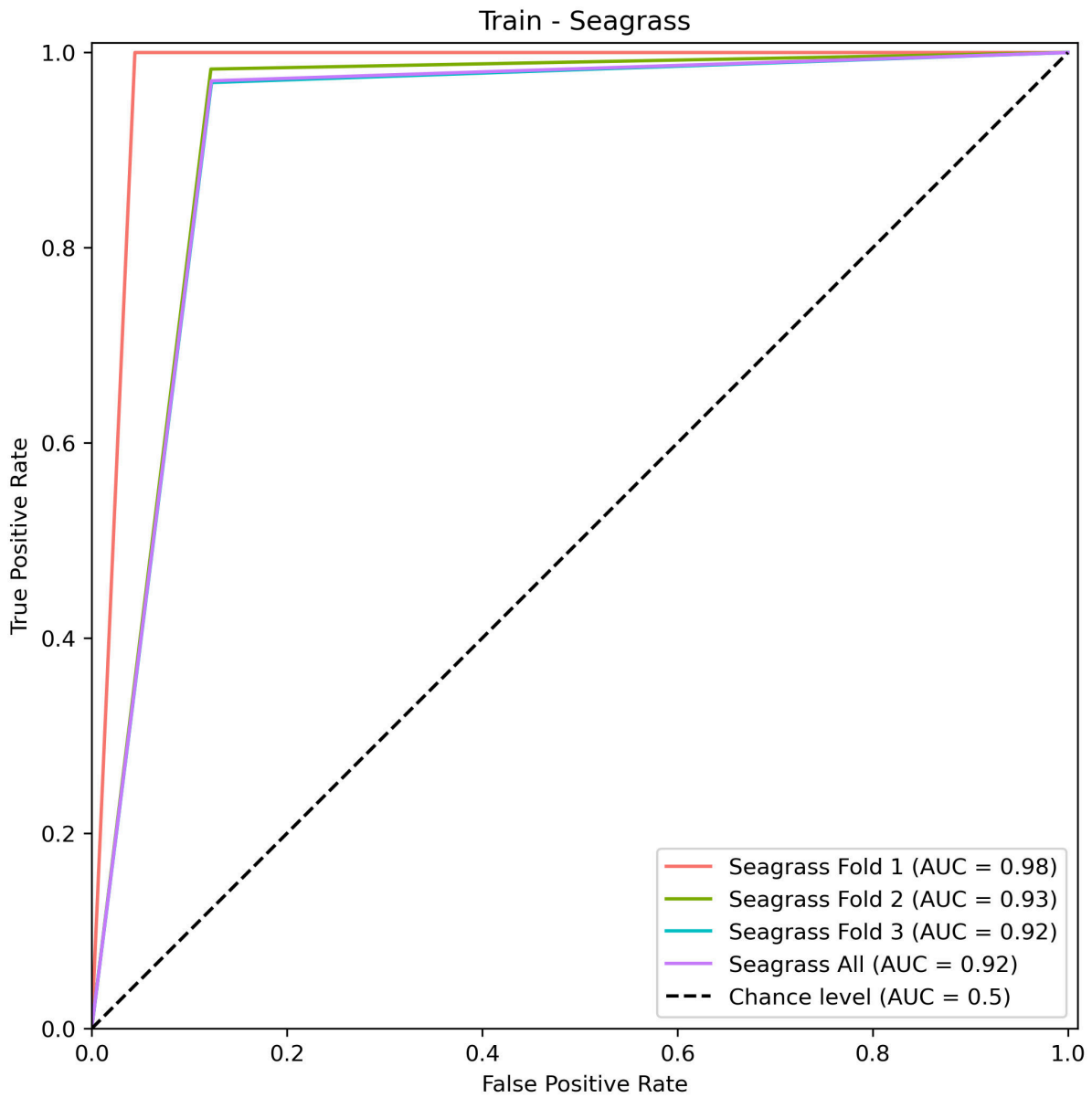


Figure S18 – AUC curves of the explanatory power for the Seagrass.

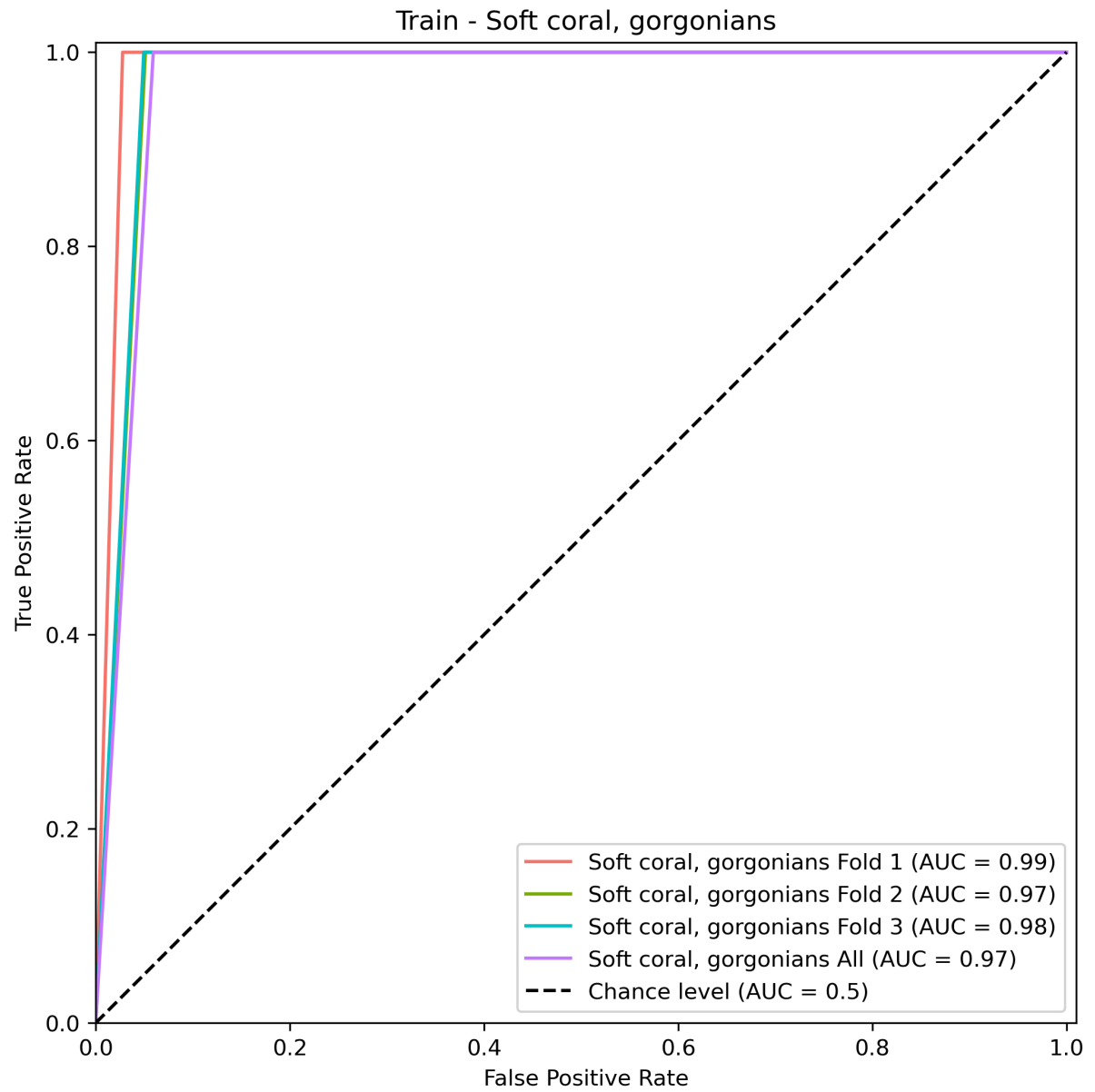


Figure S19 – *AUC curves of the explanatory power for the Soft coral and gorgonians.*

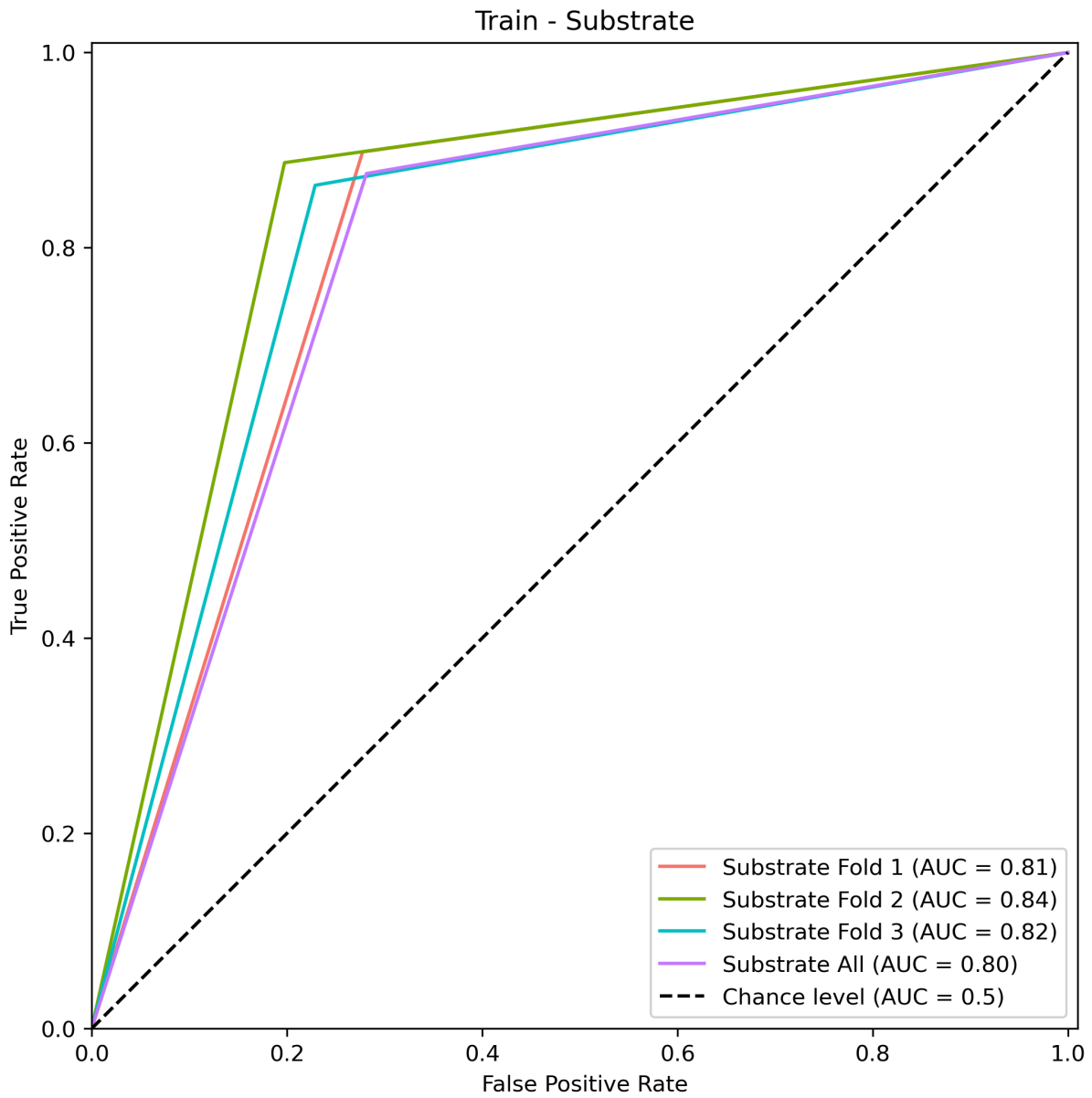


Figure S20 – AUC curves of the explanatory power for the Substrate.

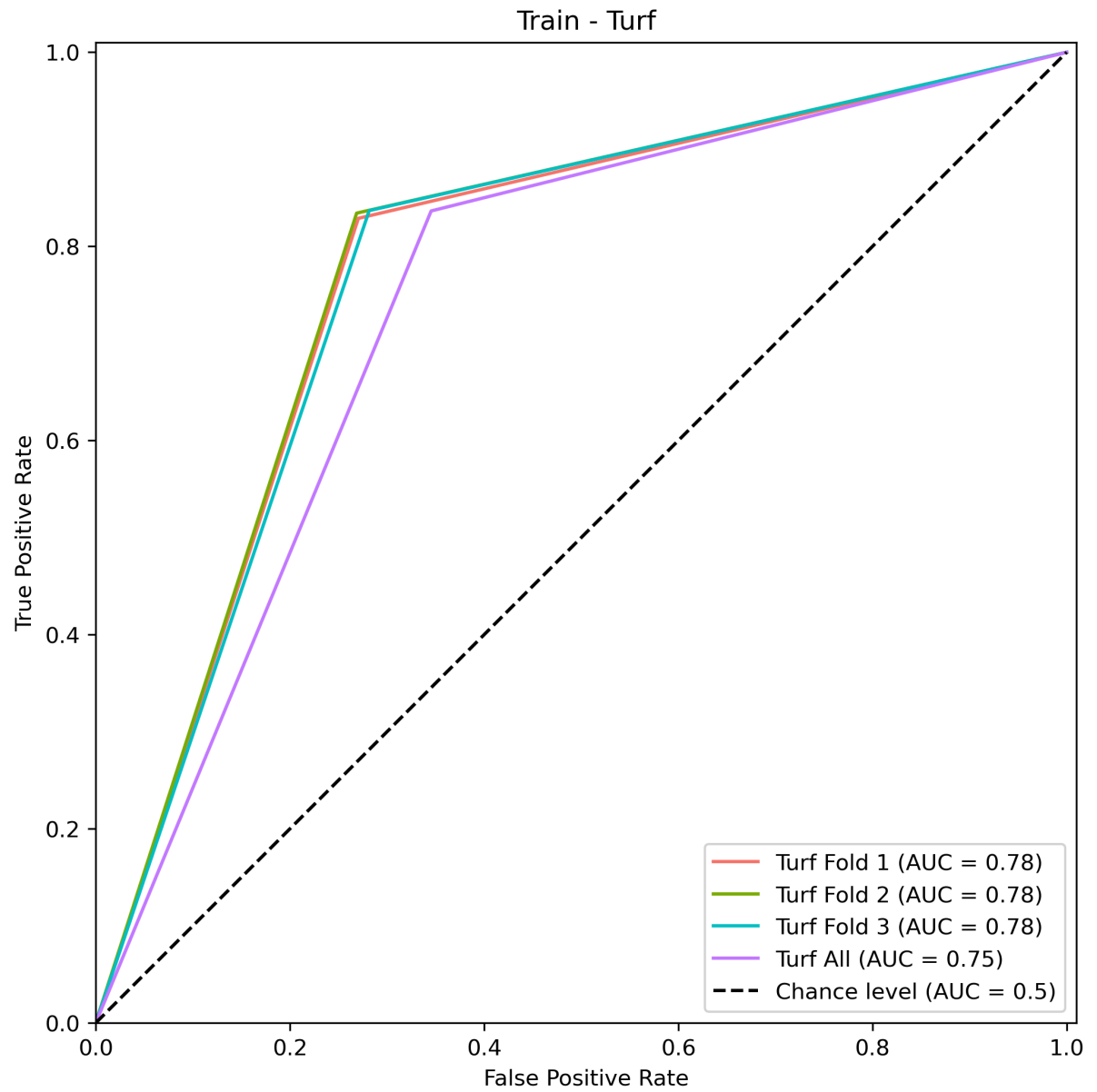


Figure S21 – *AUC curves of the explanatory power for the Turf.*

AUPRC Curves

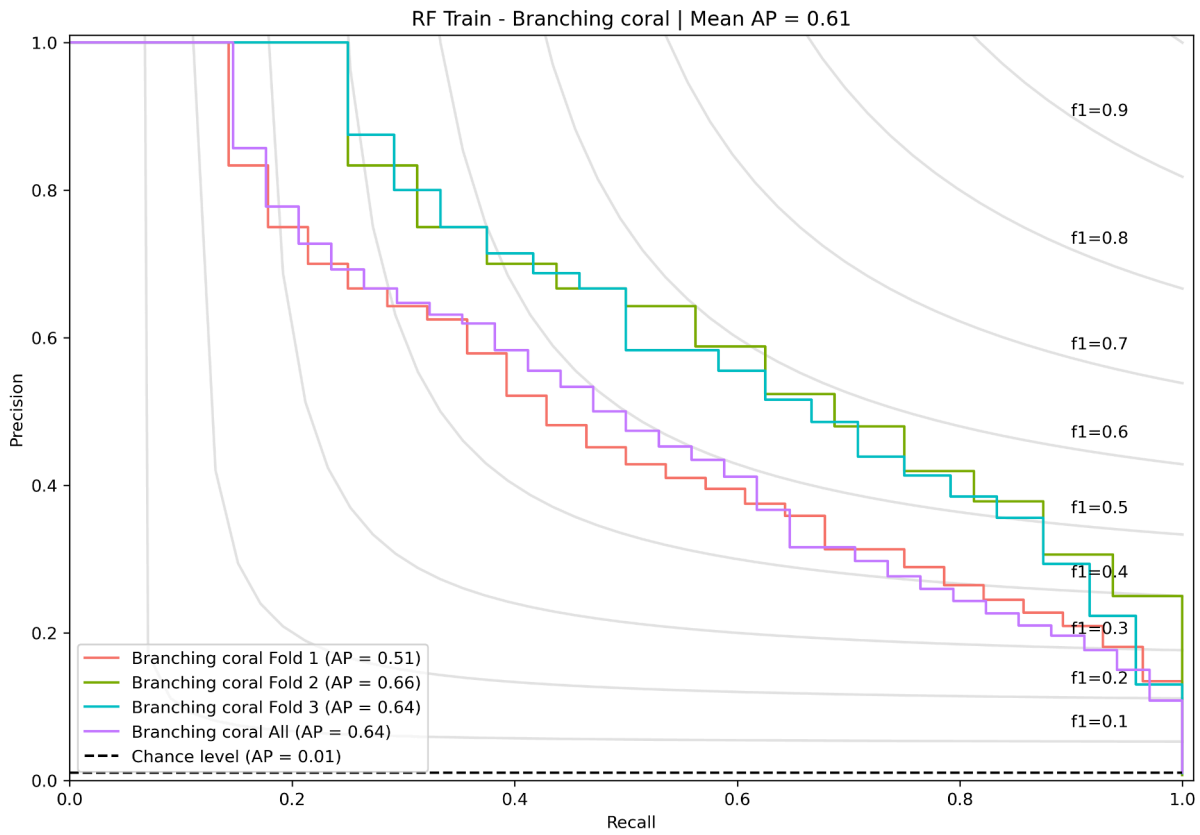


Figure S22 – AUPRC curves of the explanatory power for the Branching coral. The grey lines represent isolines of F1-score.

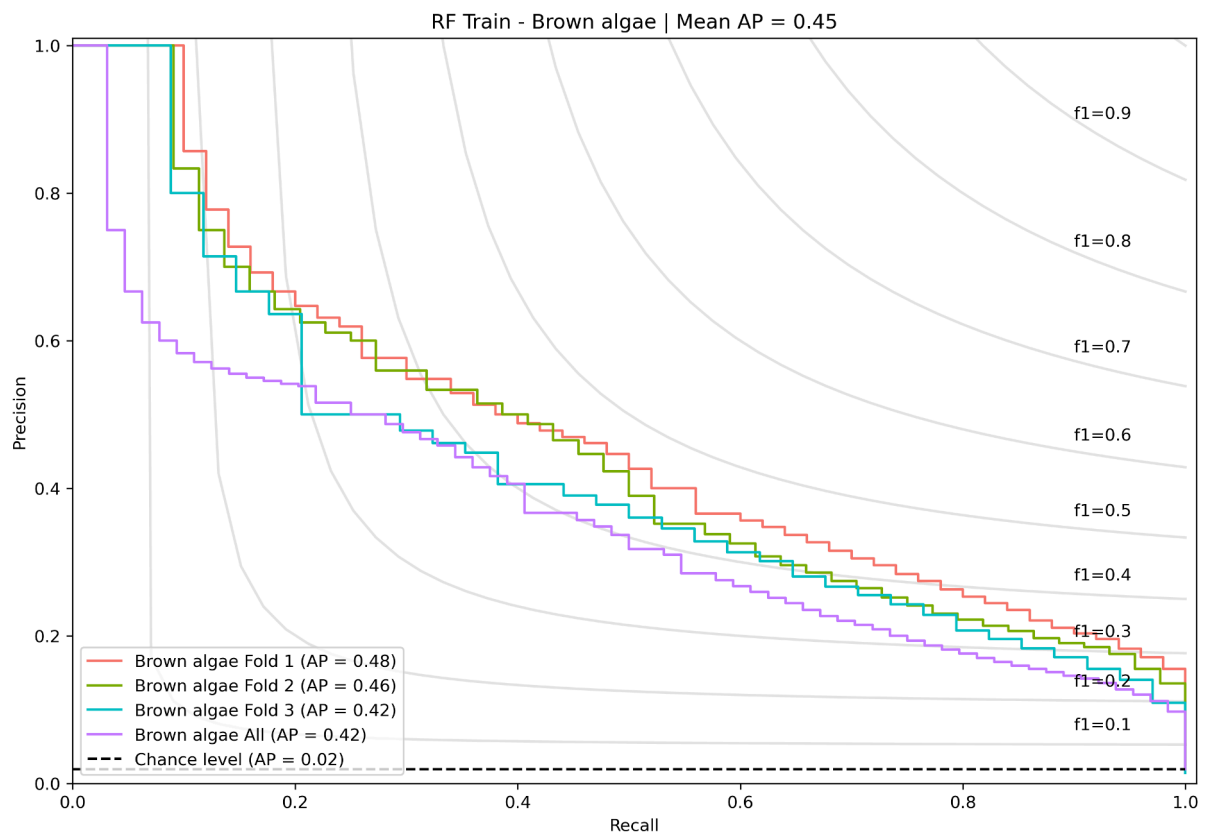


Figure S23 – AUPRC curves of the explanatory power for the Brown algae. The grey lines represent isolines of F1-score.

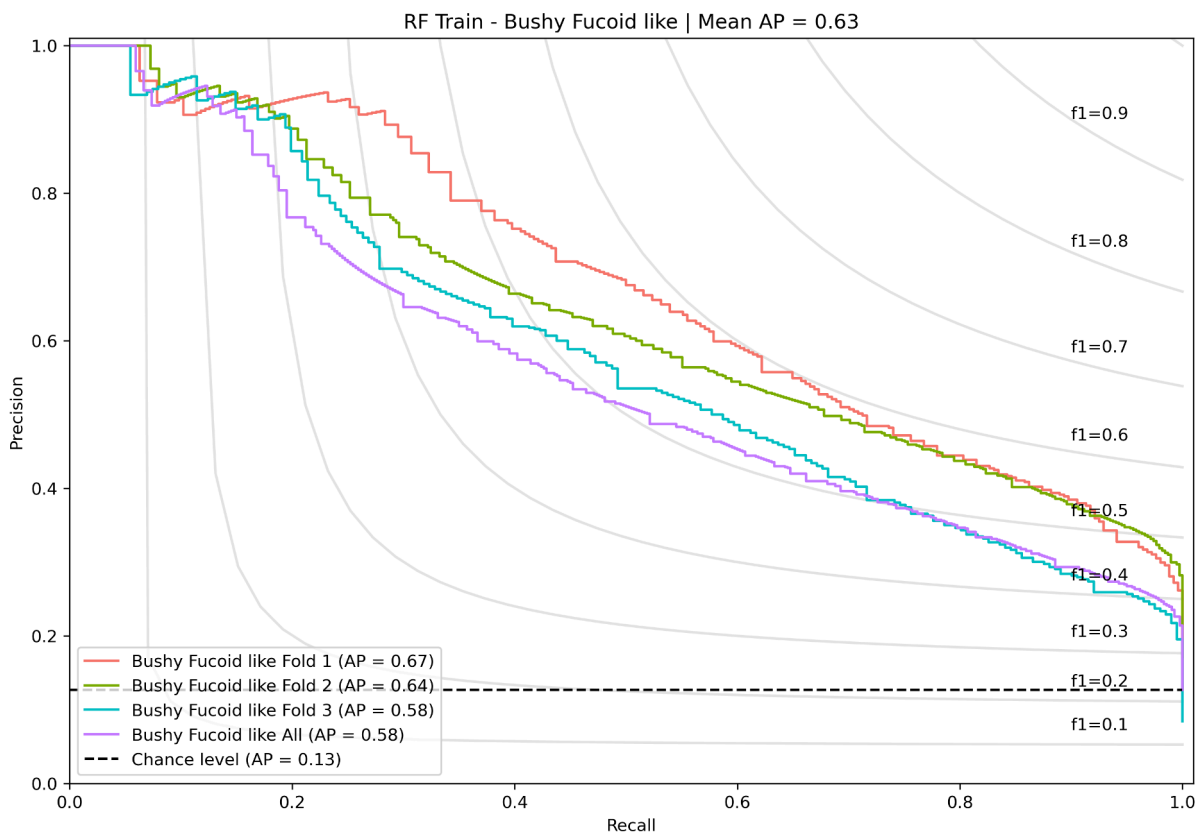


Figure S24 – AUPRC curves of the explanatory power for the Bushy Furoid like. The grey lines represent isolines of F1-score.

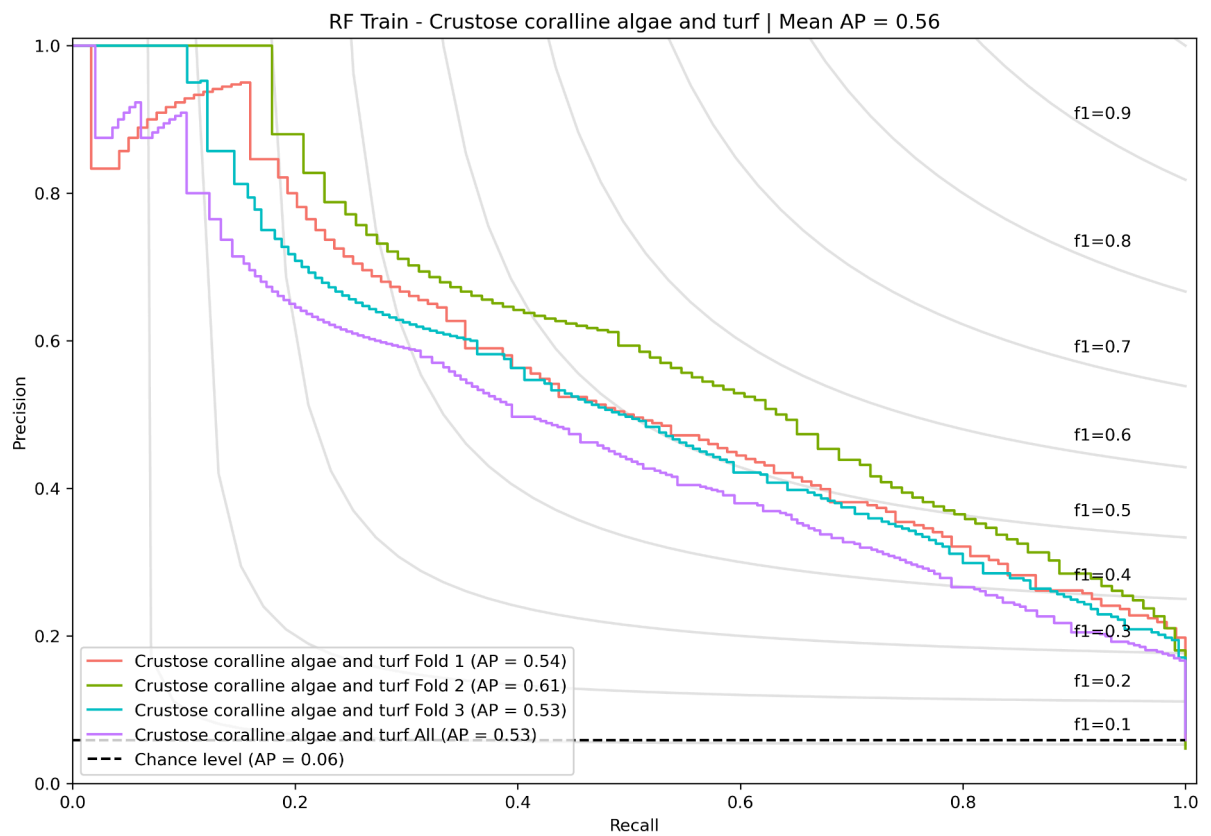


Figure S25 – AUPRC curves of the explanatory power for the Crustose coralline algae and turf. The grey lines represent isolines of F1-score.

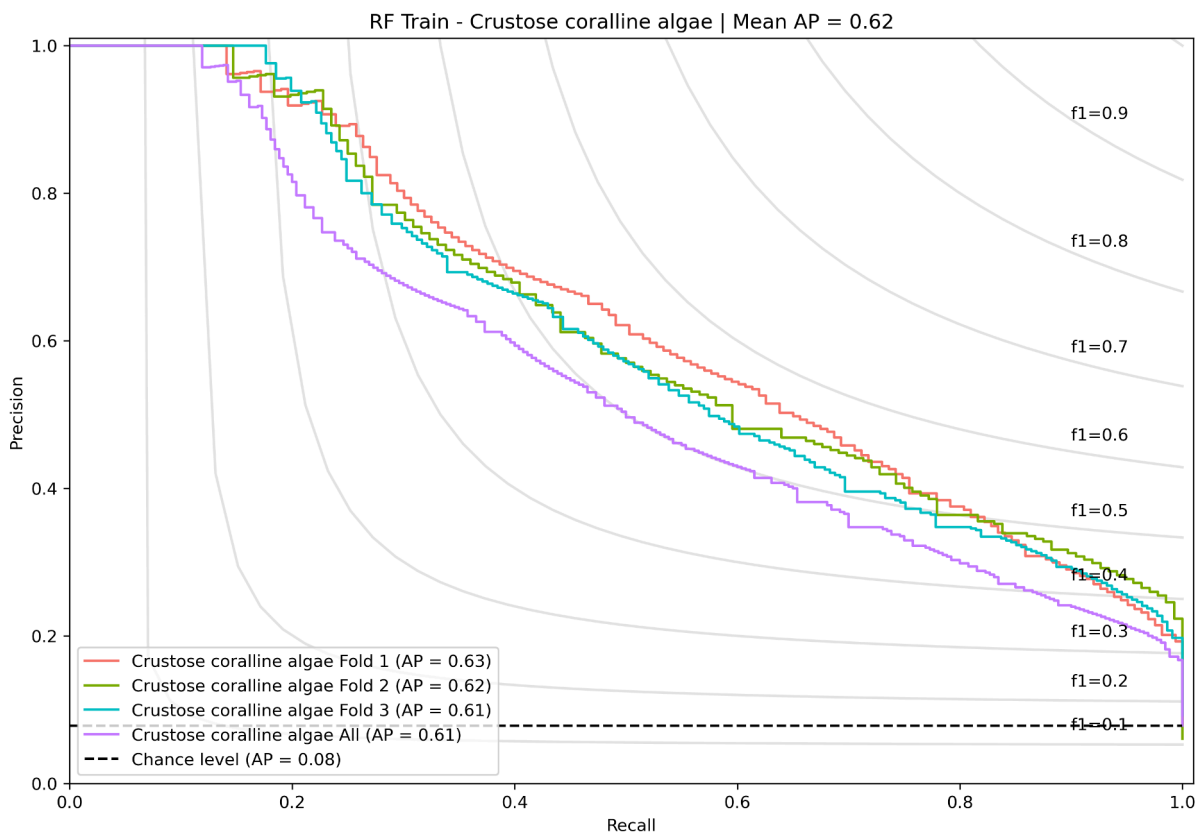


Figure S26 – AUPRC curves of the explanatory power for the Crustose coralline algae. The grey lines represent isolines of F1-score.

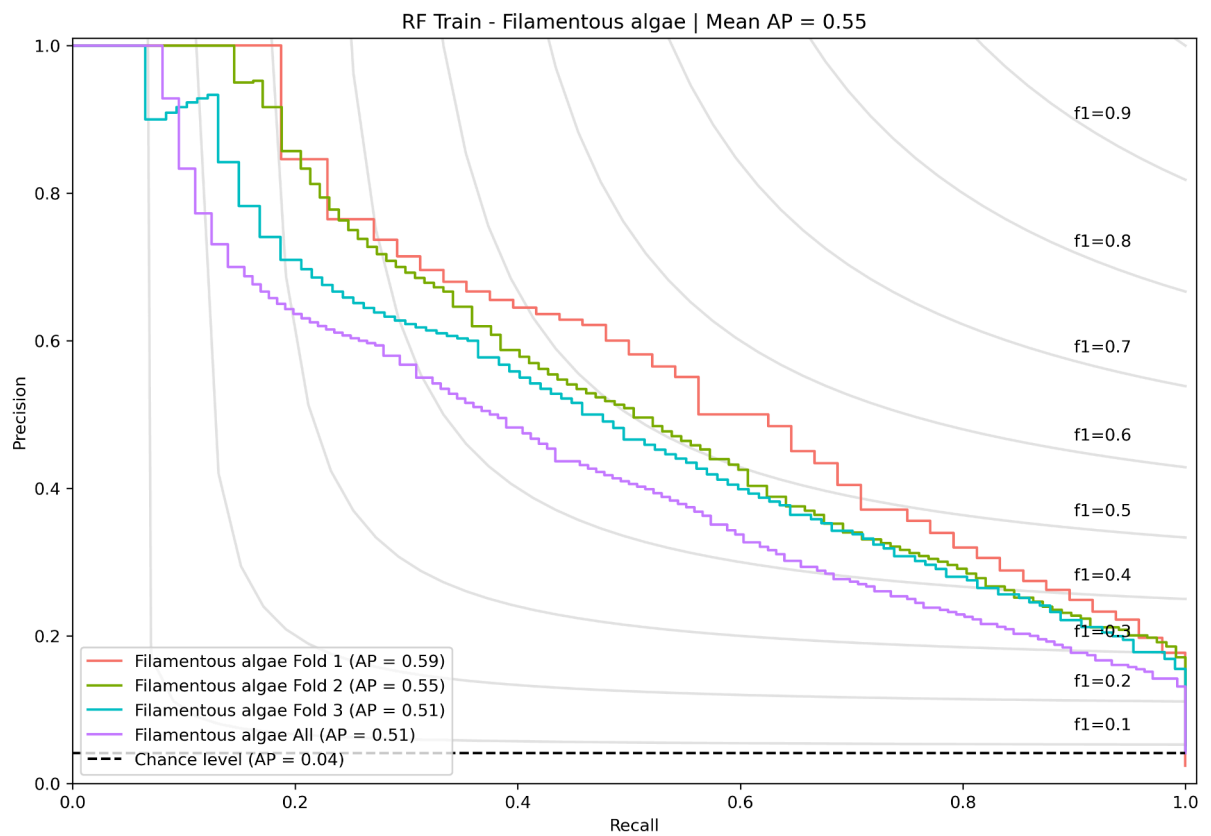


Figure S27 – *AUPRC curves of the explanatory power for the Filamentous algae. The grey lines represent isolines of F1-score.*

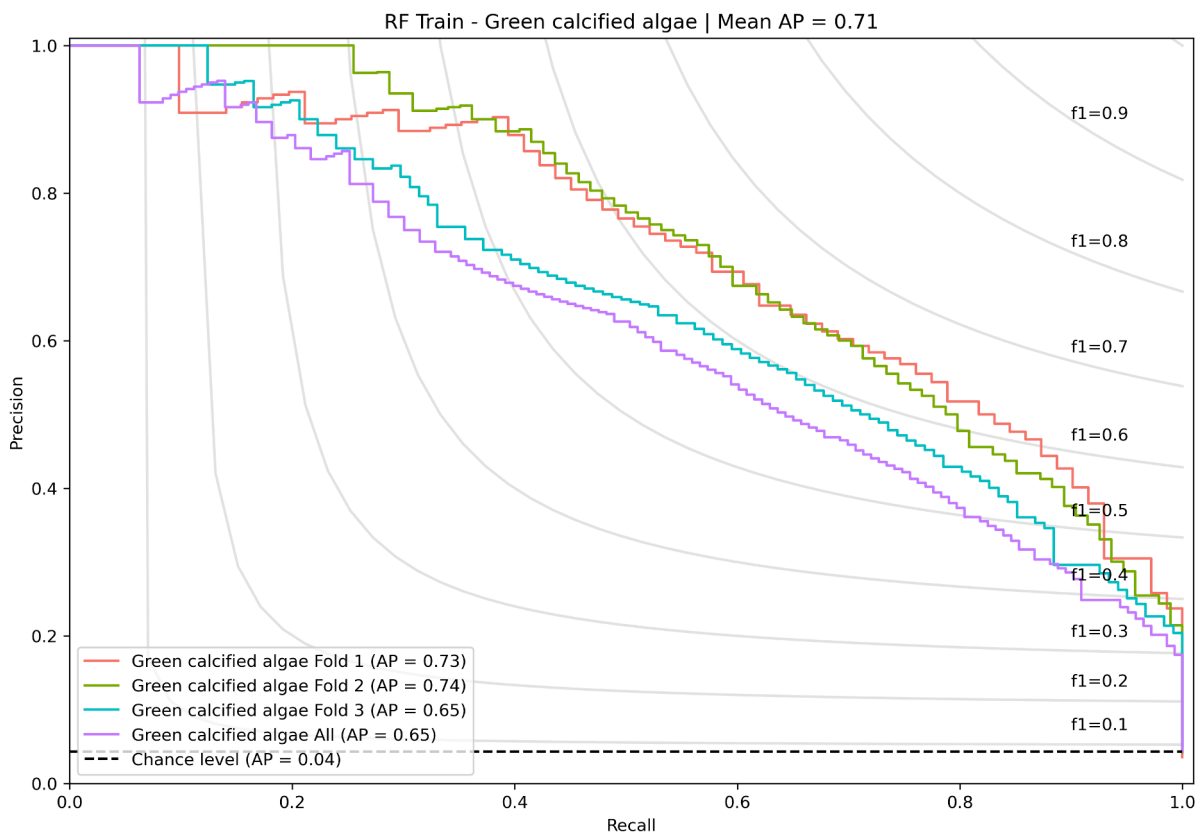


Figure S28 – AUPRC curves of the explanatory power for the Green Calcified algae. The grey lines represent isolines of F1-score.

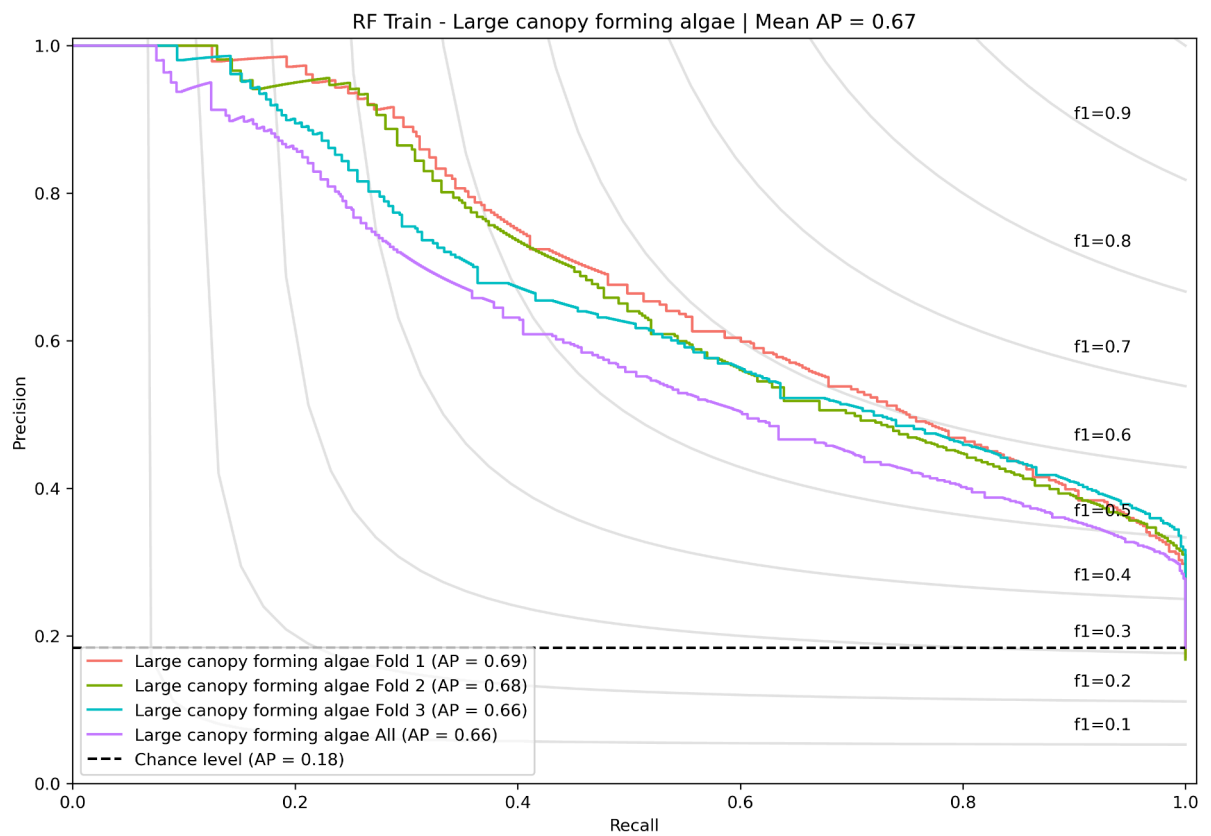


Figure S29 – AUPRC curves of the explanatory power for the Large canopy forming algae. The grey lines represent isolines of F1-score.

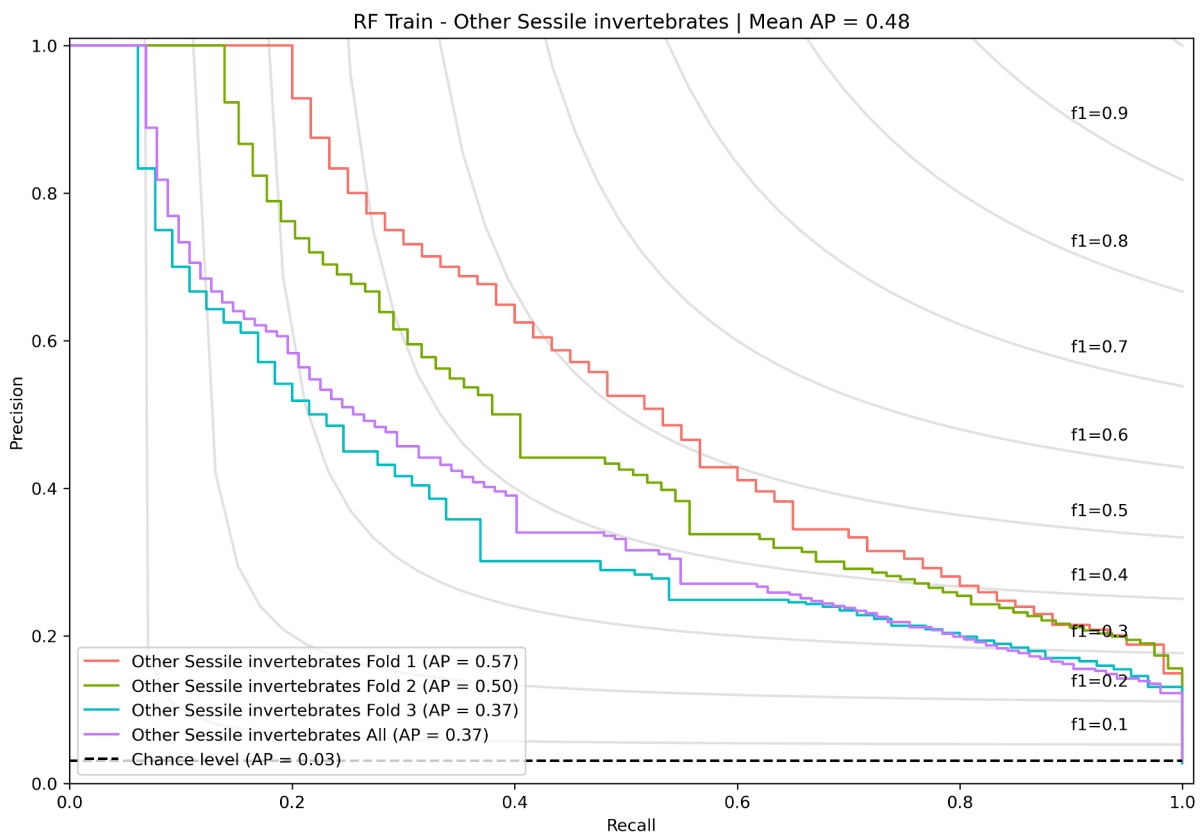


Figure S30 – AUPRC curves of the explanatory power for the Other Sessile invertebrates. The grey lines represent isolines of F1-score.

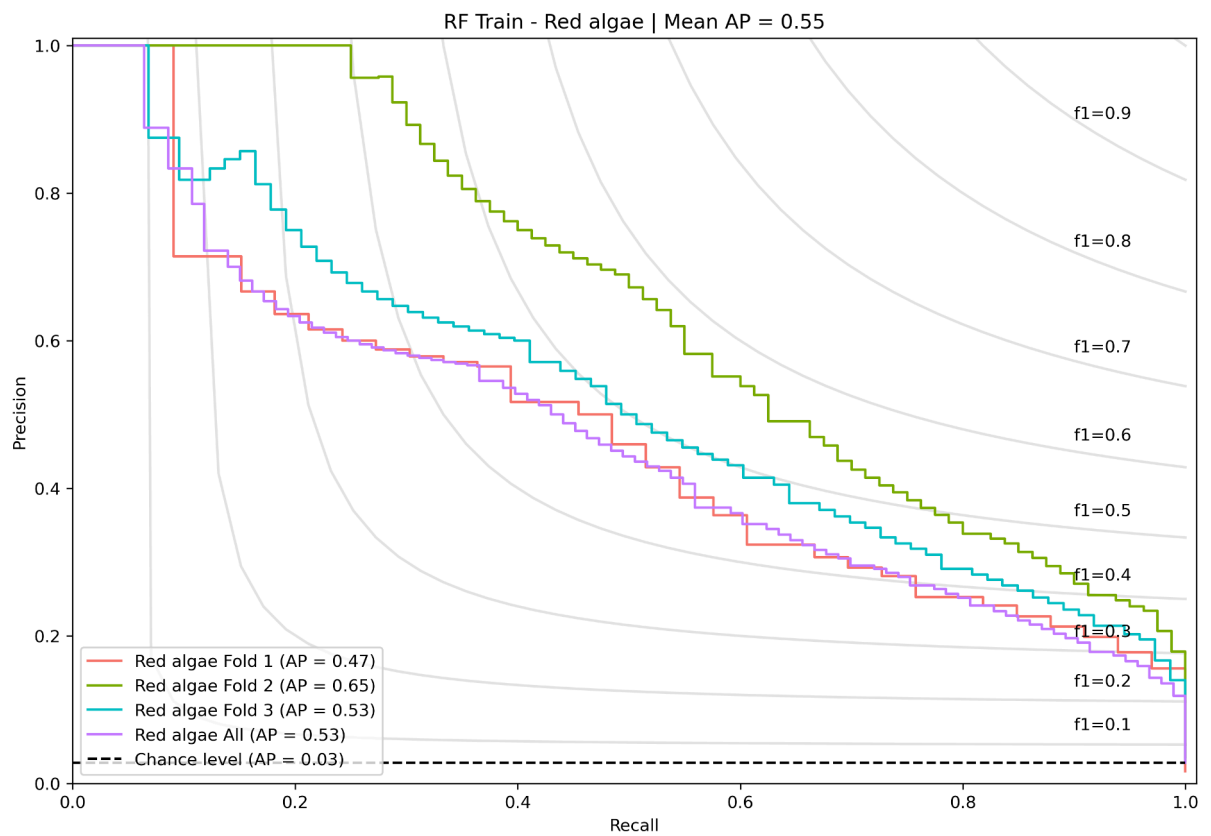


Figure S31 – AUPRC curves of the explanatory power for the Red algae. The grey lines represent isolines of F1-score.

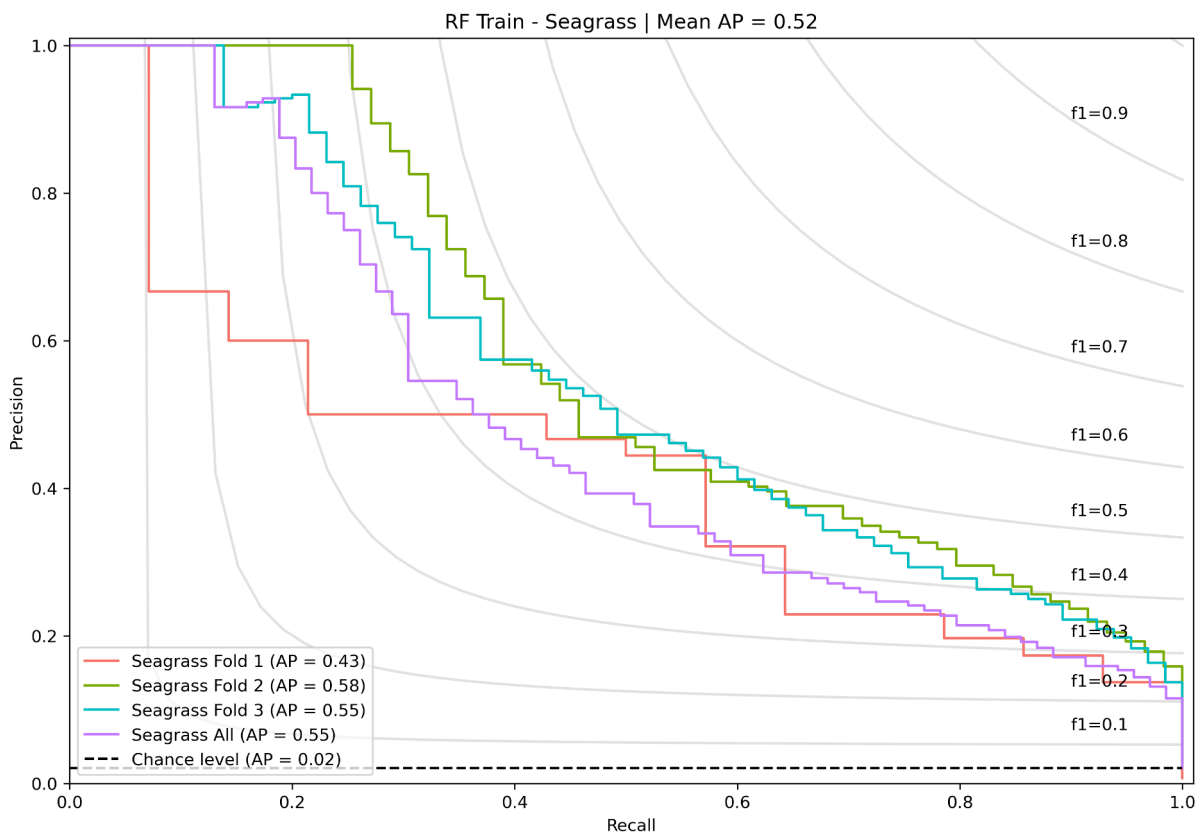


Figure S32 – AUPRC curves of the explanatory power for the Seagrass. The grey lines represent isolines of F1-score.

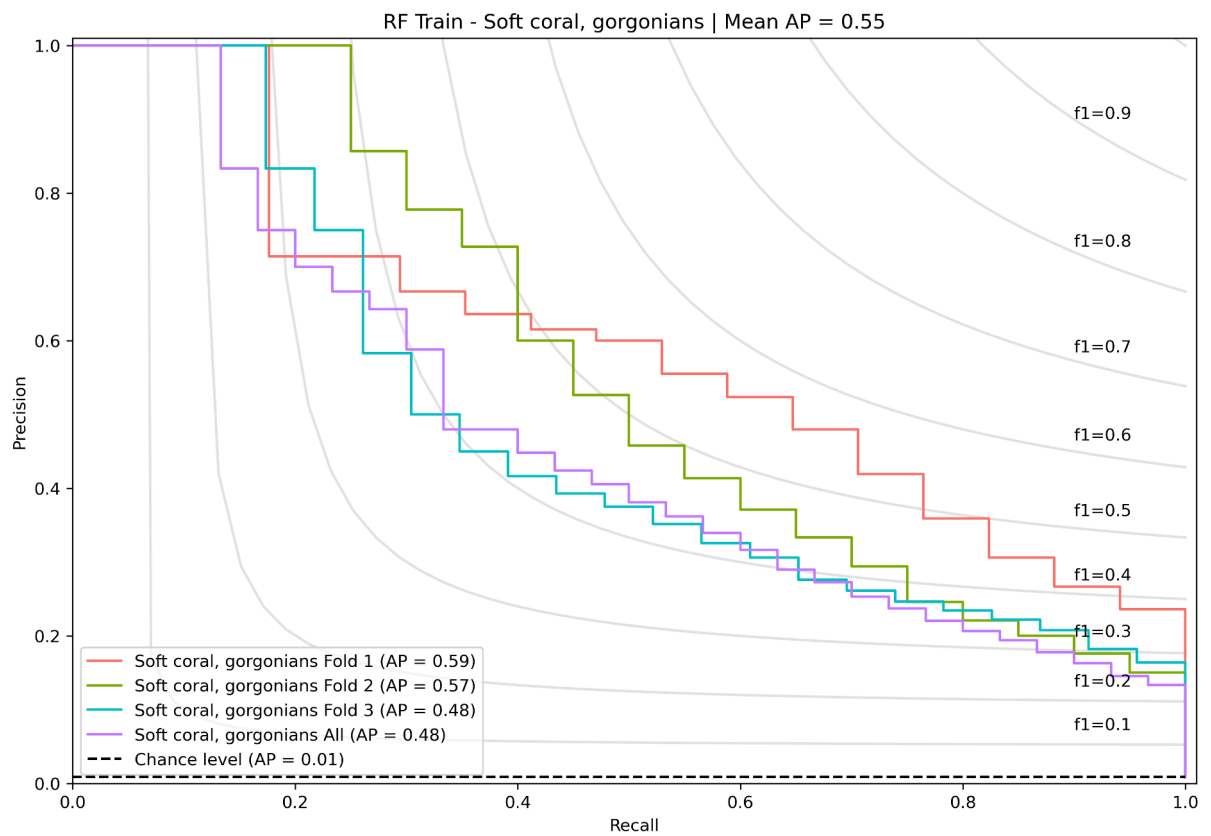


Figure S33 – AUPRC curves of the explanatory power for the Soft coral and gorgonians. The grey lines represent isolines of F1-score.

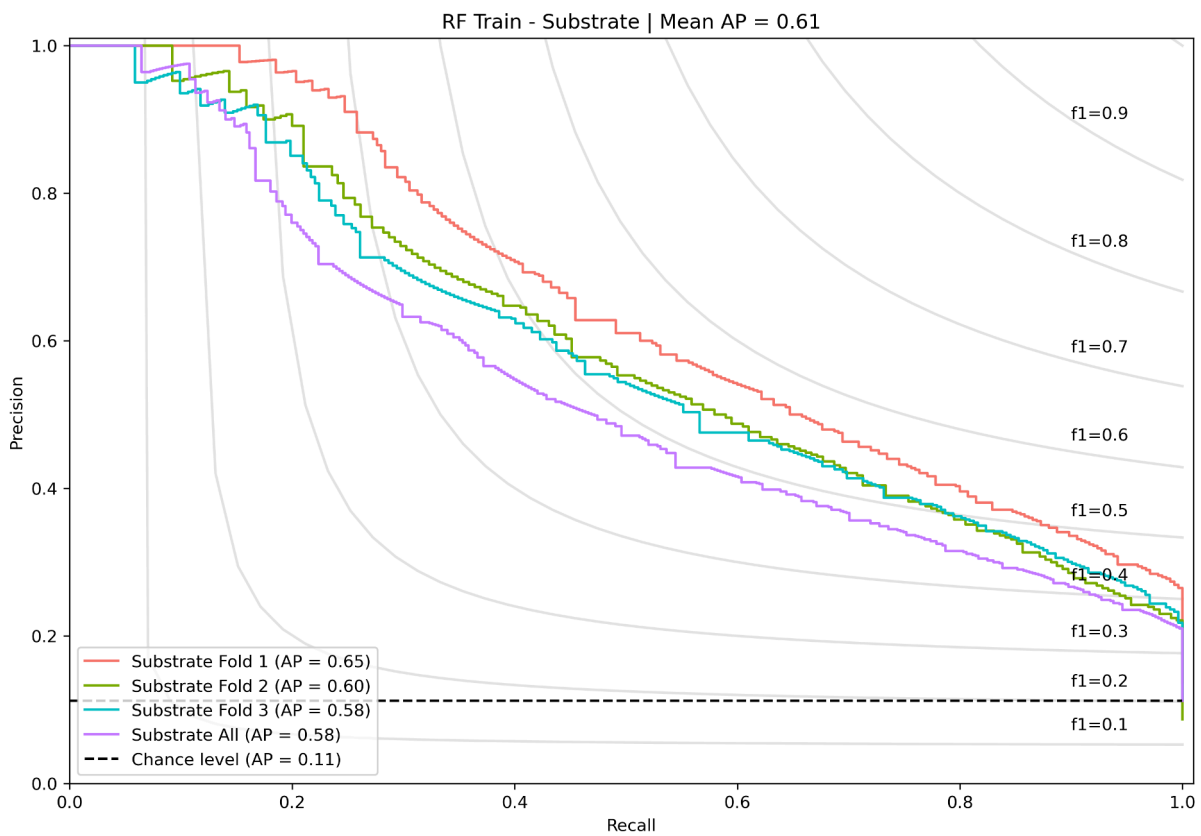


Figure S34 – AUPRC curves of the explanatory power for the Substrate. The grey lines represent isolines of F1-score.

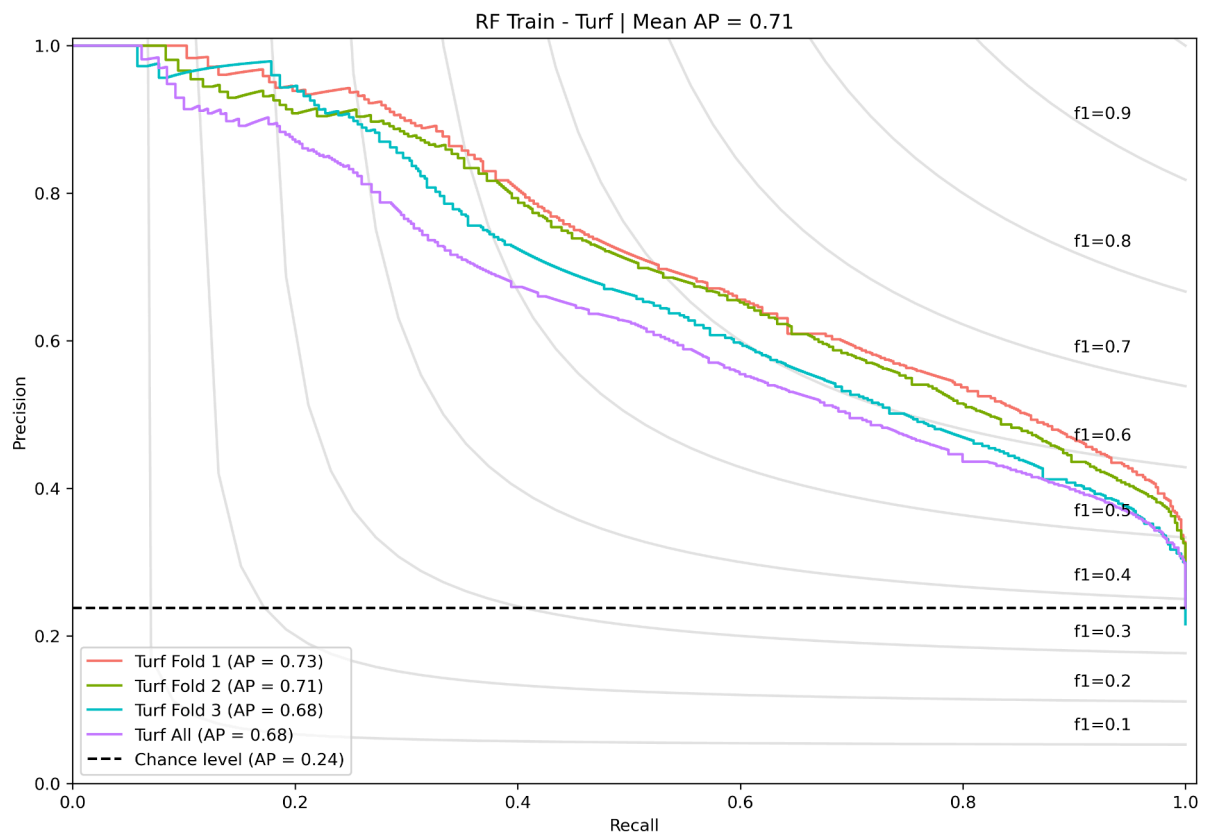


Figure S35 – *AUPRC* curves of the explanatory power for the Turf. The grey lines represent isolines of *F1*-score.

Appendix C - Model Explanation

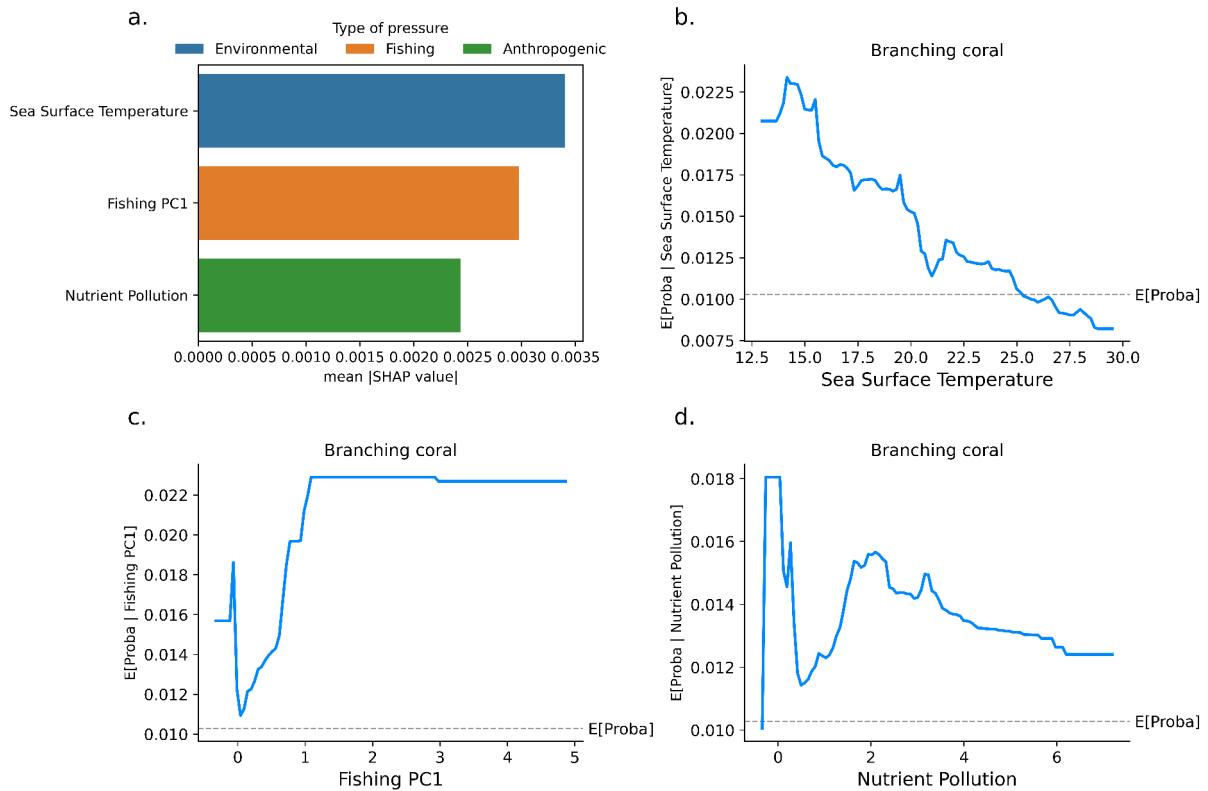


Figure S36 – a. Bar plot of the three most important variables according to the SHAP framework for Branching coral habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

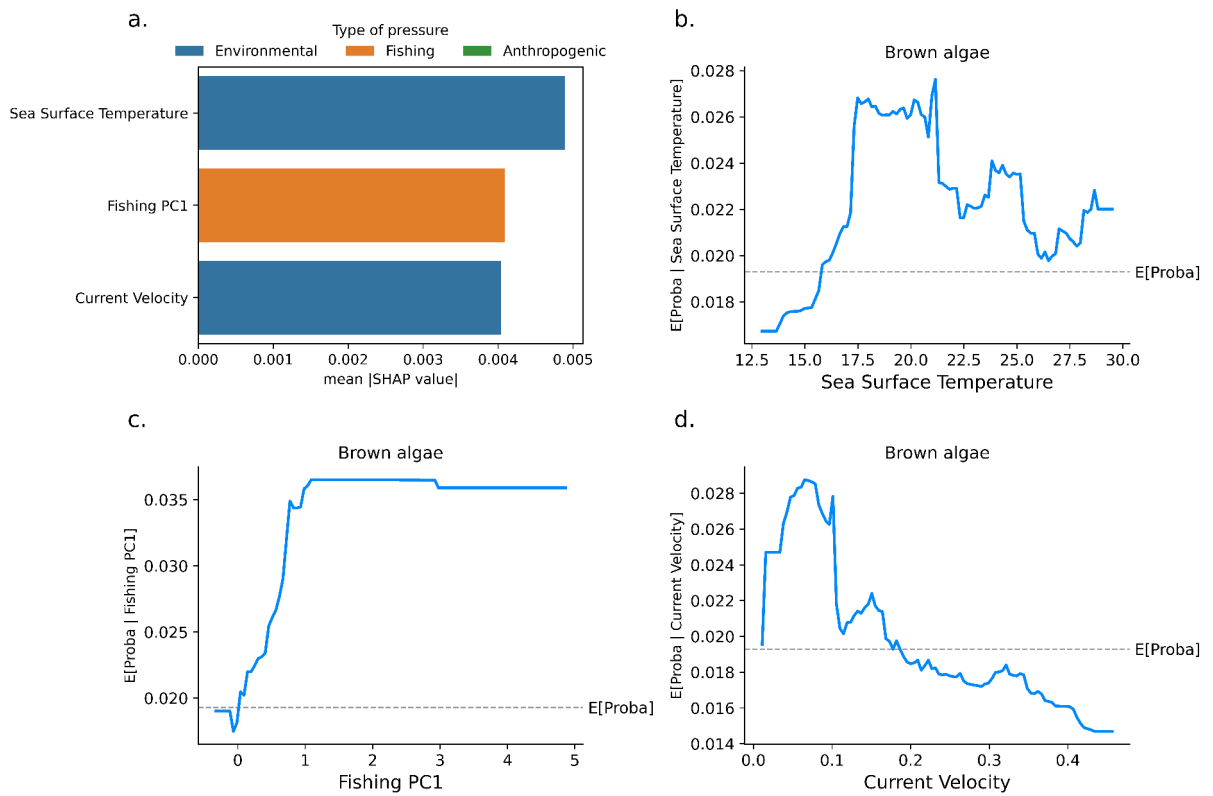


Figure S37 – a. Bar plot of the three most important variables according to the SHAP framework for Brown algae habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

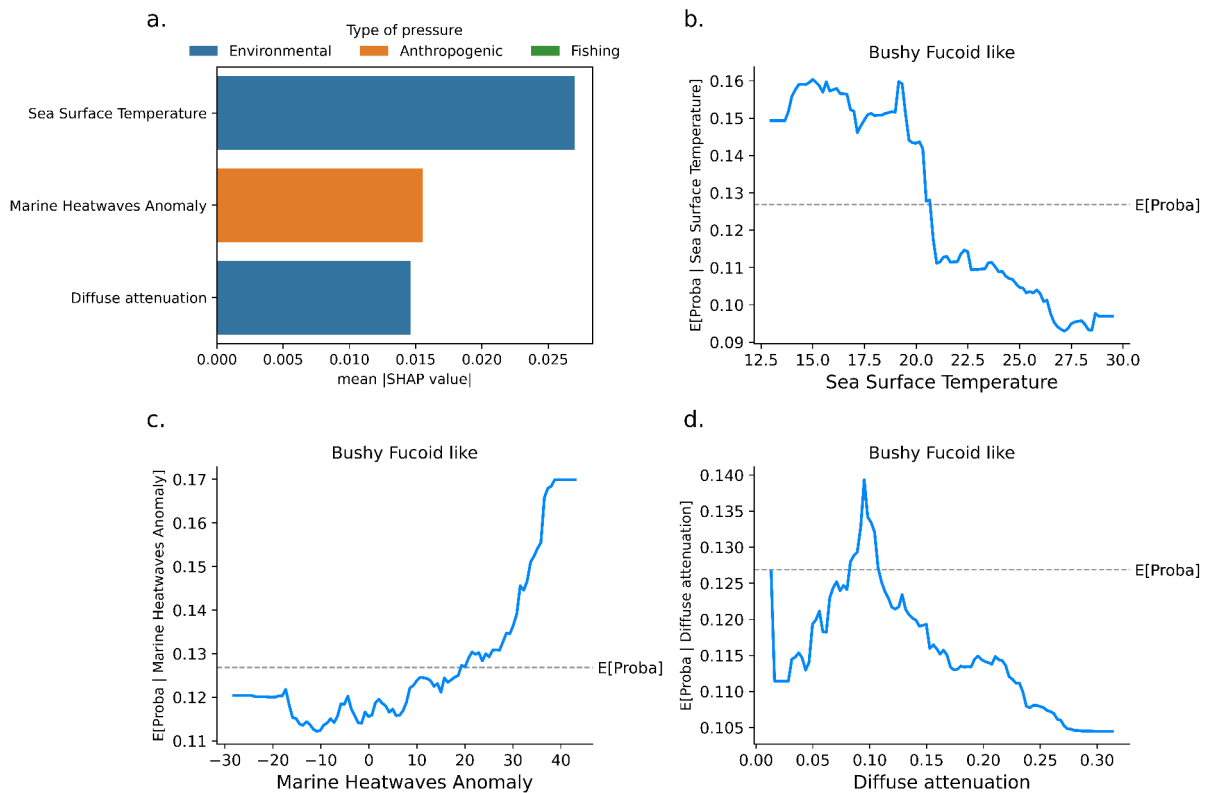


Figure S38 – a. Bar plot of the three most important variables according to the SHAP framework for Bushy Furoid like habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

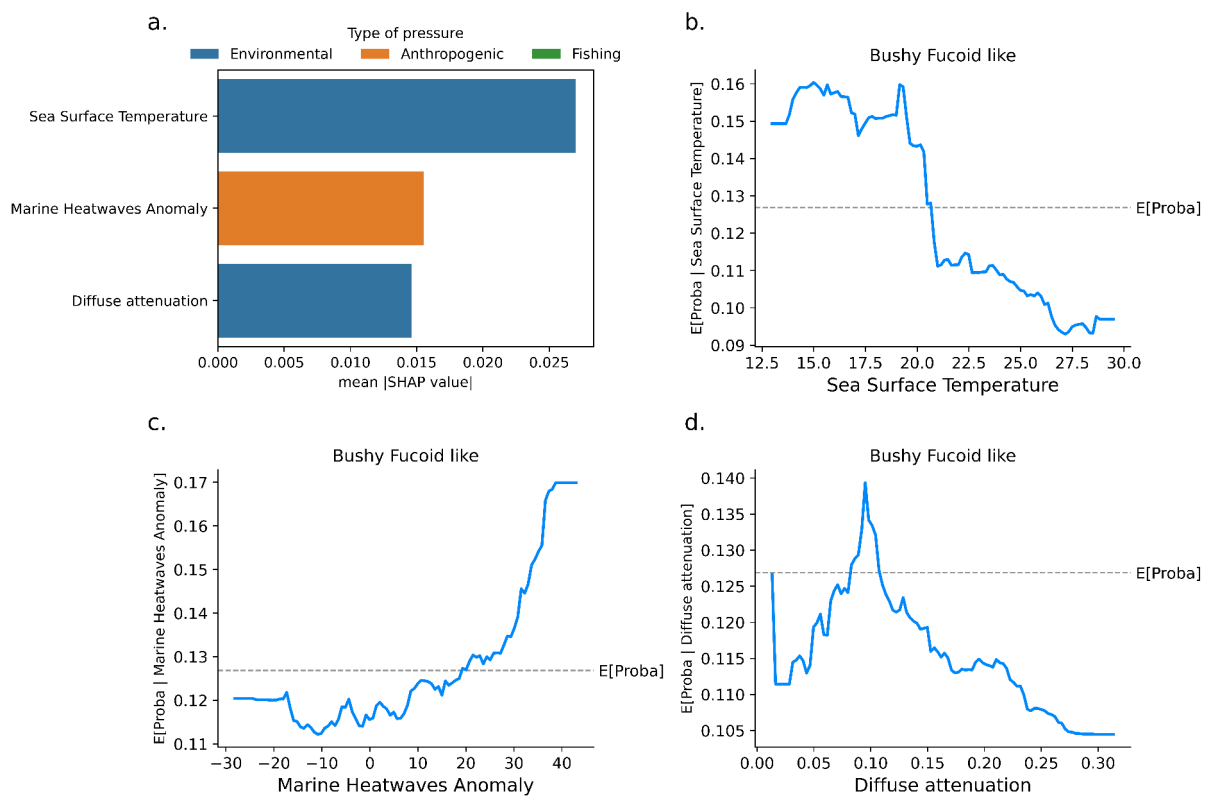


Figure S39 – a. Bar plot of the three most important variables according to the SHAP framework for Bushy Furoid like habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

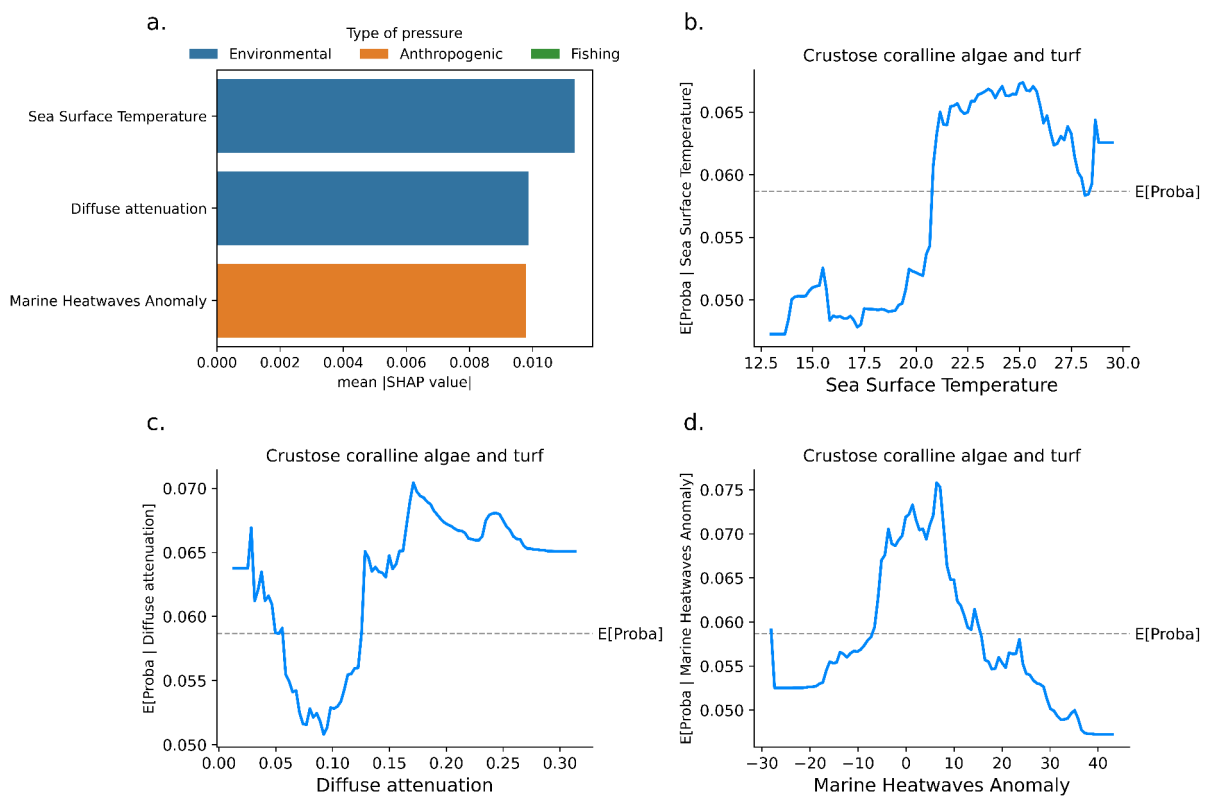


Figure S40 – a. Bar plot of the three most important variables according to the SHAP framework for Crustose coralline algae and turf habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

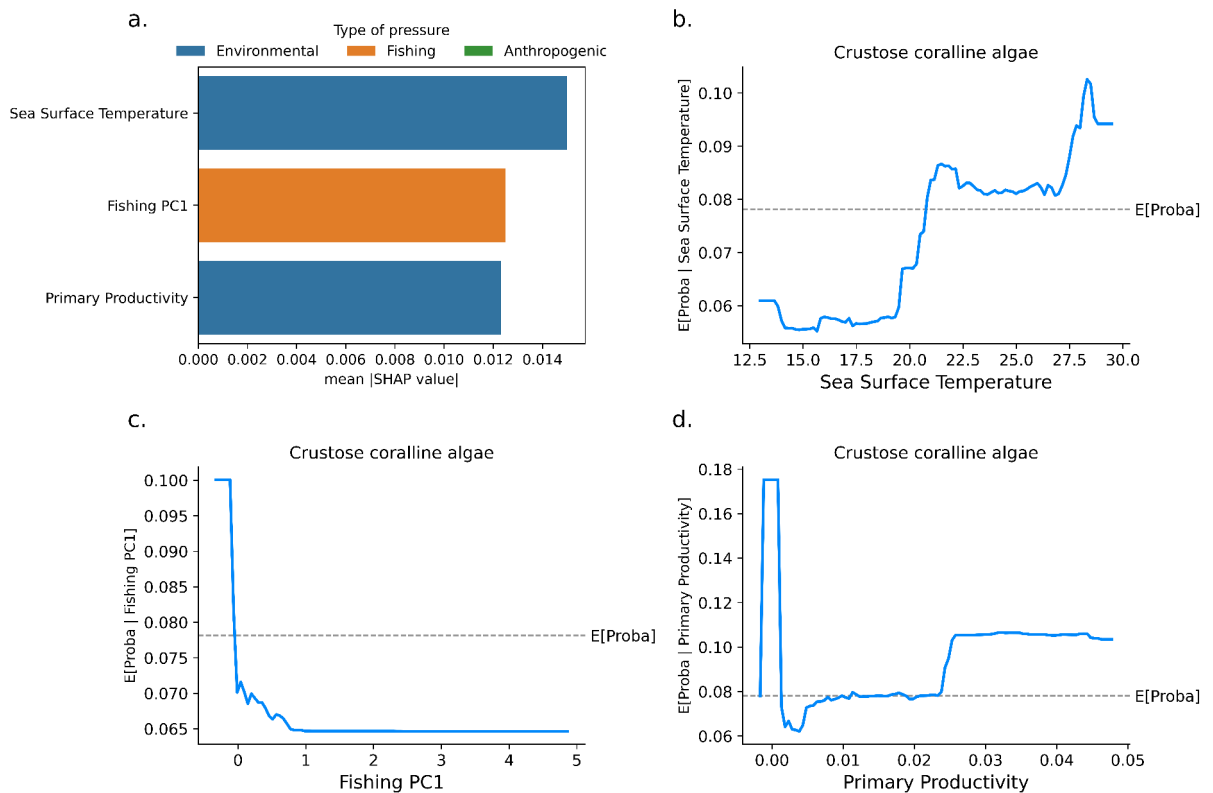


Figure S41 – a. Bar plot of the three most important variables according to the SHAP framework for Crustose coralline algae habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

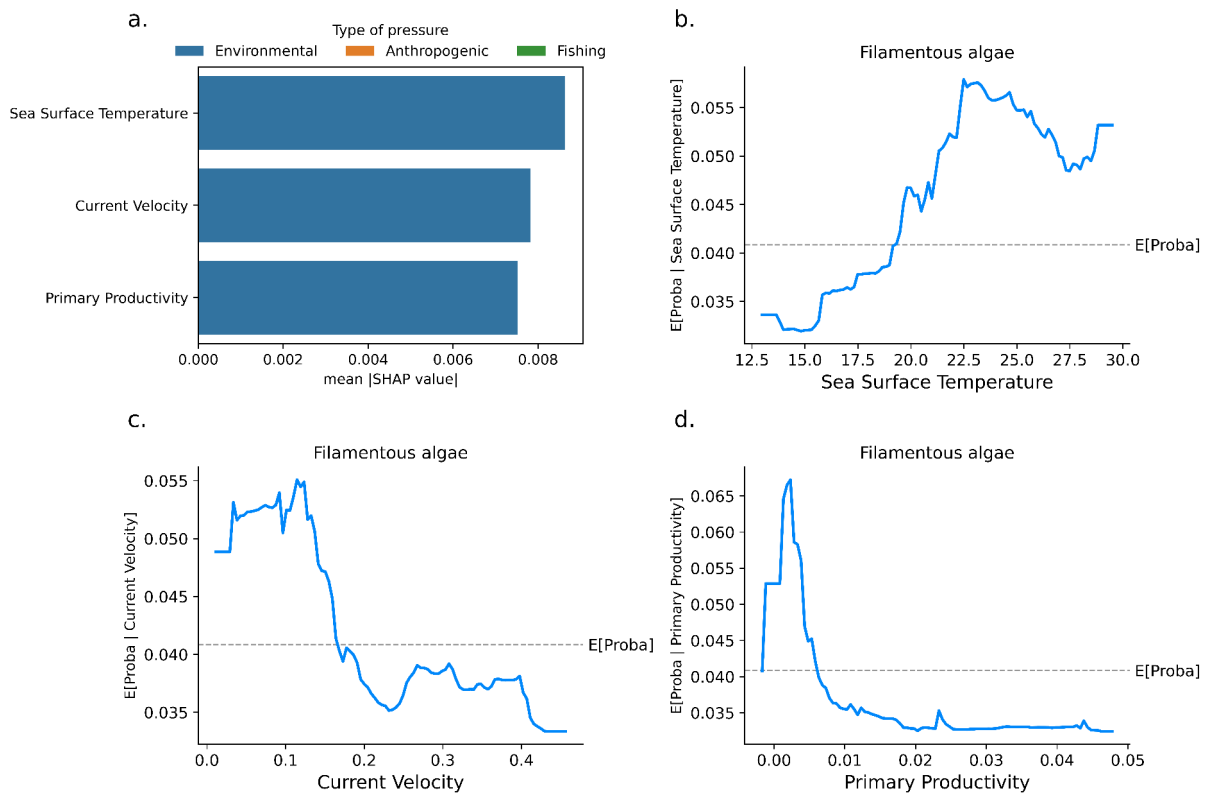


Figure S42 – a. Bar plot of the three most important variables according to the SHAP framework for Filamentous algae habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

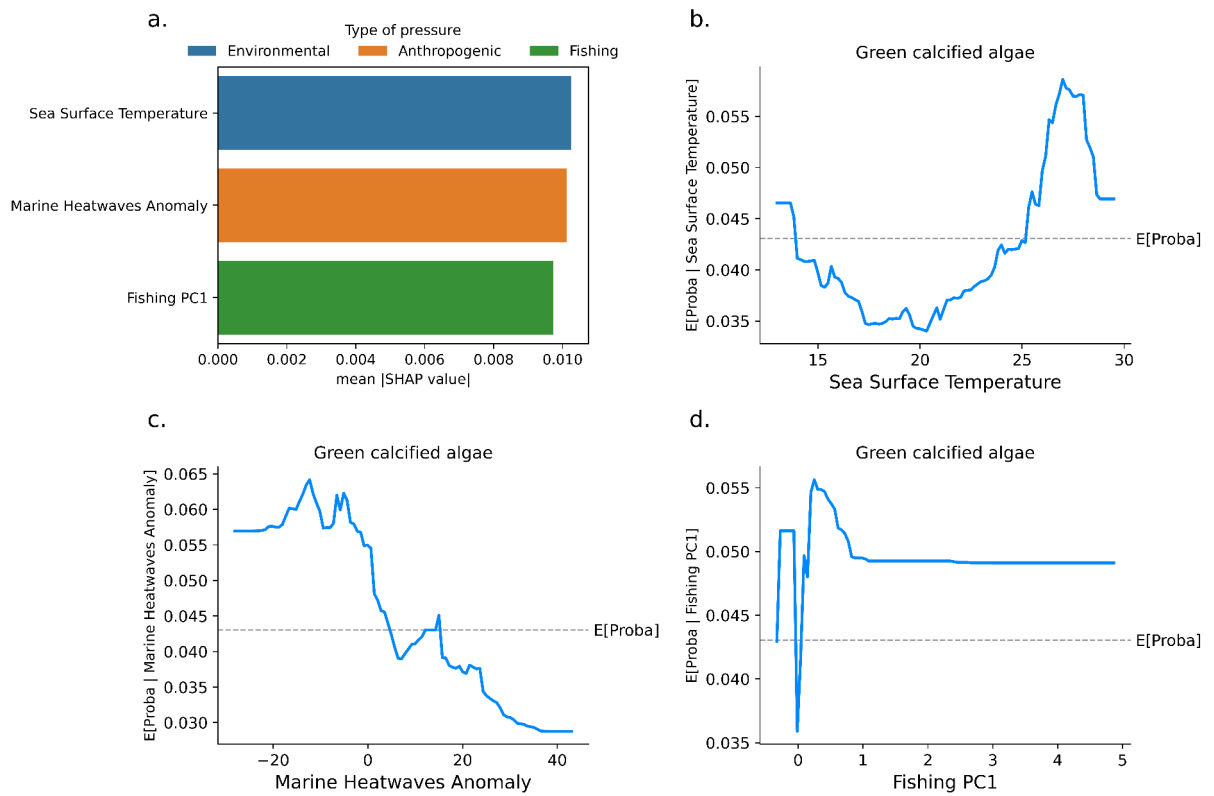


Figure S43 – a. Bar plot of the three most important variables according to the SHAP framework for Green calcified algae habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

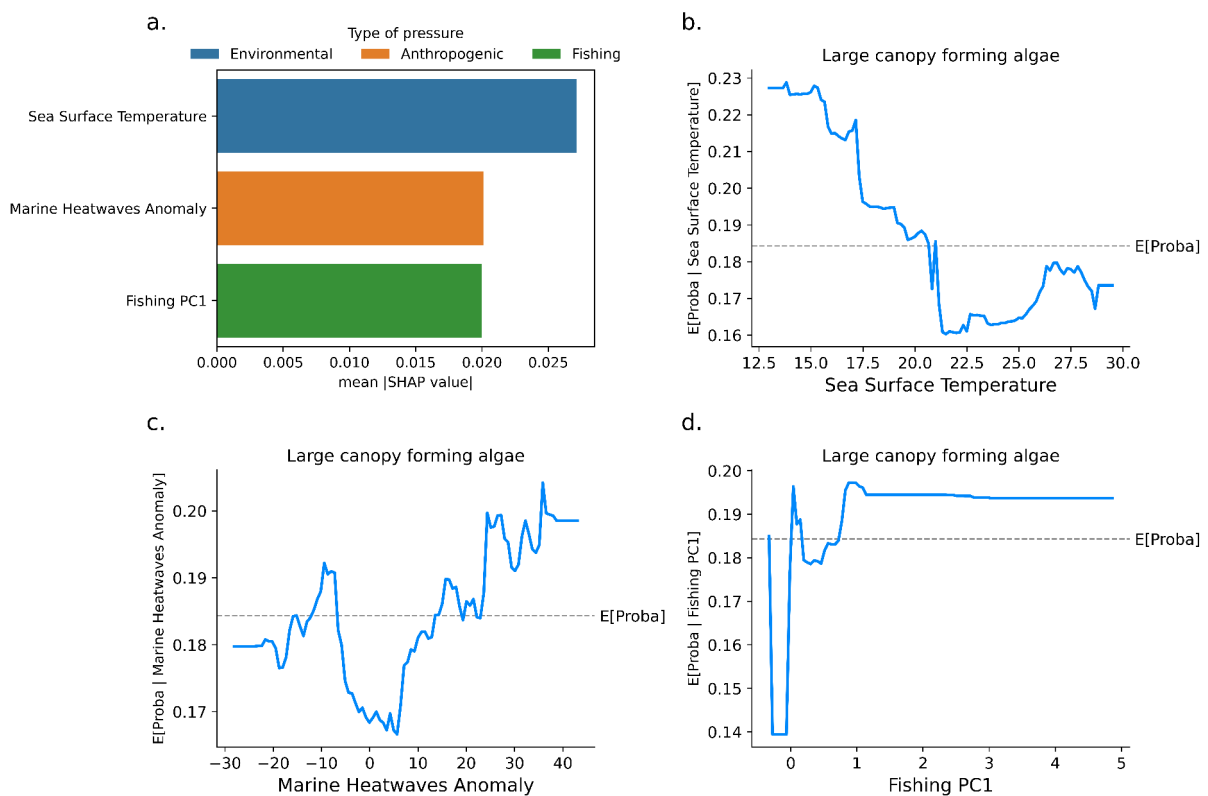


Figure S44 – a. Bar plot of the three most important variables according to the SHAP framework for Large canopy forming algae habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

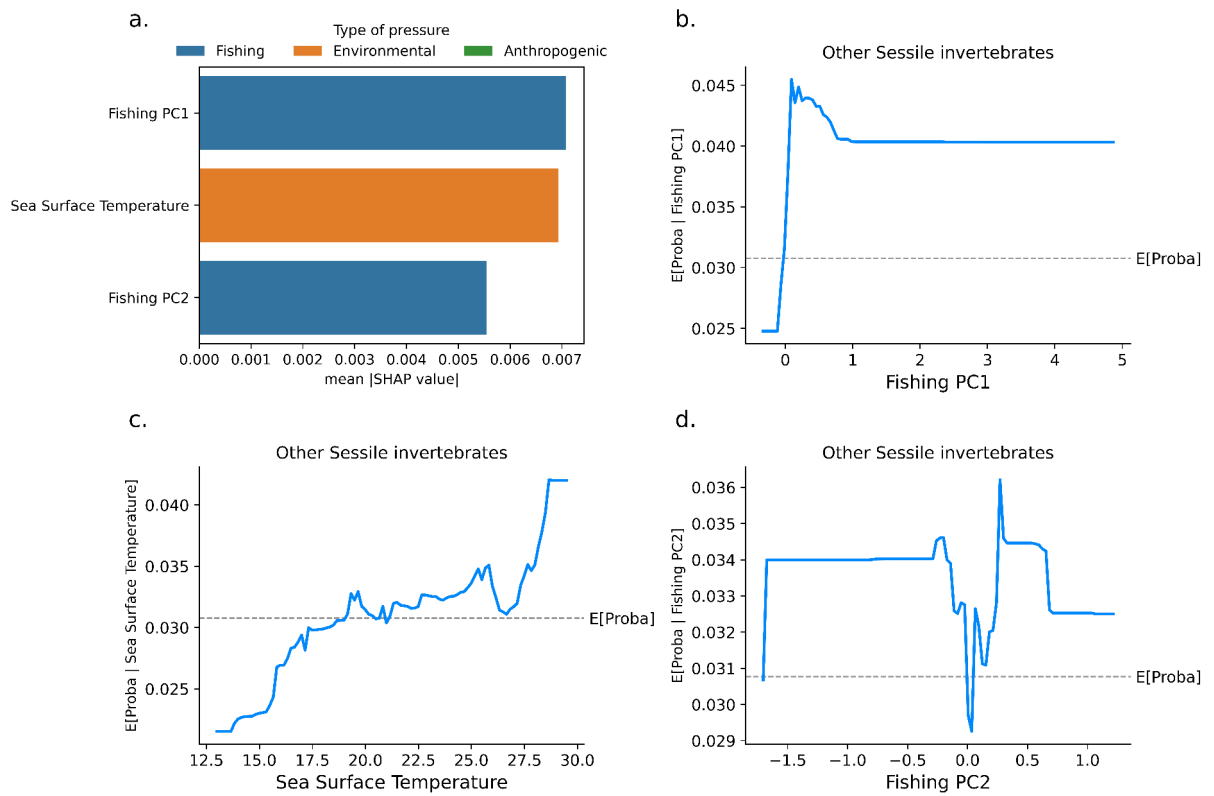


Figure S45 – a. Bar plot of the three most important variables according to the SHAP framework for Other Sessile invertebrates habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

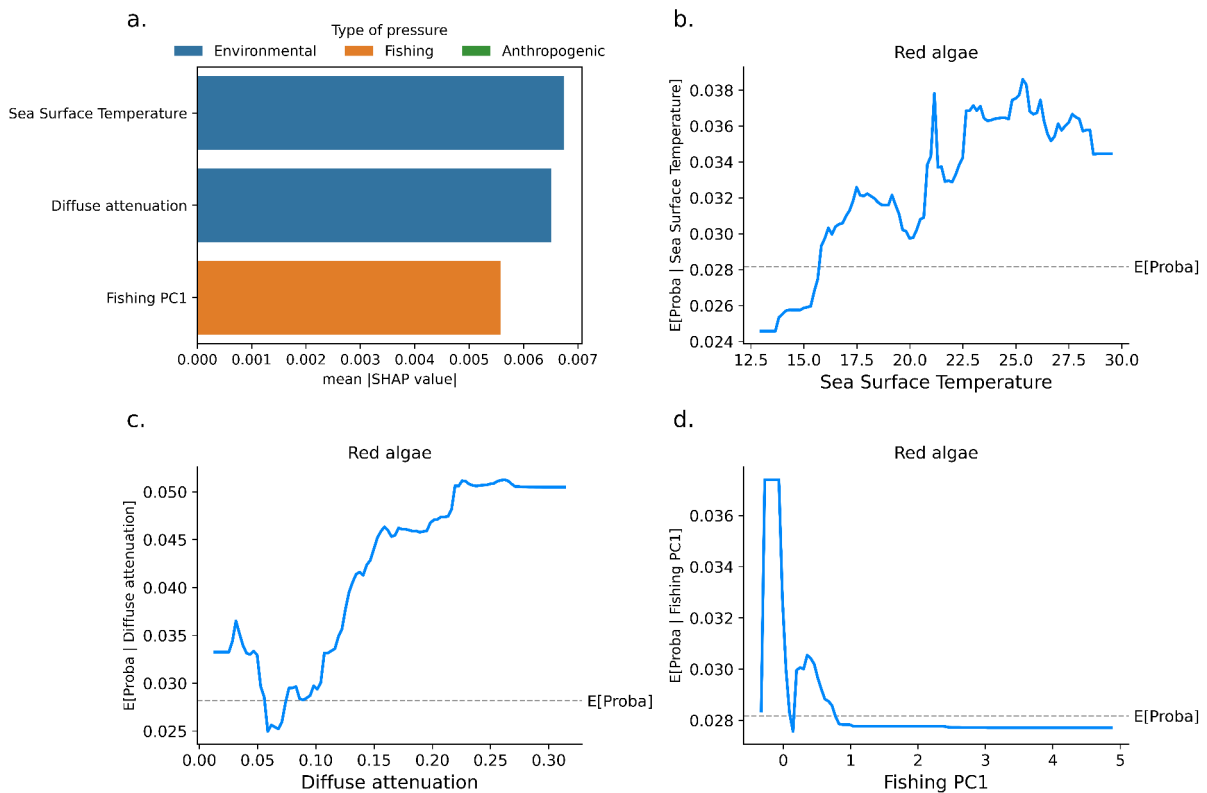


Figure S46 – a. Bar plot of the three most important variables according to the SHAP framework for Red algae habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

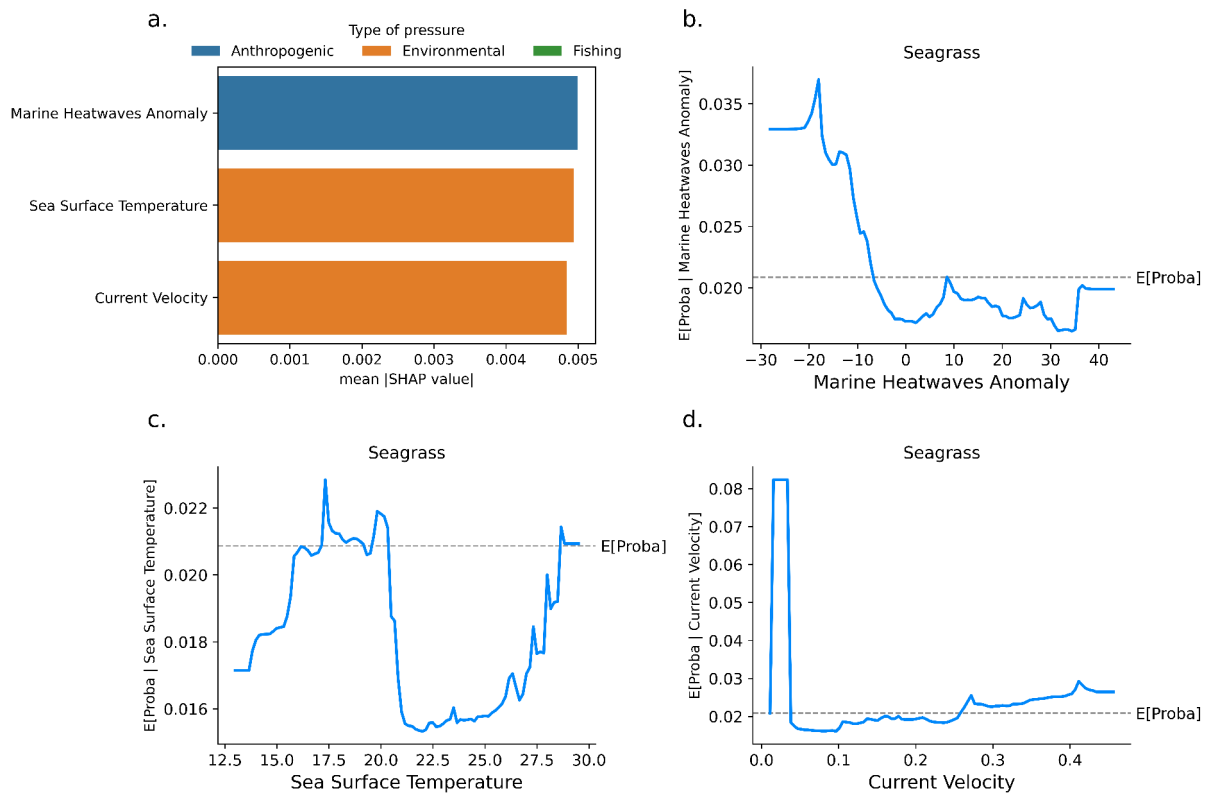


Figure S47 – a. Bar plot of the three most important variables according to the SHAP framework for Seagrass habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

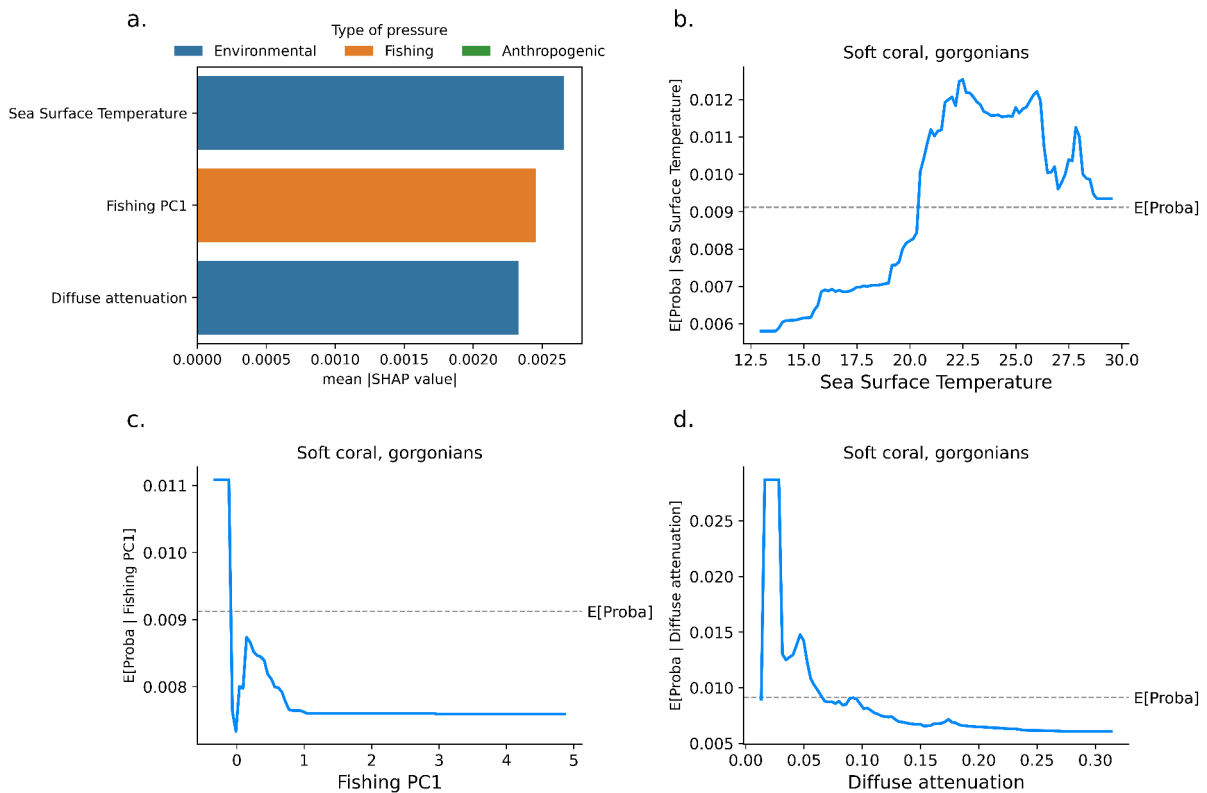


Figure S48 – a. Bar plot of the three most important variables according to the SHAP framework for Soft coral and gorgonians habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

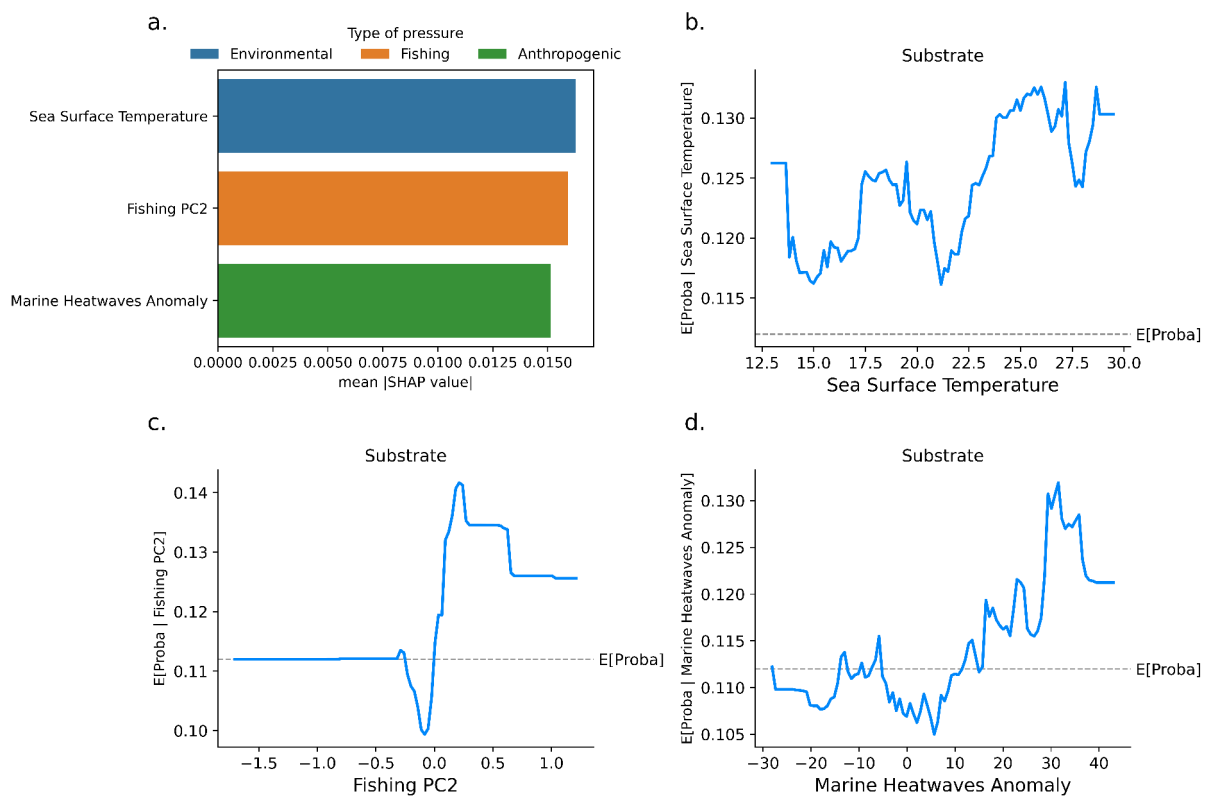


Figure S49 – a. Bar plot of the three most important variables according to the SHAP framework for Substrate habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

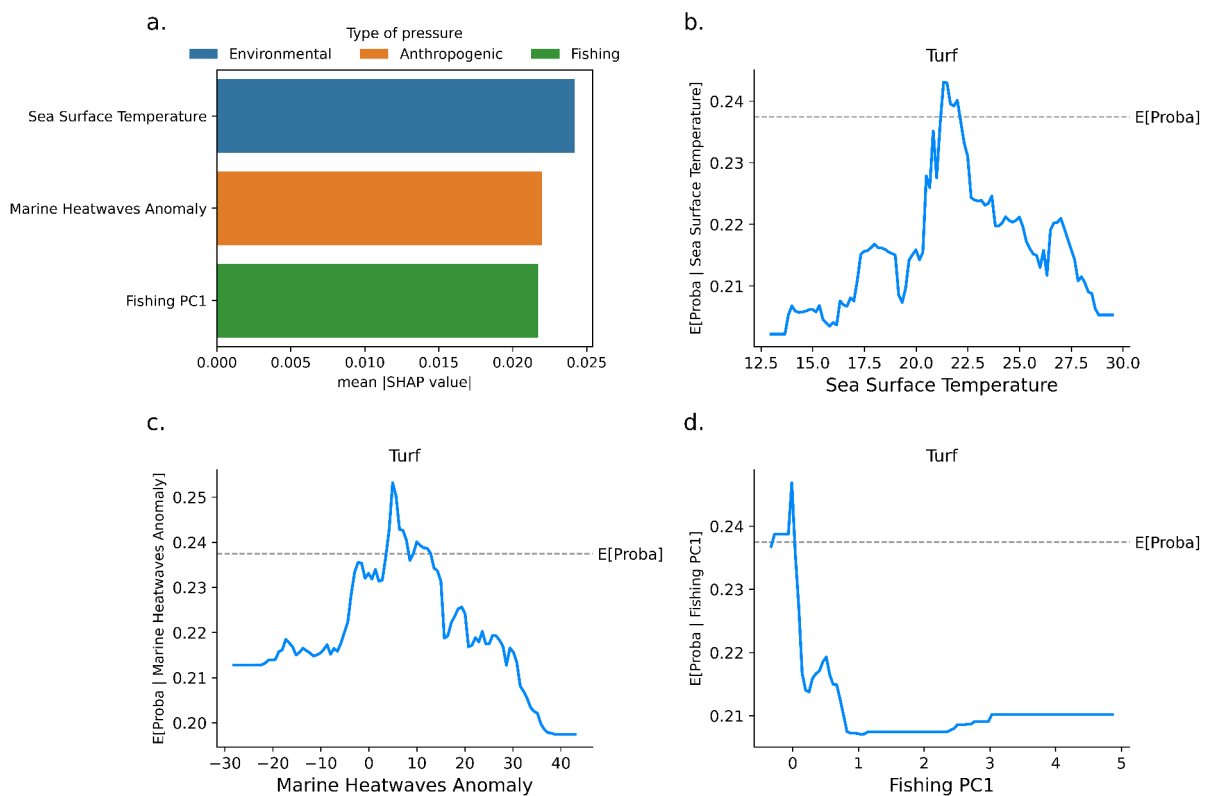


Figure S50 – a. Bar plot of the three most important variables according to the SHAP framework for Turf habitat state. The bars are coloured according to the type of pressure. b. to d. Partial Dependence plot of these three most important variables.

Bibliographie

- BOYÉ, A., (2019), *Diversité taxinomique et fonctionnelle des habitats benthiques dans l'espace et dans le temps : une perspective régionale et décennale* [thèse de doct.].
- BOYÉ, A., THIÉBAUT, É., GRALL, J., LEGENDRE, P., BROUDIN, C., HOUBIN, C., GARREC, V. L., MAGUER, M., DROUAL, G., & GAUTHIER, O., (2019), Trait-based approach to monitoring marine benthic data along 500 km of coastline, *Diversity and Distributions*, 2512, 1879-1896.
- BROOKS, S. P., & GELMAN, A., (1998), General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, 74, 434-455.
- CRESSWELL, A. K., EDGAR, G. J., STUART-SMITH, R. D., THOMSON, R. J., BARRETT, N. S., & JOHNSON, C. R., (2017), Translating local benthic community structure to national biogenic reef habitat types, *Global Ecology and Biogeography*, 2610, 1112-1125.
- LAZURE, P., & DUMAS, F., (2008), An external–internal mode coupling for a 3D hydrodynamical model for applications at regional scale (MARS), *Advances in Water Resources*, 312, 233-250.
- LECORNU, F., & ROECK, Y.-H. D., (2009), PREVIMER - Coastal observations and forecasts, *La Houille Blanche*, 951, 60-63.
- OVASKAINEN, O., & ABREGO, N., (2020), *Joint Species Distribution Modelling : With Applications in R*, Cambridge University Press.
- RIGAL, S., DEVICTOR, V., & DAKOS, V., (2020), A method for classifying and comparing non-linear trajectories of ecological variables, *Ecological Indicators*, 112, 106113.

Discussion Générale et Perspectives

L'objectif général de ce travail de thèse est d'améliorer notre compréhension des processus régissant la distribution des habitats benthiques biogéniques et de leurs communautés associées. Pour ce faire, dans le premier chapitre de ce manuscrit, j'ai étudié les règles d'assemblage dans une communauté benthique structurée par deux habitats distincts ; dans le second chapitre, j'ai identifié et caractérisé des états d'habitats biogéniques à l'échelle du globe et enfin, dans le dernier chapitre j'ai identifié les facteurs qui contrôlent la distribution spatiale de ces états d'habitats. Etant donné le rôle important de ces habitats pour maintenir la biodiversité des écosystèmes côtiers, ces travaux permettent de mieux envisager leur réponse face aux changements environnementaux en cours et les effets cascades qui peuvent en découler sur la biodiversité et le fonctionnement des écosystèmes côtiers.

5.1 Améliorer les Capacités Prédictives des Modèles de Distribution d'Espèces

L'un des principaux objectifs de cette thèse était de comprendre et de prédire les réponses des communautés benthiques côtières. À cet effet, j'ai employé deux méthodes distinctes de prédictions. Dans le premier chapitre, j'ai adopté l'approche "*prédire et agréger simultanément*" selon FERRIER et GUIBAN (2006) en utilisant un *jSDM* (WARTON et al., 2015). Dans le deuxième et troisième chapitre, j'ai employé une approche "*grouper, puis prédire*". Ces stratégies ont chacune montré leurs avantages et leurs limites. Les résultats du premier chapitre ont indiqué que l'intégration d'informations supplémentaires, comme la phylogénie ou d'autres espèces associées, renforçait la performance prédictive. Cependant, l'ajout de traits fonctionnels n'a pas consolidé les performances prédictives de

l’approche, allant au contraire de ce qui est attendu dans ce genre de modèle (OVASKAINEN et al., 2017 ; OVASKAINEN & ABREGO, 2020). Cela semble suggérer que les traits sélectionnés pour la communauté de polychètes n’étaient pas suffisamment en lien avec les conditions environnementales du modèle et que par conséquent, les liens traits-environnement ne permettaient pas d’informer suffisamment le modèle pour améliorer ses prédictions. Ainsi, savoir si l’on a utilisé les “bons” traits d’espèces reste encore une question cruciale pour obtenir le meilleur pouvoir prédictif d’un *SDM* ou *jSDM* (VESK et al., 2021).

Face à des résultats prédictifs qui ne répondaient pas à mes attentes et à une interprétabilité limitée par le nombre d’espèces en jeu, j’ai formulé une nouvelle méthode dans les chapitres 2 et 3. Au lieu de chercher à prédire la totalité de la communauté, j’ai considéré l’utilisation de proxys pour améliorer le pouvoir prédictif des modèles. J’ai ainsi choisi l’approche “*grouper, puis prédire*” de FERRIER et GUIBAN (2006) en me servant des habitats biogéniques et de leurs états possibles comme proxys des communautés benthiques et de la biodiversité côtière (PELLETIER et al., 2020). L’étendue des habitats biogéniques est depuis longtemps utilisée comme proxy de l’état des écosystèmes marins, à l’échelle globale (BUTCHART et al., 2010), mais leur étendue seule n’est pas suffisante, car, selon leur état, ces habitats n’abritent pas le même niveau de diversité (SUNDAY et al., 2017). J’ai donc exploré la prédictibilité et la réponse aux changements environnementaux de ces habitats biogéniques et de leurs différents états possibles. Cependant, cette tentative n’a pas non plus rencontré le succès escompté. Au-delà des questions soulevées par le chapitre 2 quant aux méthodes pour définir et classifier les habitats benthiques et leurs états, le chapitre 3 a montré que la prédictibilité de ces états d’habitats reste faible. Le chapitre 3 discute des facteurs qui peuvent limiter la prédictibilité d’états écologiques, mais il convient de revenir sur des points transverses avec le chapitre 1, car il existe des facteurs qui limitent la prédictibilité et transférabilité de tous modèles de distribution d’espèces (YATES et al., 2018).

Prédire en écologie s’avère complexe, notamment lors de l’étude de communautés entières ou à de vastes échelles spatio-temporelles. Néanmoins, cette thèse a mis en lumière diverses pistes de réflexion pour optimiser les prédictions. Une préoccupation majeure est la résolution des données environnementales relativement à l’écologie des espèces modélisées. En effet, les données environnementales utilisées comme co-variables sont souvent disponibles à des échelles grossières et trop larges relativement à la taille des organismes étudiés (POTTER et al., 2013). Par exemple, les données environnementales terrestres béné-

ficient d'une résolution d'environ 1 km, à l'instar de la base de données *Chelsea* (KARGER et al., 2017), cette résolution n'est que d'environ 9 km pour Bio-Oracle en milieu marin (ASSIS et al., 2018), alors que l'ordre de grandeur des espèces associées aux écosystèmes étudiées dans cette thèse ont des ordres de grandeur inférieurs au mètre. Cette discordance limite les capacités prédictives des modèles en réduisant la variabilité spatiale des conditions environnementales qu'expérimentent réellement ces espèces. Ainsi, lors de l'entraînement des modèles de distribution d'espèces, nous n'avons accès qu'à une partie tronquée de leurs niches réalisées. Une piste prometteuse serait d'intégrer les conditions microclimatiques qui expérimentent les communautés. Pour ce faire, il est possible d'envisager des loggers *in situ* (LEMBRECHTS et al., 2019), mais cette solution n'est pas réalisable pour les suivis à grande échelle spatiale. Des approches numériques pourraient améliorer notre appréhension des variations climatiques à une échelle spatiale bien plus fine : à l'aide d'approches numériques, via des approches mécanistiques telles que celles présentées par MACLEAN et al. (2019) ou des approches par réseaux de neurones telles que celles présentées par HU et al. (2023), il est possible de reconstruire les conditions climatiques à fine échelle, de l'ordre de 100 m², voir jusqu'à 10 m² (MACLEAN, 2020). Ces approches ont déjà été testées sur certains écosystèmes terrestres (STARK & FRIDLEY, 2022) et semblent une piste prometteuse pour améliorer nos capacités de prédictions (LEMBRECHTS & LENOIR, 2020) et pourraient être appliquées aux deux approches de modélisation mise en oeuvre dans cette thèse.

Pour l'approche "*grouper, puis prédire*", plusieurs méthodes peuvent optimiser le pouvoir prédictif. Une stratégie efficace consisterait à réviser la création des états d'habitats. Les travaux de cette thèse ont démontré que ces états sont actuellement basés sur des morphotypes. Cependant, cette classification regroupe parfois des espèces aux niches environnementales distinctes — par exemple, certaines laminaires des zones tempérées froides et des sargasses tropicales — dans une même catégorie, malgré des niches écologiques clairement séparées, notamment en termes de température de l'eau. Cette généralisation élargit artificiellement les niches environnementales, conduisant à des prédictions inexactes. Pour affiner cette classification, il serait pertinent d'intégrer des informations sur les communautés associées, en s'éloignant de la simple identification basée sur la couverture habitat pour prendre en compte les assemblages benthiques, comme l'a proposé DONOVAN et al. (2018). Cette approche d'état couvrant différents compartiments de l'écosystème se rapprocherait ainsi de la définition de régimes écologiques (DONOVAN et al., 2018). La méthode *UMAP*, utilisée dans cette thèse, est particulièrement adaptée à cette tâche. En effet, *UMAP*

génère une représentation topologique avant de procéder à la réduction dimensionnelle (McINNES et al., 2020). En combinant plusieurs représentations *UMAP* d'un même objet — dans ce cas, une représentation de l'écosystème basée sur les couvertures d'habitats et une basée sur la faune associée (p. ex. abondances spécifiques ou diversité fonctionnelle) —, il est possible de réaliser des opérations d'union ou d'intersection des graphes associés pour ne produire qu'une seule et même représentation en faible dimension représentant les principales variations observées lors de la combinaison de l'habitat et la faune associée (McINNES et al., 2018). Combinée à l'algorithme de groupement *HDBSCAN*, cela permettrait de découvrir des groupes plus “fins” en termes de représentation écologique, comme pour une même configuration de couverture d'habitat qui abriterait pour autant une faune très différente. Cela permettrait par exemple d'illustrer les récifs coralliens dont le fonctionnement est très différent, car la diversité fonctionnelle de poissons associés y est tout autre (BELLWOOD et al., 2019), ce qui ne ressort pas dans nos états basés seulement sur les couvertures d'habitats. Bien que cette nouvelle représentation ne simplifie pas nécessairement la modélisation dans l'ensemble, les régimes alors identifiés auraient probablement un lien plus étroit avec les conditions environnementales ainsi que des niches environnementales mieux séparées, ce qui améliorerait la modélisation d'un plus grand nombre de régimes.

Une étude complémentaire qui prendrait en compte les traits fonctionnels des communautés associées pourrait offrir une perspective enrichissante pour mieux comprendre leur état de santé (VILLÉGER et al., 2010) et définir des régimes écologiques incorporant les deux aspects de l'écosystème. Dans cette optique, les travaux de DONOVAN et al. (2018) représentent un cas d'étude pertinent. En se concentrant sur les coralliens de l'archipel d'Hawaï, ces travaux ont mis en lumière plusieurs régimes écologiques distincts dont trois sur les cinq nous intéressent précisément dans le cadre de cet exemple. Certains récifs présentent des caractéristiques non dégradées (c.-à-d. caractérisées par une forte couverture en corail et forte biomasse de poisson), d'autres sont dans un état intermédiaire (c.-à-d. forte couverture de corail, biomasse de poisson moyenne ou couverture moyenne de corail et forte biomasse de poisson) et la troisième catégorie est celle des récifs dégradés (c.-à-d. caractérisées par une forte couverture en algue gazonnante et une faible biomasse de poisson) (DONOVAN et al., 2018). Ainsi, même dans le scénario le moins favorable où les capacités prédictives resteraient mauvaises avec une plus fine granularité de régime écologique, cette approche *UMAP-HDBSCAN* nous offre le cadre idéal pour (1) caractériser les différents régimes écologiques à l'échelle du globe (2) identifier les états où

l'habitat est dégradé, mais les communautés associées intactes et vice-versa et enfin (3) identifier au long terme les transitions entre les différents régimes écologiques. Le point 2 est particulièrement intéressant, car il nous permettrait d'identifier les compartiments de la biocénose qui pourraient être des candidats à l'identification de signaux précoces de changements de régime.

5.2 Modélisation des Niches Environnementales : Entre Attentes et Réalités

Les chapitres 1 et 3 grâce à l'utilisation de méthodes de distribution d'espèces nous ont permis d'estimer une portion de la niche réalisée des objets d'études : les espèces de polychètes associées à deux habitats dans le cas du chapitre 1 et les états d'habitats dans le chapitre 3. Dans les deux approches, les niches estimées étaient cohérentes avec notre connaissance de ces écosystèmes. Dans l'approche de modélisation par *jSDM* mis en place dans le chapitre 1, nous nous sommes particulièrement intéressés à la variabilité des niches environnementales estimées par les différentes configurations du modèle. Contrairement à nos hypothèses, la majorité du signal que les différentes formulations des modèles arrivaient à capturer ne provenait pas des variables environnementales incluses dans les modèles, puisqu'en étudiant la forme des courbes de réponse, je me suis rendu compte que parmi les neuf formes de courbes de réponses possibles décrites par RIGAL et al. (2020), la majorité de celles que j'observais était plate. Une grande partie de la variance expliquée par les différents modèles provenait en réalité des effets aléatoires. Or parmi les effets aléatoires, l'importance des herbiers de zostères restait faible, notamment comparée aux variations spatiales (Fig. 5.1). Ce manque d'effet structurant des herbiers de zostères sur les communautés benthiques semble surprenant au regard de la littérature (BOSTRÖM et al., 2006). Cet effet pourrait être expliqué par l'importance dans cette zone des effets neutres dans la structuration des communautés de polychètes (BOYÉ et al., 2019). Une autre raison pour le manque d'effet des variables environnementales sur les espèces est la raison citée ci-dessus, à savoir le manque de résolution des variables environnementales pour caractériser la niche écologique des espèces modélisées. Le chapitre 3 quant à lui a permis une bonne caractérisation des niches des états si l'on se fie au bon pouvoir explicatif du modèle. Cela pourrait s'expliquer grâce (1) à l'importante couverture spatiale des données du *RLS* et (2) à la force des relations non linéaires que peut apprendre une méthode non paramétrique comme les *Random Forest* (VAYSSIÈRES et al., 2000 ; VALAVI et

al., 2021). Cependant, le pouvoir prédictif limité et la capacité à bien prédire les absences laissent supposer que là encore les niches environnementales estimées sont tronquées, ce qui peut expliquer pourquoi le modèle prédisait mieux les absences que les présences.

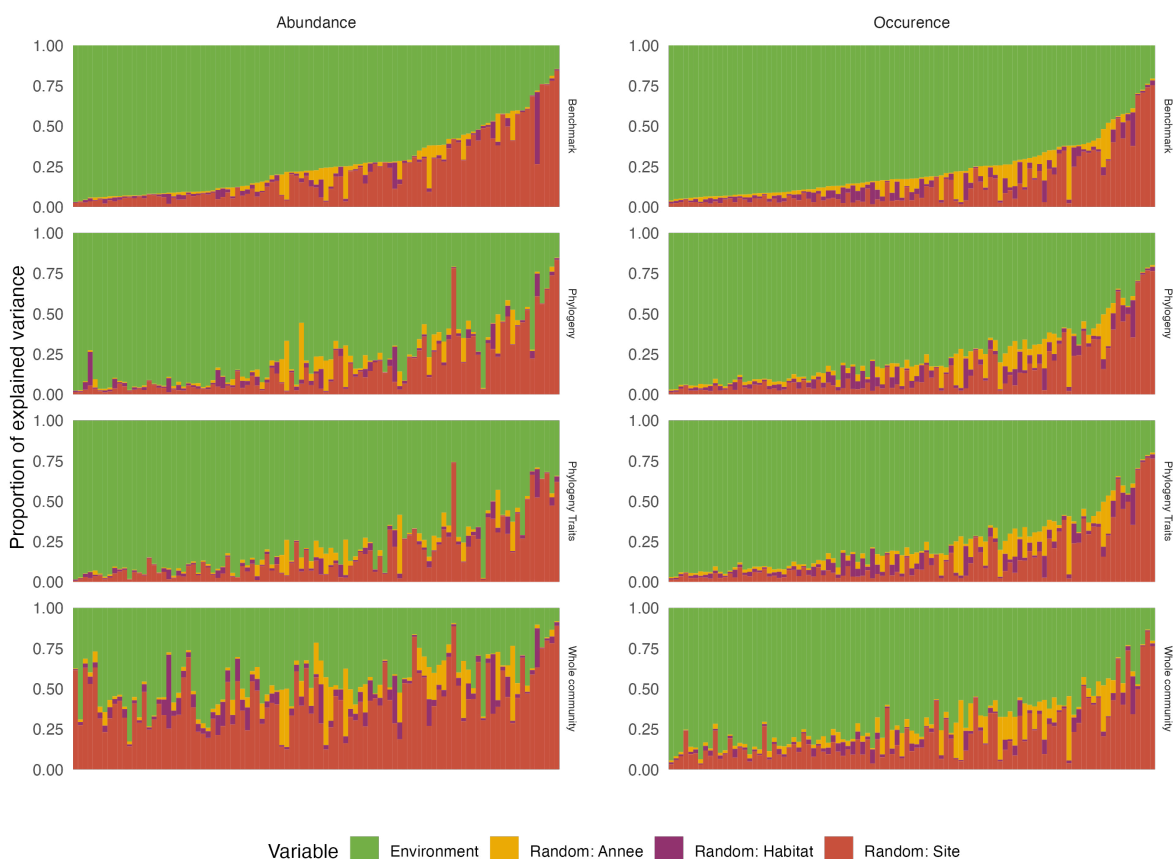


Figure 5.1 – *Comparaison entre les quatre structures de modèle alternatives étudiées dans le premier chapitre de cette thèse. Le graphique présente pour chaque espèce (le long de l'axe des x) la variance qui est expliquée (le long de l'axe y) par (1) les variables environnementales (Environnement) et (2) les trois effets aléatoires. Les résultats sont présentés pour les modèles ajustés aux données d'abondance (à gauche) et de présence/absence (à droite). Les espèces sont classées par ordre décroissant de variance expliquée par l'environnement pour le modèle de référence.*

5.3 Analyse des états d'habitats marins à l'échelle du globe : avancées et perspectives

L'identification des états d'habitats à l'échelle du globe constitue une avancée en écologie en nous permettant de mieux caractériser les habitats à ces grandes échelles et nous donnant une indication sur leur état de santé. Cependant, l'approche mise en place dans cette thèse n'éclaire pas sur l'état de santé des communautés que ces états d'habitats abritent. En exploitant une approche de groupement, *HDBSCAN* (MCINNES et al., 2017), associée à une nouvelle technique de réduction de dimension *UMAP* (MCINNES et al., 2020), j'ai discerné plusieurs ensembles distincts, que j'ai définis comme des "états d'habitat". Ces états d'habitats se distinguent par leur composition variée en espèces végétales (fucoides, laminaires, algues brunes, rouges, etc.) et animales (invertébrés sessiles, coraux branchus, coraux mous, gorgones...), ainsi que par la présence de divers substrats nus (comme le sable, le gravier ou la roche nue). Ces états d'habitats apportent un éclairage nuancé sur l'état de santé des zones dans lesquelles on les trouve, car, parmi ces ensembles, certains correspondent à des habitats emblématiques (tels que les coraux branchus ou les forêts de laminaires), tandis que d'autres représentent des états dégradés ou des états alternatifs (Fig. 3.7). Nos recherches, élaborées dans les chapitres 2 et 3, révèlent que les régions de transition entre les zones tempérées et tropicales, en particulier le long des côtes est et ouest de l'Australie, exhibent une plus forte richesse d'états d'habitats. Ces zones, actuellement en cours de tropicalisation (processus où les espèces typiques des eaux tempérées cèdent la place à des espèces tropicales; VERGÉS et al. (2014); VERGÉS et al. (2019)), pourraient être à l'aube d'un changement de régime. Ces changements de régime sont probablement en train de se produire actuellement, puisque nos observations indiquent une homogénéisation progressive des fonds marins, caractérisée par la prédominance croissante de l'état d'habitat "turf" depuis le milieu des années 2010, un phénomène qui corrobore les études antérieures menées sur les côtes australiennes (PESSARRODONA et al., 2022), et les tendances en cours à échelle globale (FILBEE-DEXTER & WERNBERG, 2018).

5.4 Implications des travaux entrepris dans cette thèse

Les travaux réalisés au cours de cette thèse ont non seulement marqué certaines avancées méthodologiques, mais ont également enrichi notre compréhension des habitats benthiques

à l'ère de l'Anthropocène. Dans le premier chapitre, j'ai élaboré un cadre d'analyse adaptée aux modèles *SDM* et *jSDM*, facilitant et standardisant l'analyse de ces modèles et de leur capacité à inférer les déterminants et à prédire les multiples facettes de la biodiversité. Appliqué à un jeu de données de suivi long-terme de la macrofaune au sein de deux habitats marins côtiers à une échelle régionale, cette analyse a clarifié les mécanismes d'assemblages de ces communautés. Les différents modèles réalisés ont confirmé une influence limitée des filtres environnementaux sur ces communautés (peu d'influence des variables évaluées, peu de lien avec les traits des espèces) et mis en lumière pour la première fois dans ces habitats un rôle important des co-occurrences entre espèces, qu'il conviendra d'examiner plus en détail. Cette analyse a aussi mis en avant les limites de ce cadre de modélisation pour des communautés riches, dont les principaux déterminants ne semblent pas être les filtres abiotiques, mais des processus neutres (BOYÉ et al., 2019) ou des processus biotiques (interactions entre espèces). Poursuivant cet élan, les chapitres suivants ont introduit des méthodes novatrices de réduction de dimension des données et de groupement pour analyser les données en écologie et décrypter la complexité des habitats, notamment en étudiant leurs patrons de distribution à l'échelle du globe. Cette exploration a levé le voile sur les facteurs environnementaux et anthropiques qui influencent la distribution des habitats et de leurs états, fournissant ainsi des clés pour interpréter et projeter les transformations futures de ces habitats biogéniques dans un contexte d'environnement changeant dû à l'anthropocène. Les résultats de ces recherches indiquent notamment une homogénéisation des habitats avec un accroissement de la couverture du "turf".

En somme, cette thèse souligne le rôle crucial des habitats biogéniques pour les écosystèmes marins et fournit de nouveaux outils pour mieux comprendre et prédire à la fois leur réponse aux changements environnementaux et les impacts en cascade que cela peut générer sur la faune qui leur est associée. Considérant les services essentiels que ces systèmes fournissent à l'humanité, ainsi que la valeur écologique intrinsèque des habitats biogéniques, les résultats de ces recherches mettent en relief la dégradation et la disparition de ces habitats biogéniques comme une très forte menace pesant sur les écosystèmes (IPBES, 2019). En réponse, l'urgence de programmes de surveillance étendus, tant dans l'espace que dans le temps, est indéniable pour mieux comprendre le fonctionnement des écosystèmes marins et l'importance des habitats sur la structure des communautés, notamment pour mieux appréhender les perturbations écologiques futures liées à l'impact anthropique.

Bibliographie

- 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*, 273, 277-284.
- BELLWOOD, D. R., STREIT, R. P., BRANDL, S. J., & TEBBETT, S. B., (2019), The meaning of the term ‘function’ in ecology : A coral reef perspective, *Functional Ecology*, 336, 948-961.
- BOSTRÖM, C., JACKSON, E. L., & SIMENSTAD, C. A., (2006), Seagrass landscapes and their effects on associated fauna : A review, *Estuarine, Coastal and Shelf Science*, 683, 383-403.
- BOYÉ, A., THIÉBAUT, É., GRALL, J., LEGENDRE, P., BROUDIN, C., HOUBIN, C., GARREC, V. L., MAGUER, M., DROUAL, G., & GAUTHIER, O., (2019), Trait-based approach to monitoring marine benthic data along 500 km of coastline, *Diversity and Distributions*, 2512, 1879-1896.
- BUTCHART, S. H. M., WALPOLE, M., COLLEN, B., van STRIEN, A., SCHARLEMANN, J. P. W., ALMOND, R. E. A., BAILLIE, J. E. M., BOMHARD, B., BROWN, C., BRUNO, J., CARPENTER, K. E., CARR, G. M., CHANSON, J., CHENERY, A. M., CSIRKE, J., DAVIDSON, N. C., DENTENER, F., FOSTER, M., GALLI, A., ... WATSON, R., (2010), Global Biodiversity : Indicators of Recent Declines, *Science*, 3285982, 1164-1168.
- DONOVAN, M. K., FRIEDLANDER, A. M., LECKY, J., JOUFFRAY, J.-B., WILLIAMS, G. J., WEDDING, L. M., CROWDER, L. B., ERICKSON, A. L., GRAHAM, N. A. J., GOVE, J. M., KAPPEL, C. V., KARR, K., KITTINGER, J. N., NORSTRÖM, A. V., NYSTRÖM, M., OLESON, K. L. L., STAMOULIS, K. A., WHITE, C., WILLIAMS, I. D., & SELKOE, K. A., (2018), Combining fish and benthic communities into multiple regimes reveals complex reef dynamics, *Scientific Reports*, 81, 16943.
- FERRIER, S., & GUISAN, A., (2006), Spatial modelling of biodiversity at the community level, *Journal of Applied Ecology*, 433, 393-404.
- FILBEE-DEXTER, K., & WERNBERG, T., (2018), Rise of Turfs : A New Battlefront for Globally Declining Kelp Forests, *BioScience*, 682, 64-76.
- HU, Y., MA, L., ZHANG, Y., WU, Z., WU, J., ZHANG, J., & ZHANG, X., (2023), Research on High-Resolution Reconstruction of Marine Environmental Parameters Using Deep Learning Model, *Remote Sensing*, 1513.
- IPBES, (2019), *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (E. BRONDIZIO, S. DIAZ, J. SETTELE & H. T. a. NGO, Éd.).
- KARGER, D. N., CONRAD, O., BÖHNER, J., KAWOHL, T., KREFT, H., SORIA-AUZA, R. W., ZIMMERMANN, N. E., LINDER, H. P., & KESSLER, M., (2017), Climatologies at high resolution for the earth’s land surface areas, *Scientific Data*, 41, 170122.
- LEMBRECHTS, J. J., & LENOIR, J., (2020), Microclimatic conditions anywhere at any time!, *Global Change Biology*, 262, 337-339.
- LEMBRECHTS, J. J., NIJS, I., & LENOIR, J., (2019), Incorporating microclimate into species distribution models, *Ecography*, 427, 1267-1279.
- MACLEAN, I. M. D., (2020), Predicting future climate at high spatial and temporal resolution, *Global Change Biology*, 262, 1003-1011.

- MACLEAN, I. M. D., MOSEDALE, J. R., & BENNIE, J. J., (2019), Microclima : An r package for modelling meso- and microclimate, *Methods in Ecology and Evolution*, 102, 280-290.
- MCINNES, L., HEALY, J., & ASTELS, S., (2017), hdbscan : Hierarchical density based clustering, *The Journal of Open Source Software*, 211.
- MCINNES, L., HEALY, J., & MELVILLE, J., (2020), UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction, *arXiv*.
- MCINNES, L., HEALY, J., SAUL, N., & GROSSBERGER, L., (2018), UMAP : Uniform Manifold Approximation and Projection, *The Journal of Open Source Software*, 329, 861.
- OVASKAINEN, O., & ABREGO, N., (2020), *Joint Species Distribution Modelling : With Applications in R*, Cambridge University Press.
- OVASKAINEN, O., TIKHONOV, G., NORBERG, A., GUILLAUME BLANCHET, F., DUAN, L., DUNSON, D., ROSLIN, T., & ABREGO, N., (2017), How to make more out of community data? A conceptual framework and its implementation as models and software, *Ecology Letters*, 205, 561-576.
- PELLETIER, D., SELMAOUI-FOLCHER, N., BOCKEL, T., & SCHOHN, T., (2020), A regionally scalable habitat typology for assessing benthic habitats and fish communities : Application to New Caledonia reefs and lagoons, *Ecology and Evolution*, 1014, 7021-7049.
- PESSARRODONA, A., VERGÉS, A., BOSCH, N. E., BELL, S., SMITH, S., SGARLATTÀ, M. P., & WERNBERG, T., (2022), Tropicalization unlocks novel trophic pathways and enhances secondary productivity in temperate reefs, *Functional Ecology*, 363, 659-673.
- POTTER, K. A., ARTHUR WOODS, H., & PINCEBOURDE, S., (2013), Microclimatic challenges in global change biology, *Global Change Biology*, 1910, 2932-2939.
- RIGAL, S., DEVICTOR, V., & DAKOS, V., (2020), A method for classifying and comparing non-linear trajectories of ecological variables, *Ecological Indicators*, 112, 106113.
- STARK, J. R., & FRIDLEY, J. D., (2022), Microclimate-based species distribution models in complex forested terrain indicate widespread cryptic refugia under climate change, *Global Ecology and Biogeography*, 313, 562-575.
- SUNDAY, J. M., FABRICIUS, K. E., KROEKER, K. J., ANDERSON, K. M., BROWN, N. E., BARRY, J. P., CONNELL, S. D., DUPONT, S., GAYLORD, B., HALL-SPENCER, J. M., KLINGER, T., MILAZZO, M., MUNDAY, P. L., RUSSELL, B. D., SANFORD, E., THIYAGARAJAN, V., VAUGHAN, M. L. H., WIDDICOMBE, S., & HARLEY, C. D. G., (2017), Ocean acidification can mediate biodiversity shifts by changing biogenic habitat, *Nature Climate Change*, 71, 81-85.
- VALAVI, R., ELITH, J., LAHOZ-MONFORT, J. J., & GUILLERA-ARROITA, G., (2021), Modelling species presence-only data with random forests, *Ecography*, 4412, 1731-1742.
- VAYSSIÈRES, M. P., PLANT, R. E., & ALLEN-DIAZ, B. H., (2000), Classification trees : An alternative non-parametric approach for predicting species distributions, *Journal of Vegetation Science*, 115, 679-694.
- VERGÉS, A., MCCOSKER, E., MAYER-PINTO, M., COLEMAN, M. A., WERNBERG, T., AINSWORTH, T., & STEINBERG, P. D., (2019), Tropicalisation of temperate reefs : Implications for ecosystem functions and management actions, *Functional Ecology*, 336, 1000-1013.

-
- VERGÉS, A., TOMAS, F., CEBRIAN, E., BALLESTEROS, E., KIZILKAYA, Z., DENDRINOS, P., KARAMANLIDIS, A. A., SPIEGEL, D., & SALA, E., (2014), Tropical rabbitfish and the deforestation of a warming temperate sea, *Journal of Ecology*, 1026, 1518-1527.
- VESK, P. A., MORRIS, W. K., NEAL, W. C., MOKANY, K., & POLLOCK, L. J., (2021), Transferability of trait-based species distribution models, *Ecography*, 441, 134-147.
- VILLÉGER, S., MIRANDA, J. R., HERNÁNDEZ, D. F., & MOUILLOT, D., (2010), Contrasting changes in taxonomic vs. functional diversity of tropical fish communities after habitat degradation, *Ecological Applications*, 206, 1512-1522.
- WARTON, D. I., BLANCHET, F. G., O'HARA, R. B., OVASKAINEN, O., TASKINEN, S., WALKER, S. C., & HUI, F. K., (2015), So many variables : joint modeling in community ecology, *Trends in Ecology & Evolution*, 3012, 766-779.
- YATES, K. L., BOUCHET, P. J., CALEY, M. J., MENGERSEN, K., RANDIN, C. F., PARNELL, S., FIELDING, A. H., BAMFORD, A. J., BAN, S., BARBOSA, A. M., DORMANN, C. F., ELITH, J., EMBLING, C. B., ERVIN, G. N., FISHER, R., GOULD, S., GRAF, R. F., GREGR, E. J., HALPIN, P. N., ... SEQUEIRA, A. M., (2018), Outstanding Challenges in the Transferability of Ecological Models, *Trends in Ecology & Evolution*, 3310, 790-802.

Titre : Approches quantitatives pour comprendre et prédire l'écologie, la distribution et la biodiversité des habitats benthiques dans l'Anthropocène

Mot clés : Ecologie des communautés, Ecologie numérique, Habitats benthiques, Modélisation, Anthropocène

Résumé : L'objectif de cette thèse est de mieux comprendre et prédire la biodiversité benthique et le rôle des habitats biogéniques, dans le maintien de la structure et des fonctions des écosystèmes côtiers. Cette thèse a exploré différents outils numériques et des pipelines innovants et complémentaires pour répondre à ces objectifs à différentes échelles : 1) la modélisation jointe de la distribution des espèces dans deux habitats biogéniques à une échelle régionale et 2) la définition et la modélisation, via des approches de Machine Learning, de la distribution de l'état d'habitats benthiques à une échelle globale et nationale. Ces approches complémentaires contribuent à une meilleure quantification de l'influence re-

lative des facteurs environnementaux et anthropiques (notamment épisodes de canicules marines et pression de pêche) qui déterminent la biodiversité côtière et l'état des habitats benthiques. Si dans les deux cas d'étude, la prédictibilité des espèces considérées ou des états était faible, ces travaux ont mis en évidence des stratégies pour optimiser l'inférence et la prédiction des modèles explorés. Ainsi, cette thèse apporte un point de vue critique sur les approches permettant d'étudier et de caractériser la biodiversité côtière, et sur les développements nécessaires pour mieux anticiper les réponses écologiques futures liées aux impacts anthropiques.

Title: Quantitative approaches to understand and predict the ecology, distribution and biodiversity of benthic habitats in the Anthropocene

Keywords: Community ecology, Numerical ecology, Benthic habitats, Modelling, Anthropocene

Abstract: This thesis aims at better understanding and predicting coastal benthic biodiversity with a specific focus on the role of biogenic habitats in maintaining ecosystem structure and functioning. This thesis explored how different innovative and complementary numeric tools and pipelines can address these objectives at different scales: 1) joint species distribution modelling across two biogenic habitats at a regional scale, and 2) using Machine Learning approaches, defining and modelling the distribution of benthic habitats states at a global and at a national scale. These complementary approaches quantify the relative influence of the environmental and

anthropogenic factors (including marine heatwaves and fishing intensity) that determine coastal biodiversity and the state of benthic habitats. While in both case studies the predictability of the considered species or states was low, these studies have identified future avenues to optimise models inference and prediction of benthic communities. Thus, this thesis provides a critical perspective on existing approaches available to study and characterise coastal biodiversity; and on the future developments required to better anticipate future ecological responses related to anthropogenic impacts.