



HAL
open science

Aide à la décision pour la planification spatiale marine : l'optimisation en nombres entiers au service de la conception des aires marines protégées

Adrien Brunel

► To cite this version:

Adrien Brunel. Aide à la décision pour la planification spatiale marine : l'optimisation en nombres entiers au service de la conception des aires marines protégées. Océanographie. Université de Montpellier, 2022. Français. NNT : 2022UMONG086 . tel-04067190

HAL Id: tel-04067190

<https://theses.hal.science/tel-04067190>

Submitted on 13 Apr 2023

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 POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En : Sciences de la mer

École doctorale : GAIA

Unité de recherche : UMR MARBEC, IRD

Aide à la décision pour la planification spatiale
marine : l'optimisation en nombres entiers au
service de la conception des aires marines
protégées.

Présentée par Adrien BRUNEL

Le 12 décembre 2022

Sous la direction de Sophie LANCO BERTRAND et Jérémy
OMER

Devant le jury composé de :

Alain BILLIONNET, Professeur, ENSIE	Rapporteur
Patrick MEYER, Professeur, IMT Atlantique	Rapporteur
Sandrine VAZ, Cadre de recherche, IFREMER	Examinatrice
Sophie LANCO BERTRAND, Directrice de Recherche, IRD	Directrice de thèse
Jérémy OMER, Maître de conférence, INSA	Co-directeur de thèse
Sébastien THORIN, Responsable R&D, Créocéan	Co-encadrant de thèse



UNIVERSITÉ
DE MONTPELLIER

Contents

Document de synthèse en français	4
Abstract	18
1 Introduction	25
1.1 Context	26
1.1.1 Marine spatial planning	26
1.1.2 Decision support tools	26
1.1.3 Protected areas	27
1.2 Conservation science	28
1.2.1 Environmental ethics	28
1.2.2 Emergence of conservation within science and institutions	31
1.2.3 Missions and motives	32
1.2.4 Early works (1960-1980)	33
1.3 Operations research	36
1.3.1 General overview	36
1.3.2 Integer linear programming	37
1.3.3 Illustration examples	39
1.4 Systematic reserve site selection	40
1.4.1 History	40
1.4.2 General formulation	41
1.5 Study cases	43
1.5.1 Observed data	43
1.5.2 Generated data	46
1.6 Research questions	48
2 Opening the black box of decision support tools in marine spatial planning: analysis of reserve site selection algorithms	49
2.1 Introduction	50
2.2 Material and methods	51
2.2.1 Optimisation	51
2.2.2 Input data	52
2.3 Results	54
2.3.1 Sensitivity of reserve solutions to structural elements	54
2.3.2 Parameters influence	56
2.4 Discussion	60
2.4.1 Sensitivity of reserve solutions to structural elements	60
2.4.2 Input data	61
2.4.3 Parameters influence	62
2.5 Conclusion	64

3	Producing a diverse set of near-optimal reserve solutions with exact optimisation	67
3.1	Introduction	68
3.2	Methods	70
3.2.1	Measuring the dissimilarity between two reserve solutions	70
3.2.2	Computing a presentation set	71
3.2.3	Illustration of the approaches for computing a presentation set	73
3.3	Case study	75
3.3.1	Presentation set imposing an objective value interval	75
3.3.2	Presentation set imposing a dissimilarity measure	75
3.3.3	Presentation set maximising the dissimilarity measure	76
3.3.4	Optimality gap versus dissimilarity measure	78
3.4	Discussion	79
4	Risk-averse optimisation for reserve site selection with uncertain non-binary data	83
4.1	Introduction	84
4.2	Methods	87
4.2.1	Robust framework	87
4.2.2	Chance constraint framework	88
4.2.3	Conceptual diagrams	89
4.3	Results	90
4.3.1	Robust framework	90
4.3.2	Chance constraint framework	92
4.3.3	Comparison between the two approaches	92
4.4	Discussion	95
5	Explicit design of compact, connected and gap-free reserves	99
5.1	Introduction	100
5.2	Models	102
5.2.1	General formulation of the reserve site selection problem	102
5.2.2	Connectivity of the reserve	103
5.2.3	Gap-free reserve	104
5.2.4	Compactness of the reserve	104
5.2.5	Improvements of the model	105
5.3	Numerical experiments	108
5.3.1	Illustration on a real-world instance	108
5.3.2	Feasibility assessment on generated instances	111
5.3.3	Assessment of the compactness models	112
5.4	Discussion	117
6	Discussion	119
6.1	Synthesis of the developments	120
6.2	Findings and limitations	120
6.3	Perspectives	124
6.3.1	A different example of human use allocation: site selection of offshore wind farms	124
6.3.2	Toward a spatial zoning of human activities	125
6.4	Ethical positioning	127
	References	129

Document de synthèse en français

Introduction

La crise de la biodiversité et le changement climatique en cours ont conduit à un relatif consensus sur la nécessité de préserver l'intégrité et le fonctionnement des écosystèmes exploités. Bien que déjà au cœur de multiples pressions anthropiques (pêche, aquaculture, routes maritimes, exploitation des fonds marins, activités récréatives, énergies renouvelables et fossiles, *etc.*), les milieux marins sont aujourd'hui considérés comme des espaces privilégiés pour le déploiement de "l'économie bleue" de demain. Dans ce contexte, la planification spatiale marine (PSM) vise à organiser rationnellement l'utilisation de l'espace et des ressources marines afin de réduire les tensions entre les activités humaines et les écosystèmes marins, ainsi qu'entre les acteurs de l'océan. De ce fait, la PSM s'est largement diffusée au sein des institutions en charge de la gestion et est finalement devenue le cadre privilégié dans la perspective d'un développement durable. Fruit de mobilisations collectives, la PSM est par nature au carrefour de diverses disciplines (écologie, économie, sociologie, droit, *etc.*), ce qui en fait un cadre difficile à mettre en place. Parce qu'elle se veut rationnelle et fondée sur des preuves, la PSM utilise de nombreux outils d'aide à la décision tout au long de son processus afin d'éclairer la prise de décision de manière systématique et transparente. Il existe une profusion de tels outils (Marxan, PrioritizR, SeaSketch, Atlantis, Coastal Resilience, *etc.*). Ces outils sont plus ou moins ouverts, libres, complémentaires, poursuivent des objectifs différents, mobilisent des méthodes et des connaissances variées.

En général, la PSM cherche un découpage de l'océan afin d'allouer dans le temps et l'espace les activités humaines en mer en accord avec un développement durable. Un cas particulier de la PSM est l'allocation de l'espace marin à des fins de conservation. En d'autres termes, nous nous concentrons sur les outils et les méthodes de désignation des zones protégées, un sujet particulièrement d'actualité. Cela donne à la science de la conservation une place de choix dans les débats internationaux. En effet, les aires protégées sont au cœur des politiques internationales actuelles visant à atténuer l'érosion de la biodiversité. Par exemple, les membres (États, ONG, agences) de l'Union Internationale pour la Conservation de la Nature (UICN) ont convenu de viser un niveau de protection de 30% de la surface de chaque écorégion marine d'ici à 2030, contre moins de 8% aujourd'hui. Plus récemment, le Green Deal européen fixe un objectif de 30% de la surface des eaux européennes à couvrir par des aires marines protégées (AMP) d'ici à 2030. Les AMP apparaissent donc comme une part essentielle des solutions pour assurer la préservation des écosystèmes marins. La surface concernée est immense, tout comme les dommages potentiels aux écosystèmes ou les contraintes inutiles aux activités humaines. Par conséquent, même une petite limitation de ces outils ne peut être négligée en raison de l'amplitude de l'impact potentiel. L'objectif principal de ce travail est donc d'améliorer les outils existants impliqués dans la PSM car ces outils connaissent plusieurs limitations qui doivent être abordées. Notre travail cherche à ouvrir la boîte noire de ces outils et nous espérons apporter plus de transparence et d'équité dans les processus de PSM.

Les travaux menés dans cette thèse explorent spécifiquement trois axes d'amélioration :

1. Deux classes d'algorithmes peuvent résoudre le problème de sélection de sites de réserve : les algorithmes métaheuristiques (tels que le recuit simulé, couramment mis en œuvre dans Marxan) et l'optimisation exacte (programmation en nombres entiers, couramment mise en œuvre dans PrioritizR). Bien que les approches exactes soient désormais capables de résoudre des problèmes à grande échelle, les métaheuristiques restent largement utilisées. L'une des raisons est que les logiciels basés sur des métaheuristiques fournissent un ensemble de solutions de réserve sous-optimales au lieu d'une seule. Ces solutions alternatives sont généralement bien accueillies par les parties prenantes car

elles constituent une meilleure base de négociation entre des objectifs potentiellement conflictuels. Les algorithmes métaheuristiques utilisent des procédures aléatoires pour explorer l'espace des solutions de réserve sous-optimales. Par conséquent, ils peuvent produire une grande quantité de solutions alternatives similaires, donc peu informatives, ce qui nécessite généralement un post-traitement statistique lourd. Il manque des méthodes efficaces pour générer un ensemble diversifié de solutions proches de l'optimum à l'aide de l'optimisation exacte. Nous présentons ici deux nouvelles approches pour résoudre ce problème. Nos algorithmes contrôlent explicitement à la fois l'écart d'optimalité et la dissimilarité entre les solutions de réserve alternatives. Ils permettent l'identification d'un ensemble de solutions de réserve parcimonieux, mais significatif. Les algorithmes présentés ici pourraient augmenter l'adoption de l'optimisation exacte par les utilisateurs finaux. Ces méthodes devraient contribuer à des discussions moins confuses et plus transparentes dans la conception des politiques de conservation. Ceci est l'objet du Chapitre 3.

2. La plupart des outils d'aide à la décision largement utilisés (*e.g.* Marxan, PrioritizR) ne permettent généralement pas de prendre en compte explicitement l'incertitude. Or, ces incertitudes peuvent conduire à des solutions de réserve inefficace, à des dommages potentiellement irréversibles pour les écosystèmes et à des contraintes inutiles pour les parties prenantes. Les approches probabilistes ont été appliquées avec succès avec des données de présence/absence mais ce cadre est trop restrictif lorsque des données non-binaires sont disponibles. Nous proposons deux approches averse au risque incorporant une incertitude paramétrique dans les modèles de sélection de sites de réserve : (1) un cadre d'optimisation robuste dans lequel un nombre donné de paramètres est toléré de dévier d'une valeur nominale, (2) un problème d'optimisation avec contrainte en probabilité résolu en utilisant une approche d'approximation par échantillonnage. Pour les deux modèles, un paramètre de niveau de risque permet aux utilisateurs finaux d'établir le risque qu'ils sont prêts à prendre. Ce travail leur fournit des outils pour concevoir des solutions de réserve qui sont robustes à l'incertitude afin de faire face aux changements globaux actuels. Ceci est l'objet du Chapitre 4.
3. Pour des raisons écologiques, d'application pratique et de gestion, une réserve sans cohérence spatiale a peu de chances d'être mise en œuvre. La plupart des outils d'aide à la décision largement utilisés pour la sélection de sites de réserve (*e.g.* Marxan, PrioritizR) n'incluent qu'une contrainte spatiale grossière dans leurs modèles d'optimisation par le biais d'un paramètre de compacité. La compacité d'une solution est appliquée empiriquement en pénalisant le périmètre global dans la fonction objectif. Plusieurs modèles d'optimisation ont tenté de surmonter cette difficulté et de tenir compte d'une propriété spatiale donnée : fragmentation limitée, connectivité des sites sélectionnés, zones tampons autour des sites sélectionnés, *etc.* Mais, à ce jour, il n'existe pas de modèle de sélection de sites de réserve qui garantisse la production d'une réserve connectée, compacte et sans trous. Nous proposons un programme linéaire en nombres entiers utilisant la théorie des graphes pour construire explicitement une réserve qui soit connectée, compacte et sans trous. Pour ce faire, nous utilisons une approche de flots multi-commodités qui incorpore ces attributs spatiaux dans un modèle de sélection de sites de réserve. Nous avons testé la faisabilité numérique de notre modèle sur des instances générées et sur le cas réel de Fernando de Noronha. Nos résultats montrent que nous pouvons faire respecter la compacité, la connectivité et l'absence de trous en utilisant un seul modèle. Ce travail fournit aux utilisateurs finaux un modèle d'optimisation pour concevoir des solutions de réserve présentant des propriétés spatiales souhaitables, ce qui augmente leurs chances d'être mises en œuvre. Ceci est l'objet du Chapitre 5.

Bien que la sélection de sites de réserve puisse sembler être un aspect très spécifique de la PSM,

comprendre et améliorer ces outils d'aide à la décision est une étape nécessaire avant d'aborder la question plus globale de la PSM : comment trouver un zonage spatial équitable et durable des activités humaines ? En permettant de comprendre précisément pourquoi et comment le résultat a été obtenu, les axes présentés ici devraient, nous l'espérons, ouvrir la boîte noire des outils d'aide à la décision et finalement contribuer à un meilleur processus de planification.

L'étude de cas utilisée comme illustration numérique des méthodes développées dans la thèse est l'archipel brésilien de Fernando de Noronha dans l'Atlantique tropical. Pour des raisons méthodologiques, nous avons également utilisé un jeu de données généré.

Produire un ensemble de solutions de réserve différente et proche de l'optimalité par l'optimisation exacte

Un ensemble de solutions de réserve alternatives est généralement nécessaire pour créer des réserves naturelles efficaces, car elles favorisent de meilleures négociations entre les différentes parties prenantes. Dans un cadre décisionnel, les contraintes et les objectifs liés à la conception d'une réserve peuvent être difficiles à exprimer dans un formalisme d'optimisation numérique spatialement explicite. Tout phénomène non modélisé susceptible d'influencer la décision peut entraîner des divergences avec les solutions proposées par l'optimisation. Par exemple, des mécanismes sociaux complexes régissent la décision finale d'implantation des parcs éoliens (Bell et al. 2005; Virtanen et al. 2022). Le processus de décision, qui repose finalement sur des négociations, nécessite donc une certaine latitude sur les solutions possibles à envisager. La génération de solutions alternatives donne aux utilisateurs la possibilité de trouver une solution qui pourrait être plus satisfaisante par rapport à ces objectifs non modélisés. Pour ces raisons, la capacité des outils d'aide à la décision à produire un éventail de solutions, au lieu d'une seule, est fréquemment mise en avant dans la littérature (Pressey et al. 1993; Possingham et al. 2000; Possingham et al. 2006; Sarkar 2012; Ferretti et al. 2019). Par conséquent, les outils de sélection de sites de réserve ont besoin d'options permettant de générer différentes alternatives proches de l'optimalité.

La capacité à produire des solutions alternatives a souvent été présentée comme un atout majeur des algorithmes métaheuristiques par rapport aux approches d'optimisation exactes (Pressey et al. 1993; Possingham et al. 2000; Possingham et al. 2006; Sarkar 2012). Dans un sondage réalisé auprès des utilisateurs de Marxan (Ardron et al. 2010), "la génération de solutions multiples était de loin le point fort le plus souvent relevé de Marxan" par rapport aux autres algorithmes de sélection de sites de réserve. Pourtant, des avancées récentes ont rendu les méthodes d'optimisation exactes plus attrayantes pour les utilisateurs (Schuster et al. 2020; Hanson et al. 2019; Beyer et al. 2016; Rodrigues and Gaston 2002) car elles fournissent la solution optimale en un temps raisonnable, même pour des grandes instances. La capacité à générer de multiples solutions semble donc être le dernier argument restant en faveur des algorithmes métaheuristiques. Techniquement parlant, les métaheuristiques s'appuient sur des processus aléatoires pour créer une diversité implicite au sein de l'ensemble des solutions (voir l'annexe B-2.1 dans (Serra-Sogas et al. 2020)). Au contraire, les méthodes de résolution exactes produisent généralement, par construction, une seule solution optimale et ne sont pas conçues pour produire un éventail de solutions différentes. Cette limitation majeure restreint considérablement la capacité des méthodes de résolution exacte à informer les problèmes de conservation du monde réel. Cependant, en l'absence de critères explicites, les approches métaheuristiques ne fournissent aucun contrôle sur les alternatives générées et ne garantissent pas non plus l'obtention de solutions réellement différentes. La recherche de solutions alternatives proches de l'optimalité avec des méthodes de résolution exactes a commencé à être discutée

dans (Fischer and Church 2005) et a été explorée dans le contexte général de la recherche opérationnelle (Chang et al. 1982a; Brill et al. 1982; Chang et al. 1982b; Makowski et al. 2000). L'algorithme développé dans (Arthur et al. 1997) calcule l'ensemble exhaustif des solutions optimales d'un problème de sélection de sites de réserve. Dans le même esprit, un algorithme de sélection par séparation et évaluation a montré comment des solutions sous-optimales peuvent être déduites avec des méthodes exactes (Önal 2004). L'outil de sélection de sites de réserve *PrioritizR* offre également des fonctions supplémentaires permettant aux utilisateurs de construire un portefeuille de solutions alternatives¹. Cependant, la production de solutions alternatives basées sur leur distance à l'optimalité avec des méthodes de résolution exactes ne garantit toujours pas l'obtention de solutions différentes. Cela nous a motivé à introduire explicitement une mesure de dissimilarité dans la recherche de solutions alternatives.

Dans ce travail, nous proposons deux algorithmes itératifs incorporant un critère de dissimilarité explicite pour construire un ensemble de solutions proche de l'optimalité significativement différentes les unes des autres avec des méthodes de résolution exactes. Les solutions sont sélectionnées en fonction d'une dégradation contrôlée de la fonction objectif et en utilisant une mesure de dissimilarité explicitement formulée. Nous avons observé que l'utilisation de la distance naturelle comme mesure de dissimilarité conduit à des solutions alternatives qui incluent strictement la solution optimale. En considérant qu'il ne s'agissait pas d'une alternative valable mais seulement d'une solution dégradée, notre mesure de dissimilarité a permis de discriminer de tels cas, ce qui est nouveau en science de la conservation. L'Algorithme *MinDegradation* fournit les solutions alternatives les moins coûteuses qui sont suffisamment différentes les unes des autres selon un seuil de dissimilarité donné. L'Algorithme *MaxDissimilarity* fournit les solutions les plus différentes les unes des autres pour une dégradation de l'objectif fixé. Ces procédures impliquent la formulation de programmes linéaires en nombres entiers mixtes résolus par des méthodes exactes. Une autre contribution importante est l'analyse comparative de ces deux procédures entre elles et avec les méthodes existantes.

Nos résultats montrent que la génération de solutions alternatives en fonction de l'intervalle de la valeur objective peut entraîner une faible variabilité entre les solutions, car elles sont très similaires les unes aux autres. Ces solutions qui ne diffèrent que de quelques unités de planification sont assez peu informatives. Elles peuvent difficilement être considérées comme des alternatives. Pire encore, plus loin de la valeur optimale, la variabilité entre les solutions alternatives semble peu pertinente car la procédure augmente artificiellement la valeur objective en incluant des unités de planification inutiles. En tant que telle, elle répond mal au besoin de solutions alternatives à la fois bonnes et différentes. La mesure de dissimilarité que nous incorporons nous permet de surmonter cette limitation. Les algorithmes proposés cherchent explicitement à générer une dissimilarité entre les solutions de réserve et à fournir de véritables alternatives. De manière similaire à (Chang et al. 1982b; Brill et al. 1982), la mesure de dissimilarité que nous définissons permet d'éviter que les réserves alternatives n'englobent la solution optimale. Un autre écueil, particulièrement frappant dans les approches métaheuristiques, est la nécessité de générer de nombreuses solutions alternatives afin d'explorer largement l'espace des solutions. Cette grande quantité de solutions alternatives nécessite un post-traitement statistique pour identifier quelques solutions distinctes. Cela nécessite souvent des analyses statistiques supplémentaires, comme par exemple la fréquence de sélection des sites de réserve ou l'analyse de regroupement. En revanche, nos méthodes fournissent directement un ensemble de présentation composé de solutions significativement distinctes. Quelques solutions qui sont à la fois bonnes et différentes les unes des autres peuvent donc suffire.

¹Plus de détails sur <https://prioritizr.net/reference/portfolios.html>

En résumé, la force de ce travail réside dans le fait que seules quelques itérations sont nécessaires pour générer un ensemble de présentation de solutions vraiment différentes. De plus, les méthodes développées sont hautement personnalisables. Par exemple, d'autres mesures de dissimilarité pourraient être utilisées dans nos procédures itératives pour évaluer les différences entre les solutions (Makowski et al. 2000). Ces différences ne dépendent que de la définition d'une mesure de dissimilarité, et peuvent être adaptées en fonction du cas d'application. Un autre avantage de ce type d'approche est de pouvoir quantifier explicitement la qualité des solutions alternatives générées. Comme la recherche de solutions alternatives est effectuée par des méthodes de résolution exactes, nous connaissons l'écart à l'optimum, ce qui donne plus de contrôle à l'utilisateur final. Enfin, la production de l'ensemble de présentation est entièrement contrôlée par deux paramètres. L'utilisateur peut alors choisir exactement le compromis entre la diversité de l'ensemble des solutions alternatives et l'écart à l'optimum. En ce qui concerne les points faibles, les approches proposées sont principalement limitées par le temps de calcul nécessaire. Ce temps peut être important pour certaines instances et il augmente avec le nombre d'alternatives demandées de par la nature itérative de l'approche. Cependant, nous n'avons pas cherché à améliorer le temps de calcul dans ce travail. Dans l'état actuel des algorithmes, nous pouvons fournir des ordres de grandeur pour le temps de calcul avec un ordinateur personnel (Intel Core i7-8850H CPU @ 2.60GHz) lorsque 4 alternatives sont demandées. Nous avons observé les ordres de grandeurs suivants pour le temps de calcul sur plusieurs instances générées :

- environ 2-3 minutes pour 500 unités de planification et 3 éléments de conservation avec l'Algorithm MaxDissimilarity
- environ 10-60 minutes pour 1000 unités de planification et 5 éléments de conservation avec l'Algorithm MaxDissimilarity.
- environ 10-20 secondes pour 500 unités de planification et 3 éléments de conservation avec l'Algorithm MinDegradation.
- environ 2-15 minutes pour 1000 unités de planification et 5 éléments de conservation avec l'Algorithm MinDegradation

Ces temps de calcul doivent être relativisés. Si nous ne cherchons pas nécessairement une preuve d'optimalité, ils peuvent être beaucoup plus faibles. Nos algorithmes nous permettent de fournir rapidement des solutions intéressantes et réalisables si nous décidons de conserver la solution courante après un temps maximum donné.

En conclusion, contrairement à ce qui était couramment affirmé dans la littérature (Possingham et al. 2000; Ardron et al. 2010), notre travail a montré que les méthodes d'optimisation exactes utilisées pour le problème de sélection de sites de réserve peuvent également être avantageuses pour produire un ensemble de solutions alternatives. Par conséquent, il n'est pas vrai que les métaheuristiques sont les seules méthodes capables de produire un ensemble de présentation. En outre, l'inclusion d'un critère de dissimilarité explicite directement dans le modèle d'optimisation a permis de construire un ensemble de présentation plus contrôlé et transparent. En recherchant des solutions sensiblement différentes, nous avons augmenté les chances d'aborder des objectifs qui ne sont pas nécessairement modélisés, tels que des objectifs sociopolitiques ou de gestion. Le faible nombre d'alternatives nécessaires avec nos méthodes peut éviter un bruit inutile dans le processus de prise de décision. En d'autres termes, les algorithmes proposés peuvent renforcer le pouvoir des utilisateurs en leur donnant plus de contrôle sur les alternatives produites et en supprimant l'analyse de post-traitement habituellement nécessaire. Nous espérons que ces méthodes pourront au moins jeter un nouvel éclairage sur les discussions relatives à la conservation et, à terme, apporter plus de succès aux décisions de conservation dans la pratique.

Optimisation averse au risque pour la sélection de réserve avec des données non-binaires incertaines

Les outils de planification spatiale de la conservation visent à trouver les meilleurs sites de réserve à partir des connaissances disponibles afin d'assurer la persistance de la biodiversité à long terme et, éventuellement, des services écosystémiques. Mais l'incertitude inhérente aux connaissances disponibles peut entraîner une mauvaise prise de décision et conduire à des solutions de réserve inefficaces. Cette inefficacité peut causer des dommages irréversibles aux écosystèmes, des contraintes inutiles sur les usages humains, et plus généralement un gaspillage des ressources de conservation déjà limitées. Il apparaît donc crucial de fournir aux décideurs des méthodes permettant de quantifier le compromis entre le risque et le coût d'une décision. Dans un contexte de changement global, les méthodes de sélection des réserves qui tiennent compte des incertitudes peuvent conduire à produire des réserves qui ont plus de chances d'atteindre leurs objectifs de conservation. De plus, la connaissance incomplète et imparfaite est à la base de la science de la conservation (Soule 1985). C'est pourquoi la prise en compte de l'incertitude a été identifiée comme une lacune importante de la science de la conservation (Margules and Pressey 2000; Foley et al. 2010; Reside et al. 2018). Le type d'incertitude qui peut intervenir dans les problèmes de sélection de réserves est détaillé dans (Regan et al. 2009). Dans ce travail, nous avons considéré une incertitude épistémique, qu'il s'agisse de variabilité (source naturelle telle que le changement climatique) ou d'imprécision (du modèle ou de la mesure). En pratique, cette incertitude affecte les quantités non-binaires des éléments de conservation du problème d'optimisation de la sélection des sites de réserve.

Le principal cadre utilisé dans les méthodes de sélection de sites de réserve pour intégrer explicitement l'incertitude sont les formulations avec des contraintes en probabilité (Polasky et al. 2000; Haight et al. 2000; Araújo and Williams 2000; Williams and Araújo 2000; Sarkar et al. 2004; Cabeza et al. 2004; Tole 2006; Moilanen et al. 2006a; Moilanen et al. 2006b). Une formulation de contraintes en probabilité vise à s'assurer que la probabilité de satisfaire des contraintes données est supérieure à un certain seuil. Ces travaux sont basés sur des données binaires puisqu'ils considèrent la présence/absence des éléments de conservation. Le modèle probabiliste développé dans (Haight et al. 2000), linéarisé dans (Polasky et al. 2000) en introduisant un niveau de risque toléré, a constitué une contribution méthodologique majeure. Elle a permis des développements ultérieurs dans le même cadre (Araújo and Williams 2000; Williams and Araújo 2000; Sarkar et al. 2004; Cabeza et al. 2004; Tole 2006; Moilanen et al. 2006a; Moilanen et al. 2006b). Ils sont basés sur la connaissance *a priori* de la probabilité de présence (ou de persistance dans (Araújo and Williams 2000; Williams and Araújo 2000)) de chaque élément de conservation à l'intérieur d'une unité de planification. Ces probabilités sont censées être connues ou déduites de données environnementales (Araújo and Williams 2000; Tole 2006). Elles sont considérées comme incertaines en raison des imprécisions du modèle et une analyse du manque d'information est appliquée dans (Moilanen et al. 2006a; Moilanen et al. 2006b). Les problèmes d'optimisation résultants ont été résolus principalement à l'aide d'heuristiques, bien qu'une comparaison avec des méthodes de résolution exactes ait également été effectuée dans (Sarkar et al. 2004). La clé de ces cadres stochastiques est que la variable aléatoire associée à une quantité d'éléments de conservation dans une unité de planification a une distribution de probabilité binaire (présence ou absence, persistance ou disparition). De cette façon, la probabilité globale de présence d'un élément de conservation dans la solution de la réserve peut être exprimée analytiquement. Cette expression permet ensuite d'exprimer un programme linéaire en nombres entiers déterministe. Des détails théoriques supplémentaires peuvent être trouvés dans (Beraldi and Ruszczyński 2002). Cette approche n'est pas possible lorsque des données non-binaires sont disponibles. Dans notre cas, nous avons considéré des

mesures d'abondance : la quantité de l'élément de conservation dans une unité de planification est non-binaire, puisqu'elle peut prendre n'importe quelle valeur positive. L'obtention d'une expression analytique similaire à celle de (Polasky et al. 2000) n'était pas possible avec des données non-binaires. D'où la nécessité de développer un autre modèle pour prendre en compte l'incertitude sur les données non-binaires, quitte à résoudre une approximation. Une autre approche similaire, mais compatible avec des données non-binaires, est déployée dans MarProb², *i.e.* Marxan avec probabilités (Game et al. 2008; Carvalho et al. 2011; Tulloch et al. 2013). La probabilité qu'une unité de planification soit détruite dans le futur est incluse dans la description du problème. La probabilité qu'un élément de conservation n'atteigne pas son objectif est calculée à l'aide d'une table statistique sous l'hypothèse d'une distribution gaussienne. Dans le même esprit que l'approche Marxan, cette probabilité a été incluse dans la fonction objectif avec une pénalité de manque de représentation, puis résolue à l'aide d'un algorithme méta-heuristique. Enfin, un indice de robustesse est proposé dans (Beech et al. 2008) en utilisant une stratégie de Monte-Carlo. Une distribution de probabilité prédéfinie permet de générer des échantillons de chaque quantité de caractéristiques de conservation. L'indice de robustesse est simplement la proportion de fois où une unité de planification est sélectionnée parmi les solutions de réserve calculées pour chaque instance dérivée des échantillons.

L'optimisation incertaine fait apparaître deux philosophies opposées quant à l'attitude à l'égard du risque : l'approche neutre vis-à-vis du risque et l'approche averse au risque. Une attitude neutre vis-à-vis du risque est souvent utilisée lorsque le résultat d'une décision est répété de nombreuses fois. De cette façon, la décision est conçue pour optimiser une quantité attendue au fil du temps. En revanche, lorsque le résultat d'une décision est le fruit d'une seule réalisation, nous pouvons ne pas tolérer le risque encouru. En sciences de la conservation, une décision de découpage spatial est prise une fois et le risque associé à une mauvaise décision peut entraîner la destruction d'habitats et l'extinction d'espèces. C'est pourquoi, notre préférence pour le risque s'est portée sur les approches averses au risque. Dans ce travail, nous montrons aux utilisateurs comment incorporer l'incertitude affectant les quantités non-binaires des éléments de conservation dans le problème de sélection des sites de réserve en utilisant des cadres d'optimisation averses au risque. Nous proposons deux modèles d'optimisation qui tiennent compte explicitement du risque qu'un utilisateur est prêt à prendre. Dans la première approche, nous avons considéré une incertitude paramétrique non-probabiliste : les quantités des éléments de conservation peuvent prendre une gamme de valeurs possibles au lieu d'une valeur particulière. Pour résoudre ce problème, nous proposons une formulation d'optimisation robuste. En d'autres termes, nous avons cherché à trouver la meilleure solution réalisable quelle que soit la réalisation de l'incertitude au sein de chaque ensemble de paramètres incertains. Mais, afin d'éviter la solution du pire cas trop conservatrice, nous avons introduit un ensemble d'incertitude budgétisé (Bertsimas and Sim 2004). Dans ce cadre, nous avons cherché à trouver la meilleure solution réalisable en considérant un nombre de paramètres définis par l'utilisateur qui sont tolérés à dévier de leur valeur nominale. L'optimisation qui en résulte est un programme linéaire mixte en nombres entiers déterministe. Enfin, une analyse de sensibilité du paramètre de niveau de risque fournit un moyen simple de représenter la robustesse de la solution de réserve de la solution nominale au pire cas. Dans la deuxième approche, nous avons considéré une formulation averse au risque avec des contraintes de probabilités. Comme nous ne disposons pas d'une distribution de probabilité binaire, le calcul analytique de la quantité globale probable de chaque élément de conservation dans la réserve n'est pas disponible. C'est pourquoi, nous avons abordé ce problème d'optimisation en utilisant une approche d'approximation par échantillonnage (Luedtke and Ahmed 2008). L'idée est de discrétiser l'espace de probabilité en utilisant un ensemble d'échantillons probables. De cette façon, la probabilité peut être estimée par une

²<https://marxansolutions.org/software/>

proportion parmi les échantillons considérés. Pour concevoir la solution de réserve, nous avons généré 100 échantillons probables en utilisant les géostatistiques. Nous avons généré des échantillons conditionnellement à ce qui a été observé aux points de mesure. Au lieu de développer un indice de robustesse comme dans (Beech et al. 2008), nous avons pris en compte les échantillons de manière explicite dans la formulation d'un programme linéaire en nombres entiers. Nous avons calculé la solution de réserve de la formulation de la contrainte de probabilités pour plusieurs valeurs de niveau de risque. Nous avons comparé le niveau de risque avec une probabilité estimée plus précise sur 1000 échantillons.

Nos résultats ont montré ce qu'il en coûte de produire des solutions de réserve averse au risque qui sont robustes à l'incertitude et à quel point elles le sont. Alors que le formalisme robuste traitait des quantités surestimées des éléments de conservation, le formalisme de la contrainte de probabilités traitait d'une incertitude plus probabiliste contenue dans les échantillons. Les modèles proposés dans ce travail sont nouveaux par rapport aux modèles classiques présentés dans (Polasky et al. 2000; Haight et al. 2000) qui ne sont compatibles qu'avec des données binaires. Nos approches ont permis de tenir compte d'une incertitude qui affecte les quantités d'éléments de conservation non-binaires, ce qui est souhaitable dans de nombreux cas. En général, la solution de réserve robuste est apparue coûteuse par rapport à la solution nominale. Par exemple, en considérant que la moitié des paramètres est surestimée, la réserve robuste coûterait plus de 50% par rapport à la solution nominale lorsque le paramètre de déviation est fixé à 30%. Lorsque nous avons évalué les solutions robustes sur les 1000 échantillons générés, bien que les coûts impliqués soient importants, les objectifs de couverture sont presque toujours atteints. Les solutions du cadre robuste sont un moyen sûr d'être immunisé contre l'incertitude, même pour de petits budgets d'incertitude, mais elles peuvent être très coûteuses. À l'inverse, les solutions du formalisme de contrainte de probabilités fournissent des solutions moins chères, seulement quelques pourcents au-dessus de la solution nominale, qui sont plus ou moins immunisées contre le risque selon le choix de l'utilisateur. Cela semble permettre un réglage plus fin du risque qu'un décideur est prêt à prendre. Nous avons également observé que des solutions robustes pour de petites valeurs du budget d'incertitude permettaient d'obtenir les mêmes résultats. Cependant, la relation entre le budget d'incertitude et la probabilité d'atteindre les cibles n'est pas aussi simple que dans le formalisme des contraintes de probabilités. Il peut être plus difficile d'utiliser ce cadre pour concevoir une solution de réserve qui soit immunisée contre une incertitude contenue dans les échantillons.

Nos résultats ont également montré que la comparaison de la solution du problème nominal avec nos solutions averse au risque permet d'identifier dans quelle mesure des unités de planification données contribuent à la robustesse de la solution. Les unités de planification en dehors de la solution nominale et qui sont souvent sélectionnées dans les solutions averse au risque sont susceptibles d'augmenter la robustesse de la solution vis-à-vis de l'incertitude considérée. À l'inverse, les unités de planification de la solution nominale qui ne sont pas sélectionnées dans les solutions averse au risque lorsque nous augmentons l'incertitude sont susceptibles d'être perdues lorsque l'incertitude est réalisée. Nous observons que ces unités de planification ne sont pas les mêmes dans les cadres de contraintes robustes et de probabilités. Ce type d'observations était à la base de l'indice de robustesse développé dans (Beech et al. 2008). Dans ce travail, nous sommes allés plus loin, puisque nous tenons compte *a priori* des échantillons dans le modèle d'optimisation dans le formalisme des contraintes de probabilités.

Le cadre robuste fournit un moyen d'incorporer l'incertitude sans aucune distribution de probabilité. Le seul prix à payer est d'ordre computationnel car il implique un programme linéaire en nombres entiers plus important par rapport au problème nominal. Le formalisme des con-

traintes de probabilités est une approche plus probabiliste. Notre travail a montré comment nous pouvions résoudre efficacement une approximation de ce problème difficile en utilisant des échantillons générés. Les limites du formalisme des contraintes de probabilités peuvent provenir du nombre d'échantillons nécessaires. Comme la taille du problème est liée au nombre d'échantillons, nous ne pouvons pas augmenter le nombre d'échantillons sans causer des difficultés de calcul. Notre expérience utilisant les géostatistiques pour générer des échantillons a illustré une autre limite : le paramètre de niveau de risque du formalisme des contraintes de probabilités ne représente pas strictement le risque réel. Pourtant, l'intérêt des décideurs est de garantir une solution de réserve pour se protéger contre un niveau de risque réel prédéfini. Puisqu'il y avait une corrélation apparente entre ces deux quantités, nous aurions pu fixer empiriquement le niveau de risque du modèle pour atteindre un niveau de risque réel prédéfini. Cependant, la relation entre le nombre d'échantillons et la probabilité réelle d'atteindre les objectifs devrait être étudiée plus en détail. Dans ce travail, nous n'avons considéré que l'incertitude relative à un élément de conservation unique, ce qui limite l'évaluation de l'effet de couplage. Nous avons fait ce choix pour des raisons de simplicité, mais cela mériterait d'autres explorations pour établir des conclusions plus générales. Une autre limite observée est que nos solutions averses au risque sont assez similaires malgré l'augmentation des paramètres du niveau de risque. Cela s'explique par l'étude de cas considérée : les éléments de conservation sont principalement distribués autour de l'île principale, tout comme les solutions.

Nos méthodes averses au risque fournissent aux utilisateurs des moyens simples de déduire une relation univoque entre la robustesse d'une solution de réserve et le coût associé. Ces cadres permettent également d'identifier les zones susceptibles d'être perdues lorsque l'incertitude se réalise, ce qui constitue une information précieuse dans un contexte de prise de décision. Le choix du formalisme à privilégier dépend de la modélisation de l'incertitude : distribution de probabilité ou intervalles de valeurs possibles. Le compromis entre le niveau de risque et le coût de la robustesse doit être fait par le décideur dans les deux cadres.

Conception explicite de réserves compactes, connectées et sans trous

La résolution du problème d'optimisation de sélection des sites de réserve aboutit souvent à la sélection de sites de réserve dispersés. Pourtant, la conception de réserves compactes, connectées et sans trous est généralement nécessaire pour des raisons écologiques, de gestion et d'application de la loi (Diamond 1975). Une réserve est connectée si l'on peut se déplacer n'importe où à l'intérieur sans avoir à la quitter. Un trou dans la réserve est une zone extérieure à la réserve que l'on ne peut quitter sans traverser la réserve. Actuellement, les attributs spatiaux des réserves sont peu pris en compte dans les outils d'aide à la décision utilisés pour la sélection des réserves. Dans les outils d'aide à la décision largement utilisés pour la sélection des réserves (*e.g.* Marxan, PrioritizR), le seul attribut spatial explicitement considéré est la compacité globale d'une solution (Ball and Possingham 2000; Ball et al. 2009; Watts et al. 2009; Hanson et al. 2020). La compacité d'une solution est imposée en pénalisant directement le périmètre global de la réserve dans la fonction objectif du problème d'optimisation traité. Cette approche soulève plusieurs problèmes. Tout d'abord, l'expression du périmètre est quadratique par rapport aux variables de décision. La linéarisation de l'expression du périmètre implique l'ajout de nombreuses variables de décision et contraintes (Billionnet 2013; Beyer et al. 2016) qui peuvent être coûteuses en calcul dans un contexte de programmation en nombres entiers. De plus, cette approche transforme le problème en un problème multi-objectif où le coût d'une solution et son périmètre sont implicitement en concurrence. En pratique, le multiplicateur de

compacité est déterminé de manière empirique jusqu'à ce qu'une solution réponde aux exigences spatiales jugées satisfaisantes. Cela affaiblit la nature systématique de l'approche de conception des réserves, bien qu'une définition plus systématique du multiplicateur de compacité soit proposée dans (McDonnell et al. 2002). Des améliorations utilisant à la fois le périmètre et la surface de la réserve dans l'objectif ont été proposées dans (McDonnell et al. 2002) pour renforcer la compacité de la réserve. Dans le même ordre d'idées, une combinaison pondérée des mesures de compacité et de connectivité est incluse dans l'objectif et résolue à l'aide de métaheuristiques dans (Nalle et al. 2002). Dans tous les cas, la connectivité et l'absence de trous ne sont pas garantis, mais plutôt susceptibles d'apparaître avec l'application empirique de la compacité de la réserve.

Dans un contexte de recherche opérationnelle, plusieurs modèles d'optimisation ont été proposés pour tenir compte explicitement de propriétés spatiales spécifiques (Williams et al. 2004; Williams et al. 2005; Billionnet 2013; Billionnet 2016; Billionnet 2021). Par exemple, les modèles d'optimisation ont pour but de sélectionner un noyau de réserve avec une zone tampon autour (Williams and ReVelle 1998; Clemens et al. 1999). Mais ces modèles n'obligent pas la réserve à être connectée et sans trous, bien qu'une telle réserve puisse en émerger. Une grande famille de modèles tire parti des possibilités de modélisation offertes par l'utilisation des distances entre les sites candidats. La minimisation de la somme des distances deux à deux ou de la distance maximale entre tous les sites de la réserve (Önal and Briers 2002) favorise les réserves compactes, mais ne garantit pas que la réserve soit connectée et sans trous. Il en va de même pour les modèles qui contraignent deux sites distincts contenant le même élément de conservation à être plus proches qu'une distance seuil prédéfinie (Williams 2006). Une autre grande famille de modèles d'optimisation tire parti de la théorie des graphes (Önal and Briers 2006; Wang and Önal 2011; Wang and Önal 2013; Jafari and Hearne 2013; Billionnet 2016; Shirabe 2005) notamment pour assurer explicitement la connectivité de la réserve. Cependant, la sélection des sites peut encore aboutir à l'inclusion de trous dans la solution de réserve. Une réserve perforée par des trous ne peut pas être utilisée dans une conception de réserve à grande échelle. Si des trous apparaissent dans une solution proposée par un outil d'aide à la décision, elles seront soit incorporées arbitrairement dans la réserve, soit connectées artificiellement à l'extérieur (dans les deux cas, cela conduira souvent à l'utilisation de solutions sous-optimales), soit la solution fournie sera ignorée. L'utilisation de modèles imposant la connectivité et favorisant la compacité est susceptible de favoriser les réserves sans trous, mais cela n'est pas garanti. Par exemple, il peut être nécessaire de concevoir des réserves naturelles autour de zones qui ne peuvent pas être incluses dans la réserve, comme un port ou une route commerciale. Ces zones ne peuvent pas être entourées par la réserve et doivent rester accessibles de l'extérieur. Les modèles les plus récents fournissent souvent une solution de réserve avec des trous dans ces cas. La prise en compte des trous dans les réserves est rarement abordée dans la littérature. L'absence de trous dans la réserve peut être obtenue par la recherche itérative d'une réserve sans trous parmi des solutions légèrement sous-optimales ((Billionnet 2016)). Ce modèle n'empêche pas *a priori* les trous d'être inclus dans la réserve, mais espère plutôt qu'une telle solution existe même si la valeur objective se retrouve dégradée. Une telle procédure est intéressante mais ne garantit pas l'obtention d'une réserve connectée, compacte et sans trous avec la meilleure valeur objective. Un modèle sélectionnant des régions cellulièrement convexes (également dans des grilles régulières) qui sont ainsi connectées et sans trous est donné dans (Williams 2003). Un tel modèle peut être adapté pour éviter les réserves avec des trous, mais l'exigence de convexité peut négliger certaines solutions admissibles connectées et sans trous si elles ne sont pas cellulièrement convexes.

Dans ce travail, nous avons proposé un programme linéaire en nombres entiers qui produit

des réserves compactes, connectées et sans trous. Un modèle d'optimisation faisant appel à la théorie des graphes impose la connectivité à la fois de la réserve et de la non-réserve ce qui permet d'obtenir des réserves connectées et sans trous. Nous avons utilisé un modèle de flux multi-commodités pour imposer la connectivité. Nous avons choisi un modèle de flux multi-commodités plutôt qu'un modèle de flux unique afin d'exprimer les contraintes de flux comme des contraintes paresseuses. L'utilisation de contraintes dures à la place de contraintes paresseuses prend beaucoup plus de temps voire même rend impossible de trouver des solutions dans la limite de temps de 1000 secondes dans la plupart des cas. La compacité globale de la réserve est définie en spécifiant un rayon ou un périmètre maximum de la réserve. Nous avons également fourni une réduction du problème, en profitant de la structure rectangulaire de la grille. En effet, en raison de la forme du graphe induit par une grille rectangulaire, nous avons appliqué un "échiquier" pour séparer les nœuds en deux ensembles. En fonction de sa position sur l'échiquier, un nœud est étiqueté noir ou blanc et appartient à l'ensemble correspondant. Ensuite, nous avons appliqué le modèle de flux multi-commodités uniquement aux nœuds noirs pour réduire la taille du problème sans aucune perte. Nous avons comparé les solutions de notre modèle avec celles de l'état de l'art, à la fois en termes de temps de calcul et de qualité des solutions par rapport aux exigences spatiales. Le code utilisé pour ce travail est libre, ouvert et disponible. Le modèle que nous avons proposé est hautement personnalisable en ce qui concerne la façon dont nous appliquons la compacité. Un bon réglage du multiplicateur de compacité, du périmètre maximum ou du rayon maximum permet de façonner finement les attributs spatiaux de la réserve. Nous avons illustré sur l'instance du monde réel de Fernando de Noronha que notre modèle pouvait déjà être utilisé dans la pratique. Nous avons évalué numériquement la généralité des approches proposées sur plusieurs instances générées composées de 300 ou 500 unités de planification et de 3 éléments de conservation. En ce qui concerne la taille des instances dans les travaux existants, le nombre d'unités de planification considéré dans ce travail est similaire aux autres travaux existants : 100 unités de planification dans (Billionnet 2012), 131 dans (Önal and Briers 2002), 225 dans (Billionnet 2021), 324 unités de planification (Williams 2006), 391 unités de planification dans (Önal and Briers 2006), 400 dans (Billionnet 2016)). Cependant, le nombre d'éléments de conservation est un ordre de grandeur supérieur à celui des travaux existants, toujours autour de 100 éléments de conservation. La comparaison de la taille des instances reste difficile à interpréter, car les instances ne sont pas générées de la même manière que dans la littérature.

Au lieu de supprimer les solutions de réserve avec des trous de l'espace de recherche (Billionnet 2016), nous avons utilisé des contraintes de flux paresseuses pour *a priori* construire une non-réserve connectée. Nous avons fait dans ce travail ce qui était préconisé dans la discussion de (Billionnet 2016) en incluant *a priori* un modèle pour empêcher la formation de trous dans les réserves. Nous nous distinguons également de (Billionnet 2016) car nous avons utilisé un modèle de flux multi-commodités et des contraintes paresseuses pour faire respecter la connectivité dans le problème de l'ensemble minimal. L'utilisation d'un flux multi-commodités dans les problèmes de sélection de sites de réserve n'est pas nouvelle et a déjà été mentionnée dans (Billionnet 2021) mais il s'agit d'une approche différente par rapport à (Billionnet 2012; Billionnet 2016; Williams 2006).

La principale limite de notre travail est la taille limitée des instances qui peuvent être résolues avec notre modèle dans un temps raisonnable. Le problème d'optimisation que nous avons proposé reste difficile à résoudre numériquement. Cela était attendu car la résolution de modèles qui prennent en compte des contraintes spatiales a tendance à être plus exigeante numériquement en général, en particulier lorsque la taille du problème augmente (Wang et al. 2018). Bien que la taille des instances considérées soit du même ordre de grandeur que ce qui est habituelle-

ment fait dans une littérature similaire, il s'agit tout de même de petites instances par rapport aux applications qui ne prennent pas en compte des exigences spatiales. D'un autre côté, cela pourrait être atténué par le fait que nous pouvons potentiellement fournir des solutions sans la preuve d'optimalité. Nos résultats montrent que l'obtention d'une réserve compacte, connectée et sans trous n'entraîne qu'une faible augmentation du coût de sélection du site par rapport aux modèles implémentés dans les outils d'aide à la décision tels que Marxan et PrioritizR. Le seul prix à payer est un temps de calcul plus important. Par conséquent, l'obtention de réserves spatialement cohérentes est davantage un défi numérique qu'une question de ressources limitées par rapport au coût considéré. Étant donné que les solutions optimales ne sont pas si coûteuses par rapport aux modèles de l'état de l'art, une solution de réserve compacte, connectée et sans trous qui est légèrement sous-optimale peut tout de même être intéressante. De plus, sélectionner une solution de réserve sous-optimale est souvent tolérée dans la littérature sur la conservation, notamment lorsque des métaheuristiques sont utilisées pour résoudre le problème.

La grande étendue et l'aspect "toile d'araignée" des réserves sont une caractéristique commune aux modèles qui imposent la connectivité de la réserve (Önal and Briers 2006; Billionnet 2012). Nous avons atténué cet écueil inhérent aux modèles de connectivité en contraignant le rayon et le périmètre de la réserve à rester sous un seuil prédéfini. Nos résultats ont également montré que l'inclusion de la compacité par pénalité du périmètre dans l'objectif ou par l'inclusion d'une contrainte de périmètre maximum aident le modèle à trouver une solution plus rapidement. Afin de conserver une approche à objectif unique, il peut être préférable de contraindre un périmètre maximal de la réserve au lieu d'utiliser une pénalité dans l'objectif, mais nos résultats ont montré que cela rend la résolution plus lente pour les instances de 500 unités de planification.

Dans ce travail, nous n'avons pas abordé le problème de la représentation des éléments de conservation dans deux composantes distinctes de la réserve, ce qui est parfois une caractéristique souhaitable pour être robuste aux événements catastrophiques (*e* épidémie, incendie). Imposer une distance minimale entre deux sites sélectionnés contenant le même élément de conservation (Williams 2006) était la méthode pour obtenir une telle propriété. L'inclusion d'une telle contrainte pourrait améliorer le modèle, ce qui peut donner lieu à des développements futurs. Une autre caractéristique que nous n'avons pas prise en compte dans ce travail est la nature potentiellement différente des unités de planification exclues *a priori* : certaines peuvent être traversées, d'autres non. Par exemple, dans la conception d'une réserve marine, une unité de planification terrestre est exclue *a priori* et ne peut être traversée. En revanche, une unité de planification située dans le port est exclue *a priori* mais peut être traversée. La principale implication de cette différence en matière de modélisation est qu'un chemin assurant la connectivité dans les modèles de flux ne peut pas traverser toutes les unités de planification exclues *a priori*. Ce point n'est pas inclus dans nos modèles actuels et pourrait donner lieu à des développements futurs.

Conclusion

Cette thèse s'est intéressée aux propriétés des outils d'aide à la décision couramment utilisés dans les négociations de la planification spatiale marine, en particulier ceux mis en œuvre pour identifier de manière optimale la localisation des aires de conservation (*i.e.* aires marines protégées). Nous avons d'abord (Chapitre 2) dressé un panorama des outils existants, des raisons de leur coexistence, de leur formalisme et de leur fonctionnement, ainsi que de leurs forces et faiblesses respectives. Un résumé graphique des outils existants est illustré dans la Figure 1. Après avoir établi que les outils d'optimisation de résolution approchés (métaheuristiques) avaient

principalement une raison d'être historique (surmonter des problèmes de calcul aujourd'hui disparus), nous avons choisi de nous focaliser sur les outils d'optimisation en résolution exacte pour développer un certain nombre de leurs fonctionnalités qui faisaient défaut jusqu'à présent. Nous avons d'abord proposé des algorithmes qui permettent de produire un ensemble diversifié de solutions de réserve. Alors que les algorithmes métaheuristiques fournissent par construction un échantillon aléatoire de l'espace des solutions, sans aucune indication sur la proximité à l'optimum, ni aucun contrôle sur la diversité réelle de ces solutions (ce qui implique d'en produire beaucoup), nous avons proposé (Chapitre 3) des algorithmes permettant de produire des solutions alternatives, tout en contrôlant explicitement leur proximité à l'optimum et en maximisant la différence entre ces solutions. Nos algorithmes permettent donc de construire un portefeuille parcimonieux de solutions (peu de solutions sont nécessaires puisqu'elles sont construites pour être réellement différentes), dont la dégradation par rapport à l'optimum peut être explicitement contrôlée. Un autre axe de développement important nous a semblé être la prise en compte de l'incertitude dans ce type de modèle. Dans les applications existantes, les données d'observation sont prises pour une vérité absolue, alors que l'on sait, par exemple, que chaque instrument présente une variance de mesure. De même, les effets de la variation saisonnière, par exemple, bien que souvent connus, sont rarement pris en compte dans ces modèles statiques. Nous avons donc proposé (Chapitre 4) des algorithmes qui permettent d'introduire explicitement les connaissances existantes sur les différents types d'incertitude affectant les données d'entrée et de construire des solutions de réserve robustes à ces variations autour des données d'observation. Enfin, nous avons pu identifier qu'une solution mathématiquement optimale n'est souvent pas forcément pratique à gérer dans sa géométrie : des problèmes de discontinuité, de non-compacité, d'enclaves peuvent apparaître dans les solutions proposées par les modèles, et sont généralement supprimés dans le monde réel par une simplification *a posteriori* de la géométrie de la réserve. Dans le Chapitre 5, nous avons proposé des solutions numériques efficaces et explicites pour éviter l'apparition de ces phénomènes si tel est le souhait des acteurs, évitant ainsi des simplifications hasardeuses et *a posteriori* de la forme de la réserve. La Figure 2 fournit un résumé graphique des développements algorithmiques produits lors de cette thèse.

Mots-clés: planification spatiale marine ; outils d'aide à la décision ; aires marines protégées ; sélection optimale de sites de réserve ; programmation linéaire en nombres entiers ; Atlantique tropicale.

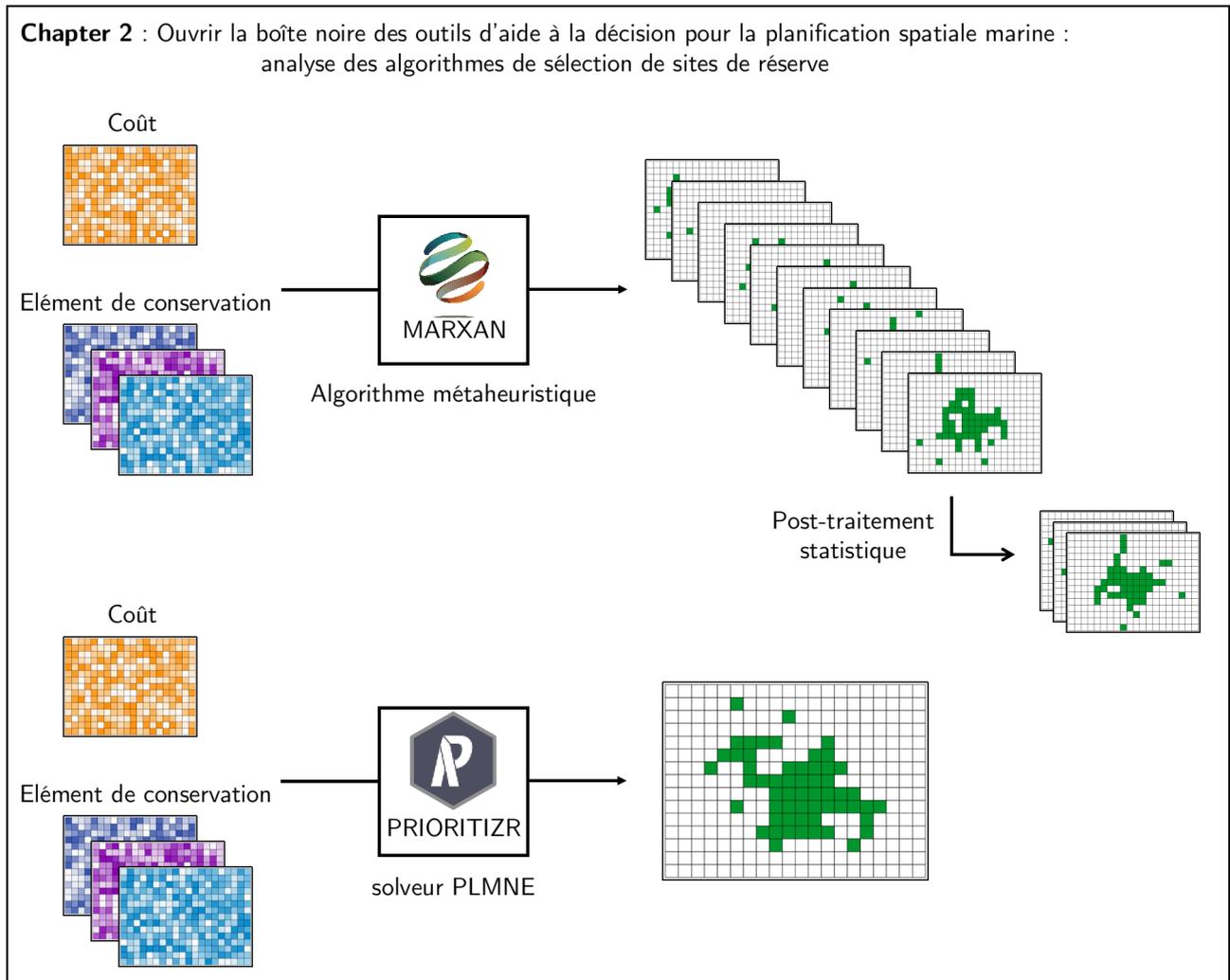
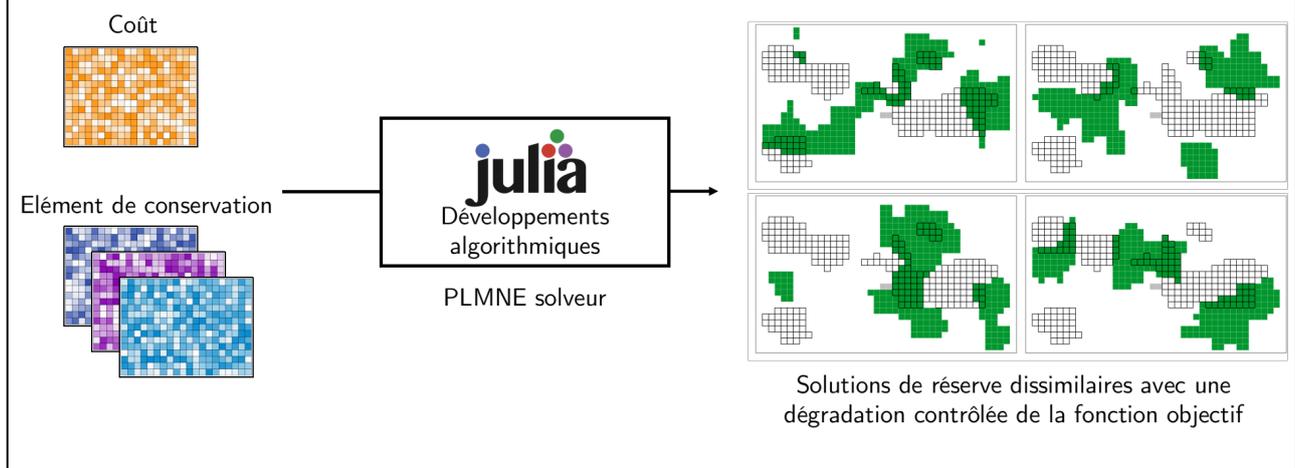
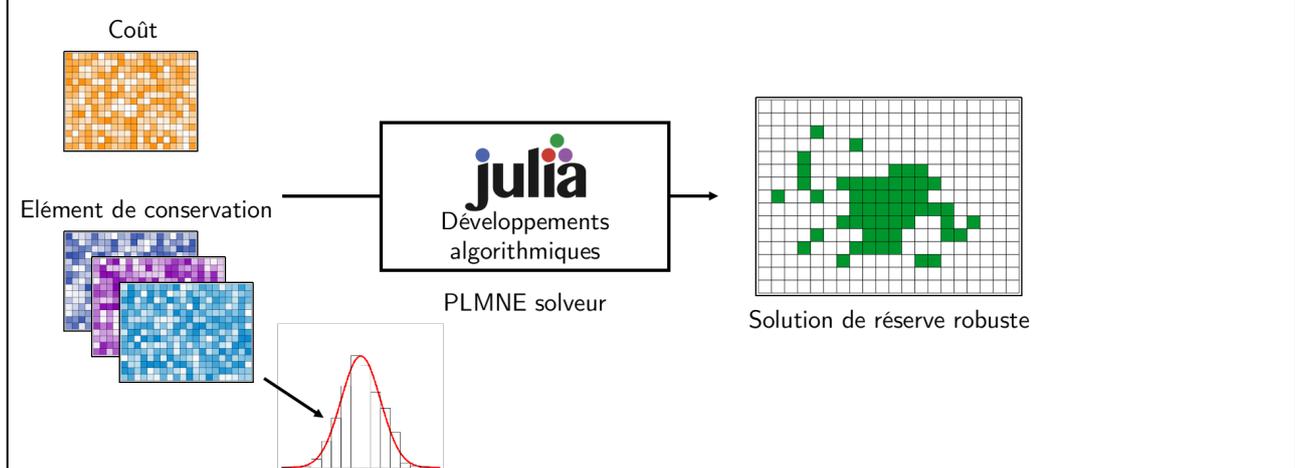


Figure 1: Résumé graphique des outils mettant en œuvre les modèles de sélection de sites de réserve de l'état de l'art. Marxan déploie un algorithme métaheuristique pour construire un nombre défini par l'utilisateur de solutions de réserve qui nécessitent un post-traitement statistique. PrioritizR déploie des méthodes d'optimisation exactes pour trouver la solution la moins coûteuse au problème de sélection des sites de réserve.

Chapitre 3 : Produire un ensemble de solutions de réserve différentes et proches de l'optimalité par l'optimisation exacte



Chapitre 4 : Optimisation averse au risque pour la sélection de réserve avec des données non-binaires incertaines



Chapitre 5 : Conception explicite de réserves compactes, connectées et sans trous

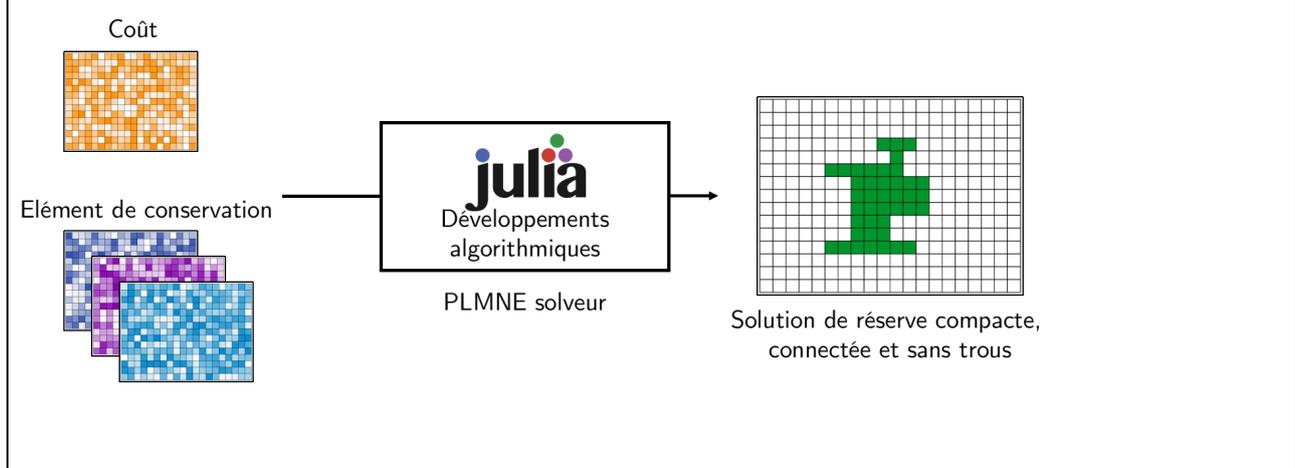


Figure 2: Résumé graphique des développements algorithmiques de la thèse. Le Chapitre 3 fournit des méthodes d'optimisation exacte pour déduire un ensemble diversifié de solutions proches de l'optimalité, dissimilaires entre elles et avec une dégradation prédéfinie de la valeur de la fonction objectif. Le Chapitre 4 fournit deux approches utilisant l'optimisation exacte pour traiter une incertitude qui affecte les données non-binaires du problème de sélection du site de réserve. Le Chapitre 5 fournit des méthodes d'optimisation exacte qui construisent des réserves compactes, connectées et sans trous. Le code Julia associé est ouvert, gratuit et disponible sur <https://github.com/AdrienBrunel>. PLMNE = programmation linéaire mixte en nombres entiers.

Abstract

The biodiversity crisis and ongoing climate change have led to a relative consensus on the need to preserve the integrity and functioning of exploited ecosystems. Although already at the heart of multiple anthropogenic pressures (fishing, aquaculture, shipping routes, seabed exploitation, recreational activities, renewable and fossil energies, etc.), marine environments are now considered as ideal spaces for the deployment of tomorrow's "blue economy". In this context, Marine Spatial Planning (MSP) aims to rationally organise the use of marine space and resources in order to reduce tensions between human activities and marine ecosystems, as well as between the stakeholders of the ocean. As a result, MSP has been widely disseminated within the institutions in charge of management and has finally become the preferred framework in the perspective of a sustainable development. Since MSP is the fruit of collective mobilisations, it is by nature at the crossroad of various disciplines (ecology, economy, sociology, law, *etc.*) making it a challenging framework to establish. Because it is intended to be rational and evidence-based, MSP makes use of numerous decision-support tools throughout its process in order to inform decision-making in a systematic and transparent manner. There is a profusion of such tools (Marxan, PrioritizR, SeaSketch, Atlantis, Coastal Resilience, *etc.*). These tools are more or less open, free, complementary, pursue different objectives, mobilise various methods and knowledge.

In general, MSP seeks an ocean zoning in order to allocate in time and space human activities at sea in accordance with a sustainable development. A particular case of MSP is the allocation of marine space for conservation use. In other words, we focus on the tools and methods for designating protected areas, a particularly hot topic at the moment. This gives the conservation science a prominent place in international debates. Indeed, protected areas are at the heart of current global policies to mitigate the erosion of biodiversity. For instance, the members (states, NGOs, agencies) of the International Union for Conservation of Nature (IUCN) have agreed to target a level of protection of 30% of the surface of each marine eco-regions by 2030, compared with less than 8% today. More recently, the European Green Deal sets a target of 30% of the surface of the European waters to be covered by marine protected areas (MPAs) by 2030. MPAs thus appear to be an essential part of the solutions for ensuring the preservation of marine ecosystems. The surface areas involved are huge, and so are the potential damages toward ecosystems or unnecessary constraints on human activities. Consequently, even a small limitation of these tools cannot be overlooked due to the amplitude of the potential impact. The main purpose of this work is thus to overcome some of the known limitations of these decision support tools involved in MSP. In particular, we seek to decipher some black box issues of the tools currently used, and bring more transparency in their parametrization so that they foster more equity in MSP processes.

The work carried out in this thesis explores specifically three axes of improvement:

1. Two classes of algorithms may solve the reserve site selection problem: metaheuristic algorithms (such as simulated annealing, commonly implemented in Marxan) and exact optimisation (*i.e.* integer programming, commonly implemented in PrioritizR). Although exact approaches are now able to solve large-scale problems, metaheuristics are still widely used. One reason is that metaheuristic-based software provides a set of suboptimal reserve solutions instead of a single one. These alternative solutions are usually welcomed by stakeholders as they provide a better basis for negotiations among potentially conflictive objectives. Metaheuristic algorithms use random procedures to explore the space of suboptimal reserve solutions. Therefore, they may produce a large amount of similar, thus uninformative, alternative solutions, which usually calls for a heavy statistical post-processing. Effective methods for generating a diverse set of near-optimal solutions using exact optimisation are lacking. Here we present two new approaches for addressing

this issue. Our algorithms explicitly control both the optimality gap and the dissimilarity between alternative reserve solutions. They allow the identification of a parsimonious, yet meaningful set of reserve solutions. The algorithms presented here could potentially increase the uptake of exact optimisation by practitioners. These methods should contribute to less noisy and more transparent discussions in the design of conservation policies. This is the subject of Chapter 3.

2. Most widely used decision support tools (*e.g.* Marxan, Prioritizr) generally do not allow to explicitly account for uncertainty. Yet, these uncertainties can lead to deprecated reserve solutions, potential irreversible damages towards ecosystems and useless constraints on stakeholders. Probabilistic approaches were successfully applied with presence/absence data but this framework is too restrictive when non-binary data are available. We propose two risk-averse approaches incorporating a parametric uncertainty within reserve site selection models: (1) a robust optimisation framework where a given amount of parameters is tolerated to deviate from a nominal value, (2) a chance constraint optimisation problem solved using a sampling approximation approach. For both models, a risk-level parameter allows conservation practitioners to establish the risk they are willing to take. This work provides conservation practitioners tools to design reserve solutions that are robust to uncertainty in order to face current global changes. This is the subject of Chapter 4.
3. For ecological, enforcement and management reasons, a reserve without any spatial consistency has limited chance of being implemented. Most widely used decision support tools for reserve site selection (*e.g.* Marxan, Prioritizr) include only a rough spatial constraint in their optimisation models through a compactness parameter. The compactness of a solution is empirically enforced by penalising the overall perimeter in the objective function. Several optimisation models tried to overcome this and accounted for a given spatial property: limited fragmentation, connectivity of selected sites, buffer zones around selected sites, *etc.* But, to this day, there is not a reserve site selection model that guarantee to produce a connected, compact and gap-free reserve. We propose a mixed-integer linear program using graph theory to explicitly build a reserve that is connected, compact and gap-free. To do so, we used a multicommodity flow approach that incorporates these spatial attributes within a reserve site selection model. We tested the computation feasibility of our model on generated instances and on the real case of Fernando de Noronha. Our results show that we can enforce compactness, connectivity and the absence of gaps using a single model. This work provides conservation practitioners an optimisation model to design reserve solutions showing desirable spatial properties, what increases their chances of being implemented. This is the subject of Chapter 5.

Although reserve site selection can seem a very specific aspect of MSP, understanding and improving these decision support tools is a needed step before addressing the more global question of MSP : how to find an equitable and sustainable spatial zoning of human uses? Allowing to understand precisely why and how the result was arrived at, the axes presented here should hopefully open the black-box of decisions support tools and eventually contribute to a better planning process.

The case study used as a numerical illustration of the methods developed in the thesis is the Brazilian archipelago of Fernando de Noronha in the tropical Atlantic. For methodological purposes, we also used generated datasets.

Keywords: marine spatial planning; decision support tools; marine protected areas; optimal reserve site selection; integer linear programming; tropical Atlantic.



Chapter 1

Introduction

1.1 Context

1.1.1 Marine spatial planning

Marine environments are frequently seen as suitable areas for the development of the so-called blue economy by institutions (*The EU Blue Economy Report 2019*; *The EU Blue Economy Report 2020*), banks or even NGOs (*WWF Briefing 2018: Principles for a Sustainable Blue Economy*). The concept of blue economy refers to the development of economic activities linked to coastal and oceanic ecosystems. This concept is particularly attractive within political spheres because it opens up new perspectives for the economic growth necessary for the survival of current capitalist models. The practical consequence of the blue economy development is the increase of the spatial occupancy of the ocean and coastline by multiple human uses: fishing, aquaculture, shipping routes, seabed exploitation, recreational activities, renewable energies, fossil energies, tourism, *etc.* Yet, these spaces are already being at the heart of multiple anthropogenic pressures, facing global changes and undergoing a biodiversity crisis. The prospect of increased exploitation of the sea may therefore seem paradoxical and represents an organisational and ecological challenge. The launch of the United Nations Decade of Ocean Science for Sustainable Development (2021-2030) at the United Nations Ocean Conference confirms the importance of this challenge at an institutional level.

Within that context, marine spatial planning (MSP) is positioning itself as a rational and collective decision-making process regulating uses of marine spaces and resources. The end goal is to reduce tensions between exploitation and conservation, as well as between ocean stakeholders. Stakeholders can be of various types (*e.g.* private companies, governments, individuals, *etc.*) and have very different interests (*e.g.* recreative activities, resource exploitation, routes, *etc.*). For example, a stakeholder can be an individual using the ocean to practice kite surfing or a private company installing offshore wind turbines. Tensions can rise between stakeholders when they share an interest in a common marine area and that their activities are not compatible. Also, the activity pursued by a stakeholder may occasion degradations to the ecosystems. According to (Ehler and Douvère 2009), "MSP implies analysing and allocating the spatial and temporal distribution of human activities in marine areas to achieve ecological, economic, and social objectives that are usually specified through a public political process". In other words, MSP aims at finding a spatial zoning of human uses that allows a sustainable exploitation of the ocean. A spatial zoning consists in deciding which human uses can be deployed at sea and in which areas they can be deployed. Finding such zoning can be assimilated to a resource allocation problem: the resource is the marine space to be allocated to each considered uses. MSP broadly diffused over the last 10-15 years to eventually emerge as the dominant marine management paradigm to reach the so-called sustainable development. It has been widely promoted by academics, practitioners, and policymakers as a process for implementing ecosystem-based management in the marine environment, reducing user conflicts, enhancing environmental protection, and facilitating maritime economies' expansion (Douvère 2008; Ehler and Douvère 2009; Brouwer et al. 2016). MSP is commonly expressed as a step-by-step process (*cf.* Figure 1.1) involving stakeholder participation to increase the chance of the MSP plan to be accepted. In this work, we mostly focused on the steps 5, 6 and 7.

1.1.2 Decision support tools

MSP strives to be a rational and evidence-based process (Pınarbaşı et al. 2017). MSP is cognitively a complex process due to the amount and heterogeneity of available information (*e.g.* oceanography, biology, economics, laws, politics, existing uses, futures uses, *etc.*). In this framework, rooted in data analysis, decision support tools (DSTs) turned out to be indispensable

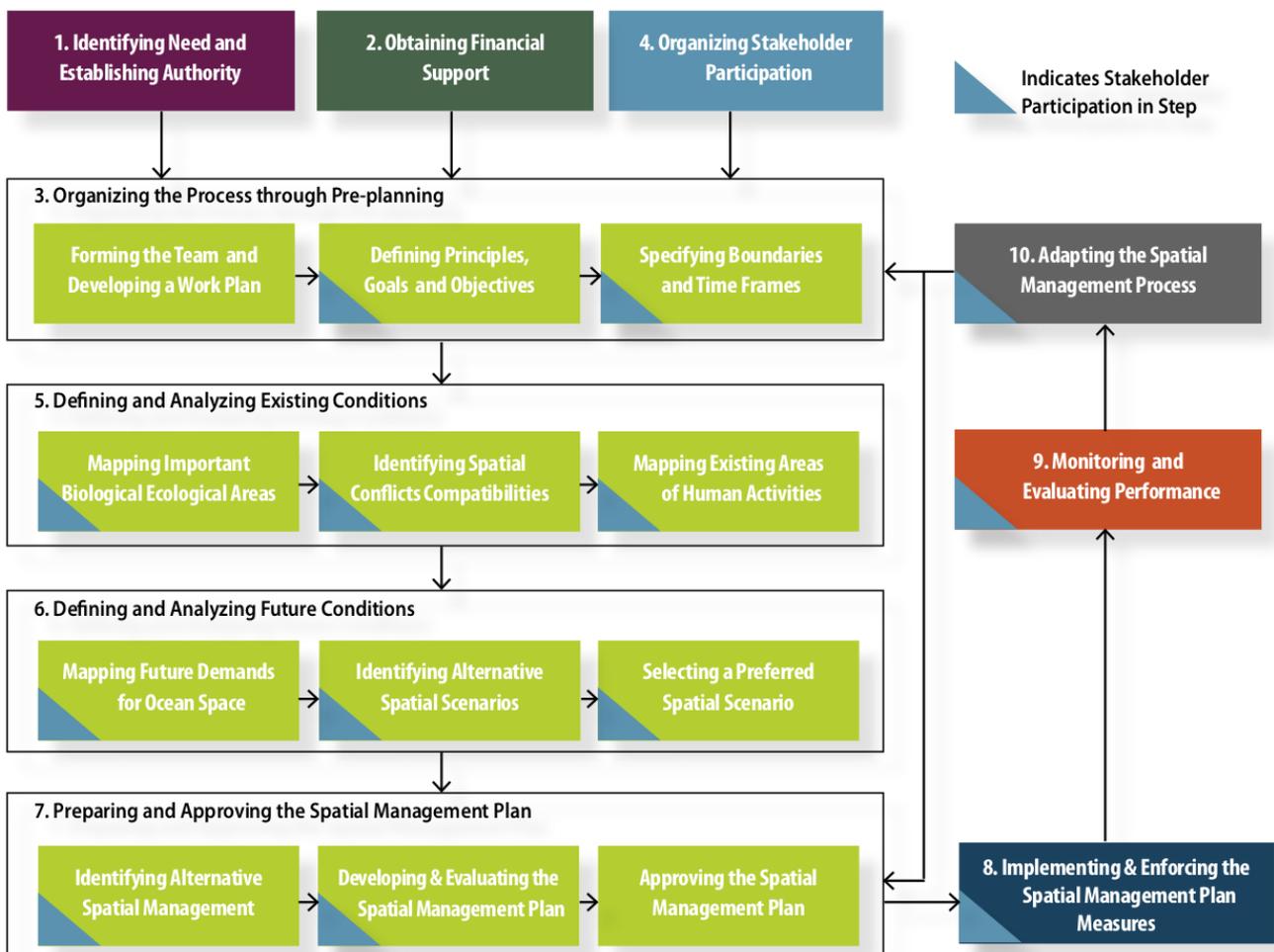


Figure 1.1: Step-by-step MSP process. This figure was extracted from (Ehler and Douvère 2009).

for rationally informing the decision-making process. DSTs take the form of spatially explicit tools, involving interactive software comprising maps, models, communication modules and additional elements that can help to solve multifaceted problems that are too complex to be solved by human intuition alone or by conventional approaches. With the help of these tools, support for decision-making can be undertaken in a more systematic and objective manner.

The number and types of DSTs have grown continuously. Main DSTs involved in MSP were catalogued in (Coleman et al. 2011) where one can find details (goal, method, input, output, etc.) about ARIES, Atlantis, Coastal Resilience, Cumulative Impacts, InVEST, MarineMap, Marxan with Zones, MIMES, Multipurpose Marine Cadastre. This useful decision guide aims to empower planners to swim among the profusion of DSTs in order to be able to choose one depending on the problem tackled. It also underlines the difficulty of developing a marine spatial planning process. That is why we choose in this work to focus on an important aspect of marine spatial planning: the systematic planning of conservation areas.

1.1.3 Protected areas

Conservation institutions identify protected areas as an essential part of the solution to ensure the resilience of marine ecosystems. For instance, the United Nations (UN) target for global ocean protection was established to 10% of the surface of coastal and marine areas in marine protected areas (MPAs) by 2020, as set forth by Aichi Target 11 under the Convention on Biological Diversity. Although we focused on MPAs, this work may still be relevant for terrestrial

reserves. The UN Sustainable Development Goal 14 reaffirms this commitment. Going further, many scientists emphasised that the 10% target was intended as a first milestone for global ocean protection, rather than an endpoint. In that respect, International Union for Conservation of Nature (IUCN) members, composed of governments, non-governmental organisations and agencies, agreed on an ambitious protection target of 30% of the surface of each marine ecoregion by 2030 (“IUCN World Parks Congress 2014 Bulletin”, “IUCN Congress 2016 Bulletin”), against less than 8% observed today and less than 2% before 2008¹. More recently, the European Green Deal aims at covering 30% of the European seas surface with MPAs, among which 10% under a strict protection. *De facto*, MPAs were proved to provide biotic communities global benefits (*e.g.* (Stolton and Dudley 2010)) especially when the reserve is strict (no access, no removal) (Claudet et al. 2020; Liu et al. 2017). The development of coastal and ocean-related human activities, the biodiversity crisis and ongoing global changes have led to a relative consensus on the need to preserve the integrity and functioning of exploited ecosystems. This gives the conservation science a prominent place to address nowadays’ challenges. DSTs for reserve design have therefore rapidly become an issue for research and use at the global level, including addressing MSP issues. Designing MPAs is a small but essential part of MSP processes (Vaughan and Agardy 2020).

1.2 Conservation science

1.2.1 Environmental ethics

Environmental ethics is a branch of philosophy which takes roots in the United-States, Norway and Australia in the early 20th. The 1960s saw the academic birth of the discipline, thus formalising previous debates. It still remains an active field in the face of current environmental crises. As defined in (Brennan and Lo 2002) and published in the Stanford Encyclopedia of Philosophy, "Environmental ethics is the discipline that studies the moral relationship of human beings to, and also the value and moral status of, the environment and its nonhuman contents". In other words, environmental ethics critically examines the relationship between mankind and other biotic and abiotic communities, seeking for ethics rules which can guide human behaviour and actions related to the environmental protection. In particular, environmental ethics address the essential questions underlying human conservation actions: why should we care for and protect the environment? What could be the moral status of nonhuman entities justifying their protection? For instance, to what extent humans are morally obliged to prevent the extinction of given endangered species, especially if it does not imply any direct interest? Do biological features hold an intrinsic value worth preserving? Is there space for a non-anthropocentric environmental ethics leading to environmental protection? These questions lie within the scope of environmental ethics and appear particularly relevant regarding conservation science if not at its very foundation. Consequently, we provide a brief overview of the various schools of thoughts of the environmental ethics and highlight the global consensus justifying and guiding the conservation science. Our review aims to enrich our comprehension of the ethical principles upon which conservation science and policies are built.

In order to deepen our understanding of the emergence of conservation actions and the underlying environmental ethics, it is necessary to go through an historical review. As depicted in (Baer 1968), humans initially nomads lived in synergy with the environment, immediate needs being satisfied thanks to the surroundings. Agriculture and animal breeding allowed humans to become sedentary and increased their survival. Their sedentary behaviour was inherently

¹<https://www.protectedplanet.net/marine>

accompanied by habitat destruction. Besides, it favoured and enhanced the vision of a controllable and exploitable nature, which was from now on considered as a distinct entity. In the late 18th, the Enlightenment movement enshrined an anthropocentric and demiurgic vision of the world. It eventually led to the 19th century climax of this vision, namely the industrial revolution and colonialism. Facing social and environmental destruction, the awareness regarding the need for nature protection emerged in western societies. Romanticism especially illustrated this trend, placing pristine nature on a pedestal as an aesthetic and spiritual source. Romanticists idealisation and personification of nature resulted in conferring it an intrinsic value, although still acknowledging the nature/human dichotomy. Paradoxically, (Baer 1968) underlines that travels and tourism probably take its roots in romanticist nature promotion, while being a great threat for the environment today. Nevertheless, a global awareness emerged from bitter observation regarding human destructive activities (monocultures, intensive breeding, urbanisation) and eventually led to first intended initiatives in nature protection as explained in (Harroy 1969) and later described in Section 1.2.2.

Let us go through a quick description of early debates among naturalist thinkers, notably depicted in (Norton 1991), which will later influence environmental ethics works. Although both inherited from the nature/human dichotomy, two major competing views emerged in the late 19th: "Preservationist" versus "Conservationist" respectively embodied by John Muir and Gifford Pinchot. On one side, preservationist Muir advocates to keep human away from pristine nature as much as possible in order to preserve its inherent aesthetic and spiritual values (Muir 1913). Muir ethics is clearly inspired by romanticists. It can also be linked with the American naturalist poet Henry David Thoreau who defended an intrinsic *raison d'être* of nature beyond economic gains, and shared his humility behind wilderness and complex life forms in a contemplative Rousseau style (Thoreau 1854). On the other side, conservationist Pinchot pled for a sustainable resource management compatible with human interests, thus promoting a more utilitarian and anthropocentric approach (Pinchot 1910). Philosophically speaking, Pinchot can be associated with Kant positions which are strongly anthropocentric and only acknowledge intrinsic value to rational agents, *i.e.*, *Homo Sapiens*. This early debate, still flowing within contemporary thoughts, illustrates both side of a continuous spectrum regarding environmental ethics guiding conservation actions. Indeed, it crystallises tensions between anthropocentric/non-anthropocentric views and perfectly illustrates the question whether to assign natural features an intrinsic value, *i.e.*, outside of human interests.

Now we have established extreme positions regarding environmental ethics, it is important to mention other main schools of thoughts that can be roughly located on the preservationist/conservationist scale. Before all, further environmental ethics developments emerged following Rachel Carson work (Carson 1962) which can be considered as the birth of the academic discipline (Brennan and Lo 2002). Indeed, Carson drew general awareness regarding ecological degradations and pointed it out as a public health topic. The "biocentrism" school of thoughts appeared in reaction to strong anthropocentric doctrines. Biocentrism rejects a supposed human moral superiority and treats all other living organisms on an equal foot. Biocentrism thus confers any living form an equal intrinsic value. Consequently, all living forms imply a moral duty and are worth protecting. However, biocentrism individualist ethics fails to apprehend ecological processes and communities equilibrium. This lack can potentially lead to ineffective conservation actions. For instance, it is sometimes compulsory to prevent by all means invasive species to exist because they could endanger the whole ecosystem integrity. That is why Aldo Leopold developed an inherently holistic position (Leopold 1949), namely its "Land Ethic" also known as "ecocentrism". Ecocentrism is inherited from contemporary theoretical developments in ecology, especially regarding species interdependence, that even *Homo Sapiens*

cannot escape. Therefore, ecocentrism places ecosystems integrity as a guiding end and can be condensed in Leopold following principle: "A thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise". Arne Næss developed his theory of "deep ecology" which repudiates the anthropocentric vision and advocates for the intrinsic valuation of the biosphere, calling for biodiversity preservation and close partnership with other living forms (Naess 1973). Deep ecology is opposed to what Næss called "shallow ecology" which limits its action in reducing industrial damages or resources depletion and considers the environmental crisis as a technical problem. Proponents of deep ecology pleads for an humble human life, limiting extractive activities to essential needs. Murray Bookchin's "social ecology" analyses the current environmental crisis as a mere consequence of other structural domination forms whether based on gender, ethnicity or social class (Bookchin 1971; Bookchin 1982). Through his words "man's obligation to dominate nature flows directly from man's domination over man", he is the first to intimately link ecological issues with political organisation. Social ecology aims at deconstructing our essentialised extractive relationship with nature, holding capitalism inherently accountable for ecological imbalances. According to Bookchin, any attempt to overcome the environmental crisis that is unable to question power distribution and political foundations is doomed to failure. Social ecology thus condemns "green capitalism" which blames individuals instead of structures. Also, Bookchin harshly criticizes deep ecology arguing this movement, although claiming its anti-capitalist origin, supports strong social hierarchies based on a social Darwinism inherited from a Malthusian tradition (Bookchin 1987). Social ecology, strongly anti-capitalist, advocates for radical social changes to overcome the environmental crisis. His ethics is accompanied by a well-defined political baseline known as "libertarian municipalism" or "communalism". Indeed, Bookchin promotes a decentralised cooperative confederation of free and self-directed communes in harmony with surroundings ecosystems, primarily focused on human welfare. In the same line, Françoise d'Eaubonne developed the "ecofeminism" in 1974 to describe an environmental ethics which apprehends ecological destruction and nature/human dichotomy through the prism of a structurally oppressive relationship. Consequently, ecofeminism directly benefits from theoretical feminist works and analyses. Ecofeminism identifies common domination mechanisms between human-centered and male-centered societies, which can help highlight legitimisation processes of the dominant group over the dominated group. More notably, male-centered societies consider women as exploitable entities similarly as anthropocentric ones with nonhuman features. Besides, it is argued patriarchy plays a particular role among other domination forms as probably the first mechanism from which fuels others, a point also supported by Bookchin's social ecology. In conclusion, several schools of thought, such as biocentrism, ecocentrism, deep ecology, social ecology and ecofeminism among others, appeared in the 20th devoted to the creation of a relevant environmental ethics and influencing civil society, governments and scientific literature.

If an historical winner of the Muir/Pinchot debate had to be named, it would certainly be Pinchot. Indeed, conservation science and policies tend to acknowledge and even essentialise our economic relationship with nature, although they aim at organising extractive activities sustainably. Consensual and politically privileged, the conceptual framework of sustainable development and ecosystem services is still today rooted in conservationist ethical positions. Therefore, the environmental ethics supporting the conservation science has a strong conservationist baseline. For example, MSP has the clear objective of organising the sustainable development of human activities at sea. An element of such organisation goes through designing nature reserves, *i.e.* areas with a minimal human influence. Paradoxically, this element is clearly from a preservationist heritage.

Debates among scientists are still quite tensed today regarding the definition of conservation science and oppose two competing views. On one side, the quite canonical and traditional conservation vision depicted in (Soule 1985). On the other side, the extremely pragmatic and anthropocentric vision carried by Kareiva, the former chief scientist of The Nature Conservancy (TNC) (Kareiva and Marvier 2012; Kareiva et al. 2011). The answers to this view (Soulé 2013; Doak et al. 2014) exposed the division among the conservation science community. Such academic debates help to maintain the global confusion about conservation actions, which is very useful to slow down any conservation process. Paradoxically, Kareiva received a huge media coverage and is depicted as an "out-of-the-box" and "provocative" "thinker" while his position exactly reflects the strong anthropocentric ethics consensual among western societies in systems wholly based on economics. Finally, since the academic birth of environmental ethics, a lot of papers addressed the question of which environmental ethics should support and justify conservation efforts. Several elements were given, especially in (Norton 1994; Norton 1984; Maris 2010) which remain still relevant today and tend to question conservationist positions. In particular, (Maris 2010) built an environmental ethics based on five potentially contradictory principles (autonomy, humility, responsibility, goodwill, diversity).

The methods addressed in this thesis are based on conservationist ethics since this work is positioned within the conservation literature. Instead of refusing these tools associated to a conservationist vision, we pragmatically focused on improving the methods in order to provide a more transparent framework and hopefully contribute to more equitable discussions between stakeholders involved in MSP processes. Besides, we chose in this thesis to deal with the methods of reserve site selection, which is strongly associated with a preservationist vision.

1.2.2 Emergence of conservation within science and institutions

In this section, we discuss how conservation science as we know it today began academically thanks to international cooperation. Although undoubtedly essential today in the light of the biodiversity crisis and climate change context (*cf.* Section 1.1), conservation science remains quite young as an academic discipline and find its roots in the late 1960s. Indeed, the creation of American *Conservation Biology* journal in 1987 enshrined it as a structured academic discipline even though European *Biological Conservation* journal existed since 1968, especially with important Australian theoretical work (Sarkar 2004; Odenbaugh 2016). Nowadays, many journals with broad topics (*Nature*, *Science*, *Trends in Ecology & Evolution*, *Frontiers in Ecology and the Environment*, *Environmental Modeling & Assessment*, *BioScience*, *Biodiversity and Conservation*, *Marine Policy*, *Environmental Conservation*, *Journal of Biosciences*, etc.) publish conservation-related papers betraying a major interest towards conservation science.

If conservation was primarily a scientific concern, it later became an important political issue motivated by various environmental ethics (*cf.* Section 1.2.1). An historical pathway can be found in (Baer 1968; Harroy 1969). Early conservation actions were unintended and uniquely motivated by private interests. For instance, hunt or forest reserves were built to ensure respectively leisure or economic monopolies. These protected domains were nonetheless quite efficient regarding conservation. In the late 19th, local and intended conservation efforts successfully took place around the world through what will be later called "National Parks" (Fontainebleau (France, 1853), Yellowstone (United-States, 1872), Royal National Park (Australia, 1879), Rocky Mountains (Canada, 1885), Tongariro (New-Zeland, 1887), Monte Vedado de El Chico (Mexico, 1898), etc.). A few coordinated international initiatives existed in the early 20th and eventually failed, weakened by world wars and economic crises. Indeed, under the impulsion of the Swiss pioneer Paul Sarasin at the 8th International Congress of Zoology

held in 1910, the principle of a transnational cooperation focused on nature protection was widely accepted and confirmed in 1913 in Bern during the "*Conférence Internationale pour la Protection de la Nature*". As mentioned before, world war 1 unfortunately aborted this promising international cooperation effort. Further attempts knew the same faith due to the 1929 economic crisis and world war 2. Although international cooperation collapsed, this period witnessed a worldwide multiplication of protection attempts at the national scale (Japan, Chile, Argentina, Indonesia, Rwanda, Congo, etc.). The first stable operational international institution devoted to nature protection only emerged after world war 2. Indeed, the International Union for Protection of Nature (IUPN) was created in 1948 during the Conference of Fontainebleau, renamed IUCN in 1956. Note the semantic sliding, from "protection" to "conservation", betrays the underlying ethical debate between preservationist and conservationist, sketched in Section 1.2.1. The IUCN missions are numerous among which we can mention government lobbying, conservation actions coordination between nations, worldwide data collection and publication, public sensitization. The second half of the 20th century witnessed the creation of other notorious international organisations such as International Waterfowl & Wetlands Research Bureau (IWRB) in 1954 and World Wildlife Fund (WWF) in 1961.

1.2.3 Missions and motives

A "Conservation Biology" definition is proposed in (Sarkar 2004) and updated in (Odenbaugh 2016) both found in the Stanford Encyclopedia of Philosophy. The work in (Soule 1985) is often considered as a canonical and kick-off paper for modern conservation science. The author attempted to answer "What is Conservation Biology?" which deserves a deeper attention. Soulé drew a guideline for future conservation scientist by providing specificity of the discipline and underlying motives. He depicted conservation science as a crisis-oriented science arguing the need for immediate action although based on incomplete and uncertain knowledge. He argued that conservation science is inherently and pragmatically holistic. This holistic vision is inherited from the theoretical developments in ecology but also through its multidisciplinary aspect. Indeed, conservation science is inherently multidisciplinary in order to support as rationally as possible the decision process requiring knowledge from ecology, geography, sociology, economy, mathematics, management policies, *etc.* The denomination "conservation biology" has today become "conservation science", expressing a broader coverage not restricted to biology. Historical journal names such as *Conservation Biology* and *Biological Conservation* are the remnants of this semantic change. Soulé divided motives behind conservation science in two categories: functional and normative postulates (Soule 1985). The functional postulates highlight species interdependence, functions importance, but also that species, habitat and processes have a functioning range. Outside this range, they irreversibly disappear. The normative postulates advocate for a protection due to diversity intrinsic value in the ecocentrist tradition, while human benefits of ecosystems are also acknowledged. This dichotomy appears when someone argues that "conservation is useless". This statement is obviously wrong, because human depends on nature in so many ways: a genetic pool for potential and actual developments of medicine, an epidemic barrier, a pristine and natural laboratory to study modification due to humans (Jenkins and Bedford 1973), an outdoor classroom for pupils and researchers, a source of inspiration for artists and scientists, a museum of evolution and diversity, a renewable food and energy stock, cleaning air and water, *etc.* But these instrumental values are hopefully not the only reasons why people contribute to conservation. The usefulness of conservation is always highlighted, especially on the political ground, due to a long conservationist ethics tradition. Yet, uselessness can also be defended on the ethical ground, which raises the question of the intrinsic value of nonhuman entities. Intrinsic values are often conferred to ecosystems and can solely justify their conservation. In a less political example, we can find a similar

statement with the famous claim "mathematics is useless". Yet, mathematics is clearly useful and indispensable to many technologies we use every day. However, mathematicians often do mathematics for the love, fun and beauty of it. This motivation dichotomy is historically illustrated: nature preservation was initially mostly justified for aesthetic and spiritual values; today, the sustainable development framework and ecosystem services vision prevail behind the political motivation and thus conservation funding. Another striking example is the systematic highlight of nature instrumental values in the introduction of academic works to justify the relevance of the work. Yet, it is possible to work in favour of conservation for non-instrumental values.

Soulé concludes about conservation biology fundamental role regarding the decrease of diversity destruction rate, which can lead to huge threats and mitigate human impacts through iterative management actions. Conservation core mission is to ensure diversity persistence of both biotic and abiotic communities in a context of strong human disturbance and associated imbalances. Diversity leads to stability at a community/ecosystem level (equilibrium dynamics, resilience of ecosystems) (Soule 1985; Odenbaugh 2016; Holling 1973; Weise et al. 2020). Conservation science thus aims at preserving long-term equilibrium of ecosystems which are subject to human disturbances. More operationally, the biodiversity persistence mission also raises the question of biodiversity. What is biodiversity? Which relevant surrogates? It is a practical vague notion, as discussed in (Sarkar 2002; Sarkar and Margules 2002; Sarkar et al. 2006).

1.2.4 Early works (1960-1980)

1.2.4.1 SLOSS debate

As explicitly stated within the name, the Single Large Or Several Small (SLOSS) debate questions whether a single large area would have a greater specie richness compared to several small areas of the same surface. This 1960s debate emerged within the island biogeography scientific field. This field benefits from perfect natural laboratories: oceanic islands. Island biogeography focused on quantifying both a static link between island area and species number, but also a supposedly existing dynamic equilibrium between species extinction and immigration rate (MacArthur and Wilson 1967). Naturally, the question about most favourable shapes of islands network created the SLOSS debate. Why it is relevant regarding conservation science? Actually, islands are thought to be perfect metaphors for nature reserves (Diamond 1975; May 1975). Indeed, it is assumed that such places can be considered as a network of safe "islands" separated by unwelcoming habitats. Consequently, island biogeography scientists historically contributed to and influenced conservation early works, especially at a time when national parks began to flourish worldwide as mentioned in Section 1.2.2. Questions about relevant reserve design such as site selection or shapes thus appeared. Therefore, Diamond (1975) and May (1975) draw important principles and conclusions for reserve design derived from biogeography studies. In particular, Diamond (1975) suggested precious geometric principles regarding nature reserve (*cf.* Figure 1.2). The purpose was not to take an active side in the SLOSS debate, but mostly recommend for close if not connected reserve network in order to increase the immigration rate between ecosystems. Besides, although (Diamond 1975; May 1975) acknowledged a single large area would be more favourable than several small areas for holding species, restricted budgets and human colonisation forbid large reserves. Even if it would be possible, (May 1975) argued it would be hazardous to put "all eggs in one basket" in the light of potential catastrophic events (epidemic, fires, *etc.*). Recommendations derived from island biogeography theory were incorporated into the World Conservation Strategy (IUCN, 1980). However, experiments did not validate the equilibrium theory of island biogeography between immigration and extinction rate (Simberloff 1976). This experimental failure raised criticisms,

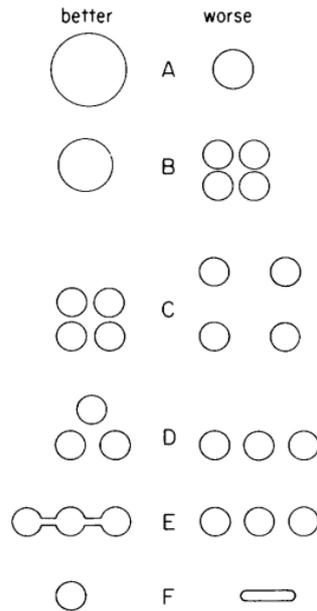


Figure 1.2: This figure extracted from (Diamond 1975) illustrates suggested geometric principles for the design of nature reserves. For each proposed shape, labelled from A to F, left design results in a lower extinction rate than the right one.

since reserve design properties were inferred using the analogy with island biogeography theories. The mere idea to search a general theory is tackled in (Simberloff and Abele 1976) as a given island network shape can favour some taxa and disadvantage other. In the continuation, (Soulé and Simberloff 1986) closed the SLOSS debate by excluding any generalisation since the answer is highly local-dependent. These questions are still central research axes in the modern conservation science.

1.2.4.2 Conservation value

Conservation science thus inherited from the island biogeography literature, especially through the SLOSS debate which stated quite valuable lessons regarding the spatial attributes of nature reserves. The question of site selection and general reserve design later emerged, because limited resources are usually available for conservation. Consequently, such research themes exposed the underlying question of "conservation value" associated with a given delineated area. For instance, a parcel dedicated to oil palm monoculture is ecologically worth less than an equal size parcel of primary forest containing hundreds of species and diverse habitats. In other words, all locations are not equal regarding the conservation interests. One of the primary scientific concern was to compute a conservation value, monetary or not, for the considered areas. The urge to value wildlife appeared as an operational response in our economy-based societies. Wildlife valuation is often defended for operational purposes, since the economic appeal is assumed to be the best motive for the public and thus conservation policy. Yet, maintaining a positive view of oneself was proved to be more efficient than economic reasons to obtain a behaviour change (Bolderdijk et al. 2012). This wildlife valuation can be imputed upon several factors: prioritize conservation actions in a limited resources context, weight quantitatively conservation with respect to other competing land-use, drive decided destruction and degradation towards the least ecologically worthy area, inform decision in a reserve selection process. The literature interest in conservation value computation illustrated once again the prevalence of Pinchot conservationist positions over preservationist's. Anyway, early effort of conservation biology in the 1970s were animated by wildlife valuation, which broadly diffused and structured literature

research questions.

Early initiatives developed a lot of effort to express a conservation value in monetary unit (Helliwell 1967; Helliwell 1969; Westman 1977) to eventually compete directly with other land-use such as agriculture, recreational activities, construction, *etc.* Although Helliwell focused on conservation monetary evaluation, he also recognized a potential use for comparative purposes and opened perspectives in the conservation literature through standard scoring systems. Indeed, he first proposed a standard although subjective approach: several factors (landscape position, light exposition, species presence, *etc.*) were scored on a 1 to 4 scale corresponding to a monetary equivalent (Helliwell 1967). The multiplication of every factor value eventually gave a price for trees and woodlands. Two years later, a more detailed process relying on the same principle is described (Helliwell 1969). In particular, wildlife benefits are divided in categories (education, recreation, actual and potential production) where seven factors (direct returns, genetic reserve, ecological balance, educational value, research, natural history interest, local character) are evaluated through a scoring system. A weighted sum allowed to eventually assess the overall conservation value. It is worth noting the author acknowledged the "horror" such process can provoke within the scientific community (Helliwell 1967) while today it is a common framework which has widely diffused among research, public and above all conservation practices (Reducing Emissions from Deforestation and Forest Degradation (REDD) policy, MSP, ecosystem services, *etc.*). Also, such ideas appeared extensively developed within an intern TNC documentation (Helliwell 1971) which shows how early this private organisation influenced conservation practices. Others took the side to avoid a monetary framework but kept the idea of systematic scoring for wildlife valuation (Tubbs and Blackwood 1971; Tans 1974; Gehlbach 1975; Goldsmith 1975; Wright 1977). Indeed, the assessment of a conservation value can serve a comparison purpose between studied areas. For example, conservation scores ranking can help decision makers to prefer a given area with respect to another when a reserve is selected or when the nature removal/damage is decided for exploitation. Obviously, the ultimate goal of systematic methods for conservation value computation is the first step to eventually feed priority ranking algorithms for reserve site selection (Tans 1974; Gehlbach 1975). Anyway, whatever the chosen unit (arbitrary or monetary), systematic scoring systems have been provided and improved accounting for various criteria in order to offer the most transparent method. In particular, the scoring systems literature review in (Wright 1977) provided a quite exhaustive list of criteria contributing to the ecological score of considered areas: species and habitat attributes (rarity, richness, diversity, *etc.*), area features (size, accessibility, availability, proximity with disturbances, threat), research and educational use, management features, *etc.* Another improvement of conservation practice is found in (Goldsmith 1975) where, motivated by a practical case study, the author brought a mapping effort as he divided the studied area in a 1 kilometre square grid. Aggregating scores through a weighted/non-weighted additive/multiplicative operation allowed to go from a multi-dimension scoring space into a scalar conservation value and occasioned debates within the community. Scientists commonly acknowledged the inherent subjectivity of such scoring systems (criteria choice and shades, relative weights, final aggregation, *etc.*). However, systematic methods, unlike arbitrary expert opinion, have the merit to provide a transparent and defensible framework opened for improvement and critics. Besides, providing a standard survey is a great tool for amateur naturalist allowing large data collection precious for conservation (Gehlbach 1975). An enlightening review on conservation value assessment can be found in (Margules and Usher 1981).

These works on conservation value, contemporary with environmental ethics developments, witnessed and enshrined an emerging conservation paradigm: nature is a service provider and

only valuable relatively to human benefits, which is why conservation is deserved. Illustrations of this paradigm can be found in (Westman 1977) as the author wondered in *Science* "How Much Are Nature's Services Worth?" but also in (Possingham and Shea 1999; Bayon and Jenkins 2010) homonym papers. Few voices pleaded otherwise, *i.e.*, in favour of defending nature intrinsic values, although such values are often highlighted in the conservation literature. Nevertheless, Ehrenfel (1976) raised relevant criticisms and objections regarding the conservation value framework. Although Ehrenfel (1976) acknowledged these valuation methods are a way to fill a gap in western policies demands bathing into an economy-based society, he affirmed a cost/benefit framework remains unsatisfactory especially towards what he called poetically "non-resource" and advocated for an intrinsic value of biological features. For example, the mere and unimpeachable species right to exist as evolutionary partners is defended. Started from the Houston toad example, the author observed that majority of biological features factually have neither directly nor indirectly demonstrated value for humans and thus fall into the non-resource category. However, the urge to protect these features exists despite motivation towards its preservation is often distrusted in western societies when in competition with another land-use. A solution could be to find a subterfuge to assign this "useless" feature an economic value and paradoxically transform it into a resource, thus opening perspectives for a protection. To do so, more or less anthropocentric reasons can be invoked (such as aesthetic, educational, research, recreational, evolutionary, avoid irreversible change, undiscovered values, *etc.*). Nevertheless, he argued such approach is not a relevant solution because it will start an inevitable cost/benefit competition with respect to other land use, not particularly favourable for conservation, which can even lead to legitimate irreversible removal. Indeed, several reasons are provided: short term gains are often superior to long term sustained yield gains, benefits highly depend on market change, conservation values are hard to identify and quantify, information imbalance disadvantages conservation. Besides, Ehrenfel advocated that developments of scoring systems and associated ranking algorithms can reverse the governance paradigm. Indeed, as such tools are available for planners, there is a risk to systematize identification of areas which can be dedicated to human exploitation, thus encouraging and even legitimating ecosystem removal. Despite these risks and warnings, these tools are used today to identify areas of human activities. Our work therefore seeks to open the black box of these tools, improve them if necessary, in order to make them as transparent as possible to an end user.

1.3 Operations research

1.3.1 General overview

Operations research is a field of mathematics covering a set of methods that seek to develop the best decisions for a given problem. Since we seek the best decisions, optimisation methods are particularly central to this branch of mathematics and are the object of our interest in this work.

It should be noted that the name "operations research" comes from the military "operations" of world war 2 when the discipline experienced a particular boom. Today, the applications of operations research are multiple and particularly suited to the capitalist issues of our societies (capital investment, production planning, commodity stock, employees management, *etc.*). Our work is not related to these issues, as we use these methods for reserve design. Indeed, in the reserve site selection problem, we apply such methods to formulate and solve an ecological problem. We seek to determine how to allocate a conservation resource, *i.e.* a space free of anthropogenic pressures, in order to best protect biodiversity.

What is optimisation? Optimisation is a branch of mathematics that seeks the value of a deci-

sion variable among possible values that minimises or maximises a given function. Thus, an optimisation problem can always be expressed as an objective function and inequality constraints. More mathematically, one can generically formulate an optimisation problem as follows:

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{X} \end{cases}$$

In other words, we seek to determine the value, under the existence hypothesis, of the decision variable $x \in \mathcal{X}$ which minimises² the objective quantity $f(x)$. Classically, it is by a set of constraints to which the variable x is subject that its set of admissible values \mathcal{X} is drawn.

Furthermore, optimisation problems are categorised according to the nature of the set \mathcal{X} and the objective function f . For example, if \mathcal{X} and f are convex, it is called convex optimisation. If the objective f is linear and \mathcal{X} can be deduced from a set of linear constraints, we speak of linear optimisation. In particular, convex or linear optimisation knows existence theorems for the optimum and relatively efficient solving methods. If the set of admissible values \mathcal{X} is discrete (respectively continuous), then we speak of integer optimisation (respectively continuous). If the set of admissible values \mathcal{X} has both continuous and discrete components, then we speak of mixed optimisation.

1.3.2 Integer linear programming

In this work, we have a specific interest for integer linear programming. When the decision variable of an optimisation problem can take integer values, we speak of integer programming by opposition to continuous programming. The "linear" adjective specifies that the objective function and the constraints involved are linear in the decision variable. In our case, the optimisation problem described later in Section 1.4.2 involves a decision variable that can take a value of 0 or 1. This binary choice reflects whether to include or not a specific delineated area within a reserve. The decision variable is $x \in \{0, 1\}^N$ where N is the number of units resulting from the study area division. Naively, we could think of this problem as simpler than continuous programming because we «only» have to compute every possibility for $x \in \{0, 1\}^N$, which is a finite number (equal to 2^N), and take the smallest value among the resulting finite set. And this would be, by nature, an impossible task impossible with a continuous decision variable. However, a finite set unfortunately does not mean computer sized. We speak of combinatorial explosion. Indeed, for $N \geq 266$, the number of evaluations of f is greater than the number of atoms in the universe ($\sim 10^{80}$). For instance, in our real-world study case, $N = 756$ which corresponds to more than 10^{227} possibilities for x . Besides, solving the associated relaxed problem (*i.e.* allowing x to explore the smallest continuous set including $\{0, 1\}^N$) and taking the closest integer solution neither theoretically nor practically guarantees to find a relevant solution (*cf.* Figure 1.3). Finally, unlike continuous programming, the derivative of f does not have any sense in discrete programming while at the foundation of continuous optimisation solvers.

For example, let us consider the following simple optimisation problem:

$$\begin{cases} \max_{x,y} & x + y \\ \text{s.t.} & 86x + 100y \leq 1155 \\ & -13x + 10y \leq 82 \\ & 11x - 50y \leq 0 \\ & x, y \in \mathbb{N} \end{cases}$$

²Without loss of generality, we speak of minimisation. Indeed, if we seek x to maximise a quantity $f(x)$, this is equivalent to minimising its opposite $-f(x)$.

An illustration of this problem can be found in Figure 1.3. Numerically, the unique continuous solution is $(\frac{385}{36}, \frac{847}{360}) \simeq (10.69, 2.35)$ with an objective value of approximately 13.05. The integer solutions are $\{(4, 8), (5, 7), (6, 6), (7, 5), (8, 4)\}$ with an objective value of 12. Importantly, the closest integer solution from the continuous solution, namely $(11, 2)$, is not even in the domain of admissible values.

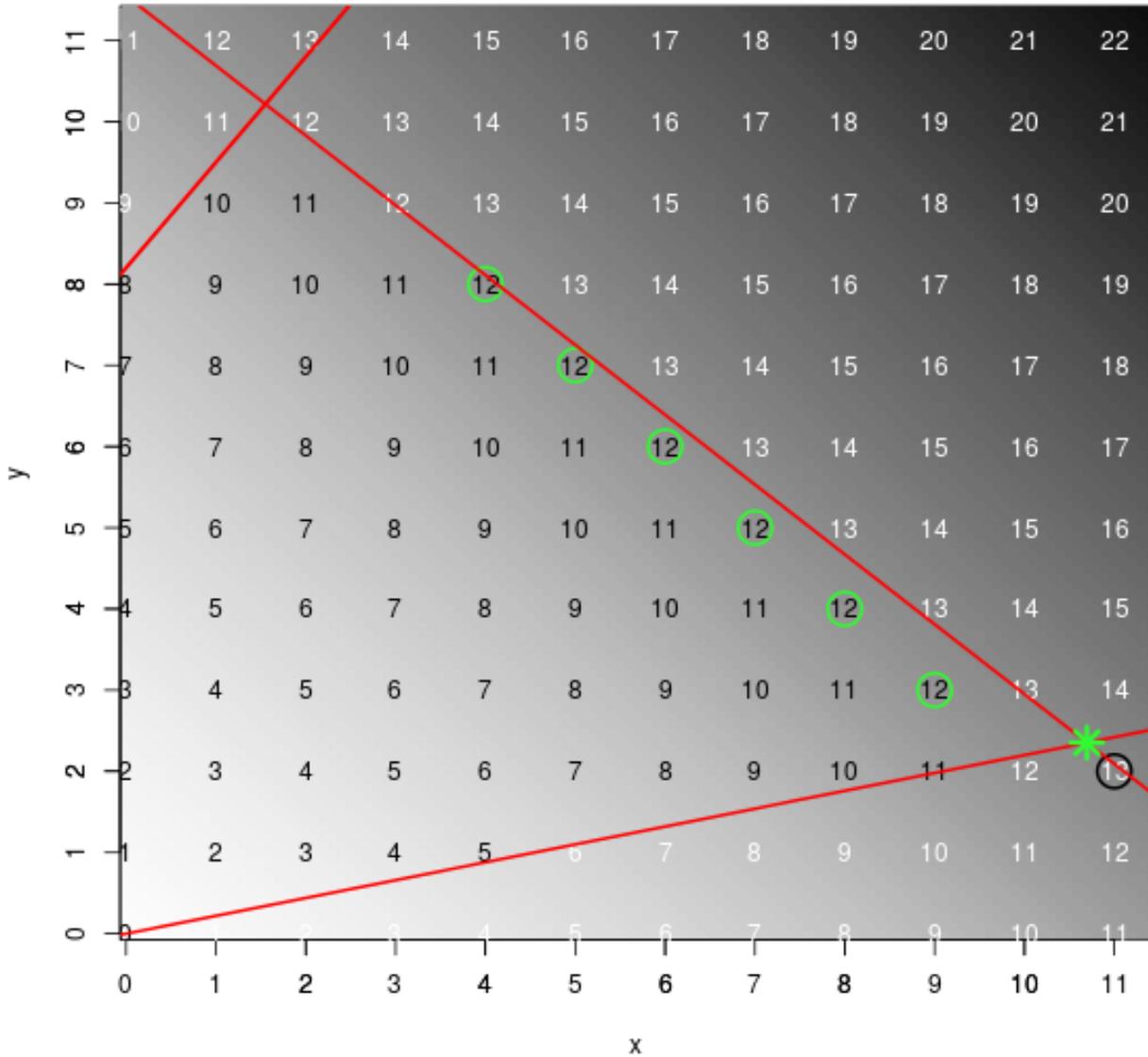


Figure 1.3: Example of a two-dimension linear optimisation problem. The values of the objective are represented with a grey shaded gradient. Admissible values for continuous decision variables x and y are delineated by linear constraints represented with red lines. The values of the objective function evaluated for integer x and y are written in black inside the admissible domain and in white outside. The unique solution of the continuous problem is represented by the green star. The closest integer solution is circled in black. The solutions of the discrete optimisation problem are circled in green.

1.3.3 Illustration examples

1.3.3.1 Hiking backpack

Let's put ourselves in the place of a hiker preparing his backpack, and more particularly his meals, in order to carry out a week in autonomy. His objective is to pack as light as possible (*cf.* Figure 1.4). But of course the hiker is obliged to cover his, supposedly known, protein, fat and carbohydrate needs for the week's hike. The hiker's decision is therefore based on whether to put each food item in his backpack. In practice, every hiker sets up a set of simple rules, more or less intuitive, based on the weight of the food and its nutritional quality until he finds a solution that satisfies him, *i.e.* an acceptable backpack weight for his hike. However, our hiker has no guarantee that this is the best solution, *i.e.* the lightest backpack. And it is to this question that mathematical optimisation tools allow us to calculate the lightest backpack that will cover his needs for the 3 nutrients considered, namely proteins, fats and carbohydrates. Selecting sites to be integrated into a protected area is exactly the same thing: we want to find the "lightest" reserve (for example the smallest, or the least expensive) that will satisfy our "nutritional needs", in this case contain/recover enough of the species we wish to protect.



Figure 1.4: Example of the integer optimisation problem for preparing a hiking backpack. Which food items to include in our backpack in order to cover our known 1-week needs for the three considered nutrients (proteins, carbohydrates, fats)?

1.3.3.2 Goats and vegetables

Let us imagine green cabbages, red cabbages and carrots growing in some goat pens. Data feeding the problem is qualified as «spatially explicit» because we can count and locate goats and vegetables. Naturally, if goats are free to access their pen, they will probably eat every vegetables. However, we would like to establish a conservation plan to protect an *a priori* defined amount of green cabbages, red cabbages and carrots. Thus, we wonder which pens are worth forbidding to goats in order to protect enough vegetables and impact the least goats. Practically, let us imagine four pens (labelled A,B,C,D) with goats and vegetables distributed as shown in Figure 1.5. We want to protect at least 3 green cabbages, 1 red cabbage and 1 carrot. As a result, it appears better to lock pen A and D instead of solely B because both achieve the targets (3 green, 1 red, 1 carrot) but respectively 1 goat is impacted instead of 3. Pen C is not worth protecting because it does not contain any vegetables while a goat is using it. In other words, systematic reserve site selection tools try to ensure conservation of a given

amount of given features (vegetables) while limiting the benefit loss associated with a given usage (goat).

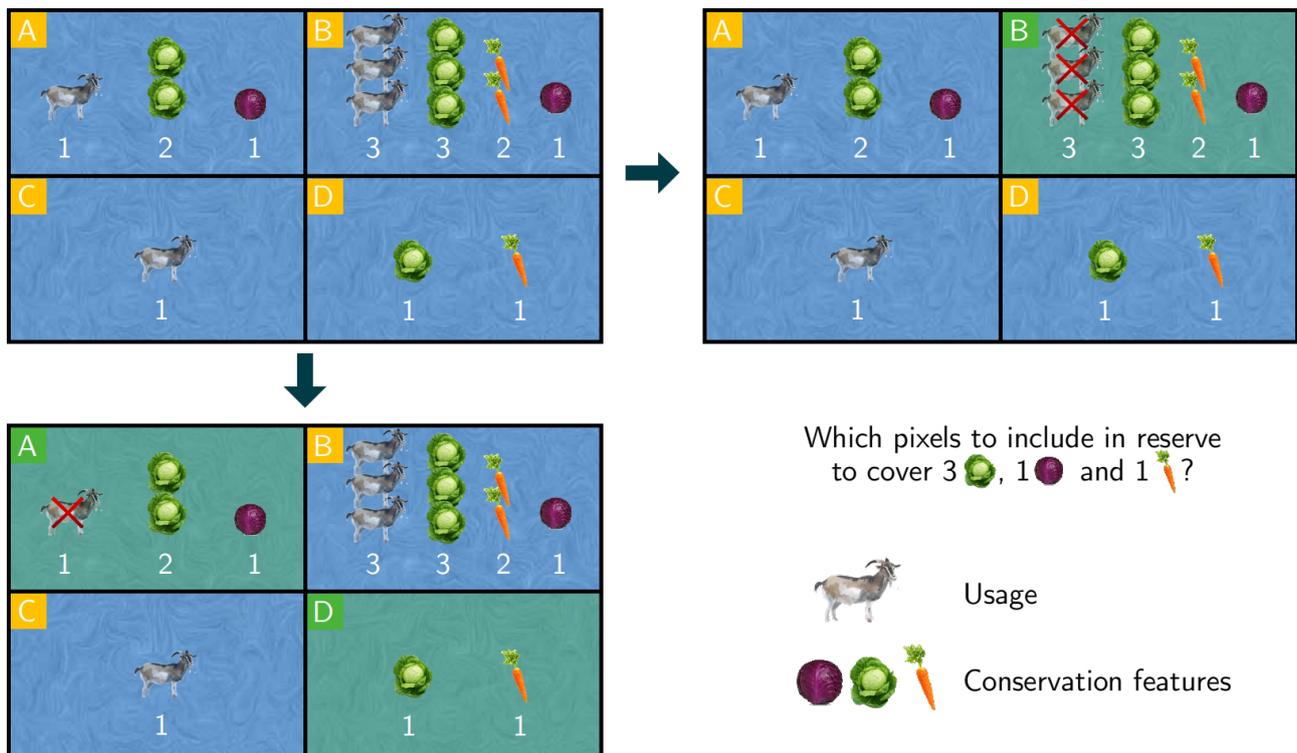


Figure 1.5: Example of an optimisation problem solved by systematic reserve site selection. Blue and green background mean pen is respectively opened and locked (*i.e.* part of the reserve). From initial situation (top left), which pen accesses to forbid in order to eventually protect 3 green cabbages, 1 red cabbage and 1 carrot while minimising impact on goats? If pen B is locked (top right), targets are achieved and 3 goats are impacted while, if pen A and D are locked (bottom left), only 1 goat is impacted and targets are still achieved.

1.4 Systematic reserve site selection

1.4.1 History

Following elaborated works dedicated to conservation value computation through detailed scoring systems (*cf.* Section 1.2.4), first developments of reserve site selection algorithms appeared and systematic planning became a major question in conservation sciences. Systematic methods for reserve site selection are a way to provide a rational choice in opposition with *ad-hoc* selection. Indeed, nature reserves were historically selected where no other land use were to be found. However, in an environmental crisis context with limited resources, *ad-hoc* reserves are unlikely to fill conservation purposes as depicted in (Pressey 1994; Pressey and Tully 1994) and systematic planning appeared as a relevant solution to guide conservation effort. The main mission of conservation science is to ensure biodiversity persistence, which is therefore the purpose followed by systematic reserve site selection algorithms. As a reminder, conservation features (species, habitat, etc.) are spatially explicit data and the purpose is to select pieces of region composing a consistent nature reserve.

Priority ranking algorithms are a first naive approach to develop a systematic reserve site selection. The principle is simply to sort considered areas based on their conservation value in order to eventually select the highest ones for priority conservation actions or until a surface budget is reached (Tans 1974; Gehlbach 1975; Rabe and Savage 1979). Obviously, reserve selection

resulting from ranking algorithms entirely depends on the chosen scoring system. For instance, output can widely vary whether conservation value is rarity or richness based. This subjectivity inherited from conservation value computation is mitigated though the selection procedure transparency and defensible character. However, such approach lacks of representativeness as ranking algorithms cannot systematically ensure a full coverage of every considered biological features. For example, if 2 specie are located within the studied area, it is likely areas with the highest conservation scores only contain 1 specie and no individuals of the other.

Since then, complementarity became a principle of reserve site selection. Consequently, an iterative ranking algorithm is developed in (Kirkpatrick 1983) aiming to compensate lack of representativeness and thus systematically ensure complementarity of nature reserve. To do so, between each site selection, scores are computed again accounting for the already covered conservation features and thus taking into account complementarity. Such iterative approaches implied the development of heuristic rules to best ensure reserve representativeness. Following the influencing Kirkpatrick's paper, several authors successfully proposed their own improved version of iterative heuristic (Margules et al. 1988; Pressey and Nicholls 1989).

Nevertheless, such "greedy" algorithms cannot mathematically ensure "efficiency" of the selected reserve, whatever the criterion behind it. Indeed, efficiency of a nature reserve can be understood in various manners, such as a reserve which represents the maximum conservation features or with a minimum area. Consequently, an optimisation framework is suggested in (Cocks and Baird 1989) in order to *a priori* account for efficiency and representativeness in conservation plans, directly expressed in objective and constraints formulation. Indeed, authors noticed the reserve site selection problem can be expressed within an integer programming (IP) field as the considered decision variable represents the choice whether to include or not a unit in the nature reserve, *i.e.* mathematically worth 1 or 0. This major paradigm shift allowed to benefit from decision theory literature and moved conservation debate towards unprecedented and still relevant questions: Which optimisation solvers realise the best trade-off between optimality and time computation? Which objective formulation best address conservation missions? Which constraints best account for ecological stakes (connectivity, representation, shapes, redundancy, etc.)? Iterative heuristic remained preferred, discarding optimisation methods for computation time reasons (Pressey et al. 1997). The optimisation framework widely diffused and became canonical within conservation literature, as further developments followed and overcame this limitation (Possingham et al. 1993; Church et al. 1996; Possingham et al. 2000; Margules and Pressey 2000; Possingham et al. 2006). In particular, the development of Marxan impulsed by TNC (demonstrating once again TNC influence) participated in the broad diffusion of the optimisation framework standard in systematic reserve site selection. An important literature review regarding systematic conservation planning can be found in (Csuti et al. 1997; Sarkar 2012).

1.4.2 General formulation

In a reserve site selection problem, the study area is discretised into a set J of planning units within which a set I of conservation features are distributed. The amount of conservation feature $i \in I$ within the planning unit $j \in J$ is denoted a_{ij} . Each planning unit has a cost c_j representing the socio-economic cost of closing this unit. One then seeks to find the least cost collection of planning units covering a sufficient amount of each conservation features. The coverage of the conservation feature i is considered sufficient if it exceeds a user-defined target t_i . The decision is about whether to include the planning unit in the reserve. Consequently, we associate the decision variables x_j with each planning unit j : $x_j = 1$ if a planning unit

j belongs to the reserve and $x_j = 0$ otherwise. The reserve compactness is the only spatial attribute incorporated in the model. Other spatial concerns, such as contiguity, minimum fragmentation, buffer zones, corridors, *etc.*, may govern the site selection but are not considered here. Since a small perimeter involves a compact reserve, the reserve perimeter is included in the objective function. The perimeter is computed as the total length of the boundaries between reserved and non-reserved planning units. To model this, the length of the shared boundary between planning units j_1 and j_2 is denoted $b_{j_1j_2}$. A multiplier, noted β , is used within the objective function to reflect the importance of compactness relatively to the total cost of site selection. Mathematically speaking, the general problem of reserve site selection is expressed as the following combinatorial optimisation problem P_0 :

$$P_0 : \begin{cases} \min_x & \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1j_2} x_{j_1} (1 - x_{j_2}) \\ \text{s.t.} & \sum_{j \in J} a_{ij} x_j \geq t_i & \forall i \in I \\ & x_j \in \{0, 1\} & \forall j \in J \end{cases}$$

We can linearise the quadratic term of the objective function when decision variables are binary (Billionnet 2013; Beyer et al. 2016). Sets of planning units *a priori* excluded or included in the reserve are respectively noted \mathcal{LO} and \mathcal{LI} . Considering this linearisation but also locked-in and locked-out planning units, we write the full mathematical optimisation problem P_0^f of reserve site selection:

$$P_0^f : \begin{cases} \min_{x,z} & \sum_{j \in J} c_j x_j + \beta \left(\sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1j_2} z_{j_1j_2} + \sum_{j \in J} x_j b_{j,N+1}^* \right) \\ \text{s.t.} & \sum_{j \in J} a_{ij} x_j \geq t_i & \forall i \in I \\ & z_{j_1j_2} \leq x_{j_1} & \forall j_1 \in J, \forall j_2 \in J \\ & z_{j_1j_2} \leq x_{j_2} & \forall j_1 \in J, \forall j_2 \in J \\ & z_{j_1j_2} \geq x_{j_1} + x_{j_2} - 1 & \forall j_1 \in J, \forall j_2 \in J \\ & x_j = 0 & \forall j \in \mathcal{LO} \\ & x_j = 1 & \forall j \in \mathcal{LI} \\ & x_j \in \{0, 1\} & \forall j \in J \\ & z_{j_1j_2} \in \{0, 1\} & \forall j_1 \in J, \forall j_2 \in J \end{cases}$$

This problem is a generalisation of the minimum set cover problem, which is known to be NP-hard (Garey and Johnson 1979). It is a non-convex problem due to the binary nature of the decision variables. Yet, it can be expressed as an integer linear program and available solvers can solve realistic instances in a reasonable time. Another approach to solve this problem is to use metaheuristic algorithms to quickly identify suboptimal solutions. This is the strategy deployed in Marxan. Practically speaking, in this work we will use both CbC and Gurobi to have the exact solution and the simulated annealing procedure of Marxan when we use a metaheuristic approach. CbC is a free and open-source solver (Forrest et al. 2018) from the COIN-OR³ project (Lougee-Heimer 2003). Gurobi is a commercial solver used under a free academic licence. The code developed in Julia language (Bezanson et al. 2012; Bezanson et al. 2015) using the JuMP optimisation library (Dunning et al. 2017) is available on the GitHub repository <https://github.com/AdrienBrunel>. Julia was our preferred coding language, although we also used R to be able to communicate towards more ecology-oriented disciplines. The numerical experiments were all realised on a personal computer (Intel Core i7-8850H CPU @ 2.60GHz).

³Common Optimization INterface for Operations Research

1.5 Study cases

1.5.1 Observed data

1.5.1.1 Fernando de Noronha

Fernando de Noronha is a small oceanic archipelago in the western tropical Atlantic, made up of 21 islands, islets and rocks with a total surface area of 26 km^2 , and constituting a genuine Brazilian natural and cultural heritage. Its distance from the coasts⁴ has allowed it to preserve until today a relatively wild coastline where a great marine biodiversity evolves in clear waters and constitutes on land a refuge for native fauna and flora. The main island, 10 km long and 3.5 km wide, is the only one inhabited by man. Fernando de Noronha hosts a small-scale fishery composed of approximately 10 artisanal and recreational vessels. In 2001, the archipelago was listed as a World Natural Heritage Site by UNESCO. An oasis of marine life in a relatively barren and open ocean, these islands play a key role in the process of reproduction, dispersion and colonisation by marine organisms in the entire tropical South Atlantic. The productive waters of the archipelago provide an important feeding ground for species such as tuna, billfish, cetaceans, sharks and sea turtles when they migrate to the African coast. These islands also contain the largest concentration of tropical seabirds in the western Atlantic, and include the only examples of the Atlantic island forest and oceanic mangrove in the South Atlantic. The Dolphin Bay is home to an exceptional population of resident dolphins. The Fernando de Noronha ecosystem is legally protected by a number of federal laws and state regulations, including a marine national park since 1988. For all these reasons, Fernando de Noronha is a conservation showcase in Brazil while facing many interests, such as oil industry, tourism intensification and fisheries, resulting in an open sky laboratory for marine spatial planning. In the frame of the EU RISE Paddle project and related projects (FAROFA surveys, LMI Tapioca, Mafalda project, *etc.*), a series of field research surveys were conducted since 2015, providing the spatially explicit data on fish and fisheries used hereafter.

1.5.1.2 Grid

Prior to any work, Fernando de Noronha study area was divided into planning units, *i.e.* our conservation resource soon to be allocated. We built a grid made of regular rectangular pixels, with longitude and latitude respectively in $[32.65^\circ\text{W}, 32.30^\circ\text{W}]$ and $[3.95^\circ\text{S}, 3.75^\circ\text{S}]$ ranges. We chose a 0.01° resolution which results in considering $N=36 \times 21=756$ planning units. Both our boundaries and resolution choices were justified to properly capture data feeding our case study. This discretisation process allowed us to transform the input geographical layers into vectors and matrices. This operation was required to fit in the optimisation framework and thus tackle mathematically the reserve site selection problem. Pixels located in Fernando de Noronha land and harbour were *a priori* excluded from potential reserve site candidate (see transparent grey pixels in Figure 1.7). In other words, these *locked-out* planning units were not authorised to be included in a reserve. Fernando de Noronha harbour is the archipelago nerve centre, what justified it could not be included in a strict reserve. Regarding the exclusion of land, the purpose was to avoid a fictive bridge between two non-connected marine areas.

1.5.1.3 Fish biomass

Recent acoustic surveys around Fernando de Noronha collected *in situ* data on fish distribution (Bertrand 2019). Acoustic raw data were processed to synthesise the collection of fish echoes as a nautical area back-scattering strength (Maclennan et al. 2002), *i.e.* s_A , summed over the water

⁴360 km from Natal, 545 km from Recife

column. Figure 1.6 displays the s_A raw spatial distribution as purple circles along sampling transects (solid black lines). We used s_A as a proxy for fish biomass (Simmonds and MacLennan 2005). We chose to treat fish biomass as categories, assigning each observation to its quartile prior to the interpolation. A fifth category was added to account for null densities. Interpolating between sample data allowed producing a continuous 2D-view of fish biomass distribution within the sampling area. Outside this area, as the reserve site selection optimisation models cannot deal with absent data, we set values to 0, although we did not know the actual fish distribution there. The interpolation consisted in indicator co-kriging where each indicator variable was coding for a given category (Bez and Braham 2014; Chiles and Delfiner 2012). Finally, as acoustic data resolution was finer than our grid, we selected the most frequent class of s_A values within each planning unit as a conservation feature surrogate hereafter. Results of this process are presented in Panel B in Figure 1.7.

1.5.1.4 Habitats

Bathymetric data were collected from GEneral Bathymetric Chart of the Ocean (GEBCO) online platform⁵. GEBCO 2014 was preferred over 2020 update because *in situ* depth measurement from recent surveys (see above) were closer to 2014 than 2020 interpolation. Data resolution is 30 arcseconds (*i.e.*, 0.0083°) both for latitude and longitude. Such resolution was consistent with our 0.01° (*i.e.*, 36 arcseconds) grid resolution. Continental shelf and shelf break can be considered as two separate suitable habitats for benthic and demersal fish, both worth protecting. A GEBCO point was discriminated as continental shelf or shelf break respectively for depth within [0m, 50m] and [50m, 200m] ranges. Finally, according to the majority of point states (*i.e.*, continental shelf or shelf break habitat) within each planning unit, the predominant conservation feature took the value 1 while the other 0. For instance, we assigned a value of 1 for the continental shelf and 0 to the shelf break feature if there were more continental shelf than shelf break points within the planning unit. Note, in case of equality, we assigned the pixel to continental shelf. Results of this process were illustrated in Panel C and D in Figure 1.7. Note that the continental shelf or shelf break habitat distributions did not overlap in our gridded data due to the very nature of the data processing described above.

1.5.1.5 Fishery

Fishery data were composed of 69 GPS tracks from fishers' trips collected *in situ* through the 5 past years at Fernando de Noronha. Fishery activity in Fernando de Nornoha is performed daily by 4-10 vessels. Although the sampling did not cover the entire fleet, it is reasonable to assume we have a significant insight on the fishery activity. An hidden Markov segmentation model was applied (Tatiana Beltrão Alves Da Costa personal communication) (Joo et al. 2013; McClintock et al. 2020) to the fishery data in order to classify each trajectory segment into one behavioural state: fishing or travelling. We considered the amount of points in the fishing state as a quantitative proxy representing pressure due to fishing activities. Consequently, in order to build a socio-economic cost for each planning unit, we counted the number of points in the fishing state falling within each planning unit and called this quantity FC for «Fishing Count». The socio-economic cost is intended from a manager perspective. For instance, selecting a planning unit with a high fishing points concentration in the reserve would be a cost for humans despite being a pressure relief for biodiversity. Grid boundaries were chosen to capture fishermen's interests in the extreme west of Fernando de Noronha. Results of this process were illustrated in Panel A in Figure 1.7.

⁵<https://download.gebco.net/>

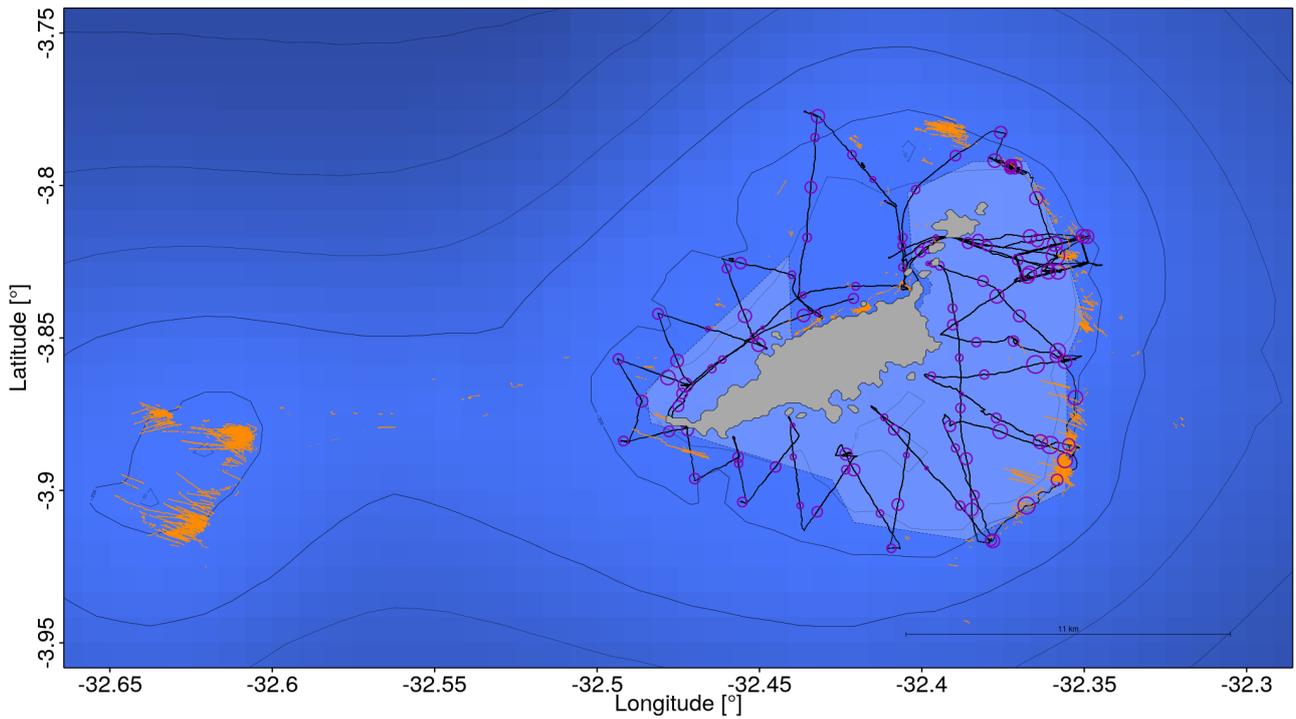


Figure 1.6: Raw input data feeding the conservation problem. Bathymetric raw data (GEBCO 2014) is represented by a light to deep blue colour gradient with appended iso-depth solid thin black lines (50m, 200m, 1000m, 2000m, 3000m, 4000m). Fishermen's boats GPS points estimated in a fishing state are illustrated with orange dots. Acoustic raw data are depicted by purple circles whose radius is proportional to $\sqrt{s_A}$ value along line boat transects represented with a solid thick black line. Light grey polygon shows the limits of the existing marine park.

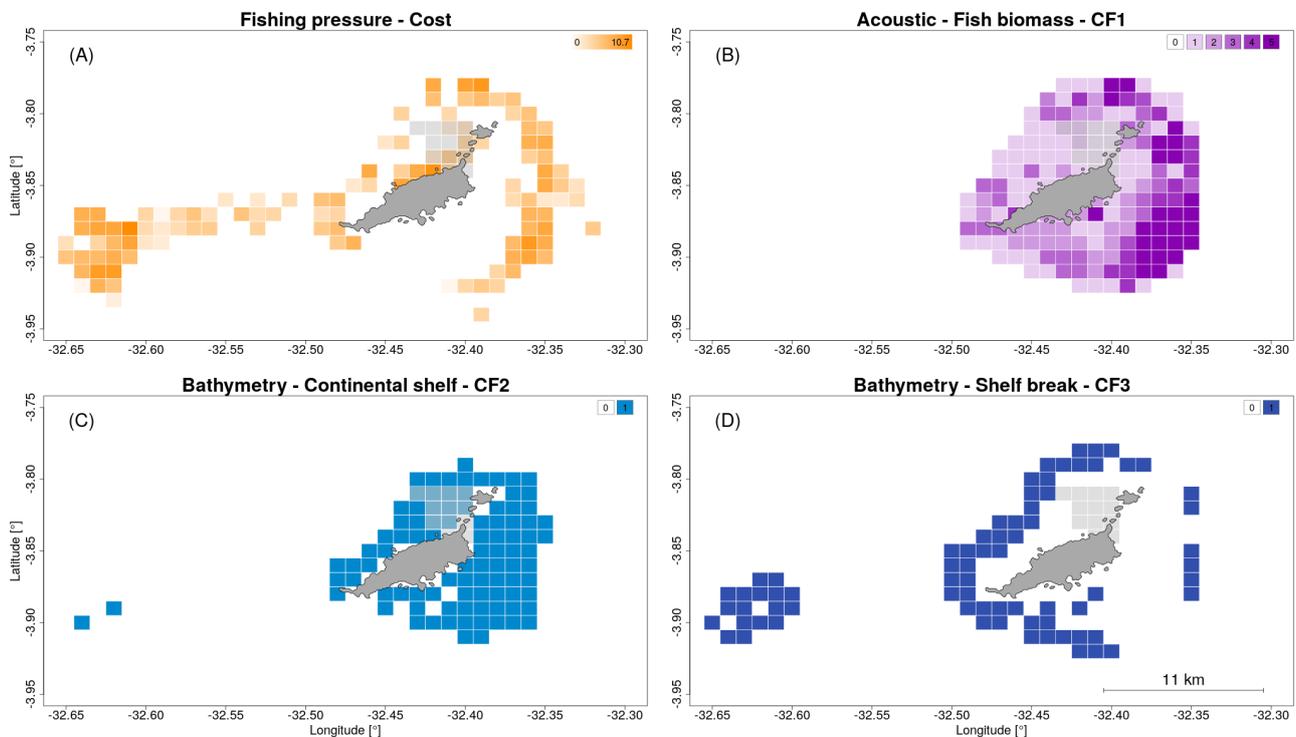


Figure 1.7: Data used for the reserve site selection optimisation problem. (A) Fishery-based cost layer in a continuous orange colour gradient. (B) Fish Biomass conservation feature surrogate in a discrete purple colour gradient. (C) Continental shelf and (D) Shelf break habitat conservation feature surrogates in light and deep blue respectively. Transparent grey pixels are the planning units a priori excluded from the solution.

1.5.2 Generated data

For testing purposes, we developed a systematic way of building user-defined scenarios for reserve site selection optimisation problems. The idea is to provide the conservation literature tools to facilitate benchmarks of developed methods in conservation planning. Therefore, the main ambition is to generate realistic discrete spatial distributions of the considered conservation features.

Technically speaking, we arbitrarily chose to compute the amount a_{ij} of a conservation feature $i \in I$ in a planning unit $j \in J$ by randomly drawing this value in a Gaussian distribution.

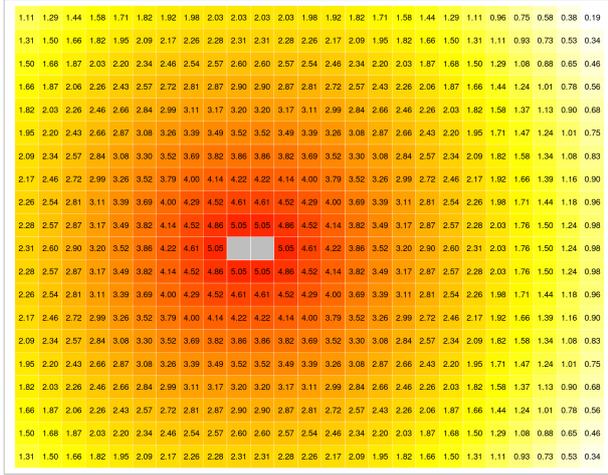
$$a_{ij} \sim \mathcal{N}(m_{ij}, \sigma_{ij}^2)$$

The mean value m_{ij} of the Gaussian distribution only depends on the distance d_{ij} to the closest (chosen or randomly drawn) N_{epi} epicentres associated to the conservation feature $i \in I$. To be more precise, the mean value m_{ij} depends on $d_{ij}^{\alpha_i}$, where α_i is a predefined parameter for each conservation feature $i \in I$ controlling the dispersion of the mean values distribution relatively to the epicentres.

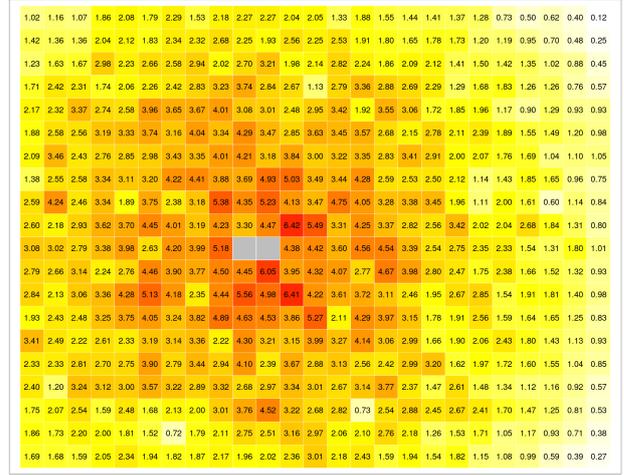
$$m_{ij} = \mu_i \left[1 - \left(\frac{d_{ij}}{d_{max}} \right)^{\alpha_i} \right]$$

The maximum mean value, *i.e.* the mean value at the epicentres, is a chosen parameter μ_i for each conservation feature $i \in I$. If no epicentres are provided, the mean value of the Gaussian distribution depends on the distance to the locked-out planning units supposed to represent a shoreline. The standard deviation σ_{ij} of the Gaussian distribution is such as $\sigma_{ij} = \sigma_i m_{ij}$ where σ_i is a chosen parameter for each conservation feature $i \in I$. The code used to generate data is available in open access⁶. The instance is characterised by the rectangular grid size N_x and N_y and the number of conservation features N_{cf} . We show the generated spatial distributions of two conservation features resulting from the data generation procedure in Figure 1.8.

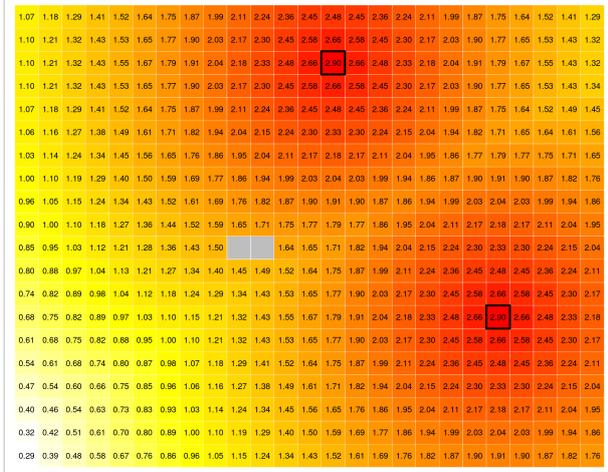
⁶https://github.com/AdrienBrunel/data_generation



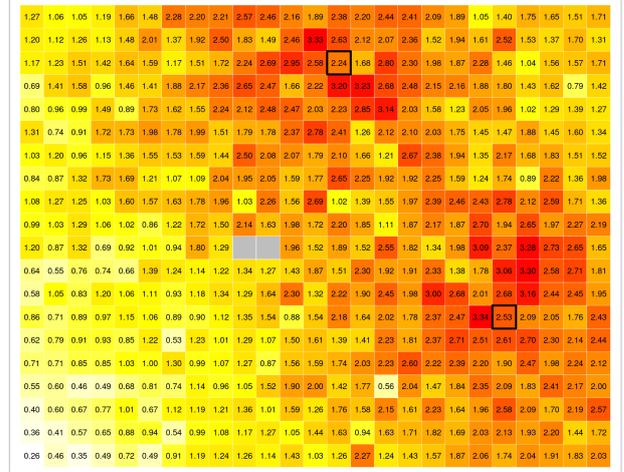
(a) Spatial distribution of the mean value m_{ij} when epicentres correspond to the locked-out planning units. The maximum mean value is 5.7.



(b) Random drawing from the Gaussian distribution, where the mean values are shown in Panel 1.8a. Dispersion coefficient is $\alpha_i = 0.75$.



(c) Spatial distribution of the mean value m_{ij} where 2 epicentres are randomly drawn among planning units. The maximum mean value is 2.9.



(d) Random drawing from the Gaussian distribution where the mean values are shown in Panel 1.8c. Dispersion coefficient is $\alpha_i = 0.78$.

Figure 1.8: Example of the generated spatial distribution for two different conservation features in a 25×20 rectangular grid. The amounts of considered conservation feature are shown with a yellow to red gradient. The corresponding numerical values are written in black inside the planning units. Locked-out planning units are represented in grey. We chose $\sigma_i = 0.20$.

1.6 Research questions

Implementing an MSP approach is a large and complex task: the multidisciplinary nature of the approach, the heterogeneity of stakeholders' interests, the absence of complete knowledge to fully support a decision, the imbalance of quantity and quality of data informing a problem, the profusion of DSTs *etc.* In particular, determining the best spatial zoning of human activities at sea based on the available ecological and socio-economic knowledge can be difficult and overwhelming. To better understand the tools and methods involved in such processes, we address the more specific case of reserve site selection. Addressing the reserve site selection optimisation problem may even lead us to develop ideas and methods that could be of great help for the more difficult problem of finding a global spatial zoning for MSP. In addition, the current agenda of international conservation policies made the reserve site selection problem a today's problem.

Yet, there are still research gaps and questions in the methods used to solve this problem, despite being already widely applied in real-world cases. In particular, building a relevant set of alternative solutions using exact optimisation methods is needed. Uncertainty is poorly taken into account, while it is inherent to input data of such problems. Spatial attributes of reserves are often advocated to discard a reserve solution and should be accounted explicitly. In this thesis, we provide (1) an analysis of existing reserve site selection tools, (2) two iterative procedure to build a set of truly dissimilar alternative reserve solutions using exact optimisation, (3) two models to account for an uncertainty affecting non-binary parameters using exact optimisation, (4) a global integer programming model to build compact, connected and without holes reserve solutions. More generally, our research involves the improvement of the methods used for reserve site selection and eventually in MSP.

The end goal of this work is to bring more equity in MSP processes by increasing the understanding and transparency of the methods involved. We thus aim to empower stakeholders and reduce the epistemic injustice. Although we acknowledge these tools are a result of environmental ethics we question, we still pragmatically want to improve the methods involved. Our work aims to make the decision more transparent and accountable.

Chapter 2

Opening the black box of decision support tools in marine spatial planning: analysis of reserve site selection algorithms

2.1 Introduction

Conservation planning tools such as Marxan have been widely used as DSTs in MSP processes. For instance, those tools were mobilised in 40% of the MSP procedures that implemented a formal analysis tool in (Pınarbaşı et al. 2017) meta-analysis; and they were likely to be handled by a variety of users (*e.g.*, scientists, NGOs, planners). While this mathematical formalisation of the reserve site selection problem has provided great advances in solving complex problems fed with highly heterogeneous data, it also came with a series of limitations. In particular, Pınarbaşı et al. (2017) identified the following: the limited functionalities of each DST, especially in the later stages of MSP, leading to coupling the use of several DSTs, the limited lifespan of DSTs due to the lack of updating, the fact that DSTs are mostly used for environmental issues, the cost of DSTs and last but not least, the fact these DSTs introduced a high technicality in the reserve site selection process. Here we argue that technical choices required by these DSTs, too often not made explicit, may introduce pitfalls in MSP discussion tables and convey the risk of dispossessing part of the stakeholders involved in MSP of their critical expertise on the solution provided by the algorithms. In that case, the original quest for transparency may turn out to produce new *black box* effects. Given the importance of data at almost every stage of its implementation, informational questions are at the very core of MSP (Trouillet 2019; Trouillet 2020). By being likely related to the rationalist and quantitative model (Kidd and Ellis 2012), MSP participates in the return of evidence-based planning and favour a certain revival of positivism (Faludi and Waterhout 2006). In this logic, DSTs and other tools mobilised by geodesign (Goodchild 2010) require greater attention. Such an approach is in line with the critical current on MSP (Flannery et al. 2020), which has been developing in recent years.

In that context, our work aims at informing practitioners about stakes, possibilities and limitations of MSP approach through reserve site selection tools. The purposes of this chapter are to (1) detail and question the mathematical framework of reserve site selection DSTs to end-users through illustrations of a simple case study, (2) provide guidelines for the use of optimisation-based reserve site selection tools, (3) draw global awareness of stakeholders on reserve site selection DSTs by deciphering the effects data and parameterisation options may have on the final solutions and thus avoid blind trust in a decision-making process or misinterpretation. Our case study for scenario simulations takes place in Fernando de Noronha archipelago in the Brazilian tropical Atlantic.

2.2 Material and methods

The general optimisation framework for reserve site selection is detailed in Section 1.4.2. The study site of Fernando de Nornoha was described in Section 1.5.1.1. The grid used to discretise the study area and express an optimisation problem was described in Section 1.5.1.2.

2.2.1 Optimisation

2.2.1.1 Formulation

Various optimisation models emerged from the encounter between operations research and conservation science. The two main generic models, namely the *minimum set* and *maximum coverage* formulations, are detailed in Table 2.1.

Minimum set formulation	Maximum coverage formulation
$\begin{cases} \min_{x \in \{0,1\}^N} & c^T x + \beta x^T B(1 - x) \\ \text{s.t.} & Ax \geq t \end{cases} \quad (2.1)$	$\begin{cases} \max_{x \in \{0,1\}^N} & \omega^T Ax \\ \text{s.t.} & c^T x + \beta x^T B(1 - x) \leq b \end{cases} \quad (2.2)$

Table 2.1: Minimum set and maximum coverage formulations modelling the reserve site selection optimisation problem. Let M species be distributed among N planning units. Cost $c \in \mathbb{R}^N$, conservation feature distribution $A \in \mathbb{R}^{M \times N}$, common boundary length of planning unit $B \in \mathbb{R}^{N \times N}$, targets $t \in \mathbb{R}^M$, conservation feature relative weight $\omega \in \mathbb{R}^M$, budget $b \in \mathbb{R}$, compactness parameter $\beta \in \mathbb{R}_+$, planning unit status $x \in \{0, 1\}^N$.

In the minimum set formulation (*cf.* Equation (2.1)), we seek for the collection of planning units that meet predefined conservation targets at a minimum socio-economic cost. In the maximum coverage problem (*cf.* Equation (2.2)), we seek for the planning unit collection that cover the maximum amount of conservation features within the limits of a predefined socio-economic cost budget. In both formulations, the compactness parameter β set the relative importance of the reserve perimeter with respect to the reserve cost. The bigger the compactness parameter, the more spatially aggregated the computed reserve.

2.2.1.2 Solving methods

In an integer programming framework, the solving method choice is essential as it directly influences the solution output. Two main families exist to solve the same optimisation problem: metaheuristics and exact solving methods. Metaheuristic solvers, *e.g.* simulated annealing algorithm used in Marxan, output a user-defined number of suboptimal reserve solutions which are interpreted as alternative solutions by practitioners in the decision process. Exact methods give a single optimal solution. Finally, metaheuristics do not face any restriction in the optimisation formulation nature, while exact solvers can only deal with linear problems. In our binary programming context, the quadratic element $x^T B(1 - x)$ in Table 2.1 can be linearised (Billionnet 2007). Practically speaking, many solutions exist to solve the reserve site selection optimisation problem embedded in various software. In this work, regarding exact integer linear programming algorithms, we used free and open-source CbC solver (Forrest et al. 2018) from COIN-OR project (Lougee-Heimer 2003) called through a dedicated code¹ developed in Julia language (Bezanson et al. 2012; Bezanson et al. 2015) using JuMP optimisation library (Dunning et al. 2017). The Julia language allowed us to directly express and customise the

¹<https://github.com/AdrienBrunel/reserve-site-selection>

optimisation problem according to a specific need. For a less technical audience, the same solutions can be found thanks to the newcomer PrioritizR R package (Hanson et al. 2020) based on COIN-OR project free and open source Symphony solver (Harter et al. 2017; Ralphs et al. 2019). For metaheuristic solvers, we used the simulated annealing algorithm of Marxan (Ball et al. 2009; Game and Grantham 2008; Ardron et al. 2010).

2.2.2 Input data

It is essential to understand that data entering reserve site selection DSTs should ideally provide detailed and true spatial distribution of every considered biodiversity features (species, habitats, ecological processes, *etc.*) and human activities. Practically, we only have access to a measured surrogate dataset for these spatially explicit layers. For example, the GPS tracks of several equipped birds could be a relevant proxy representing the spatial distributions of the species. Similarly, ocean depth can be used as a habitat surrogate. We rarely have direct access to true spatial distributions of the variables of interest (*e.g.* number of individuals of a given species, ecological niche location, allele distribution within a species, detailed fishing catches, *etc.*) whether we represent a conservation feature or the cost layer. Consequently, we often need to derive this piece of information through an indirect although more accessible source of data, *i.e.* a measure and estimation of a surrogate data distribution. The conservation feature or cost distribution are sensitive elements as the whole optimisation process is based upon it.

The input data used for this chapter were described in Section 1.5.1. Section 1.5.1.3 detailed how we used the acoustic data as a surrogate for the fish biomass conservation feature. Section 1.5.1.4 detailed how we used the bathymetry data as a surrogate for two binary conservation features: the continental shelf and shelf break habitats. Section 1.5.1.5 detailed the fishery data we used as a cost layer. Illustrations of the raw and processed data are respectively shown in Figure 1.6 and Figure 1.7.

In order to have a better grasp on input data influence, we focused on how we processed the cost distribution. Consequently, several cost layers, directly involved in the optimisation objective expression, were considered resulting in 5 different scenarios described below :

- **Cost₁ = 1**
Simple and constant cost, worth 1 for every planning unit. In first approximation, it is a common and relevant approach to consider equally every pixel.
- **Cost₂ = 1 + FC**
Raw use of our fishing points count, namely FC except that we added 1 in order to avoid planning unit worth 0. Indeed, "free" planning units can greatly pollute research space and solution interpretation.
- **Cost₃ = 1 + ln(1 + FC)**
A natural logarithm was applied to 1+FC (we added 1 to force consistency with the logarithm definition domain). We once again added 1 to the global expression to avoid "free" planning units, for the exact same reasons detailed above.
- **Cost₄ = FC_{1→10} scale**
According to FC value, we projected the cost on a 1 to 10 scale. This process can be understood as a grade and has the huge benefit to be computed whatever the data feeding cost representation.
- **Cost₅ = FC_{1→100} scale**
Idem as above but with a 1 to 100 scale to observe the scale resolution influence.

We assessed the impact of the shape of input data on output results by comparing the maps of reserve solutions. If we consider two spatial distributions (cost or solution) as independent random variables X and Y , the statistical correlation coefficient between X and Y appeared as a reasonable metric for sensitivity. In particular, a correlation coefficient of 1 means maps are strictly identical. When there is no variability in the studied distribution (for example Cost_1 is constant through all planning units), the standard deviation σ_X is worth 0 and correlation coefficient is logically not defined. As we compared several scenarios, we had a collection of correlation coefficients composing cost and solution correlation (symmetric) matrices. In conclusion, we had a simple quantitative comparison index between gridded maps provided by the correlation matrices between cost distributions, where the line/column number corresponds to the considered scenario number.

2.3 Results

2.3.1 Sensitivity of reserve solutions to structural elements

2.3.1.1 Formulation

We performed sensitivity analyses on cost and conservation feature coverage for both maximum coverage and minimum set formulations. More precisely, the sensitivity analysis was performed on the cost parameter regarding maximum coverage and conservation feature targets for the minimum set (we considered equal targets for the three conservation features). Results were synthesised in Figure 2.1. First, our approach showed the bijection between the reserve cost and conservation features protection levels with both formulations. Indeed, at one reserve cost corresponded one protection level for each conservation feature. Furthermore, when looking closer at the maximum coverage results, the curve corresponding to the continental shelf (light blue circles) was the highest while the one corresponding to the shelf break was the lowest (deep blue circles). So the continental shelf was the feature participating the most to the global coverage score while shelf break the least. Thus, the continental shelf habitat was numerically easier to represent than shelf break in optimal reserve solution. We can explain this as a combination of two reasons thanks to Figure 1.7 : planning units including the continental shelf conservation feature were cheaper than the planning units including the shelf break but also included more significant amount of fish biomass.

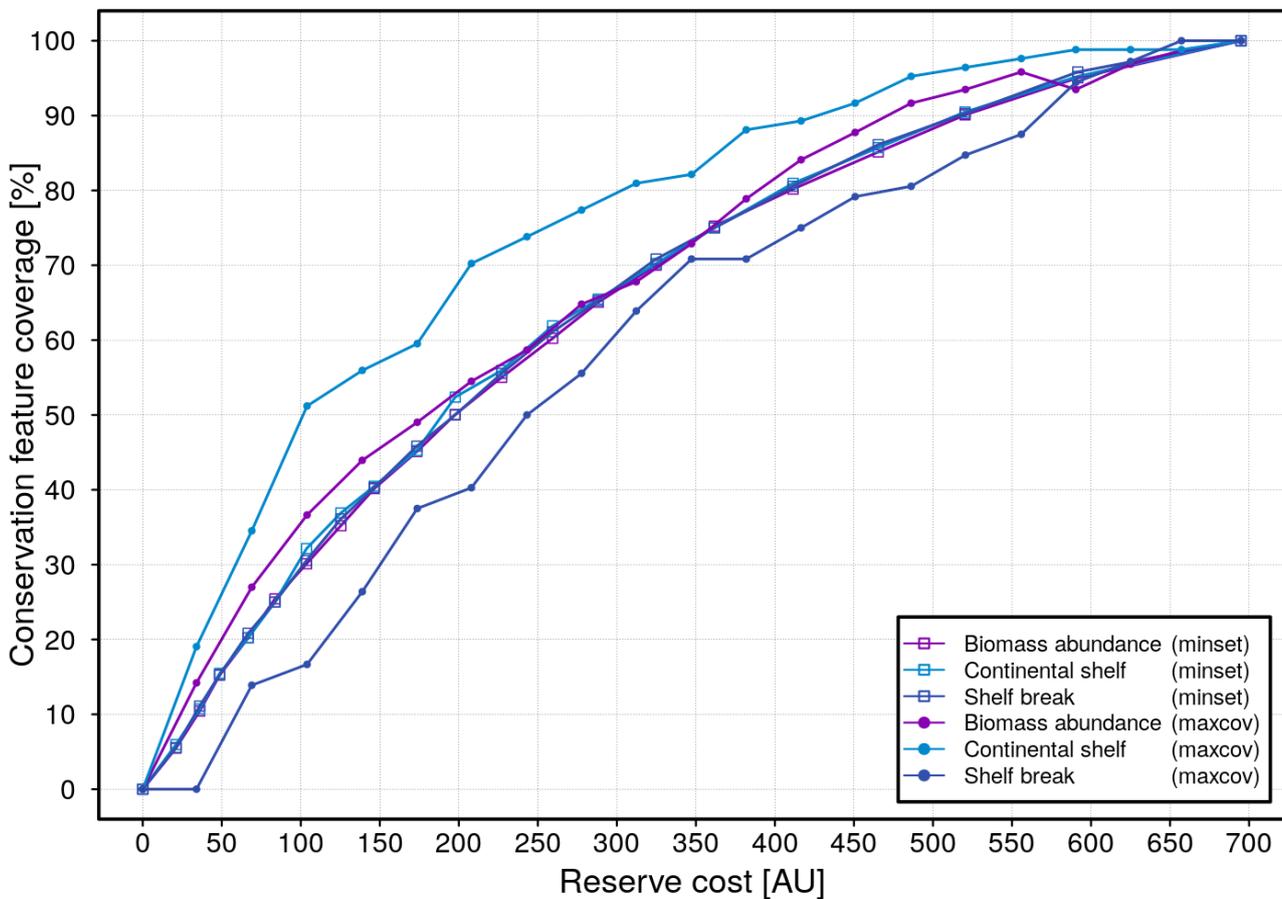


Figure 2.1: Reserve cost in arbitrary units versus conservation feature coverage in % for both minimum set and maximum coverage formulations. The 3 conservation features coverage are shown in purple (biomass abundance), light blue (continental shelf) and deep blue (shelf break) while associated formulation is depicted through full circles (maxcov) versus empty squares (minset) on the curve. The scenario considered here included a cost layer worth $1 + \ln(1 + FC)$, a compactness parameter $\beta = 1$. Exact solving is performed thanks to CbC solver.

2.3.1.2 Solving method

We here illustrated (*cf.* Figure 2.2) results provided by reserve site selection DSTs computed with both solving approaches (see Section 2.2.1.2). The metaheuristic results were represented by a green colour gradient representing the Marxan selection frequency. The white number within planning units indicates how many times it was selected among 100 Marxan runs. Planning units with a red border depict the reserve derived thanks to the exact integer programming algorithm. We first observed a visual difference between the metaheuristic and optimal solutions. Metaheuristic results spread more in space what makes sense as it explored many suboptimal solutions and thus more planning units. In particular, low depth isolated pixels in the extreme west of the study area and eastern pixels were sometimes selected by metaheuristics while they did not belong to the optimal solution. It can be explained as they had an important fishing cost as we can see in Figure 1.7. The aggregated aspect was due to the active compactness penalty (see Section 2.3.2.2 for details). Note locked-out pixels were not included in the reserve solution as expected. Furthermore, we can observe reserve solutions were centered around the main island which is simply explained by the fact most of the conservation features to be covered lied there as depicted in Figure 1.7. Also, in this small size study case, a 5% gap was derived between optimal and mean metaheuristic solution scores (among 100 Marxan runs).

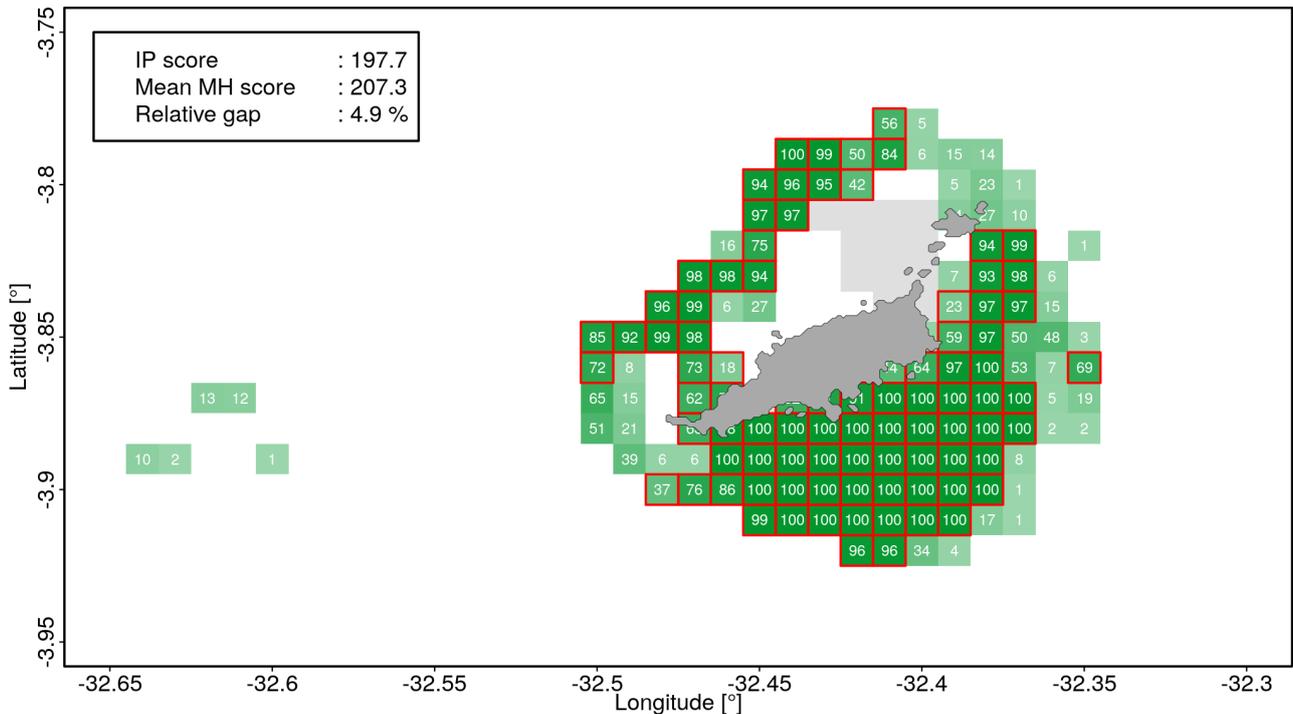


Figure 2.2: Metaheuristic versus exact integer programming reserve solutions. Minimum set formulation, 3 conservation features each represented with a 50% protection target, considered cost layer worth $1 + \ln(1 + FC)$ and compactness parameter $\beta = 1$. Selected planning units within optimal reserve solution by Marxan (using Simulated Annealing metaheuristic algorithm) are represented with a green colour gradient according to selection frequency among 100 Marxan runs (white number inside planning unit). Red border around planning unit indicates selection by integer programming exact (free open-source) solver CbC.

2.3.1.3 Influence of cost expression

Table 2.2 depicts the link between cost distribution (orange figures) and associated reserve solution (green figures) correlation coefficients. For example, the correlation coefficient between $Cost_5$ and $Cost_2$ was greater than 0.998 because cost distributions were almost identical. It could be expected as $FC_{1 \rightarrow 100}$ scale, due to the thin scale choice, well captured FC spatial distribution. Conversely, the use of a natural logarithm ($Cost_3$) implied a way smaller cor-

relation coefficient of 0.55 when compared to Cost_2 . Note dashes in the first line of the cost correlation matrix corresponded to undefined correlation coefficient because Cost_1 involved a constant distribution and thus a standard deviation of 0. Now, the remaining question is what were the implication of such cost distribution differences in the computed optimal reserve? Did correlated distribution imply a correlated reserve solution? Did a completely different cost end in a completely different reserve? In order to lead our analysis upon the cost expression, we considered 3 conservation features with each a 50% target and set $\beta = 0$. We did not account for the compactness parameter because a given $\beta > 0$ would involve a different quantitative share of compactness penalty in the objective as cost term ranges greatly change with the way we compute it (*e.g.* more than 10 000 with Cost_2 , less than 10 in a scenario with Cost_4). First, we quantitatively observed a weak but existing correlation between solutions. It can simply be explained by the fact every scenario shared the exact same conservation feature spatial distributions for feeding the optimisation problem formulation, logically reflected in similar reserve solutions. Also, despite the logarithm application, the reserve solutions obtained with Cost_2 and Cost_3 were quite alike (correlation coefficient of 0.93). Table 2.2 illustrates that similar cost distribution can end up in a different reserve solution (see Cost_3 and Cost_5 cost and solution correlation coefficients) while different costs can lead to a similar reserve solution (see Cost_2 Cost_3 cost and solution correlation coefficients). Also, in order to study the effect of a data gap, we simulated a scenario where we removed the biomass abundance data layer and only kept habitat data (continental shelf and shelf break). Note solutions were computed with $\beta = 0$ for relevance purpose as we wanted to observe only the effect of a data gap without any compactness considerations. We obtained a correlation coefficient of 0.75 between scenarios with and without biomass abundance data. As expected, we observed a notable difference between reserve solutions as it did not have to cover abundance biomass data anymore. Although, both scenarios had most input in common what justified why global results were concentrated around Fernando de Noronha archipelago and showed common selected planning units.

	Cost ₁		Cost ₂		Cost ₃		Cost ₄		Cost ₅	
	1		1 + FC		1 + ln(1 + FC)		FC _{1→10} scale		FC _{1→100} scale	
	Cost	Solution	Cost	Solution	Cost	Solution	Cost	Solution	Cost	Solution
1	1		-	0.40	-	0.41	-	0.47	-	0.47
1 + FC	*		1		0.55	0.93	0.85	0.79	1.00	0.82
1 + ln(1 + FC)	*		*		1		0.83	0.84	0.58	0.87
FC _{1→10} scale	*		*		*		1		0.87	0.89
FC _{1→100} scale	*		*		*		*		1	

Table 2.2: Cost layer and solution correlation matrices. Correlations coefficient between cost (orange) and solution (green) spatial distributions from one scenario to another. Correlation coefficient for Cost_1 does not exist (because cost distribution is constant). Star symbols indicate symmetric coefficients. For those simulations, we fixed a minimum set formulation, 3 conservation features each represented with a 50% protection target and compactness parameter $\beta = 0$.

2.3.2 Parameters influence

2.3.2.1 Coverage targets

Figure 2.1 presents the results of a sensitivity analysis over coverage targets. The targets were simultaneously and equally increased. Regardless the formulation, we observed a non-linear and concave progression of coverage with respect to the reserve cost. It thus implied increasing conservation feature coverage is more and more expensive.

2.3.2.2 Compactness parameter

Sensitivity analysis over the compactness parameter β is shown in Figure 2.3. As we can see, a smooth decreasing trend appeared when we plotted the reserve perimeter $x^T B(1-x)$ versus the compactness parameters β . It made sense as β was the penalty directly applied to the reserve outside perimeter within the objective function (see Equation (2.1)). Therefore, the greater the penalty, the smaller the perimeter. We can also see the decrease seemed to quickly ease and eventually reached an equilibrium before decreasing again for way bigger values. However, this second decrease (for $\beta \geq 8$) is fictive as solutions included pixels at the border (see the reserve solution of the right panel) of the study area with an artificial 0 boundary value (as no neighbours exist). It is a purely numerical edge effect but this common mistake can be observed in published research, attesting it is not a well-known pitfall, *e.g.*, (Delavenne et al. 2012; Beyer et al. 2016; Magris et al. 2021). Again, a sensitivity analysis can easily show when this kind of odd behaviour of the solution appears. More generally, as soon as $\beta \geq 0$, a planning unit at the edge of the study area is more likely included in the reserve solution.

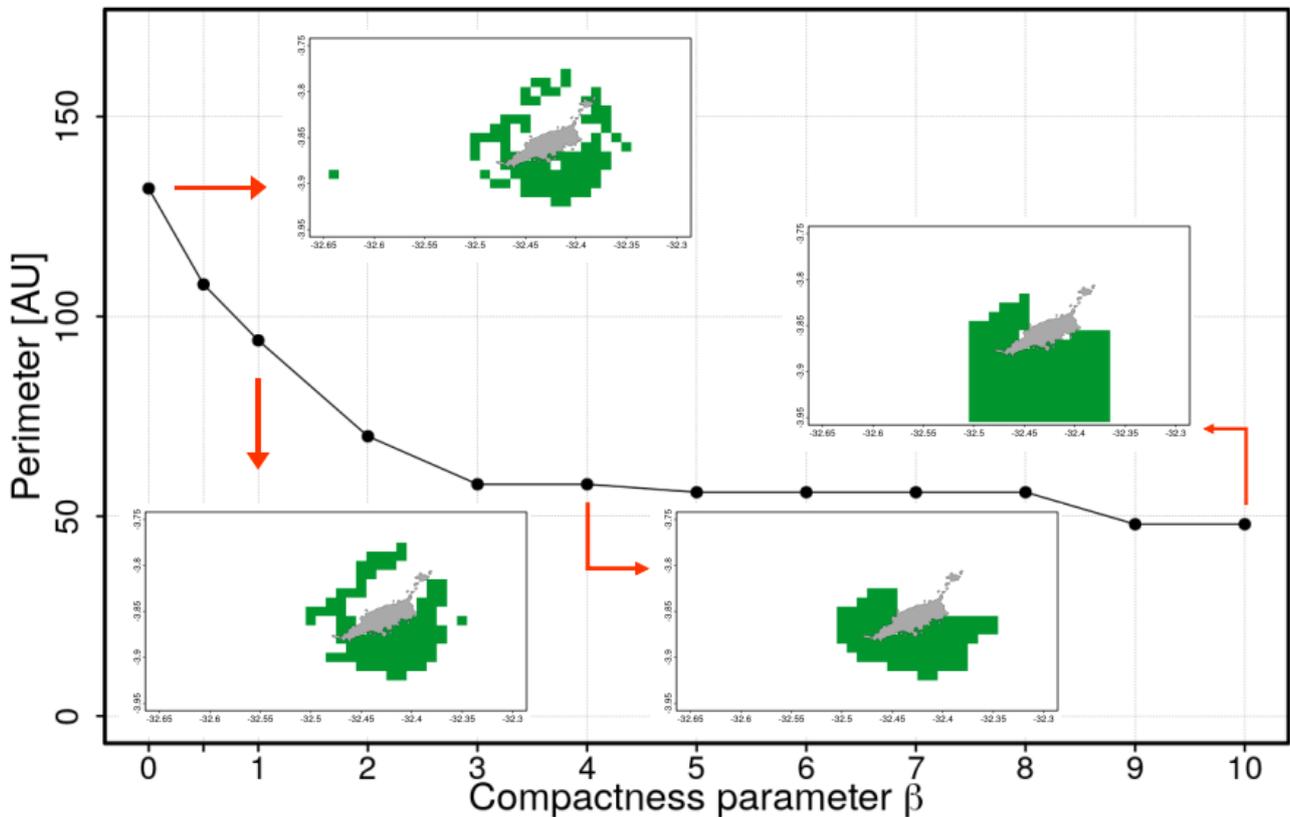


Figure 2.3: Perimeter (in arbitrary unit) versus compactness parameter. Sensitivity analysis performed on the compactness parameter β with respect to the reserve solution perimeter. Minimum set formulation, 3 conservation features represented, considered cost layer worth $1 + \ln(1 + FC)$, exact solving thanks to CbC solver.

2.3.2.3 Compactness parameter correction

We observed in Figure 2.3 but also in other publications, *e.g.* (Delavenne et al. 2012; Beyer et al. 2016; Magris et al. 2021), that planning units at the edge of the study area are more likely to be included in the reserve solution. This unwanted edge effect is explained by the fact the planning units at the edge have less common frontiers with surrounding pixels due to their position at the border. These planning units thus have artificially a smaller weight in the reserve perimeter computation. For instance, in a regular grid, a middle planning unit has 4 neighbours while a pixel at the border has 3 and a corner has only 2. Starting from

this observation, we provided a simple correction : we added one fictive planning unit which shared a boundary with every planning units located at the edge of the grid (see Figure 2.4). The length of this boundary depended on what was missing to reach an equal weight for the perimeter computation. Indeed, a planning unit at the corner misses 2 edges, while another pixel at the border only misses 1. The fictive pixel was locked-out (*i.e.* never included in the reserve solution) thus leaving the rest of the optimisation problem undisturbed.

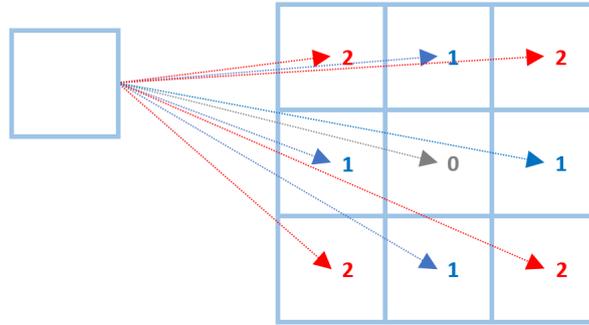


Figure 2.4: Principle of the correction involving the addition of a fictive pixel connected to the external edge of the grid. The newly added fictive pixel is invisible in the original model (necessarily locked-out), but allows the outer edge of the study area to be taken into account within the computation of the reserve solution perimeter.

Mathematically speaking, we considered an augmented boundary matrix B^* now including the boundary lengths of the fictive pixel and every other planning units. Consequently, B^* was composed of the previous matrix B used so far, completed by an additional column and row. For consistency purpose, the decision variable vector was also augmented with a component associated with the aforementioned fictive pixel. Since it would never be selected in the reserve, this component was *a priori* set to 0). The detailed expression of the augmented matrix $B^* \in \mathbb{R}^{(N+1) \times (N+1)}$ and the vector $x^* \in \mathbb{R}^{(N+1)}$ are given in (2.3).

$$B^* = \begin{pmatrix} & & & b_{1,N+1}^* \\ & B & & \vdots \\ & & & b_{N,N+1}^* \\ \cdots & b_{N+1,1}^* & \cdots & b_{N+1,N}^* & 0 \end{pmatrix} \quad x^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_N^* \\ x_{N+1}^* \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \\ 0 \end{pmatrix} \quad (2.3)$$

The additional coefficients $b_{i,j}^*$ of the matrix B^* were used to indicate how many sides each pixel $i \in \{1, \dots, N\}$ shared with the outer boundary and thus with the fictive pixel. Hence, those extra coefficients were defined as follows :

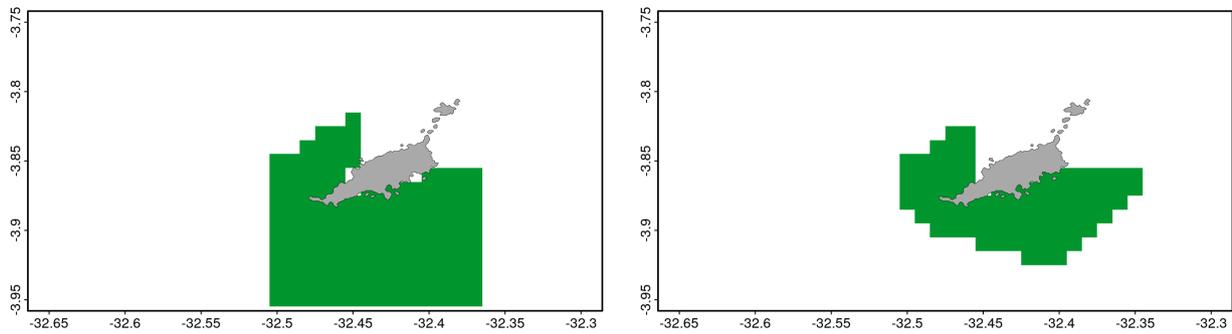
$$\forall i \in \{1, \dots, N\},$$

$$b_{i,N+1}^* = b_{N+1,i}^* = \begin{cases} 1, & \text{if pixel } i \text{ shares a single side with the outer boundary} \\ 2, & \text{if pixel } i \text{ shares 2 sides with the outer boundary (i.e. located at a corner)} \\ 0, & \text{otherwise} \end{cases}$$

Note the last diagonal coefficient $b_{N+1,N+1}^*$ was set to 0 (like the other diagonal coefficients of the matrix B) since the planning units were not connected to themselves. Considering the prior changes, the reserve perimeter was calculated as follows :

$$\begin{aligned}
\mathbf{x}^{*T} \mathbf{B}^* (\mathbf{1} - \mathbf{x}^*) &= \sum_{i=1}^{N+1} \sum_{j=1}^{N+1} x_i^* b_{i,j}^* (1 - x_j^*) \\
&= \sum_{i=1}^N \sum_{j=1}^N x_i b_{i,j} (1 - x_j) + \sum_{i=1}^N x_i b_{i,N+1}^* \\
&= \mathbf{x}^T \mathbf{B} (\mathbf{1} - \mathbf{x}) + \mathbf{x}^T \mathbf{b}^*
\end{aligned} \tag{2.4}$$

Denoting $\mathbf{b}^* = (b_{1,N+1}^*, \dots, b_{N,N+1}^*)^T$ in Equation (2.4), we can see the new perimeter calculation was composed of two terms : the known quadratic term $\mathbf{x}^T \mathbf{B} (\mathbf{1} - \mathbf{x})$ used previously to calculate the reserve perimeter but also $\mathbf{x}^T \mathbf{b}^*$ which represented the contribution of the outer boundary of the study area to the perimeter computation. Thus, the addition of a fictive pixel only involved the addition of this new term in the model. The extra row in matrix \mathbf{B}^* appeared unnecessary in the perimeter calculation since the decision variable x_{N+1} associated with the fictive pixel was always set to 0. However, the presence of this row allowed the \mathbf{B}^* to remain a square and symmetric matrix eventually allowing to write the model in a compact form.



(a) Reserve solution with $\beta = 10$ without implementing the compactness parameter correction.

(b) Reserve solution implementing the compactness parameter correction.

Figure 2.5: Reserve solutions obtained with and without the compactness parameter correction. Minimum set formulation, 3 conservation features each represented with a 50% protection target, considered cost layer worth $1 + \ln(1 + FC)$, compactness parameter $\beta = 10$, exact solving provided by CbC solver.

The modifications described above were added to both the minimum set and the maximum coverage original formulations. We performed some computational experiments with the updated minimum set formulation and compared it with the original one. Results can be found in Figure 2.5 which shows side by side the solutions obtained with and without the proposed correction. It can be seen in Figure 2.5b that the selected reserve did not extend to the edge of the area like in the original model (Figure 2.5a) and the perimeter was now, as expected, correctly derived by the model. Moreover, the CPU time required to solve this instance with the new formulation turned out to be of the same order of magnitude than the original formulation.

2.4 Discussion

2.4.1 Sensitivity of reserve solutions to structural elements

We clarified features structuring reserve site selection procedure especially underlying mathematics. Indeed, an optimisation framework implied technical choices, not always made explicit, such as the formulation and the optimisation solving method to choose to address the conservation problem. In particular, DSTs with embedded formulations and solvers (*e.g.* Marxan, Prioritizr), although comfortable to use, made impossible to customise the problem.

2.4.1.1 Minimum set vs Maximum coverage

Minimum set formulation betrays a strong economy-centered vision which could itself be discussed. Indeed, this vision, consensual following Marxan developments, aimed at minimising impacts on human activities. The main concern was thus primarily to preserve an human activity. Maximum coverage formulation, more biodiversity-aimed, can also be studied to better enlighten the problem. Computing solutions from both formulations clearly allowed to draw a more complete and balanced picture of the problem. It could also lead to better numerical interpretation. Different formulations shaped different solutions so both need to be studied in order to efficiently inform and support the decision making process. In particular, Figure 2.1 shows the kind of information decision makers could be interested in and extract. The link between reserve cost and conservation features protection level with both formulations can help decision makers to understand implication in nature protection.

2.4.1.2 Single-objective view

Optimisation formalism is inherently single-objective which means only one human activity can be properly accounted for in the reserve design process in the minimum set formulation (see Equation 2.1). Consequently, it can poorly represent several stakeholders which is a pity in the frame of MSP. We could certainly create a global anthropogenic index by combining several human activity information although it should be avoided to keep visible the competition between socio-economic costs. Multiplying single-objective computation is a better practice to highlight contradictions between stakeholders interests and leave the final arbitration to decision makers. In this sense, exact optimisation methods are more adapted as scenario multiplication is advised as described in Section 2.2.1.2. Similarly, in the maximum coverage formulation, users had to referee between conservation features weights due to the single-objective nature. Anyway, subjectivity is part of the conservation-based planning process which requires transparency in return. Finally, multi-objective optimisation field could provide better answer elements around the notion of Pareto front which could be addressed to deepen global understanding. Note it is required to use exact solving methods to draw a relevant Pareto front. Furthermore, the initial formulation in Equation (2.1) was already multi-objective as we included both the reserve cost (represented by $c^T x$) but also the reserve perimeter length (represented by $x^T B(1 - x)$) within the objective. We did not mention it explicitly, but there was an invisible competition between these two elements which could lead to misinterpretations.

2.4.1.3 Solving method

Although metaheuristic were historically preferred due to Marxan developments, exact solvers should prevail in the future. In our case study (see Section 2.3.1.2), the gap between both methods was acceptable due to the small size of the problem. A further analysis was detailed in (Schuster et al. 2020) highlighting bigger performance gaps on a wider, more meaningful application both in optimality and time computation. These matters were extensively discussed

in the past (Church et al. 1996; Pressey et al. 1996; Önal 2004; Vanderkam et al. 2007). Then, the linearisation requirement associated with exact methods was not a problem in reserve site selection as the linearisation is possible (Billionnet 2007). Finally, the number of solutions provided by metaheuristics were often seen as an advantage (Ardron et al. 2010) while interpreting many suboptimal solutions is a challenge by itself. First, this numerical so-called flexibility is questionable as we do not know how far from optimum solutions are. In addition, a given stakeholder could easily find among these, a solution suiting its own interests. It could eventually leave the final choice to the most influential lobby and be the breeding ground of ocean grabbing (Queffelec et al. 2021). Unlike metaheuristics, exact solvers provides a single optimal solution, greatly encouraging the multiplication of scenarios to better enlighten the conservation question. Yet, if alternative solutions were really needed, it could also be achieved with exact solvers. A simple procedure could consist in excluding iteratively solutions found and thus derive the optimal solution exhaustive set. We could even introduce a relaxation parameter to explore suboptimal solutions with an *a priori* given optimal gap.

2.4.2 Input data

We highlighted biases due to the input data feeding a conservation-based planning approach. Indeed, these DSTs are data driven, restricted to a spatially explicit nature, which means outcome can only reflect geographical layers input quality. Besides, as we cannot access true spatial distribution, we must use surrogate data. It implies to process data in order to be as close as possible to an unknown reality and thus express a relevant optimisation problem. However, as illustrated in Section 2.2.2, data processing involved many modelling choices. It appeared other choices could have been made with the same level of relevance but with a potential different reserve outcome. Such ideas must be clearly stated and understood by practitioners. Below, we provided a list of elements regarding data feeding DSTs that needed to be questioned by stakeholders as they can greatly influence reserve design results.

2.4.2.1 Surrogacy

The way we estimate a quantitative index from a surrogate dataset is a sensitive step with respect to the reserve output as we demonstrated in Section 2.3.1.3. Such technical step, although done honestly, can eventually lead to imbalance between stakeholders involved in the MSP or towards biodiversity conservation. The way we go from the raw material to a processed and gridded input data can lead to quite different solutions. For example, arbitrary application of a logarithm function to smoothen raw data can advantage a given stakeholder without anyone noticing. To avoid this pitfall, only sensitivity analyses and transparency on the transformations applied to the raw material can deepen user's grasp on data processing influence. Few works dealt with data processing influence (Drira et al. 2019; Visconti et al. 2013; Carvalho et al. 2010; Fiorella et al. 2010). Besides, a measure is at some point guided by reality (biological, economical, geographical accessibility of the surrogate measure) which implies a natural bias towards accessible data. For example, megafauna is potentially over-represented while other smaller species can be underestimated if not voluntarily forgotten due to the lack of surrogate measure. Only large data collection surveys and data gathering can mitigate this effect. Finally, we can also argue that the mere fact to whether or not include a given feature in the reserve design process is a first and essential bias. Considering a given feature obviously implies it will be accounted for in the procedure but also mean other features will be completely forgotten (by choice, lack of data or even knowledge). Therefore, considered features inherently imbalance stakeholders and biodiversity interests. For example, a stakeholder struggling to provide data will be under-represented and thus potentially harmed through the MSP process. Similarly, a

species impossible to track is not accounted for in the reserve design and can suffer from the MSP process.

2.4.2.2 Data type

Data feeding reserve site selection DSTs are necessarily spatially explicit, *i.e.* quantitatively located in space, allowing us to associate each planning unit cost and conservation feature amounts. For example, a conservation feature can be a quantity like a biomass or a number of items. However, nature of data involved in MSP is not always spatially explicit and conversion can be difficult if not impossible. Indeed, data can be purely qualitative or at best semi-qualitative. Consequently, such data cannot be accounted for in reserve site selection tools and can be removed from the input dataset and potentially weaken a given stakeholder. For example, deriving a map of the diving activity is hard as quantifying this activity can be at best done thanks to shade of diving pressure from "low number of visits" to "diving hotspot". One way to mitigate such weakness in first approximation is to transform the best we could qualitative data into semi-quantitative one with level of intensity.

2.4.2.3 Quantity and quality

Stakeholders providing great data both in quality and quantity is likely to be favoured through DSTs as their interests will be well represented and not forgotten within the site selection procedure and even potentially protected. Such DSTs follow the "garbage in, garbage out" rule which underlines their strong data dependency. Indeed, solutions can only be as good as the input data feeding the optimisation model. Once again, only large data collection surveys of every feature can enhance quality and equity of the reserve design process.

2.4.2.4 Uncertainty

Data is considered certain in most reserve site selection algorithms as uncertainty is difficult to handle within an optimisation framework. Yet, uncertainty is everywhere due to measure itself but also inherent to ecological processes. For example, a value of 0 is algorithmically equivalent to a certain absence while it can practically mean a lack of data sampling. Accounting for uncertainty in reserve selection procedure is a great deal and several approaches help to mitigate this lack (Monte-Carlo approach, robust optimisation, chance constraints, stochastic optimisation, *etc.*). (Regan et al. 2009; Reside et al. 2018)

2.4.3 Parameters influence

Through the illustration provided in our work, we detailed to what extent parameters (conservation feature coverage targets, compactness parameter) choices can widely shape the results of a reserve site selection procedure. Such statement appears more than logical in a parametric model, however it is important to establish a quantitative link between parameters and outputs. Deciphering parameters influence can also avoid imbalances in the MSP process in favour of more technical users.

2.4.3.1 Coverage targets

If not ecologically guided, coverage targets can be used as tuning parameters. Indeed, as demonstrated in Section 2.3.1.1, these parameters were directly linked with the reserve cost (*cf.* Figure 2.1) and should be manipulated with great care as it could lead to imbalance in the marine spatial planning process. If any ecological information is available, a sensitivity analysis

on the coverage targets is the minimum that should be realised to best inform the reserve design process.

2.4.3.2 Compactness parameter

Examples in the published literature showed that an unwanted edge effect due to the compactness parameter was not particularly highlighted. Formulation should be modified to avoid such traps (locked-out a fictive pixel linked to every pixel at the border for instance, see Appendix 2.3.2.3 for detailed formulation). Finally, $x^T B(1 - x)$ compactness share is historically included in the objective. Yet, another legitimate approach could be to directly constrain the outside perimeter with a given boundary budget b_p leading to the constraint $x^T B(1 - x) \leq b_p$ (*e.g.* see Equation (2.2)). Such approach would be more straightforward and avoid invisible multi-objective competition between compactness and coverage. Besides, a blind setting of the compactness parameter potentially provides unwanted numerical effects but can also lead to a change of "regime" in the solution, *i.e.* a completely different solution because the compactness demand overcome the original objective. In particular, a given regime can favour a stakeholder with respect to another so a great care must be observed. Performing a sensitivity analysis on the compactness parameter is the least we can do to have a better grasp on its influence (see Section 2.3.2.2).

2.5 Conclusion

Few works had already pointed out some effects from input data initial formats (Carvalho et al. 2010; Visconti et al. 2013), interpolation transformation (Drira et al. 2019) and weighting (Fiorella et al. 2010). The example developed in this article builds upon and goes further by systematically clarifying the mathematical functioning of each step of reserve selection DSTs to end-users through numerical and graphical illustrations. We deciphered the effects data and parameterisation options may have on the final solutions and showed that DSTs present at least two points of attention. The first confirmed the tricky issue of input data (bathymetry, fishing, proxies used, *etc.*) which significantly influenced the DST results. Similarly, the absence of data may penalise certain stakes without this always being spelled out. The second concerns the numerous technical choices made throughout the process by the DSTs users and designers: from the definition of the grid playing as spatial referential to the processing of the data, including the "minimum set vs. maximum coverage" choices, *etc.* Based on our case study, we provide specific guidelines for mitigating to some extent these technical pitfalls:

- Perform sensitivity analyses on parameters to enhance numerical understanding
- Compute both the minimum set and the maximum coverage formulation to better enlighten the conservation problem
- Document with transparency every modelling choice, in particular regarding the construction of the objective function which implies inherent subjectivity (*e.g.*, how the cost is built)

More generally, we illustrated that the informational questions are spread over the entire geographic information chain, from data production to its use for management purposes. In this sense, this study finally raised fundamental questions about the place and role of the data producers, the technicians who process it and the decision-makers who use it. As roles become blurred (Goodchild 2009), it is necessary to try to take into account the needs of the end users the most upstream in this geographic information chain, either by involving them in each of the stages, or by making each of these stages and the associated issues more understandable and accessible by them. Consequently, a better knowledge of the issues at stake throughout this geographic information chain will foster a better understanding of the various biases noted in this example, thus allowing to avoid most of the traps, and *in fine* limits the risk of ocean grabbing (Queffelec et al. 2021) and favour equitable MSP negotiations.

Transition

In this chapter, we analysed several aspects of the optimisation models for reserve site selection. We opened the black-box of such tools by performing sensitivity analyses toward several parameters, questioned the optimisation models, addressed the solving methods, in order to draw a clearer picture of the issues involved. During the analysis of the solving methods, we were confronted to the difficulty of analysing an important number of reserve solutions by using metaheuristics algorithm. We were surprised that this was considered as an advantageous feature of metaheuristics within the conservation literature. Yet, it appeared to us as an additional complexity involving an heavy post-processing. That is why, in the next chapter, we focused on producing a small amount of truly different reserve solutions using exact optimisation methods.



Chapter 3

Producing a diverse set of near-optimal reserve solutions with exact optimisation

3.1 Introduction

Biodiversity and habitats are threatened worldwide (IPBES 2019). Building comprehensive networks of nature reserves has become a popular conservation solution (Ticco 1995; Tundi Agardy 1994; Le Saout et al. 2013) and was shown to bring significant benefits (Claudet et al. 2008; Stolton and Dudley 2010; Liu et al. 2017). At sea, for instance, current political objectives are to cover 30% of the marine spaces under jurisdiction by 2030 with marine protected areas (IUCN 2014; IUCN 2016; Commission 2020). Similar concerns also exist on land (Baillie and Zhang 2018; Dinerstein et al. 2019). Several methods (often embedded within a decision support tool software, *e.g.* Marxan or PrioritizR) select reserve sites given constraints and objectives.

A set of alternative reserve solutions is usually necessary to create effective nature reserves as they support better negotiations between different stakeholders. In a decision-making framework, constraints and objectives of reserve design may be difficult to formulate in the context of a spatially-explicit numerical optimisation. Any unmodeled phenomenon that may influence the decision can lead to divergences with the proposed solutions. For example, complex social mechanisms govern the final decision of wind farm locations (Bell et al. 2005; Virtanen et al. 2022). The decision process, based eventually on negotiations, thus requires some latitude on the possible solutions to be considered. Generating alternative solutions gives conservation practitioners the possibility of finding a solution that could be more satisfactory with respect to these unmodeled objectives. For these reasons, the ability of decision support tools to produce a range of solutions instead of a single one, has been put forward frequently in conservation literature (Pressey et al. 1993; Possingham et al. 2000; Possingham et al. 2006; Sarkar 2012; Ferretti et al. 2019). Consequently, reserve site selection tools do need options for generating different, near-optimal alternatives.

The ability to produce alternative solutions has often been presented as a key strength of metaheuristic algorithms over exact optimisation approaches (Pressey et al. 1993; Possingham et al. 2000; Possingham et al. 2006; Sarkar 2012). In a survey realised among Marxan users (Ardrón et al. 2010), "generating multiple solutions was by far the most commonly noted strength of Marxan" over other reserve site selection algorithms. Yet, recent advances made exact optimisation methods more attractive for conservation practitioners (Schuster et al. 2020; Hanson et al. 2019; Beyer et al. 2016; Rodrigues and Gaston 2002) because they provide the optimal solution even for large-scale instances in reasonable time. The ability of generating multiple solutions thus seems to be the last argument remaining in favour of metaheuristic algorithms. Technically speaking, metaheuristics rely on random processes to create an implicit diversity within the set of solutions (see Appendix B-2.1 in (Serra-Sogas et al. 2020)). On the contrary, exact solving methods usually produce, by construction, a single optimal solution and are not designed for producing a range of different solutions. This major limitation severely restricts the ability of exact solving methods to inform real-world conservation problems. However, in the absence of explicit criteria, metaheuristic approaches do not provide any control over the alternatives generated, nor do they guarantee to have truly different solutions. The search for near-optimal alternative solutions with exact solving methods began to be discussed in (Fischer and Church 2005) and was explored in the general context of operational research (Chang et al. 1982a; Brill et al. 1982; Chang et al. 1982b; Makowski et al. 2000). The algorithm developed in (Arthur et al. 1997) computes the exhaustive set of optimal solutions of a reserve site selection problem. In the same line, a branch and bound screening algorithm showed how suboptimal solutions can be derived with exact methods (Önal 2004). The reserve site selection tool PrioritizR also provides additional functions allowing users to build a portfolio

of alternative solutions¹. Though, producing alternative solutions based on their distance to optimality with exact solving methods still does not guarantee to provide different solutions. This motivated us to explicitly introduce a dissimilarity measure in the search for alternative solutions.

In this work, we propose two iterative algorithms incorporating an explicit dissimilarity criterion to build a range of near-optimal solutions significantly different from each other with exact solving methods. Solutions are selected based on a controlled objective value degradation and using an explicit dissimilarity measure. We discuss and compare two metrics for the dissimilarity criterion. Our results show that generating alternative solutions according to the objective value interval can result in a low variability among solutions, as they are very similar to each other. These solutions which only differ by a few planning units are quite uninformative. They can hardly be considered as alternatives. Even worse, further from the optimal value, the variability among alternative solutions appears irrelevant because the procedure artificially increases the objective value by including pointless planning units. As such, it poorly answers the need for both good and different alternative solutions. The dissimilarity measure we incorporate allows us to overcome this limitation. The proposed algorithms explicitly seek to generate dissimilarity between reserve solutions and provide true alternatives. Similarly to (Chang et al. 1982b; Brill et al. 1982), the dissimilarity measure we define allows to avoid alternative reserves embedding the optimal one. Another pitfall, particularly striking in metaheuristic approaches, is the need to generate numerous alternative solutions in order to widely explore the solution space. This large amount of alternative solutions requires a statistical post-processing to identify a few distinct solutions. It often requires additional statistical analyses, *e.g.* the selection frequency of reserve sites or clustering analysis (Ardrón et al. 2010; Harris et al. 2014; Linke et al. 2011). By contrast, our methods directly provide a presentation set composed of significantly distinct solutions. A few alternatives that are both good and different from each other can thus be sufficient.

At this point, it is necessary to precise the terminology chosen to dissipate any ambiguity about the generation of alternative solutions of an optimisation problem. Indeed, this concept is covered by different terminologies in conservation biology literature. For example, the term "flexibility" was used in (Ardrón et al. 2010) for designating this feature of Marxan. In other works, this term referred to the ability of an optimisation model to be easily tuned (Rodrigues et al. 2000; Cocks and Baird 1989). We thus avoided the term flexibility. In PrioritizR, the term "portfolio" is used for functions generating a set of alternative solutions. In this context, a portfolio means a collection of solutions. In other works (*e.g.* (Fischer and Church 2005; Ardrón et al. 2010; Delavenne et al. 2012; Metcalfe et al. 2015) among others), portfolio was used for designating a collection of sites, *i.e.* one reserve solution. To dissipate this ambiguity, we also avoided the term portfolio. Finally, in the operations research literature, the term chosen is "presentation set". This term makes explicit the fact that these alternative solutions are intended to be presented to decision-makers and decided upon. In this work, we used "presentation set" to name the collection of alternative solutions of an optimisation problem.

¹More details at <https://prioritizr.net/reference/portfolios.html>

3.2 Methods

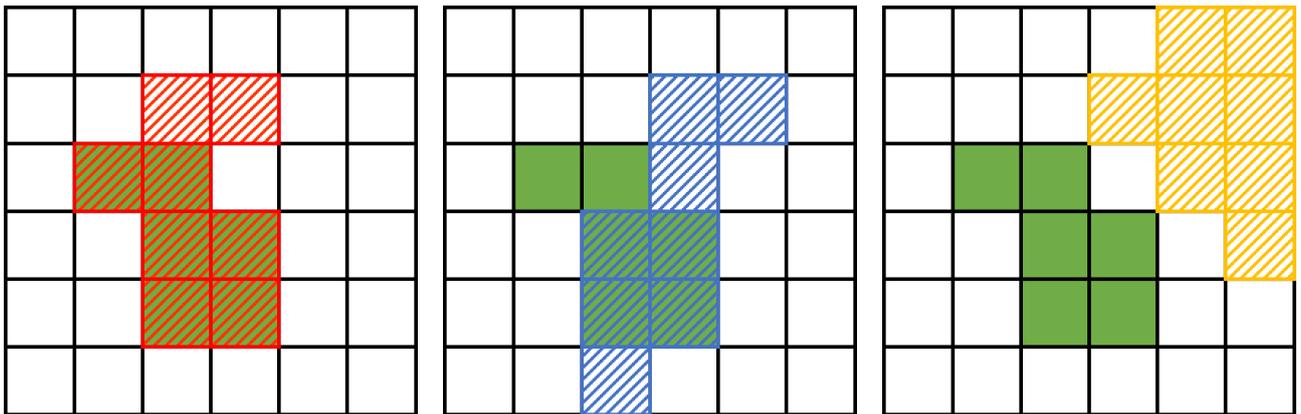
In this section, we present the dissimilarity measure we used for quantifying the differences between two reserve solutions. Then, we provide the mathematical formulation of new approaches incorporating our dissimilarity measure for generating presentation sets. The general formulation of the reserve site selection considered is detailed in Section 1.4.2.

3.2.1 Measuring the dissimilarity between two reserve solutions

For providing a diverse presentation set, we first need a function characterising the dissimilarity between two solutions. Consequently, when $x, y \in \{0, 1\}^N$, we defined the following dissimilarity measure:

$$d(x, y) = \sum_{j \in J} x_j(1 - y_j),$$

Dissimilarity measure d counts the number of planning units selected in x and not in y . This measure is rather a pseudo-distance than a distance, because it does not meet the separation property. Indeed, $d(x, y) = 0$ does not imply that $x = y$. Instead, $d(x, y) = 0$ whenever the reserve defined by x is included in that defined by y . This is actually an enjoyable property for the production of alternative solutions, because there is no real point in presenting an alternative solution that would strictly include an optimal solution x^* . As an illustration, Figure 3.1 displays three examples where the dissimilarity measure d is compared with the distance $D(x, y) = d(x, y) + d(y, x)$, which is equal to 0 only if $x = y$. For instance, on the leftmost example, we see that the red solution is simply a worse alternative to the green solution, so we wanted to avoid this case. This type of measure was proposed in the context of land-use planning through the 4 steps of the "Hop Skip and Jump" procedure (Brill et al. 1982).



(a) The green reserve is strictly included in the red reserve. We have $d(x, y) = 0$ and $D(x, y) = 2$.

(b) The blue reserve has planning units both inside and outside the green reserve. We have $d(x, y) = 2$ and $D(x, y) = 6$.

(c) The green reserve has an empty intersection with the yellow reserve. We have $d(x, y) = 6$ and $D(x, y) = 14$.

Figure 3.1: Numerical examples of the dissimilarity measure d and distance D . The reserve x depicted in green includes 6 planning units. Other reserves y , hatched in red, blue and yellow, include 8 planning units. According to the dissimilarity measure d , the green and red reserve are the same. The dissimilarity measure characterises differences between two reserves as much as distance D as illustrated with the blue and yellow reserve examples.

3.2.2 Computing a presentation set

In this section, we describe our two methods that compute a presentation set using the dissimilarity measure d . The approach proposed by the `add_gap_portfolio` function² of PrioritizR is referred to as the Algorithm `AddGapPortfolio`.

3.2.2.1 Imposing dissimilarity between alternative solutions

For some predefined dissimilarity threshold δ and k feasible solutions x^0, \dots, x^{k-1} of P_0 , we may impose that a new alternative solution x differs sufficiently from x^l for every $l \in \llbracket 0, k-1 \rrbracket$ by constraining $d(x^l, x)$ to be at least equal to δ . More formally, this can be achieved by adding the following constraints to the initial optimisation problem P_0 :

$$c_d(x^l, \delta) : d(x^l, x) = \sum_{j \in J} x_j^l (1 - x_j) \geq \delta, \quad \forall l \in \llbracket 0, k-1 \rrbracket.$$

The integer linear program solved at iteration $k \geq 1$ is P_1^k such as:

$$P_1^k : \begin{cases} \min_x & \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} x_{j_1} (1 - x_{j_2}) \\ \text{s.t.} & \sum_{j \in J} a_{ij} x_j \geq t_i & \forall i \in I \\ & d(x^l, x) \geq \delta & \forall l \in \llbracket 0, k-1 \rrbracket \\ & x_j \in \{0, 1\} & \forall j \in J \end{cases}$$

Algorithm `MinDegradation` details the pseudocode of the iterative procedure we implemented to produce the presentation set. The procedure stops if the problem becomes infeasible or the maximum number of iterations n is reached. Infeasibility is reached when no alternative satisfying the dissimilarity constraints can be found. If the user wants a larger presentation set, they may choose a smaller threshold δ .

Algorithm `MinDegradation` Iterative search of n alternative solutions of P_0 with at least δ dissimilarities to the past solutions.

Require: P_0, x^*, n, δ

Ensure: x^1, \dots, x^k

- 1: $k \leftarrow 0; P \leftarrow P_0; x^0 \leftarrow x^*$ ▷ initialisation
 - 2: **while** (P is feasible & $k < n$) **do** ▷ stop when infeasible or enough solutions
 - 3: $k \leftarrow k + 1$
 - 4: add $c_d(x^{k-1}, \delta)$ to P ▷ impose dissimilarities to the past solutions
 - 5: solve P ▷ get an optimal solution x^k or detect that P is infeasible
 - 6: **end while**
-

3.2.2.2 Maximising dissimilarity between alternative solutions

Another option we investigated was to iteratively seek for the most different solution at a user-defined extra cost relatively to the optimal value. More formally, assume that $k-1$ alternative solutions x^1, \dots, x^{k-1} have been previously computed. At iteration $k \geq 1$, we seek for an alternative solution x^k that maximises $\Delta(x) = \min \{d(x^l, x) : l \in \llbracket 0, k-1 \rrbracket\}$ among the solutions that do not exceed the cost $(1 + \gamma)z^*$. Given that $\Delta(x)$ is not a linear function of x , we needed to linearize its expression. The classical method to do this uses that the minimum value among a finite set is the maximum lower bound of the set, *i.e.*, $\max_x \Delta(x) =$

²More details at https://prioritizr.net/reference/add_gap_portfolio.html.

$\max_{\Delta} \{\Delta : \Delta \leq d(x^l, x)\}$. The corresponding mixed integer linear program we solved at iteration $k \geq 1$ is given by:

$$P_2^k : \left\{ \begin{array}{ll} \max_{x, \Delta} & \Delta \\ \text{s.t.} & \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} x_{j_1} (1 - x_{j_2}) \leq (1 + \gamma) z^* \\ & \sum_{j \in J} a_{ij} x_j \geq t_i & \forall i \in I \\ & \sum_{j \in J} x_j^l (1 - x_j) \geq \Delta & \forall l \in \llbracket 0, k - 1 \rrbracket \\ & x_j \in \{0, 1\} & \forall j \in J \\ & \Delta \in \mathbb{R}_+ \end{array} \right.$$

Algorithm [MaxDissimilarity](#) details the pseudocode of the iterative procedure we implemented to produce the presentation set. We provided more numerical details for this algorithm in Section [3.3.3](#).

Algorithm MaxDissimilarity Iterative search of n alternative solutions maximising the dissimilarity to the past solutions at a relative extra cost budget γ .

Require: $P_2^1, x^*, z^*, n, \gamma$

Ensure: x^1, \dots, x^k

- 1: $k \leftarrow 0; P \leftarrow P_2^1; x^0 \leftarrow x^*$ ▷ initialisation
 - 2: **while** (P is feasible & $k < n$) **do** ▷ stop when infeasible or enough solutions
 - 3: $k \leftarrow k + 1$
 - 4: add $c_d(x^{k-1}, \Delta)$ to P ▷ impose dissimilarities to the past solutions
 - 5: solve P ▷ get an optimal solution x^k or detect that P is infeasible
 - 6: **end while**
-

3.2.2.3 Imposing an objective value interval

We show how we produced the presentation set composed of alternative solutions located at a predefined objective value interval. We here developed our own algorithm although the function `add_gap_portfolio` of `PrioritizR` allows to generate the same set of alternative solutions.

Let $\gamma_1 \in \mathbb{R}^+$ and $\gamma_2 \in \mathbb{R}^+$, such as $\gamma_1 \leq \gamma_2$, be the boundaries of the objective value interval relatively to the optimal value z^* . The constraints $c_l(\gamma_1)$ and $c_u(\gamma_2)$ are imposing the objective value to belong to the predefined interval $[(1 + \gamma_1)z^*, (1 + \gamma_2)z^*]$:

$$c_l(\gamma_1) : \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} x_{j_1} (1 - x_{j_2}) \geq (1 + \gamma_1) z^*$$

$$c_u(\gamma_2) : \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} x_{j_1} (1 - x_{j_2}) \leq (1 + \gamma_2) z^*$$

If $\gamma_1 = \gamma_2 = 0$, we explore only the optimal solutions set. For $\gamma_1 > 0$, we explore alternative solutions that are strictly suboptimal.

The constraint $c_D(y, \delta)$ impose the solution x to have at least δ different planning units with respect to y :

$$c_D(y, \delta) : D(x, y) = \sum_{j \in J} y_j (1 - x_j) + x_j (1 - y_j) \geq \delta$$

Importantly, $\delta = 1$ forbids x and y to be strictly equal.

Practically, we first add to the optimisation problem the constraints $c_l(\gamma_1)$ and $c_u(\gamma_2)$ which must be satisfied at every iteration. Then, to derive a pool of alternative solutions, we excluded

at iteration $k \geq 1$ the solution x^{k-1} derived the iteration before. The addition of constraint $c_D(x^{k-1}, 1)$ guarantee this. Indeed, this constraint prevents the searched solution at iteration $k \geq 1$ to be exactly x^{k-1} .

Practically, the integer linear program solved at iteration $k \geq 1$ is P_3^k such as:

$$P_3^k : \left\{ \begin{array}{ll} \min_x & \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} x_{j_1} (1 - x_{j_2}) \\ \text{s.t.} & \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} x_{j_1} (1 - x_{j_2}) \leq (1 + \gamma_2) z^* \\ & \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} x_{j_1} (1 - x_{j_2}) \geq (1 + \gamma_1) z^* \\ & \sum_{j \in J} a_{ij} x_j \geq t_i & \forall i \in I \\ & \sum_{j \in J} x_j (1 - x_j^l) + x_j^l (1 - x_j) \geq 1 & \forall l \in \llbracket 0, k - 1 \rrbracket \\ & x_j \in \{0, 1\} & \forall j \in J \end{array} \right.$$

The constraints $c_l(\gamma_1)$ and $c_u(\gamma_2)$ used in P_3^k are not linear. We linearised these constraints exactly as we did for the model P_0^f described in Section 1.4.2. Algorithm [AddGapPortfolio](#) details the pseudocode of the iterative procedure we implemented to produce the presentation set. The procedure stops if the problem becomes infeasible or the maximum number of iterations n is reached. Infeasibility is reached when the objective value of the alternative solution exceeds the upper bound γ_2 . If the user wants a larger presentation set, they may choose a greater threshold γ_2 . For instance, if $\gamma_1 = 0$ and γ_2 is high enough, Algorithm [AddGapPortfolio](#) returns the n solutions with the smallest objective value. If n is chosen high enough, Algorithm [AddGapPortfolio](#) returns the exhaustive set of solutions with an objective value relatively to the optimal value within the interval $[\gamma_1, \gamma_2]$. Unlike metaheuristics where the optimality gap is unknown, we *a priori* established it using this algorithm. We thus offer users more control over the presentation set provided.

Algorithm AddGapPortfolio Iterative search of n best alternative solutions whose objective values relatively to the optimal value z^* of solution x^* of problem P_0 belongs to the predefined interval $[\gamma_1, \gamma_2]$.

Require: $P_0, x^*, z^*, n, \gamma_1, \gamma_2$

Ensure: x^1, \dots, x^k

- 1: $k \leftarrow 0; P \leftarrow P_0; x^0 \leftarrow x^*$ ▷ initialisation
 - 2: add $c_l(\gamma_1)$ and $c_u(\gamma_2)$ to P
 - 3: **while** P is feasible & $k < n$ **do** ▷ stop when infeasible or enough solutions
 - 4: $k \leftarrow k + 1$
 - 5: add $c_D(x^{k-1}, 1)$ to P ▷ exclude previous solution
 - 6: solve P ▷ get an optimal solution x^k or detect that P is infeasible
 - 7: **end while**
-

3.2.3 Illustration of the approaches for computing a presentation set

Figure 3.2 illustrates schematically how the alternative solutions are selected by different methods. Solutions are mapped in a specific plane: the optimality gap versus the dissimilarity to the optimal solution of P_0 . Figure 3.2a illustrates the alternative solutions produced by repeating a metaheuristic algorithm such as simulated annealing. These alternative solutions would be scattered in the considered plane. These are neither guaranteed to be close to optimality nor to be different from the optimal solution. Algorithm [AddGapPortfolio](#) selects the alternative solutions in a given objective value interval. Figure 3.2b shows the alternative solutions that

would produce Algorithm [AddGapPortfolio](#). These solutions can be at any dissimilarity to the optimal solution, although their objective value belongs to a predefined interval by construction. Figure 3.2c shows how Algorithm [MinDegradation](#) would select the leftmost solution among the solutions above a predefined dissimilarity threshold. In other words, Algorithm [MinDegradation](#) would select the solution closest to the optimum at a fixed dissimilarity measure. Similarly, for Algorithm [MaxDissimilarity](#), the first alternative selected would be the solution with the most dissimilarity given a tolerated degradation of the objective value. In Figure 3.2d, this corresponds to the topmost solution among the solutions at the left of a predefined objective value threshold.

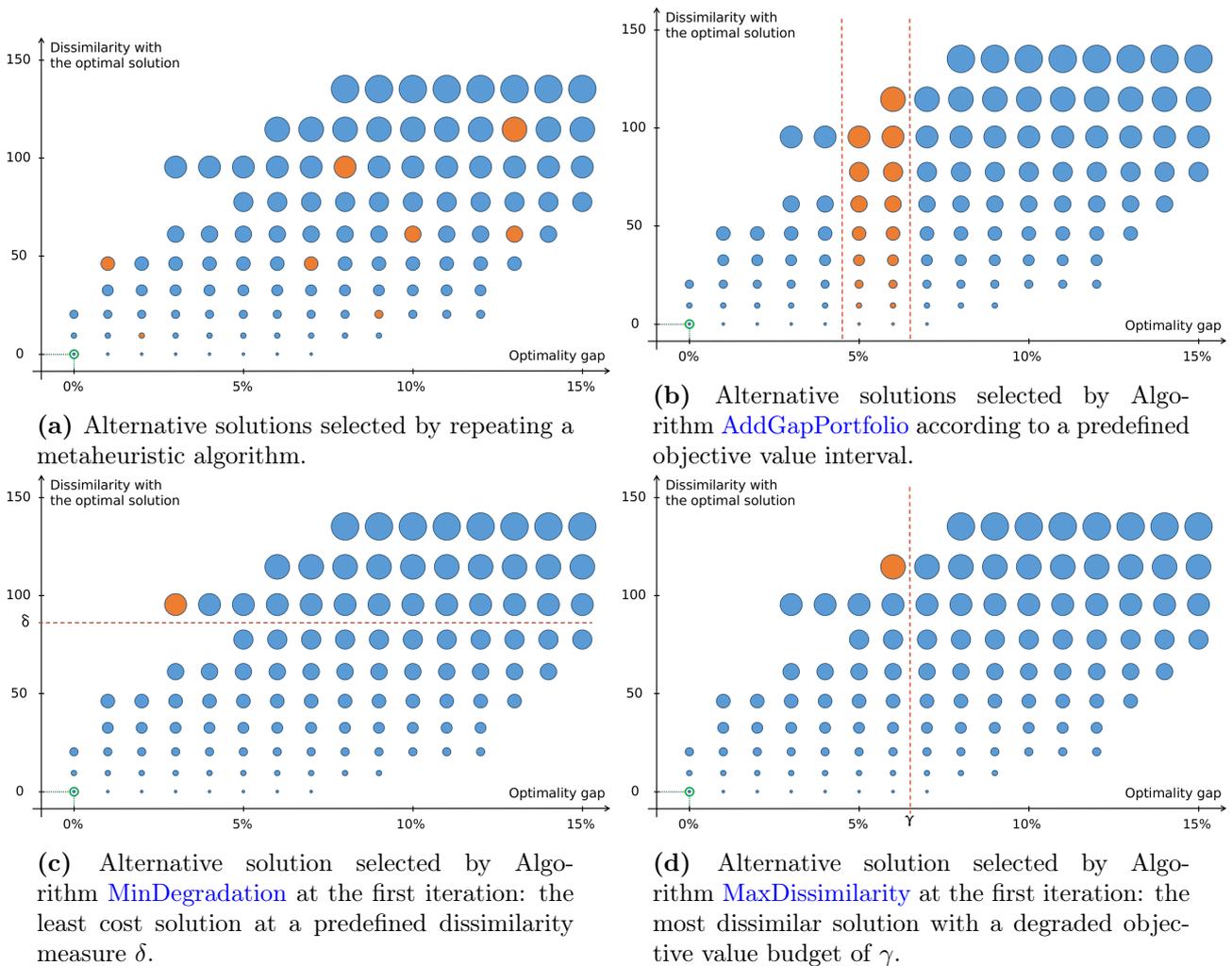


Figure 3.2: Schematic representation of alternative solutions selected by metaheuristics, Algorithm [AddGapPortfolio](#), Algorithm [MinDegradation](#) and Algorithm [MaxDissimilarity](#). Each circle represents a reserve solution. The reserve solutions are located by the optimality gap and the dissimilarity to the optimal solution $d(x^*, x)$. The optimal solution x^* is circled in green at the bottom left of this plan. The bigger the circle, the greater the dissimilarities to the optimal solution. Alternative solutions that would select an algorithm are depicted with orange circles.

3.3 Case study

The code for this work is open, free and available³. We used the Gurobi solver under a free academic licence called through a code developed in Julia language (Bezanson et al. 2012; Bezanson et al. 2015) using the JuMP optimisation library (Dunning et al. 2017). The developed methods were numerically tested on the real-world example of Fernando de Noronha composed of 3 conservation features and 756 planning units. The details of the data describing the case study can be found in Section 1.5.1.

3.3.1 Presentation set imposing an objective value interval

Exhaustive set of optimal solutions ($\gamma_1 = \gamma_2 = 0$) Algorithm [AddGapPortfolio](#) returns all the solutions whose relative optimality gaps belong to the interval $[\gamma_1, \gamma_2]$. By setting $\gamma_1 = \gamma_2 = 0$, we were thus able to compute the exhaustive set of optimal solutions. Here, we have 16 optimal solutions with an objective value $z^* = 197.71$. Panel A of Figure 3.3 shows the selection frequency among optimal solutions, *i.e.* the percentage of time a planning unit was selected among the 16 optimal solutions. We observe a low variability since 84 over 93 planning units were selected at a 100% frequency. The nine planning units with a selection frequency below 100% have all a cost of 1 and have similar amounts of conservation features. The 16 optimal solutions are composed of combinations of these nine planning units of similar characteristics that still meets targets.

Alternative solutions by increasing optimality gaps ($\gamma_1 = 0, \gamma_2 > 0$) We computed the $n = 500$ following suboptimal solutions from best to worst optimality gap. We set γ_2 to a high enough value to be certain Algorithm [AddGapPortfolio](#) produces the n alternative solutions. The objective value of the last and worst solution returned by the algorithm was 198.98, which corresponded to an optimality gaps of 0.64%. Any value of γ_2 greater than 0.64 would have led to the same 500 alternative solutions. As above, Panel B of Figure 3.3 illustrates the selection frequency of these 500 alternative solutions. Again, a low variability is observed, because many planning units have similar characteristics and are interchangeable. The visual impression is thus similar to the exhaustive set of optimal solutions.

Alternative solutions within an objective value interval ($\gamma_1 > 0, \gamma_2 > 0$) We set $\gamma_1 > 0$ to get suboptimal solutions with an optimality gap of at least γ_1 . We chose γ_2 high enough to have $n = 100$ alternative solutions. Panel C and D in Figure 3.3 respectively show results for $\gamma_1 = 0.05$ and $\gamma_1 = 0.15$. We observe a greater variability than for the two previous presentation sets. However, when comparing the conservation features distribution in Figure 1.7, many planning units selected in the alternative solutions do not increase the amount of conservation features in the reserve nor decrease its perimeter. These planning units are only selected to deteriorate the objective value and thus satisfy the constraint of the objective value interval. Although the variability appears greater in Panel C and D compared to other panels, the core of the reserve is still globally similar to the optimal solutions.

3.3.2 Presentation set imposing a dissimilarity measure

In this section, we applied Algorithm [MinDegradation](#) to our case study. Figure 3.4 shows $n = 4$ alternative solutions found by the iterative procedure for $\delta = 20$. We first observe that, as expected, at least 20 planning units selected in the optimal solution x^* are not found in the alternative solutions. The dissimilarity to the optimal solution appears more clearly than the

³GitHub repository at https://github.com/AdrienBrunel/rssp_presentation_set.git

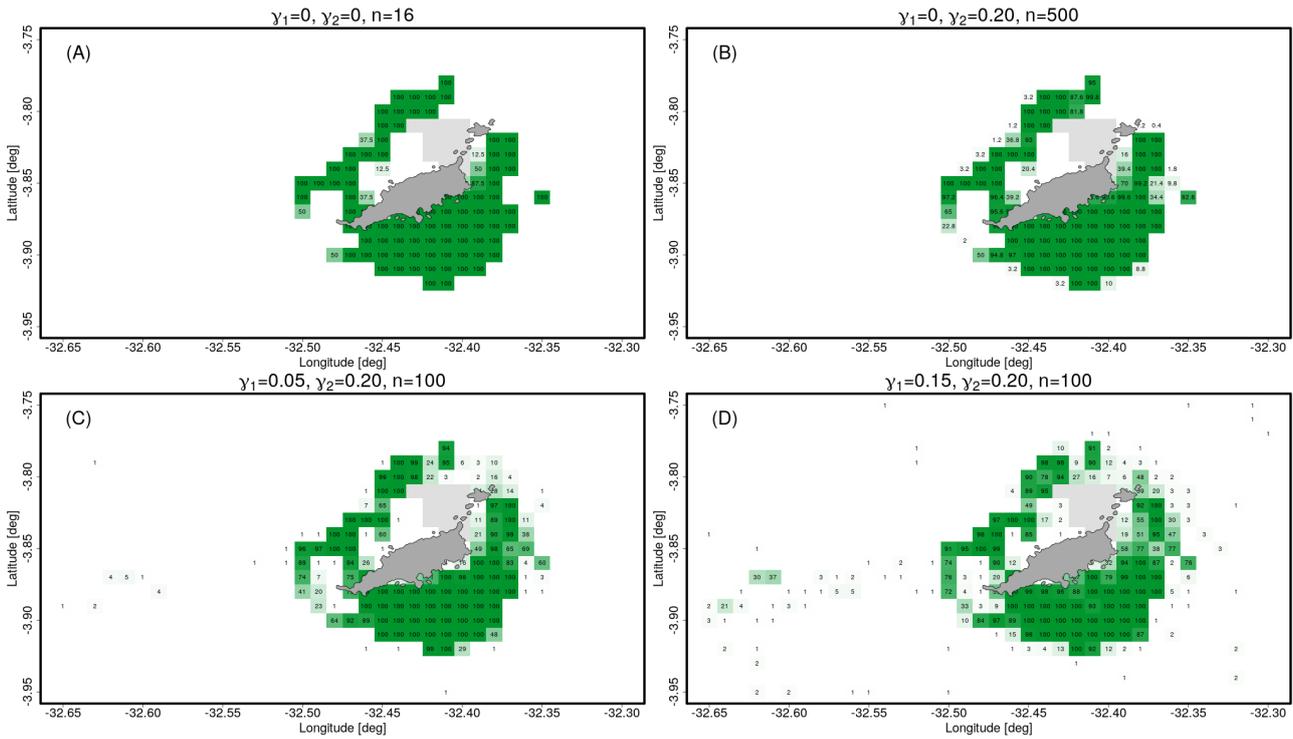


Figure 3.3: Selection frequency among alternative solutions obtained with Algorithm [AddGapPortfolio](#). Selected planning units of alternative reserve solutions are represented with a green transparency gradient according to the selection frequency expressed in percentage (black number inside planning unit).

alternatives produced by Algorithm [AddGapPortfolio](#). The alternative solutions proposed in Figure 3.4 cover different regions of the archipelago, although limited by the fact that non-zero amounts of conservation feature are aggregated around the main island. In particular, the southern region is privileged in Panel B whereas the north and east of Fernando de Noronha are preferred in Panel C. Panel A shows a solution similar to the optimal one, although two planning units are selected at the extreme west of the study area. Panel D displays a solution cut into several pieces all around the main island.

3.3.3 Presentation set maximising the dissimilarity measure

We applied Algorithm [MaxDissimilarity](#) to find the $n = 4$ alternative solutions maximising the minimum dissimilarity to the past solutions. Illustrations are presented in Figure 3.5 for a relative extra cost budget of $\gamma = 10\%$. Interestingly, a clear visual difference between the four alternative reserves appears in Figure 3.5. Panel A proposes a solution cut into 4 pieces, favouring the east of the archipelago. Panel B shows a clear preference for the south of the island. Panel C is the most similar to the optimal solution, although two planning units are found at the extreme west of the main island. Panel D presents a reserve in two pieces, one in the north and one in the south. The dissimilarity measure between the reserve in Panel A and the optimal one is equal to 27 planning units. The first iteration simply maximises the dissimilarities to the optimal solution. The minimum dissimilarity measure between the reserve in Panel B to past solutions is 22: the dissimilarity to the optimal solution and to the solution in Panel A are both equal to 22. By definition of Algorithm [MaxDissimilarity](#), there is no other solution such that the dissimilarity measure from those two past solutions are both greater than 22.

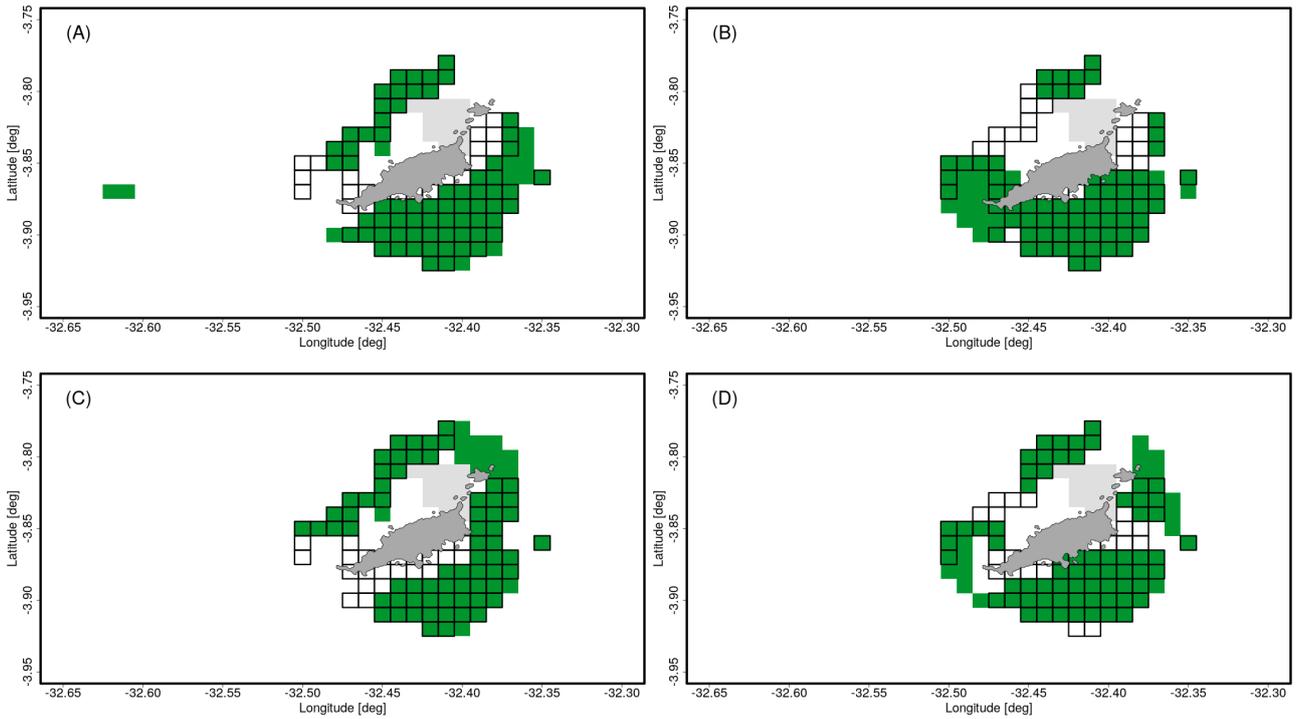


Figure 3.4: Example of alternative solutions obtained with Algorithm **MinDegradation** for $\delta = 20$. The alternative reserve solution is represented in green, while the optimal solution x^* is depicted with planning units delimited by a thick black border. The number of white planning units with a thick black border corresponds to the dissimilarity measure between the optimal solution and the alternative solution.

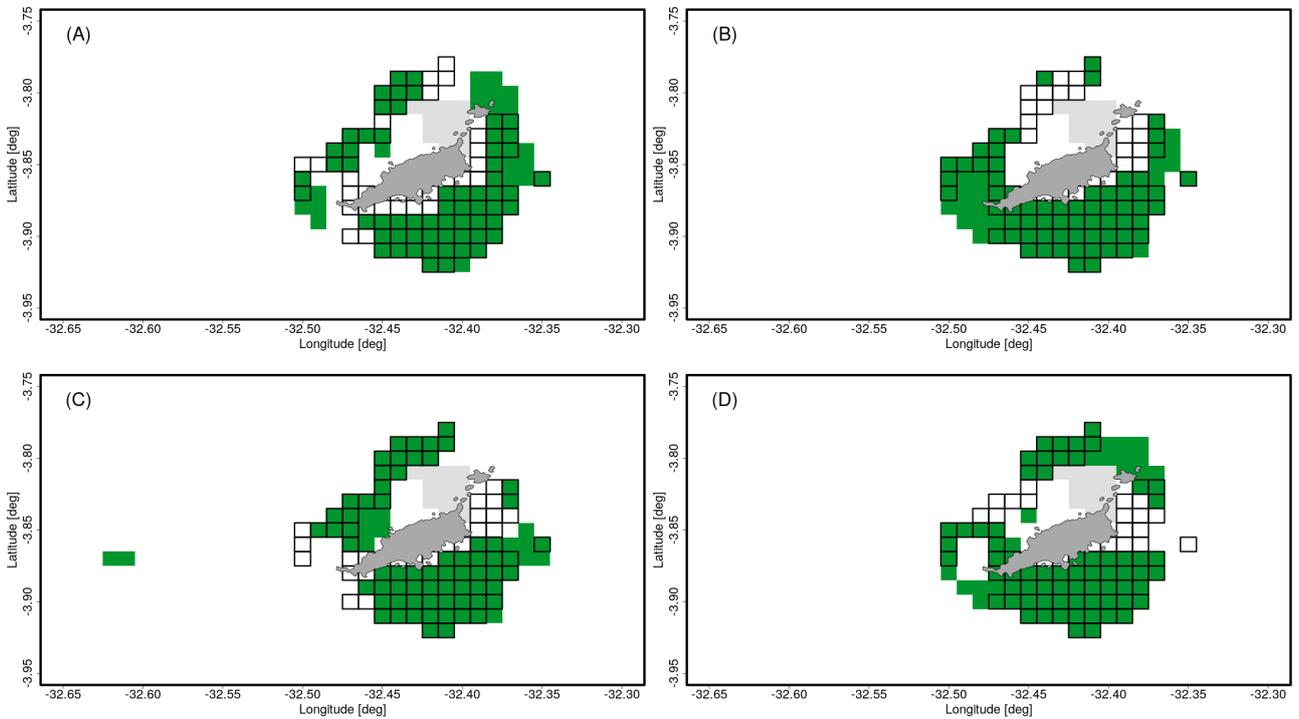


Figure 3.5: Example of alternative solutions maximising the minimum dissimilarity measure to the past solutions at a fixed extra cost. We show four successive alternatives with an extra cost budget $\gamma = 10\%$. The alternative reserve solution is represented in green, whereas the optimal solution x^* is depicted with planning units delimited by a thick black border.

3.3.4 Optimality gap versus dissimilarity measure

We compared the alternative solutions obtained with Algorithm [AddGapPortfolio](#), [MinDegradation](#) and [MaxDissimilarity](#) for various values of the parameters involved. To do so, we represented the mean optimality gap of the alternative solutions and the dissimilarity to the optimal solution. We repeatedly applied Algorithm [AddGapPortfolio](#), [MinDegradation](#) and [MaxDissimilarity](#) by respectively setting the γ_1 , δ and γ parameter to increasing values.

All curves in Figure 3.6 increase. As expected, the top curve is obtained with Algorithm [MaxDissimilarity](#) because it explicitly sought to maximise the dissimilarity to the past solutions. Since the dissimilarity measure was not considered at all in Algorithm [AddGapPortfolio](#) but only the targeted objective value interval, the corresponding curve is the lowest and is not strictly increasing. The curve obtained with Algorithm [MinDegradation](#) is in between the two others because it explicitly accounted for the dissimilarity measure but did not seek to maximise it.

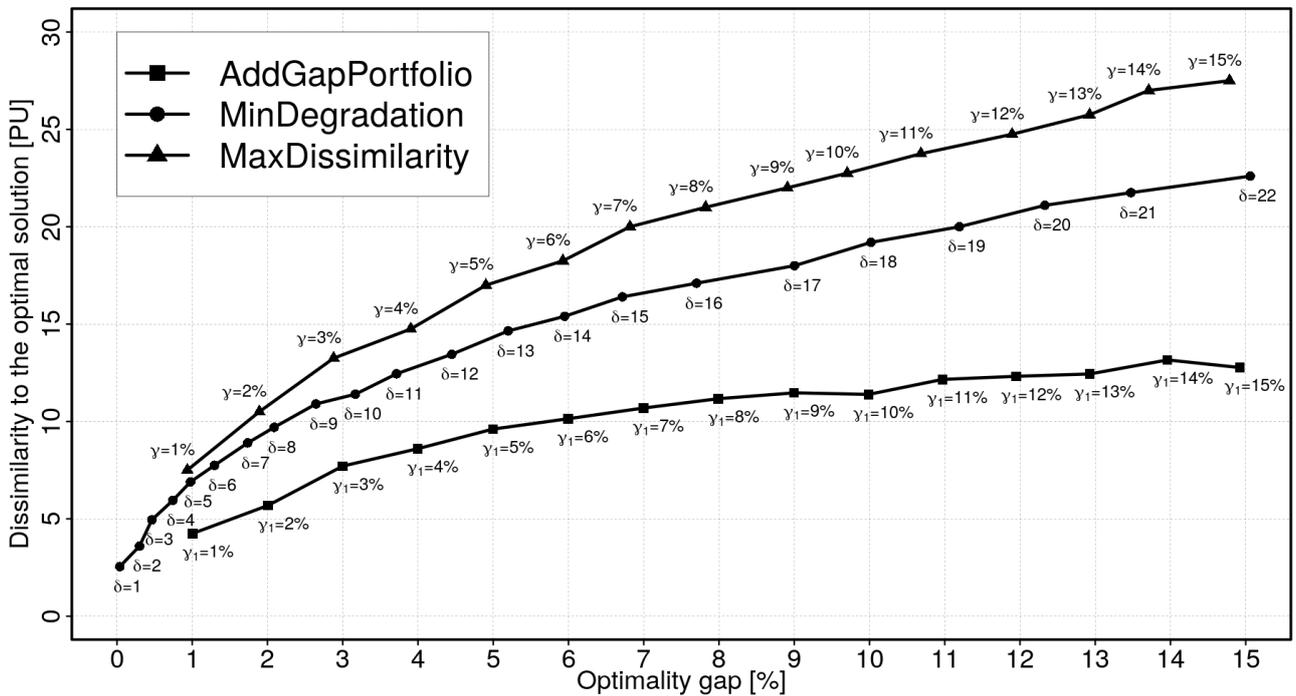


Figure 3.6: Dissimilarity to the optimal solution versus the optimality gap. Algorithm [AddGapPortfolio](#) for $\gamma_1 \in [1\%, 15\%]$ and $n = 100$, Algorithm [MinDegradation](#) for $\delta \in [1, 22]$ and $n = 20$, Algorithm [MaxDissimilarity](#) for $\gamma \in [1\%, 15\%]$ and $n = 4$. Optimality gaps and dissimilarity measures are averaged over the alternative solutions composing the presentation set obtained with the considered algorithm.

3.4 Discussion

We proposed two procedures to produce a diverse set of near-optimal solutions using exact solving methods. The presentation set was composed of alternative solutions that are not only different from the optimal solution, but also different from each other thanks to the iterative modification of the nominal optimisation problem. We observed that using the natural distance as the dissimilarity measure leads to alternative solutions that strictly includes the optimal one. Considering that it was not a valuable alternative but only a degraded solution, our dissimilarity measure allowed to discriminate such cases, which is new in conservation biology. The Algorithm [MinDegradation](#) provides the least cost alternative solutions that are sufficiently different from each other according to a given dissimilarity threshold. The Algorithm [MaxDissimilarity](#) provides the most different solutions from each other at a fixed degradation of the cost. These procedures implied the formulation of mixed integer linear programs solved using exact methods. Another important contribution is the comparative analysis of these two procedures among them and with existing methods.

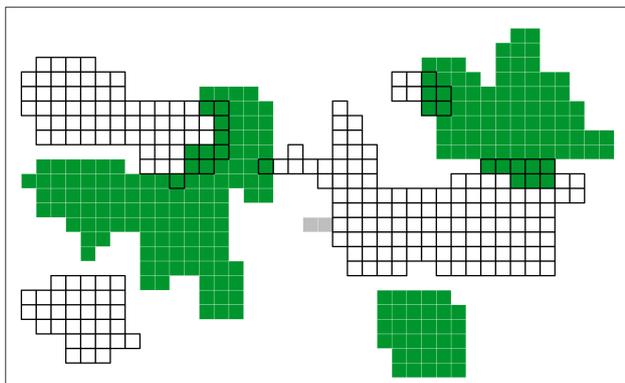
In summary, the strength of this work lies in the fact that only a few iterations are needed to generate a presentation set of truly different solutions. Moreover, the methods developed are highly customisable. For example, other dissimilarity measures could be used in our iterative procedures to assess the differences between solutions in the same line as in (Makowski et al. 2000). These differences only depends on the definition of a dissimilarity measure, and can be adapted according to the application case. Another advantage of this type of approach is to be able to explicitly quantify the quality of the alternative solutions generated. Since the search for alternative solutions is carried out by exact solving methods, we know the optimality gap which gives more control to the end user. Finally, the production of the presentation set is completely controlled by two parameters. The user can then choose exactly the trade-off between the diversity of the set of alternative solutions and the optimality gap. The sensitivity analysis conducted in Section 3.3.4 is an illustration of this trade-off for each approach. Regarding the weaknesses, the proposed approaches are mostly limited by the computation time required. It can be large for some instances and it increases with the number of alternatives requested. However, we did not focus on improving the computation time in this work. In the current state of the algorithms, we can provide orders of magnitude for the computation time with a personal computer (Intel Core i7-8850H CPU @ 2.60GHz) when 4 alternatives need to be computed. To do so, we performed tests with several instances randomly generated according to the process described in Appendix 1.5.2. An example of a presentation set is given in Figure 3.7. We observed the following computational times:

- about 2-3 minutes for 500 planning units and 3 conservation features with Algorithm [MaxDissimilarity](#)
- about 10-60 minutes for 1000 planning units and 5 conservation features with Algorithm [MaxDissimilarity](#)
- about 10-20 seconds for 500 planning units and 3 conservation features with Algorithm [MinDegradation](#)
- about 2-15 minutes for 1000 planning units and 5 conservation features with Algorithm [MinDegradation](#)

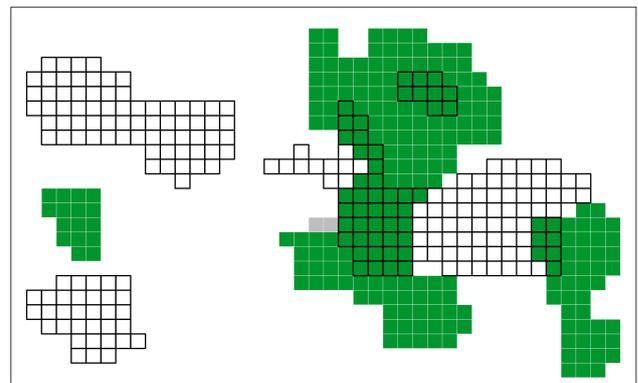
These computation times must be put into perspective. If we are not necessarily looking for a proof of optimality, they can be much lower. Our algorithms allow us to quickly provide interesting and feasible solutions if we decide to keep the current solution after a given maximum time. Finally, producing only 4 alternatives is a meaningful choice, because they are really

different alternatives that do not require additional statistical analyses.

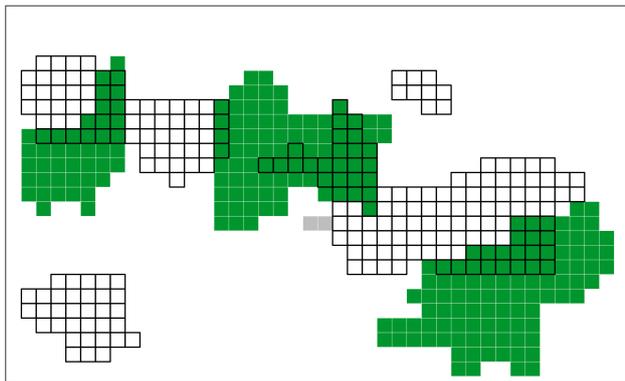
In conclusion, unlike what was commonly stated in the conservation literature (Possingham et al. 2000; Ardron et al. 2010), our work showed that exact optimisation methods used for the reserve site selection problem can also be advantageous to produce a range of alternative solutions. As a consequence, it is not true that metaheuristics are the only methods that are able to produce a presentation set. Besides, the inclusion of an explicit dissimilarity criterion directly within the optimisation model allowed to build a more controlled and transparent presentation set. By seeking significantly different solutions, we increased the chance to address objectives that are not necessarily modelled, such as socio-political or management objectives. The low number of alternatives needed with our methods may avoid unnecessary noise in the decision-making process. In other words, the proposed algorithms can potentially empower conservation practitioners by giving them more control over the alternatives produced and by removing the post-processing analysis usually needed. We hope that these methods can at least shed a new light in conservation discussions and eventually bring more success in conservation decisions in practice.



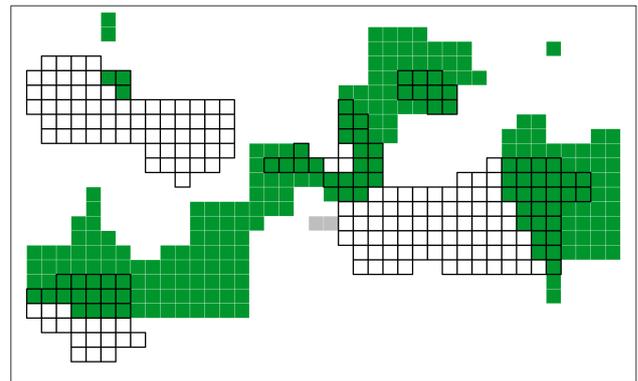
(a) Alternative reserve solution found at iteration 1 with Algorithm [MaxDissimilarity](#).



(b) Alternative reserve solution found at iteration 2 with Algorithm [MaxDissimilarity](#).



(c) Alternative reserve solution found at iteration 3 with Algorithm [MaxDissimilarity](#).



(d) Alternative reserve solution found at iteration 4 with Algorithm [MaxDissimilarity](#).

Figure 3.7: Presentation set computed with Algorithm [MaxDissimilarity](#). The considered scenario was made of 40×25 planning units and 5 conservation features. We chose an extra cost budget of $\gamma = 0.10$. Relative targets for every conservation features were set to 25%. Green planning units represents the alternative reserve solution. Planning units with a black border indicates the initial optimal solution.

Transition

In this chapter, we showed how to produce alternative reserve solutions that are dissimilar from each other using exact optimisation methods. In our models, we always assumed that the data involved was perfectly known and certain. However, it is never the case in real world problems. Indeed, data are always affected with uncertainty due to the *in-situ* collection, processing methods, *etc.* That is why we felt the need to address how to incorporate uncertainty within the reserve site selection problem.



Chapter 4

Risk-averse optimisation for reserve site selection with uncertain non-binary data

4.1 Introduction

Biodiversity and habitats are threatened worldwide (IPBES 2019). Building comprehensive networks of nature reserves is the frontline answer in the face of this situation (Ticco 1995; Tundi Agardy 1994; Le Saout et al. 2013) and it was shown to bring conservation benefits (Claudet et al. 2008; Stolton and Dudley 2010; Liu et al. 2017). At sea, for instance, current political objectives are to cover 30% of the marine spaces under jurisdiction by 2030 with marine protected areas (IUCN 2014; IUCN 2016; Commission 2020). Similar concerns also exist on land (Baillie and Zhang 2018; Dinerstein et al. 2019). Given these objectives, there is a strong demand to find the best compromises between the protection of biodiversity and the sustainability of human uses of these spaces. To analyse such problems, numerical optimisation methods are commonly implemented (Margules and Pressey 2000; Moilanen et al. 2009; Ando et al. 1998; Stewart and Possingham 2005). Such methods are often embedded within a software, *e.g.* Marxan or PrioritizR (Ball and Possingham 2000; Ball et al. 2009; Hanson et al. 2020). They are designed to systematically select reserve sites and are used as a decision support tools in real-world cases¹ (Flower et al. 2020; Fernandes et al. 2005).

Conservation spatial planning tools aim at finding the best reserve sites from the available knowledge in order to ensure the long term biodiversity persistence and eventually the provision of ecosystem services. But the uncertainty inherent to the available knowledge can result in a poor decision-making and lead to inefficient reserve solutions. This inefficiency can cause irreversible damages towards ecosystems, useless constraints on human uses, and more generally a waste of already scarce conservation resources. It thus appears crucial to provide decision makers methods allowing to quantify the tradeoff between the risk and the cost of a decision. In a global change context, reserve selection methods that account for uncertainties may lead to produce reserves that have more chances of achieving their conservation objectives. It seems even compulsory knowing that incomplete and imperfect knowledge is at the root of the conservation science (Soule 1985) That is why uncertainty within conservation planning has been identified as an important research gap (Margules and Pressey 2000; Foley et al. 2010; Reside et al. 2018). The type of uncertainty that can be involved in reserve selection problems is detailed in (Regan et al. 2009). In this work, we considered an epistemic uncertainty, whether it is variability (natural source such as climate change) or incertitude (model or measure imprecision). In practice, this uncertainty affects the non-binary conservation feature amounts of the reserve site selection optimisation problem.

The main framework used in reserve site selection methods to explicitly incorporate uncertainty are chance constraint formulations (Polasky et al. 2000; Haight et al. 2000; Araújo and Williams 2000; Williams and Araújo 2000; Sarkar et al. 2004; Cabeza et al. 2004; Tole 2006; Moilanen et al. 2006a; Moilanen et al. 2006b). A chance constraint formulation aims to ensure that the probability of satisfying given constraints is above a given threshold. These works are based on binary data since they consider presence/absence of the conservation features. The probabilistic model developed in (Haight et al. 2000), linearised in (Polasky et al. 2000) using a risk level, was a major methodological contribution. It allowed further developments in the same framework (Araújo and Williams 2000; Williams and Araújo 2000; Sarkar et al. 2004; Cabeza et al. 2004; Tole 2006; Moilanen et al. 2006a; Moilanen et al. 2006b). They are based on the *a priori* knowledge of the probability of presence (or persistence in (Araújo and Williams 2000; Williams and Araújo 2000)) of each conservation feature inside a planning unit. These probabilities are supposed to be known or derived from environmental data (Araújo and Williams

¹More case study examples can be found at <https://marxansolutions.org/community/> and https://prioritizr.net/articles/publication_record.html.

2000; Tole 2006). They are considered uncertain due to model imprecisions and an information gap analysis is applied in (Moilanen et al. 2006a; Moilanen et al. 2006b). The resulting optimisation problems were mostly solved using heuristics, although a comparison with exact solving methods was also performed in (Sarkar et al. 2004). The key behind these stochastic frameworks is that the random variable associated with a conservation feature amount inside a planning unit has a binary probability distribution (presence or absence, persistence or disappearance). This way, the overall probability of presence within the reserve solution of a conservation feature can be expressed in closed-form. This expression then allow to express a deterministic integer linear program (ILP). More theoretical details can be found in (Beraldi and Ruszczyński 2002). This approach is not possible when non-binary data are available. In our case, we considered measures of abundance: the conservation feature amount within a planning unit is non-binary, since it can take any positive value. Obtaining a similar closed-form expression as in (Polasky et al. 2000) was not possible with our non-binary data. Hence, the need to develop another model to take into account the uncertainty on non-binary data, even if it means solving an approximation. Another similar approach, compatible with non-binary data, is deployed in MarProb², *i.e.* Marxan with Probability (Game et al. 2008; Carvalho et al. 2011; Tulloch et al. 2013). The probability of a planning unit being destroyed in the future is included in the problem description. The probability that a conservation feature fails to meet its target is computed with a statistics table under the hypothesis of a Gaussian distribution. In the same line of Marxan approach, this probability was included in the objective function with a representation shortfall penalty and then solved using a metaheuristic algorithm (see Appendix B-2.1 in (Serra-Sogas et al. 2020)). In this work, we want to develop ILPs that can be solved using exact methods. Finally, a robustness index is proposed in (Beech et al. 2008) using a Monte-Carlo strategy. A predefined probability distribution allows to generate samples of each conservation feature amounts. The robustness index is simply the proportion of times a planning unit is selected among the reserve solutions computed for each instance derived from the samples.

Uncertain optimisation reveals two opposing philosophies regarding their attitude towards risk: risk-neutral and risk-averse approaches. A risk-neutral attitude is often used when the outcome of a decision is repeated many times. This way, the decision is designed to optimize an expected quantity over time. However, when the outcome of a decision is the result of one realisation, we may not tolerate the risk involved. In conservation biology, a zoning decision is taken once and the risk associated with a bad decision can lead to habitat destruction and species extinctions. That is why, our risk preference went towards the risk-averse approaches. In this work, we show conservation practitioners how to incorporate uncertainty affecting non-binary conservation feature amounts in the reserve site selection problem using risk-averse optimisation frameworks. We propose two optimisation models explicitly accounting for the risk a user is willing to take. In the first approach, we considered a non-probabilistic parametric uncertainty: the conservation features amounts can take a range of possible values instead of a particular one. To address this problem, we provide a robust optimisation formulation. In other words, we aimed at finding the best feasible solution whatever the uncertainty realisation within each parameter uncertain set. But, in order to avoid the too preservative worst-case solution, we introduced a budgeted uncertainty set (Bertsimas and Sim 2004). In this framework, we aimed at finding the best feasible solution considering a user-defined number of parameters that are tolerated to deviate from their nominal value. The resulting optimisation is a deterministic mixed integer linear program (MILP). Finally, a sensitivity analysis towards the risk-level parameter provides a simple way to represent the reserve solution robustness from the nominal solution to the worst-case. In the second approach, we considered a risk-averse formulation

²<https://marxansolutions.org/software/>

with chance constraints. Since we do not have a binary probability distribution, the analytic computation of the overall probable amount of each conservation feature in the reserve is not available. That is why, we addressed this optimisation problem using a sampling approximation approach (Luedtke and Ahmed 2008). The idea is to discretise the probability space using a set of probable samples. This way, the probability can be estimated by a proportion among the considered samples. To design the reserve solution, we generated 100 probable samples using geostatistics. We generated samples conditionally to what was observed at the measure points. Instead of developing a robustness index as in (Beech et al. 2008), we accounted for the samples explicitly in the formulation of the ILP. We computed the reserve solution of the chance constraint formulation for several risk-level values. We compared the risk-level with a more accurate estimated probability over 1000 samples.

4.2 Methods

4.2.1 Robust framework

We now assume that the amount a_{ij} of conservation feature i in the planning unit j takes its values within a continuous interval \mathcal{U}_{ij} . We characterise the interval \mathcal{U}_{ij} using a nominal value \bar{a}_{ij} and a deviation parameter $\sigma_{ij} \geq 0$. As $a_{ij} \geq 0$, we only consider adverse deviation. Formally, we have $\mathcal{U}_{ij} = [\bar{a}_{ij} - \sigma_{ij}, \bar{a}_{ij}]$. The robust optimisation problem is:

$$\begin{cases} \min_x & \sum_{j \in J} c_j x_j \\ \text{s.t.} & \sum_{j \in J} a_{ij} x_j \geq t_i \quad \forall a_{ij} \in \mathcal{U}_{ij}, \forall i \in I \\ & x_j \in \{0, 1\} \quad \forall j \in J \end{cases} \quad (4.1)$$

The problem described in (4.1) is equivalent to solve P_0 when all a_{ij} parameters take their worst values $\bar{a}_{ij} - \sigma_{ij}$. Find a solution that is immune to the worst case scenario is often too expensive. In addition, it does not provide any information about the behaviour of the reserve solution toward an inferior risk nor regarding the cost of robustness. That is why we propose a simplification of the problem by considering a budgeted uncertainty approach.

For normalisation purposes, we define the new parameter δ_{ij} such as:

$$\delta_{ij} = \frac{\bar{a}_{ij} - a_{ij}}{\sigma_{ij}} \quad (4.2)$$

This way, the deviation is inactive when $\delta_{ij} = 0$ and maximal (worth $-\sigma_{ij}$) when $\delta_{ij} = 1$. Regarding the uncertainty set, $a_{ij} \in \mathcal{U}_{ij}$ corresponds to $\delta_{ij} \in [0, 1]$.

In order to be less conservative, we only authorise $\Gamma = \sum_{i \in I} \Gamma_i$ parameters a_{ij} to deviate simultaneously. Let $\delta_i = [\delta_{i1}, \dots, \delta_{iN}]$ be a vector of $[0, 1]^N$. We define Δ^{Γ_i} , the set of δ_i within our interest:

$$\forall i \in I, \quad \Delta^{\Gamma_i} = \{\delta_i \in [0, 1]^N / \sum_{j \in J} \delta_{ij} \leq \Gamma_i\} \quad (4.3)$$

Finally, the simplified robust optimisation problem of (4.1) considering budgeted uncertainty sets Δ^{Γ_i} is:

$$\begin{cases} \min_x & \sum_{j \in J} c_j x_j \\ \text{s.t.} & \sum_{j \in J} \bar{a}_{ij} x_j - \sum_{j \in J} \delta_{ij} \sigma_{ij} x_j \geq t_i, \quad \forall \delta_i \in \Delta^{\Gamma_i}, \forall i \in I \\ & x_j \in \{0, 1\}, \quad \forall j \in J \end{cases} \quad (4.4)$$

The problem (4.4) still has an infinite set of constraints. Yet, such excess is dispensable if we impose the solution to be robust to the worst uncertainty:

$$\begin{aligned} & \sum_{j \in J} \bar{a}_{ij} x_j - \sum_{j \in J} \delta_{ij} \sigma_{ij} x_j \geq t_i \quad \forall i \in I, \forall \delta_i \in \Delta^{\Gamma_i} \\ \iff & \sum_{j \in J} \bar{a}_{ij} x_j - t_i \geq \max_{\delta_i \in \Delta^{\Gamma_i}} \left\{ \sum_{j \in J} \delta_{ij} \sigma_{ij} x_j \right\} \quad \forall i \in I \end{aligned} \quad (4.5)$$

By doing so, the right-hand side can be expressed as a linear program which represents the adversary decision regarding the most disadvantageous uncertainty:

$$\forall i \in I, \quad \sum_{j \in J} \bar{a}_{ij} x_j - t_i \geq \begin{cases} \max_{\delta_i} & \sum_{j \in J} \delta_{ij} \sigma_{ij} x_j \\ \text{s.t.} & \sum_{j \in J} \delta_{ij} \leq \Gamma_i & [v_i] \\ & \delta_{ij} \leq 1 \quad \forall j \in J & [u_{ij}] \\ & \delta_{ij} \geq 0 \quad \forall j \in J \end{cases} \quad (4.6)$$

By switching to a dual formulation of the standard linear program expressed in (4.6), the constraint becomes:

$$\forall i \in I, \quad \sum_{j \in J} \bar{a}_{ij} x_j - t_i \geq \begin{cases} \min_{u_i, v_i} & \Gamma_i v_i + \sum_{j \in J} u_{ij} \\ \text{s.t.} & v_i + u_{ij} \geq x_j \sigma_{ij} \quad \forall j \in J \\ & u_{ij} \in \mathbb{R}_+ \quad \forall j \in J \\ & v_i \in \mathbb{R}_+ \end{cases} \quad (4.7)$$

By definition, any feasible u_i and v_i gives a greater objective value than the minimum. Consequently, considering any feasible u_i and v_i is even more demanding relatively to constraint (4.7). Therefore, problem (4.4) can be equivalently formulated as the following MILP:

$$P_{rob} : \begin{cases} \min_{x, u, v} & \sum_{j \in J} c_j x_j \\ \text{s.t.} & \sum_{j \in J} \bar{a}_{ij} x_j - (\Gamma_i v_i + \sum_{j \in J} u_{ij}) \geq t_i \quad \forall i \in I \\ & v_i + u_{ij} \geq x_j \sigma_{ij} \quad \forall i \in I, \forall j \in J \\ & x_j \in \{0, 1\} \quad \forall j \in J \\ & u_{ij} \in \mathbb{R}_+ \quad \forall i \in I, \forall j \in J \\ & v_i \in \mathbb{R}_+ \quad \forall i \in I \end{cases} \quad (4.8)$$

4.2.2 Chance constraint framework

Let (Ω, \mathbb{P}) be a probability space and $\omega \in \Omega$. The probability that all targets t_i are reached by a given solution $x \in \{0, 1\}^N$ is $\mathbb{P}(\sum_{j \in J} a_{ij}(\omega) x_j \geq t_i, \forall i \in I)$. Let α be a risk-level parameter. We propose an optimisation problem seeking for the least-cost reserve solution that has a probability of α to cover all targets. In other words, given any realisation $\omega \in \Omega$ of the uncertainty, the reserve solution has a risk of $(1 - \alpha)$ to fail to cover all the targets. The optimisation problem including this joint chance constraint is:

$$\begin{cases} \min_x & \sum_{j \in J} c_j x_j \\ \text{s.t.} & \mathbb{P}(\sum_{j \in J} a_{ij}(\omega) x_j \geq t_i, \forall i \in I) \geq \alpha \\ & x_j \in \{0, 1\} \quad \forall j \in J \end{cases} \quad (4.9)$$

Problem (4.9) is intractable if the probability of meeting all targets cannot be expressed in closed-form. To solve this difficult problem, we approximate the probability by considering a sampling approximation of the probability space. Let p^s be the occurring probability of sample $s \in S$ obtained in the sample of size $|S|$. We thus introduce z^s the decision variable worth 1 if all the sample constraints are satisfied simultaneously for the sample $s \in S$ and 0 otherwise. We finally get the following sampling approximation of the optimisation problem (4.9):

$$P_{sa} : \begin{cases} \min_{x, z} & \sum_{j \in J} c_j x_j \\ \text{s.t.} & \sum_{j \in J} a_{ij}^s x_j \geq z^s t_i \quad \forall i \in I, \forall s \in S \\ & \sum_{s \in S} z^s p^s \geq \alpha \\ & x_j \in \{0, 1\} \quad \forall j \in J \\ & z^s \in \{0, 1\} \quad \forall s \in S \end{cases} \quad (4.10)$$

4.2.3 Conceptual diagrams

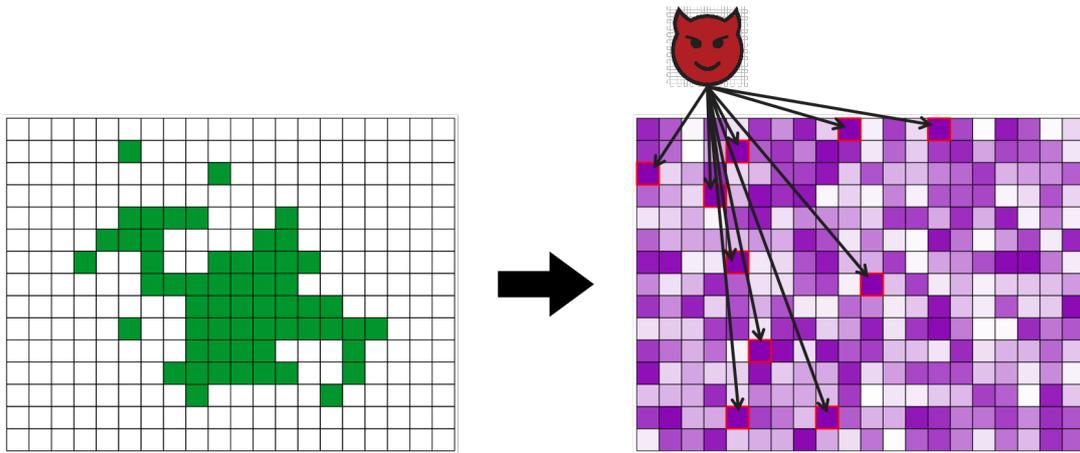


Figure 4.1: Conceptual diagram of the robust framework solved using a budgeted uncertainty set approach. On the left panel, a reserve candidate is represented with green planning units. On the right panel, the spatial distribution of a conservation feature amounts is represented with a purple transparency gradient. The adverse part chooses a predefined number (uncertainty budget) of planning units that will take their worst value in their uncertainty set. This choice is made to penalise the most the objective value of the solution considered. The robust reserve solution seeks the least-cost solution that meets the conservation feature targets.

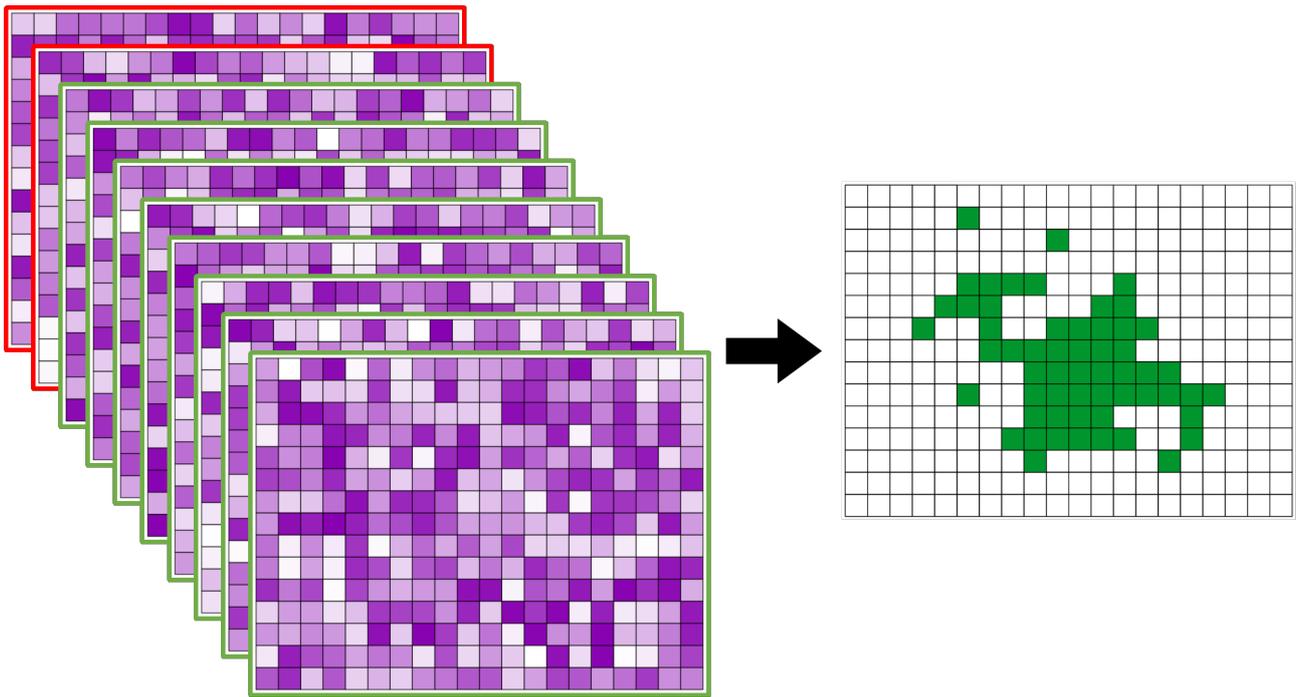


Figure 4.2: Conceptual diagram of the chance constraint framework solved using a sampling approximation approach. On the right, a reserve candidate is represented with green planning units. On the left, 10 samples for the spatial distribution of a conservation feature amounts is represented with a purple transparency gradient. The probability space is discretised with samples. The reserve solution must satisfy a predefined proportion of the considered samples. In the diagram, the reserve solution meets the target of the conservation feature for 80% of the samples.

4.3 Results

In the following, we only considered the uncertainty of the fish biomass conservation feature, since the two others can be considered certain. Regarding the robust numerical application in Section 4.3.1, we chose the same value for every deviation parameters, *i.e.* $\sigma_{ij} = \sigma$ for all $i \in I$ and $j \in J$. Several values of $\Gamma_1 = \Gamma$ are tested. The maximum value for Γ is 157 since we have 157 non-zero planning units for the fish biomass. Regarding the chance constraint framework, in Section 4.3.2, the design of the reserve solution was done over 100 generated samples. This number was chosen to have small computation times. The probability of meeting the sample constraint was assessed over 1000 samples (different from the 100 considered in P_{sa}). To produce a user-defined number of samples for the fish biomass amount distribution, we used conditional simulations (Chiles and Delfiner 2012) based on work (Bez and Braham 2014; Bez 2021) as done in (Salvetat et al. 2022). All samples were considered equiprobable.

4.3.1 Robust framework

We computed robust reserve solutions for several values of the uncertainty budget Γ and the tolerated deviation parameter σ . Figure 4.3 shows 4 robust solutions computed for $\sigma = 40\%$. Panel A shows the reserve solution of the nominal-value problem P_0^f because $\Gamma = 0$. In this scenario, the fish biomass within each planning unit takes its nominal value since they are not tolerated to deviate. As expected, the size and objective value of the robust reserve solution are increasing with Γ relatively to the solution of the nominal-value problem P_0^f . In particular, $\Gamma = 20$ and $\Gamma = 40$ respectively corresponds to an increase of 17.0% and 22.7% in size. It corresponds to an increase of 17.3% and 31.7% in objective value. Histograms in Figure 4.3 shows the coverage of each conservation feature in the nominal scenario by the robust solution. As expected, the coverage increases with Γ . In particular, $\Gamma = 20$ and $\Gamma = 40$ respectively corresponds to a coverage of 18.2% and 32.6% of the fish biomass in the nominal scenario. The coverage of the fish biomass and continental shelf conservation feature are increasing jointly because their spatial distribution is overlapping significantly. Panel D shows the worst-case solution. It is the least-cost solution that would still meet the targets even if the fish biomass takes its worst value in each planning unit. The worst-case solution involves an extra cost of 79% and an increase in size of 36.4% relatively to the solution of the nominal-value problem P_0^f . The worst-case solution corresponds to an increase of 66.9% for the coverage of fish biomass in the nominal scenario.

Figure 4.4 shows the sensitivity analysis towards the parameters involved in the robust framework. The objective value of the robust solutions are increasing with the uncertainty budget. Robustness logically involves an extra cost. The greater the deviation coefficient, the higher the objective value of the robust solutions. We also observe that a level is reached (except for $\sigma = 50\%$) when the uncertainty budget is greater than a given threshold. It means that the worst case solutions is reached even for value of $\Gamma < 157$. The lower the deviation coefficient, the further this level is reached.

Figure 4.5 shows the selection frequency of each planning unit within the set of robust solutions, where robust solutions are computed for $\sigma = 40\%$ and several values of Γ . The selection frequency allows mapping the planning unit the most and least involved in the robust solutions. Planning units selected in the solution of the nominal problem are almost selected in every robust solutions, *i.e.* 17 occurrences. However, a few planning units of the nominal solution are only selected a few times at the west of the island. These planning units thus tend to disappear from the robust solutions. At the opposite, many planning units east of the main island are

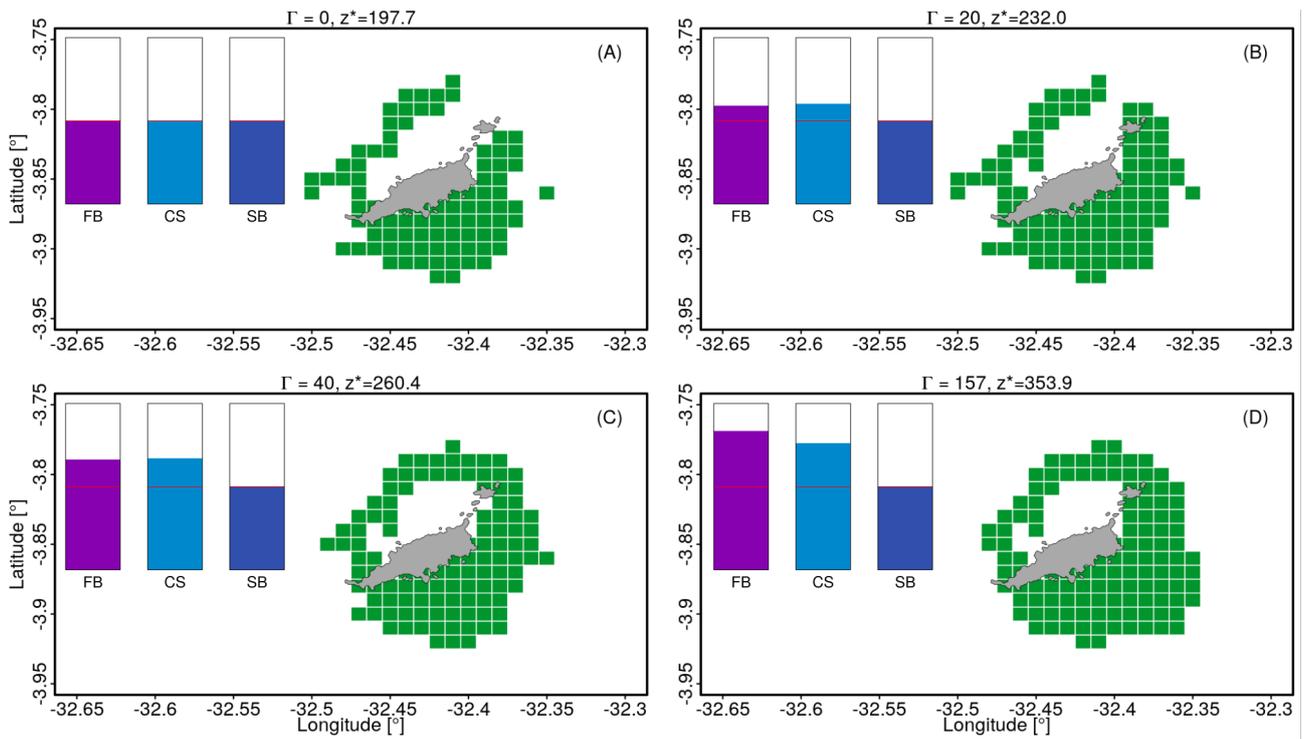


Figure 4.3: Robust reserve solutions for $\Gamma \in \{0, 20, 40, 157\}$ and $\sigma = 40\%$. Green planning units represent the reserve solution when Γ parameters are tolerated to deviate from their nominal value. Histograms at the top left of each panel shows the coverage of each conservation feature (FB=fish biomass, CS=continental shelf, SB=shelf break) by the robust solution if all parameters take their nominal value. z^* is the objective value of the considered solution.

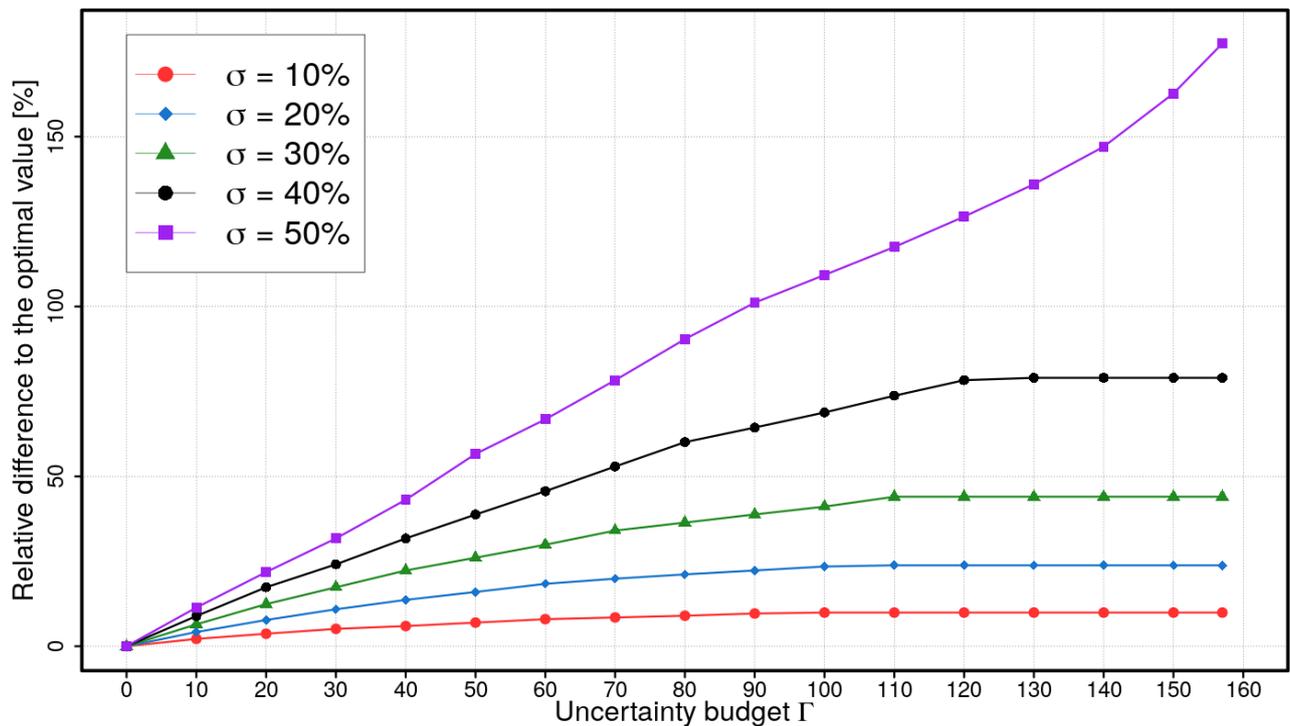


Figure 4.4: Sensitivity analysis of the deviation parameter σ . Relative difference to the optimal value of the nominal scenario computed for $\Gamma \in \{0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 157\}$ and $\sigma \in \{10\%, 20\%, 30\%, 40\%, 50\%\}$.

selected more than 10 times over the 17 simulations, although they are not included in the certain solution. These planning units are more robust to the uncertainty.

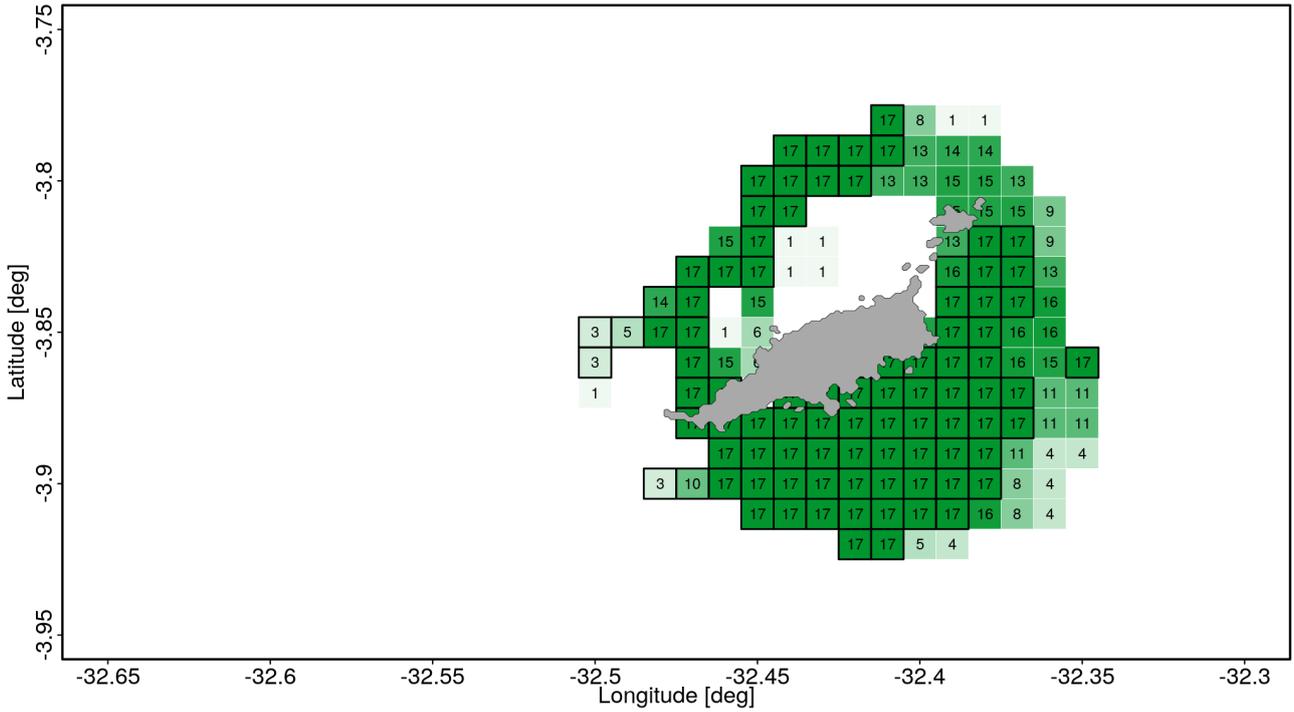


Figure 4.5: Selection frequency among robust solutions. Robust solutions are computed for $\sigma = 40\%$ and $\Gamma \in \{0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 157\}$. Selected planning units of robust solutions are represented with a green transparency gradient according to selection frequency expressed in number of occurrence (black number inside planning unit). Planning units with a black border represent the reserve solution of the nominal-value problem P_0^f .

4.3.2 Chance constraint framework

We show in Figure 4.6 the reserve solutions of P_{sa} for different values of the risk-level parameter α . To achieve the sampling approximation of the chance constraint (proportion α among the 100 samples used for design), the reserve solution is generally selecting the same planning units for different values of the risk-level α . Planning units that are selected in the nominal solution and avoided in the chance constraint formulation also remain the same planning units for different values of the risk-level α . Histograms confirm the coverage is beyond the target for the considered proportion α among the 100 design samples. The solutions approximately involve an extra 40% cost with respect to the objective value of the optimal solution of the nominal-value problem P_0^f .

We generated 1000 new samples to see how the computed solutions perform with respect to these samples. This way, we have a more accurate estimation of the performance with respect to the uncertainty. We call $\hat{\epsilon}$ the proportion among the 1000 samples where the solution of P_{sa} is feasible. Table 4.1 shows both the risk-level α and this estimated probability $\hat{\epsilon}$. We can see that there is a significant difference between the design probability α and the estimation of the true probability $\hat{\epsilon}$.

4.3.3 Comparison between the two approaches

In this section, the solutions computed using the robust and chance constraint framework are compared. These solutions' probability of meeting all the coverage targets is assessed over 1000 samples. The samples considered here did not call geostatistics and conditional simulations

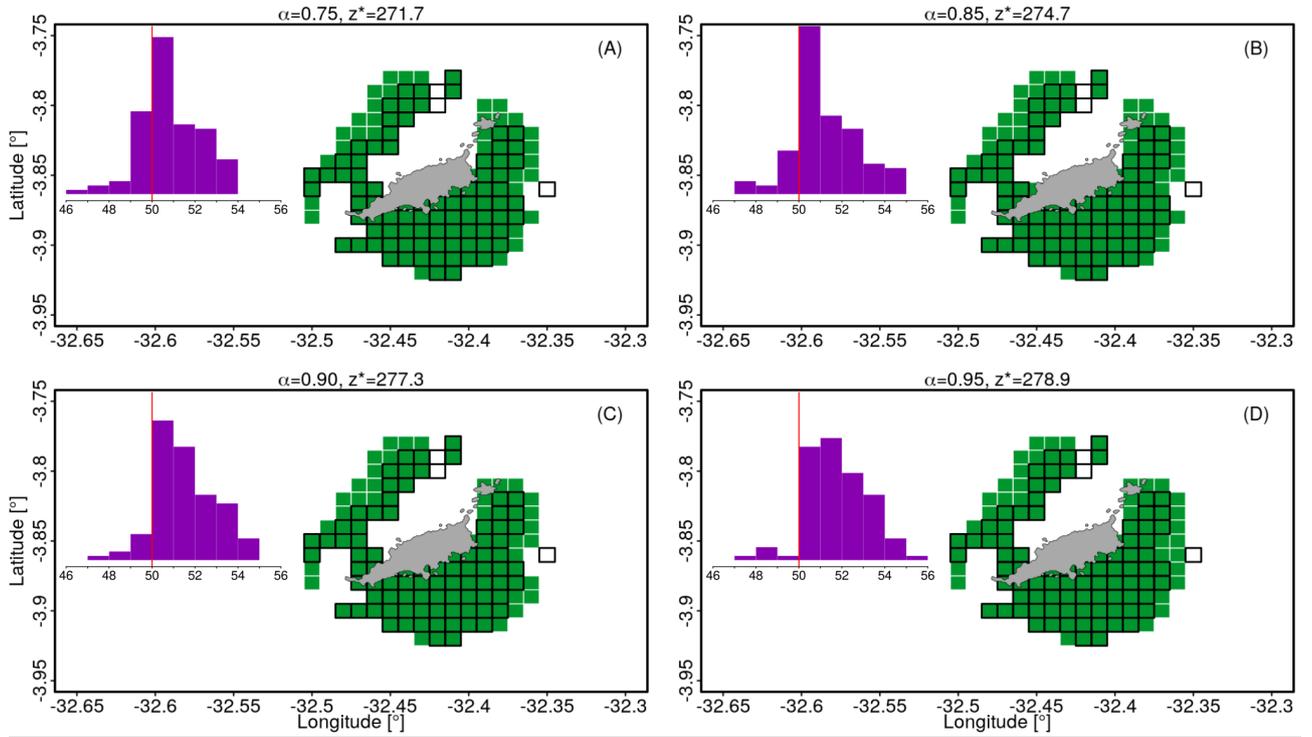


Figure 4.6: Reserve solutions satisfying a sampling approximation of a chance constraint for a risk-level $\alpha \in \{0.75, 0.85, 0.90, 0.95\}$. Green planning units represent the reserve solution. Planning units with a black border represent the reserve solution of the nominal-value problem P_0^f . Histograms at the top left of each panel show the coverage distribution for the fish biomass conservation feature for the 100 samples considered. The red line gives the fish biomass absolute target. z^* is the objective value of the considered solution.

Design risk-level α	Estimated probability $\hat{\epsilon}$	Relative difference to the nominal objective value
0.75	0.634	37%
0.80	0.656	38%
0.85	0.720	39%
0.90	0.782	40%
0.95	0.831	41%
1.00	0.993	50%

Table 4.1: Relationship between the design probability α computed using 100 samples and the estimated probability $\hat{\epsilon}$ using 1000 samples.

in order to perform a relevant comparison between the two approaches. To do so, we independently drew the amounts a_{ij} of the fish biomass conservation feature within each planning unit using a Gaussian probability distribution. The mean value of the Gaussian probability distribution is given by the nominal value \bar{a}_{ij} . The standard deviation is set to $\hat{\sigma}\bar{a}_{ij}$.

Importantly, this standard deviation is consistent with the deviation parameter σ of model P_{rob} . Let α be the risk-level parameter of model P_{sa} . The chance constraint solution is computed using 100 samples. Let Γ be the uncertainty budget parameter of model P_{rob} . Let σ be the deviation parameter describing the uncertainty set of model P_{rob} . Let $\hat{\epsilon}$ be the estimation of the true probability of satisfying the coverage constraints estimated on 1000 different samples. Let γ be the relative difference to objective value of the nominal solution and the risk-averse solution.

The nominal solution, obtained with $\Gamma = 0$, is failing to meet the targets for almost 50% of the

1000 samples for any values of $\hat{\sigma}$. Table 4.2 shows that solutions of P_{sa} involved a small increase of the objective value: around 1% for $\hat{\sigma} = 10\%$ and around 5% for $\hat{\sigma} = 50\%$. In this numerical experiment, the values of α and $\hat{\epsilon}$ are very close, even with only 100 samples used for design. At the opposite, Table 4.3 shows that robust solutions are meeting the targets for almost each of the 1000 samples. Indeed, for most values of (Γ, σ) , the proportion of feasible solutions $\hat{\epsilon}$ is above 98%. For $\Gamma \geq 20$, we have $\hat{\epsilon} = 1$. However, these reserve solutions correspond to important increase of the objective value: around 6% in average for $\hat{\sigma} = 10\%$ and around 70% for $\hat{\sigma} = 50\%$. Interestingly, for $\Gamma = 10$, the objective value increase is comparable to the values obtained with the chance constraint solutions for $\alpha \geq 0.95$.

$\hat{\sigma}$	α	$\hat{\epsilon}$	γ
10%	0.75	0.701	0.5%
10%	0.80	0.812	1.0%
10%	0.85	0.810	1.1%
10%	0.90	0.914	1.2%
10%	0.95	0.964	1.8%
10%	1.00	0.988	2.2%
20%	0.75	0.815	1.7%
20%	0.80	0.812	1.7%
20%	0.85	0.851	2.2%
20%	0.90	0.898	2.7%
20%	0.95	0.957	3.2%
20%	1.00	0.996	7.4%
30%	0.75	0.733	2.0%
30%	0.80	0.817	2.7%
30%	0.85	0.844	3.2%
30%	0.90	0.928	4.2%
30%	0.95	0.955	5.1%
30%	1.00	0.988	7.5%
40%	0.75	0.790	3.2%
40%	0.80	0.787	3.9%
40%	0.85	0.820	4.4%
40%	0.90	0.906	5.6%
40%	0.95	0.943	6.7%
40%	1.00	0.989	9.9%
50%	0.75	0.718	3.5%
50%	0.80	0.832	4.2%
50%	0.85	0.870	5.1%
50%	0.90	0.858	5.8%
50%	0.95	0.888	7.1%
50%	1.00	0.975	12.2%

Table 4.2: Chance constraint solutions results for several values of the design probability α and standard deviation σ used to generate the samples. The probability of meeting the coverage constraints $\hat{\epsilon}$ was computed using 1000 samples. Relative targets were all set to 50%. The β multiplier was set to 1.

$\hat{\sigma} = \sigma$	Γ	$\hat{\epsilon}$	γ
10%	0	0.566	0.0%
10%	10	0.988	2.2%
10%	20	1.000	3.7%
10%	40	1.000	5.9%
10%	80	1.000	9.0%
10%	157	1.000	9.9%
20%	0	0.545	0.0%
20%	10	0.985	4.2%
20%	20	1.000	7.7%
20%	40	1.000	13.6%
20%	80	1.000	21.1%
20%	157	1.000	23.8%
30%	0	0.527	0.0%
30%	10	0.980	6.4%
30%	20	1.000	12.4%
30%	40	1.000	22.3%
30%	80	1.000	36.4%
30%	157	1.000	44.0%
40%	0	0.506	0.0%
40%	10	0.981	8.9%
40%	20	1.000	17.3%
40%	40	1.000	31.7%
40%	80	1.000	60.1%
40%	157	1.000	79.0%
50%	0	0.535	0.0%
50%	10	0.980	11.4%
50%	20	1.000	21.8%
50%	40	1.000	43.1%
50%	80	1.000	90.4%
50%	157	1.000	177.4%

Table 4.3: Robust solutions results for several values of the uncertainty budget Γ , the deviation parameter $\hat{\sigma}$ used in model P_{rob} and the standard deviation σ used to generate the samples. The probability of meeting the coverage constraints $\hat{\epsilon}$ was computed using 1000 samples. Relative targets were all set to 50%. The β multiplier was set to 1.

4.4 Discussion

The main problem addressed in this work was to deal with a parametric uncertainty that affects the non-binary conservation features amounts. We considered two risk-averse optimisation frameworks to incorporate this kind of uncertainty in the reserve site selection problem. It led us to express two MILPs. First, we proposed a robust optimisation formulation using a budgeted uncertainty set approach. In this formulation, we compute the least-cost reserve solution that is immune to the deviation of conservation feature amounts within a predefined number of planning units. Then, we provided a chance constraint formulation solved using a sampling approximation. In this formulation, we compute the least-cost reserve solution that meets the conservation feature targets for a given proportion of previously generated samples.

Our results showed what it costs to produce risk-averse reserve solutions that are robust to uncertainty and how robust they are. While the robust framework dealt with overestimated conservation feature amounts, the chance constraint framework addressed a more probabilistic uncertainty contained within samples. The models proposed in this work are new with respect to the classical models presented in (Polasky et al. 2000; Haight et al. 2000) which are only compatible with binary data. Our approaches allowed to account for an uncertainty that affect non-binary conservation feature amounts which is desirable in many cases. In general, the robust reserve solution appeared expensive relatively to the nominal solution. For example, considering half of the parameters overestimated, the robust reserve would cost more than 50% relatively to the nominal solution when the deviation parameter is set to 30%. When we evaluated the robust solutions over the 1000 generated samples, although the cost involved are important, the coverage targets are always met for values of $\Gamma > 10$. The solutions of the robust framework are a sure way to be immune to uncertainty, even for small uncertainty budgets, but can be very expensive. At the opposite, the chance constraint solutions provide cheaper solutions, only a few percent above the nominal solution, that are more or less immune to the risk depending on the user choice. It seems to allow a finer tuning of the risk a decision-maker is willing to take. We also observed that robust solutions for small values of the uncertainty budget achieved the same results. However, the relationship between the uncertainty budget and the probability of meeting the targets is not as straightforward as in the chance constraint framework. It may be more difficult to use this framework to design a reserve solution that is immune to an uncertainty contained in samples.

Our results also showed that comparing the solution of the nominal problem with our risk-averse solutions allow identifying to what extent given planing units contribute to the solution robustness. Planning units outside the nominal solution and often selected in the risk-averse solutions are likely to increase the robustness of the solution toward the uncertainty considered. At the opposite, the planning units of the nominal solution that are not selected in the risk-averse solutions when we increase the uncertainty are likely to be lost when the uncertainty is realised. We observe that these planning units are not the same in the robust and chance constraint frameworks. This kind of observations were at the basis of the robustness index developed in (Beech et al. 2008). In this work, we went further, since we *a priori* accounted for the samples within the optimisation model in the chance constraint framework.

The robust framework provides a way to incorporate uncertainty without any probability distribution. The only price to pay is computational because it involves a bigger MILP with respect to the nominal problem. The chance constraint framework is a more probabilistic approach. Our work showed how we could efficiently solve an approximation of this difficult problem using generated samples. The limitations of the chance constraint framework can come

from the number of samples needed. Since the size of the MILP is related to the number of samples, we cannot increase the number of samples without causing computational difficulties. Our experiment using geostatistics for generating samples illustrated another limitation: the risk-level parameter α of the chance constraint framework did not strictly represent the actual risk, estimated with $\hat{\epsilon}$. Yet, the interest of decision-makers is to ensure a reserve solution to be protected against an actual predefined risk-level. Since there was an apparent correlation between these two quantities, we could have empirically set the risk-level α to reach a predefined $\hat{\epsilon}$. However, the relationship between the number of samples and the actual probability of meeting the targets should be further investigated. In this work, we only considered the uncertainty toward a unique conservation feature, which limits the assessment of the coupling effect. We did this choice for clarity purposes, but this would deserve further explorations to draw more general conclusions. Another limitation observed is that our risk-averse solutions are quite similar despite the increasing of the risk-level parameters. An explanation of this comes from the case study considered: the conservation features are mainly distributed around the main island, so are the solutions.

Our risk-averse methods provide conservation practitioners efficient ways to derive a one-to-one relation between the robustness of a reserve solution and the associated cost. These frameworks also allow the identification of areas that are likely to be lost when the uncertainty is realised, which is precious information in a decision-making context. The choice of the framework to privilege depends on the uncertainty modelling: probability distribution or intervals of possible values. The trade-off between the risk-level and the robustness cost is to be made by the decision-maker in both frameworks.

Transition

In this chapter, we showed two approaches that incorporate uncertainty affecting non-binary data using exact optimisation methods in the reserve site selection problem. The reserve solutions that were computed sometimes showed undesirable spatial properties such as an isolated planning unit, holes within the reserve, *etc.* Indeed, except the empirical setting of the compactness parameter β , we do not have any constraints specifying spatial attributes of our reserve solution. Yet, the spatial attribute of a reserve was originally at the core of the discipline with the SLOSS debate and is often an argument to discard a reserve solution. That is why, in the next chapter, we focused on explicitly addressing the problem of producing reserve that are compacts, connected and without holes.



Chapter 5

Explicit design of compact, connected and gap-free reserves

5.1 Introduction

Biodiversity and habitats are threatened worldwide (IPBES 2019). Building comprehensive networks of nature reserves is the frontline answer to this situation (Ticco 1995; Tundi Agardy 1994; Le Saout et al. 2013) and it has shown to bring conservation benefits (Claudet et al. 2008; Stolton and Dudley 2010; Liu et al. 2017). At sea, for instance, current political objectives are to cover 30% of the marine spaces under jurisdiction by 2030 with marine reserves (IUCN 2014; IUCN 2016; Commission 2020). Similar concerns also exist on land (Baillie and Zhang 2018; Dinerstein et al. 2019). Within this context, there is a strong demand to find the best compromises between the protection of biodiversity and the sustainability of human uses of these spaces. To analyse such problems, numerical optimisation methods are commonly implemented (Margules and Pressey 2000; Moilanen et al. 2009; Ando et al. 1998; Stewart and Possingham 2005). Such methods are often embedded within a software, *e.g.* Marxan or PrioritizR (Ball and Possingham 2000; Ball et al. 2009; Hanson et al. 2020). They are designed to systematically select reserve sites and are used as a decision support tools in real-world cases¹ (Flower et al. 2020; Fernandes et al. 2005).

Nevertheless, solving these optimisation problems often results in the selection of scattered reserve sites. Yet, designing reserves that are compact, connected, and gap-free is usually needed for ecological, management and enforcement reasons (Diamond 1975). A reserve is connected if one can move anywhere inside without having to leave it. A gap within a reserve is a zone outside the reserve you cannot leave without crossing the reserve. Currently, the spatial attributes of reserves are poorly considered in decision support tools used for reserve selection. In the widely used decision support tools for reserve selection (*e.g.* Marxan, PrioritizR), the only spatial attribute explicitly addressed is the global compactness of a solution (Ball and Possingham 2000; Ball et al. 2009; Watts et al. 2009; Hanson et al. 2020). The compactness of a solution is enforced by directly penalising the overall perimeter of the reserve in the objective function of the optimisation problem addressed. Several issues come with this approach. The linearisation of the perimeter expression involves the addition of many decision variables and constraints (Billionnet 2013; Beyer et al. 2016) which can be computationally expensive in an integer programming context. Also, this approach transforms the problem into a multi-objective problem where the cost of a solution and its perimeter are implicitly competing. In practice, the compactness multiplier is empirically determined until a solution meets the spatial requirements deemed satisfactory. This weakens the systematic nature of the reserve design approach, although a more systematic setting of the compactness multiplier is proposed in (McDonnell et al. 2002). Improvements using both the reserve perimeter and area in the objective were proposed in (McDonnell et al. 2002) to enforce the compactness of the reserve. In the same line, a weighted combination of both compactness and connectivity measure are included in the objective and solved using metaheuristics in (Nalle et al. 2002). In any case, the connectivity and the absence of gaps are not ensured but rather possibly emerging with the empirical enforcement of the reserve compactness.

In an operation research context, several optimisation models were proposed to explicitly account for specific spatial properties (Williams et al. 2004; Williams et al. 2005; Billionnet 2013; Billionnet 2016; Billionnet 2021). For instance, optimisation models aim to design a reserve core with a buffer zone (Williams and ReVelle 1998; Clemens et al. 1999). But these models do not constrain the reserve to be connected and gap-free, although such reserve can emerge from them. A large family of models takes advantage of the modelling possibilities offered by the

¹More case study examples can be found at <https://marxansolutions.org/community/> and https://prioritizr.net/articles/publication_record.html.

use of pairwise distances between candidate sites. Minimizing the sum of pairwise distances or the maximum distance between all reserve sites (Önal and Briers 2002) favour compact reserves, but do not guarantee that the reserve is connected and gap-free. The same applies for models that constrain two distinct sites containing the same conservation feature to be closer than a predefined threshold distance (Williams 2006). Another large family of optimisation model takes advantage of graph theory (Önal and Briers 2006; Wang and Önal 2011; Wang and Önal 2013; Jafari and Hearne 2013; Billionnet 2016; Shirabe 2005) in particular to explicitly ensure the reserve connectivity. However, the site selection may still result in the inclusion of gaps within the reserve solution, which we define as a set of isolated sites not assigned to the reserve and entirely disconnected from the outside (*i.e.* surrounded by the reserve). A reserve perforated by gaps cannot be used in a large-scale reserve design. If gaps appear in a solution proposed by a decision support tool, they will either be arbitrarily incorporated into the reserve, artificially connected to the outside (in either case, this will often lead to the use of suboptimal solutions), or the provided solution will be ignored. Using models imposing connectivity and promoting compactness is likely to favour gap-free reserves, but this is not guaranteed. For instance, it may be necessary to design nature reserves around areas that cannot be included in the reserve, such as a harbour or a trade route. These areas cannot be enclosed by the reserve and must remain accessible from the outside. State-of-the-art models often provide a reserve solution with gaps in these cases. Consideration of gaps within reserves is rarely addressed in the literature. Absence of gaps in the reserve can be *a posteriori* achieved by iteratively searching a gap-free reserve among slightly suboptimal solutions (Billionnet 2016). This model does not *a priori* prevent gaps to be included within the reserve, but rather hope such solution exists even if the objective value is degraded. Such procedure is interesting but does not guarantee to have the connected, compact and gap-free reserve with the best objective value. A model selecting cellularly convex regions (also in regular grids) that are thus connected and gap-free is given in (Williams 2003). Such a model can be adapted to avoid reserves with gaps, but the convexity requirement may neglect some admissible connected and gap-free solutions if they are not (cellularly) convex.

In this work, we proposed a model to build connected, compact and gap-free reserves. An optimisation model using graph theory enforces the connectivity of both the reserve and the non-reserve areas, resulting in connected and gap-free reserves. The overall compactness of the reserve is shaped by specifying a maximum radius or perimeter of the reserve. We illustrate the reserve solution provided by our approach on the real-world instance of Fernando de Noronha. We numerically assess the generality of the proposed approaches on several generated instances made of 300 or 500 planning units and 3 conservations features. Regarding the size of instances in existing works, the number of planning units considered in this work is similar to other existing works: 100 planning units in (Billionnet 2012), 131 in (Önal and Briers 2002), 225 in (Billionnet 2021), 324 planning units (Williams 2006), 391 planning units in (Önal and Briers 2006), 400 in (Billionnet 2016)). However, the number of conservation features is an order of magnitude beyond in existing work, always around 100 conservation features. The comparison of the instance sizes is still hard to interpret, because the instances are not generated in the same way.

5.2 Models

Here we present the integer linear programs for reserve site selection ensuring the reserve to be connected, compact and gap-free. Before all, we recall the general formulation of the reserve site selection problem. Then, the multicommodity flow approach using graph theory is presented for both the reserve and non-reserve areas. Finally, a reduction of the problem is proposed as well as an approach limiting the maximum radius of the graph of the reserve. The graph of the reserve is the graph induced by the planning units selected in the reserve solution.

5.2.1 General formulation of the reserve site selection problem

In a reserve site selection problem, the study area is discretised into a set of J planning units within which a set of I conservation features are distributed. The amount of conservation feature i in the planning unit j is denoted a_{ij} . Each planning unit has a cost c_j usually understood as the socio-economic cost associated with the closure of this unit. The decision is about whether to include the planning unit in the reserve. Consequently, we associate the decision variables x_j with each planning unit j : $x_j = 1$ if a planning unit j belongs to the reserve and $x_j = 0$ otherwise. One then seeks to find the least cost collection of planning units covering a sufficient amount for each conservation features. The covering of a conservation feature i is considered sufficient if it exceeds a user-defined level noted t_i .

$$\sum_{j \in J} a_{ij} x_j \geq t_i, \forall i \in I \quad (5.1)$$

Mathematically speaking, the general problem of reserve site selection is expressed as the following integer linear program P_N :

$$P_N : \begin{cases} \min_x & \sum_{j \in J} c_j x_j \\ \text{s.t.} & (5.1) \\ & x_j \in \{0, 1\} \quad \forall j \in J \end{cases}$$

In state-of-the-art models, the reserve perimeter is included in the objective function, to favour aggregated reserve solutions since a small perimeter involves a compact reserve. The perimeter is computed as the total length of the boundaries between reserved and non-reserved planning units. To model this, the length of the shared boundary between planning units j_1 and j_2 is denoted $b_{j_1 j_2}$. A parameter β is used to set the importance of compactness relatively to the total cost of site selection. The quadratic expression of the perimeter can be linearised (Billionnet 2007; Beyer et al. 2016) by replacing the product $x_{j_1} x_{j_2}$ with the new binary decision variable $z_{j_1 j_2}$ and add the following set of constraints:

$$\begin{cases} z_{j_1 j_2} - x_{j_1} \leq 0 & \forall j_1 \in J \\ z_{j_1 j_2} - x_{j_2} \leq 0 & \forall j_2 \in J \\ z_{j_1 j_2} - x_{j_1} - x_{j_2} \geq -1 & \forall j_1 \in J, \forall j_2 \in J \end{cases} \quad (5.2)$$

Finally, the general formulation of the reserve site selection problem results in the following integer linear program P_{N+COMP} :

$$P_{N+COMP} : \begin{cases} \min_{x,z} & \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} (x_{j_1} - z_{j_1 j_2}) \\ \text{s.t.} & (5.1), (5.2) \\ & x_j, z_{j_1 j_2} \in \{0, 1\} \end{cases} \quad \forall j, j_1, j_2 \in J$$

This combinatorial optimisation problem is a minimum set cover problem known to be NP-hard (Garey and Johnson 1979). It is a non-convex problem due to the binary nature of the decision variables. Yet, it can be expressed as an integer linear program and known solvers (like Gurobi or CbC) can solve it for realistic instances in a reasonable time.

5.2.2 Connectivity of the reserve

The grid is seen as a graph, where each planning unit $j \in J$ represents a node in the graph. The set of nodes is J . Planning units sharing an edge in the grid are considered neighbours and thus involve an edge $e = (e_1, e_2)$ in the graph between the nodes $e_1, e_2 \in J$. The set of edges is noted E . The corresponding directed edges ($e_1 \rightarrow e_2$) and ($e_2 \rightarrow e_1$) are called arcs. The set of arcs is noted A . We then use a multicommodity flow model developed in (Abdelmaguid 2018). The idea is to constrain every node selected in the reserve to have a flow going from the source to the sink. The source is the commodity $k \in K$, *i.e.* a selected node, and the sink is the root node of the spanning tree. Therefore, we build a path connecting every selected nodes $k \in K$ and the root node which is constrained to belong to the reserve. The reserve is thus ensured to be connected. In this model, the set of commodities is $K = J$. The selection of the root node $j \in J$ of the spanning tree associated to the reserve is represented by the binary decision variable $r_j \in \{0, 1\}$. The selection of an arc $a \in A$ in the spanning tree associated to the reserve is represented by the binary decision variable $u_a \in \{0, 1\}$. The activation of the flow of commodity $k \in K$ between the source node k and the sink node (*i.e.* root of the spanning tree) along the arc $a \in A$ is represented by the binary decision variable $f_a^k \in \{0, 1\}$. Let $V(n)$ be the set of neighbours nodes of node $n \in J$.

The selected arcs of the spanning tree must be between two nodes selected in the reserve:

$$\begin{cases} u_a \leq x_{a_1} & \forall a = (a_1, a_2) \in A \\ u_a \leq x_{a_2} & \forall a = (a_1, a_2) \in A \end{cases} \quad (5.3)$$

A maximum of one arc is activated by edge:

$$u_{(e_1 \rightarrow e_2)} + u_{(e_2 \rightarrow e_1)} \leq 1 \quad \forall e = (e_1, e_2) \in E \quad (5.4)$$

The number of arcs in the tree is equal to the number of nodes minus 1 (prevent cycle formation):

$$\sum_{a \in A} u_a = \sum_{j \in J} x_j - 1 \quad (5.5)$$

The root of the tree must be in the reserve:

$$r_j \leq x_j \quad \forall j \in J \quad (5.6)$$

There is only one root node for the spanning tree of the reserve:

$$\sum_{j \in J} r_j \leq 1 \quad (5.7)$$

If the arc is not selected, all associated flow variables are set to 0:

$$f_a^k \leq u_a \quad \forall a \in A, \forall k \in K \quad (5.8)$$

If the node is not selected, all the associated flow variables are set to 0:

$$f_a^k \leq x_k \quad \forall a \in A, \forall k \in K \quad (5.9)$$

For commodity $k \in K$, the flow at source node is 1, the flow at sink node is 0, elsewhere for selected nodes, the flow entering is the same as the flow leaving the node.

$$\begin{cases} \sum_{j \in V(n)} f_{(j \rightarrow n)}^k - \sum_{j \in V(n)} f_{(n \rightarrow j)}^k \leq r_n & \forall k \in K, \forall n \in J \setminus \{k\} \\ \sum_{j \in V(n)} f_{(j \rightarrow n)}^k - \sum_{j \in V(n)} f_{(n \rightarrow j)}^k \geq 0 & \forall k \in K, \forall n \in J \setminus \{k\} \\ \sum_{j \in V(k)} f_{(k \rightarrow j)}^k - \sum_{j \in V(k)} f_{(j \rightarrow k)}^k = x_k - r_k & \forall k \in K \end{cases} \quad (5.10)$$

Finally, the multicommodity flow model P_{CON} for the reserve is:

$$P_{CON} : \begin{cases} \min_{x,z,u,r,f} & \sum_{j \in J} c_j x_j + \beta \sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} (x_{j_1} - z_{j_1 j_2}) \\ \text{s.t.} & (5.1) - (5.10) \\ & x_j, z_{j_1 j_2}, u_a, r_j, f_a^k \in \{0, 1\} \end{cases} \quad \forall j, j_1, j_2 \in J, \forall a \in A, \forall k \in K$$

5.2.3 Gap-free reserve

We apply the same multicommodity flow model to the non-reserve to have a connected non-reserve. A connected non-reserve implies that the reserve would not have gaps within it. Thus, the term $1 - x_j$ plays the role of the term x_j . We add a fictive node α in the graph representing the area outside the studied zone. Indeed, the non-reserve must be connected to the exterior area. Note that we fix the node α to be the root of the spanning tree of the non-reserve. The selection of an arc $a \in A$ in the spanning tree associated to the non-reserve is represented by the binary decision variable $v_a \in \{0, 1\}$. The activation of the flow of commodity $k \in K$ between the source node k and the sink node α along the arc $a \in A$ is represented by the binary decision variable $g_a^k \in \{0, 1\}$. The set of edges and arcs associated with the fictive node α are respectively noted E^f and A^f . Let $E^+ = E \cup E^f$, $A^+ = A \cup A^f$, and $J^+ = J \cup \{\alpha\}$.

The added set of constraints for the non-reserve is :

$$\left\{ \begin{array}{ll} x_\alpha = 0 & \\ v_a \leq 1 - x_{a_1} & \forall a = (a_1 \rightarrow a_2) \in A^+ \\ v_a \leq 1 - x_{a_2} & \forall a = (a_1 \rightarrow a_2) \in A^+ \\ v_{(e_1 \rightarrow e_2)} + v_{(e_2 \rightarrow e_1)} \leq 1 & \forall e = (e_1, e_2) \in E^+ \\ \sum_{a \in A^+} v_a = \sum_{j \in J^+} (1 - x_j) - 1 & \\ g_a^k \leq v_a & \forall a \in A, \forall k \in K \\ g_a^k \leq 1 - x_k & \forall a \in A, \forall k \in K \\ \sum_{j \in V(\alpha)} g_{(\alpha \rightarrow j)}^k - \sum_{j \in V(\alpha)} g_{(j \rightarrow \alpha)}^k \leq 1 & \forall k \in K \\ \sum_{j \in V(n)} g_{(j \rightarrow n)}^k - \sum_{j \in V(n)} g_{(n \rightarrow j)}^k = 0 & \forall k \in K, \forall n \in J \setminus \{k\} \\ \sum_{j \in V(k)} g_{(j \rightarrow k)}^k - \sum_{j \in V(k)} g_{(k \rightarrow j)}^k = 1 - x_k & \forall k \in K \\ v_a, g_a^k \in \{0, 1\} & \forall a \in A^+, \forall k \in K \end{array} \right. \quad (5.11)$$

By adding (5.11) to P_{CON} , we get the integer linear program P_{CON+GF} that ensure connected reserve solutions to be gap-free.

5.2.4 Compactness of the reserve

5.2.4.1 Maximum radius in the graph of the reserve

We want to avoid producing connected reserve solutions that spread across the entire study area. We thus impose the radius of the graph of the reserve to remain below a predefined threshold, denoted R_{max} in the following. We have a double hope with this additional constraint. First, we will produce more compact reserves and avoid cobweb shape for the reserves. By doing so, we also limit the considered nodes when finding a path in the multicommodity flow models. By removing the nodes further than R_{max} , we also hope to increase the solving speed.

Once the graph of an incumbent reserve solution is connected, we can define the centre and the radius of the graph of the reserve. The centre is the selected node whose maximal distance

from other selected nodes is the smallest. The radius is the maximum distance in the graph between the centre and other selected nodes. Let $d(j_1, j_2)$ define the distance in the graph of the reserve between the node $j_1 \in J$ and $j_2 \in J$. This distance corresponds to the smallest path in the reserve graph from node j_1 to node j_2 . Note that the global matrix of distances between all nodes of the grid was computed outside the solving procedure. All the selected nodes of the incumbent connected reserve that are at a distance greater than R_{max} from the centre are added to the set of commodities K . Then, the following constraint is applied:

$$\sum_{\substack{j \in J \\ d(j,k) \leq R_{max}}} \sum_{n \in V(j)} f_{j \rightarrow n}^k \leq R_{max} \quad \forall k \in K \quad (5.12)$$

Finally, we impose the non-selection of nodes at a distance greater than R_{max} from the root of the tree of the reserve:

$$x_{j_1} \leq 1 - r_{j_2} \quad \forall j_1 \in J, j_2 \in J, d(j_1, j_2) > R_{max} \quad (5.13)$$

By adding (5.12) and (5.13) to P_{CON+GF} , we get the integer linear program that ensures connected and gap-free solutions to have a maximum radius of R_{max} .

5.2.4.2 Maximum perimeter of the reserve

As explained in Section 5.2.1, the compactness of a reserve in state-of-the-art models is enforced using a multi-objective approach by penalising the reserve perimeter in the objective. Rather than that, we can keep a single objective formulation and specify a maximum perimeter P_{max} the reserve should not exceed. The associated constraint is :

$$\sum_{j_1 \in J} \sum_{j_2 \in J} b_{j_1 j_2} (x_{j_1} - z_{j_1 j_2}) \leq P_{max} \quad (5.14)$$

By adding (5.12) and (5.13) and/or (5.14) to P_{CON+GF} , we get the integer linear program that ensures connected and gap-free solutions to have a maximum perimeter of P_{max} . The models that include constraints used to enforce compactness of the reserve will be named $P_{CON+GF+COMP}$. The setting of compactness parameters, *i.e.* β , R_{max} , P_{max} , will remove the ambiguity related to our choice of constraints. If β is set, we include the penalty of the reserve perimeter in the objective and add the associated linearisation constraints (5.2) to the model. If R_{max} is set, the constraints (5.12) and (5.13) are added to the model. If P_{max} is set, the constraint (5.14) is added to the model.

5.2.5 Improvements of the model

5.2.5.1 Chessboard reduction

In a rectangular grid, if we want a connected and gap-free reserve, a node in a given binary state, *i.e.* selected or unselected, cannot be surrounded by neighbouring nodes in the complementary state. The rectangular grid is thus assimilated to a chessboard, and the nodes are separated into two sets: black and white nodes. This way, the 4 neighbouring nodes of a black node are white and vice versa. Let B be the set of black nodes and W the set of white nodes. We have $J = W \cup B$. In terms of constraints, we prevent white (respectively black) nodes of the grid in a given state to be surrounded by four black (respectively white) neighbours in the same state:

$$\begin{cases} x_j \leq \sum_{i \in V(j)} x_i & \forall j \in J \\ 1 - x_j \leq \sum_{i \in V(j)} 1 - x_i & \forall j \in J \end{cases} \quad (5.15)$$

Then, we apply the multicommodity flow model only to black nodes. It means that the set of commodities is $K = B$ instead of $K = J$ in model P_{CON} or P_{CON+GF} . This is the main motivation behind this chessboard reduction: we reduce the number of expensive multicommodity flow constraints by only adding two constraints by node. This way, each node is whether associated with a commodity or satisfy constraint (5.15) and has its neighbours associated with a commodity. In the following, the chessboard reduction is systematically applied.

5.2.5.2 Lazy constraints

Enforcing flow constraints for every node can be computationally challenging in multicommodity flow models. Lazy constraints are constraints included in the model only if they are not satisfied by the incumbent solution. Since the flow constraints (5.10) in the multicommodity flow model can be separated by commodities, these constraints are implemented as lazy constraints. The motivation behind this choice is that a non-connected reserve is not a frequent case. We hope that the iterative activation of the lazy constraints is faster than considering the exhaustive set of flow constraints. In the following, the concerned constraints are systematically applied as lazy constraints.

A graph is connected if there is a path from any point to any other point in the graph. If a graph is not connected, the graph is made of two or more isolated connected subgraphs. We speak of connected components of the graph for the maximal, in terms of nodes, connected subgraphs. In our case, if the number of connected components of an incumbent reserve solution is greater than 1, the reserve is not connected and we activate the flow constraints associated with a given commodity noted $k_1 \in K$. We define the rentability of a node as the ratio between the conservation feature's total amounts within a planning unit and its cost. The commodity k_1 is chosen as the node with the highest rentability among the nodes of each connected component of the incumbent reserve solution.

We do the same for the multicommodity flow model of the non-reserve. If the number of connected components of the non-reserve graph is greater than 1, the incumbent reserve solution has a gap within it, and we activate the flow constraints associated with a given commodity noted $k_2 \in K$. The commodity k_2 is chosen as the node with the lowest rentability among each connected component of the non-reserve. Again, a gap within the reserve is hoped to be a rare case, so the lazy constraints allow a faster solving than including the exhaustive set of flow constraints (5.11).

The constraints (5.12) are also implemented as lazy constraints, and thus activated only if the incumbent reserve solution is spreading too much.

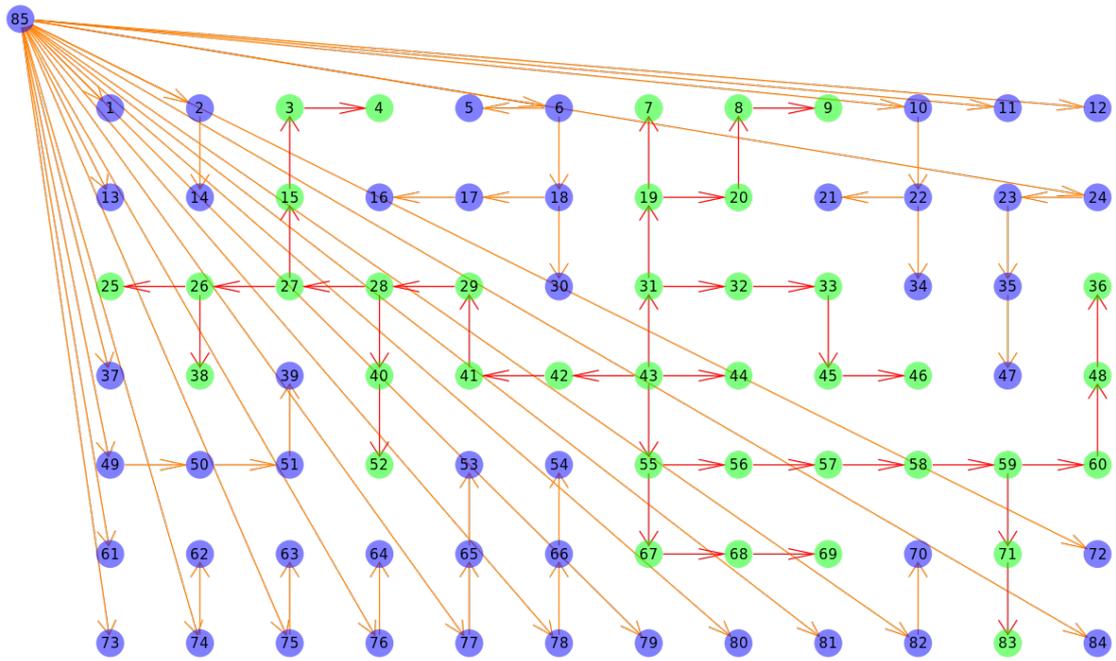


Figure 5.1: Example of the graph of a connected and gap-free reserve solution. Red arrows show the spanning tree of the reserve. The centre of the reserve tree is node 43. Radius of the graph of the reserve is 8 (reached from node 43 to 36 for example). Orange arrows show the spanning tree of the non-reserve. The centre of the non-reserve tree is node $\alpha = 85$ representing the outside area.

5.3 Numerical experiments

5.3.1 Illustration on a real-world instance

The real-world instance of Fernando de Noronha is detailed in Section 1.5.1.1 and is shown in Figure 1.7.

Table 5.1 provides the characteristics of the reserve solutions computed using the models described in Section 5.2 on the real-world instance of Fernando de Noronha. The first observation is that the spatial coherence of a reserve is not guaranteed by state-of-the-art models. Figure 5.2a and Figure 5.3a shows that the reserve site selection is really scattered for $\beta = 0$. Setting $\beta = 1$ in these models improved the global compactness of the reserve selection as illustrated in Figure 5.2b and Figure 5.3b but did not guarantee the connectivity of the reserve nor the absence of gaps within it (cf. Table 5.1). The state-of-the-art model with $\beta = 1$ shown in Figure 5.3b illustrates the problem with high covering demands and locked-out planning units: it naturally creates gaps by surrounding the locked-out planning units. When targets were set to 50%, obtaining a connected and gap-free reserve (cf. Figure 5.2c) took 19.5 seconds. It required to solve the model $P_{CON+COMP}$ with $\beta = 1$ since the reserve solution did not have any gap within it. However, when we increased the compactness demand ($\beta = 1$ and $R_{max} = 14$), a gap was appearing within the solution. We removed this by solving the model $P_{CON+GF+COMP}$. We also observe that the solving of $P_{CON+GF+COMP}$ took less time than $P_{CON+COMP}$ (179.6 against 920.5 seconds). To obtain an even more compact reserve, we directly constrained the reserve perimeter to remain below 80 instead of the 90 of the reserve solution with $\beta = 1$ and $R_{max} = 14$. The connected, compact and gap-free solution (cf. Figure 5.2d) was obtained in 366.1 seconds. When targets were set to 70%, solving the model $P_{CON+COMP}$ with $\beta = 1$ did not prevent the occurrence of gaps within the reserve (cf. Table 5.1). The compact, connected and gap-free reserve solution obtained by solving $P_{CON+GF+COMP}$ with $\beta = 1$ is shown in Figure 5.3c. Also, the selection of isolated planning units is tolerated as soon as the perimeter involved contributed less to the objective than the selection cost. Unlike the state-of-the-art models solutions, our connected and gap-free reserve solutions left a path from the harbour of Fernando de Noronha to the outside area. It took only 42.6 seconds in this example. Then we increased our compactness demands by setting $P_{max} = 90$ and $R_{max} = 16$. The corresponding solution is shown in Figure 5.3d. This reserve which looks more compact than Figure 5.3c and is still connected and gap-free.

Targets	Model	Parameters	Time	Perimeter	Cost	Radius	Components	Gaps
50%	N	$\beta = 0$	0.2	150	90.6	-	14	4
50%	N+COMP	$\beta = 1$	0.2	96	101.7	-	2	0
50%	CON+COMP	$\beta = 1$	19.5	92	106.0	16	1	0
50%	CON+COMP	$\beta = 1, R_{max} = 14$	920.5	90	112.4	14	1	1
50%	CON+COMP+GF	$\beta = 1, R_{max} = 14$	179.6	90	112.8	14	1	0
50%	CON+COMP+GF	$P_{max} = 80, R_{max} = 14$	366.1	80	125.6	14	1	0
70%	N	$\beta = 0$	0.0	156	200.2	-	7	7
70%	N+COMP	$\beta = 1$	0.1	108	216.8	-	2	2
70%	CON+COMP	$\beta = 1$	69.4	100	227.6	16	1	2
70%	CON+COMP+GF	$\beta = 1$	42.6	98	232.8	18	1	0
70%	CON+COMP+GF	$\beta = 1, R_{max} = 17$	295.1	96	235.0	17	1	0
70%	CON+COMP+GF	$\beta = 1, R_{max} = 16$	367.6	94	237.6	16	1	0
70%	CON+COMP+GF	$P_{max} = 90, R_{max} = 16$	164.2	90	243.6	16	1	0

Table 5.1: Results of the numerical experiments for 36×21 planning units and 3 conservation features for the real-world instance of Fernando de Noronha. A summary of the characteristics of the reserve solutions is provided: computation time in seconds, reserve perimeter, total cost, radius of the reserve graph, number of connected components, number of gaps. Targets are the same for the three conservation features (50% or 70%). N=nominal, COMP=compactness, CON=connectivity, GF=gap-free.

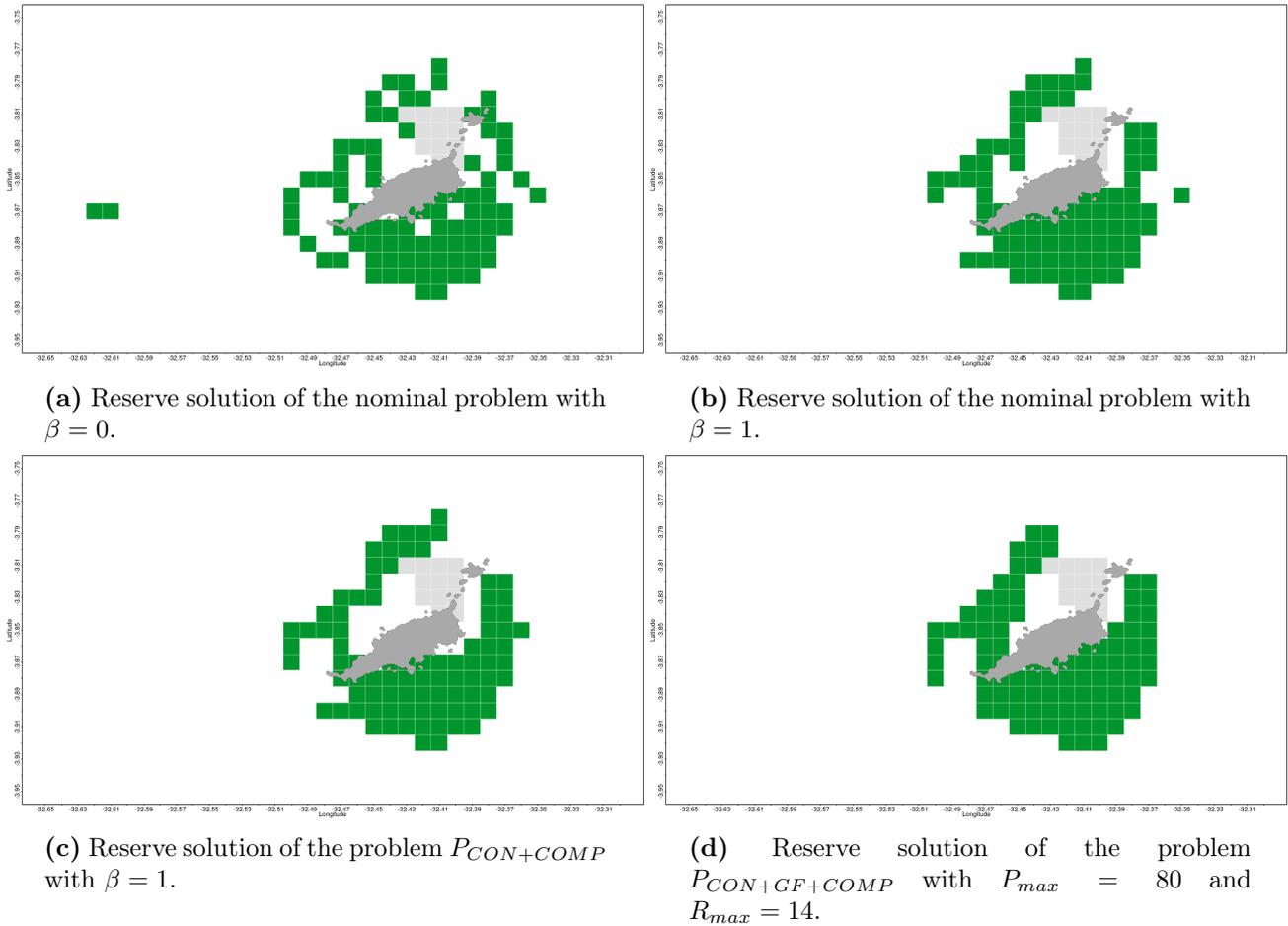


Figure 5.2: Reserve solutions of the real-world instance of Fernando de Noronha for several models. Conservation features targets are all set to 50%. Green planning units represent the reserve selection. Grey planning units are *a priori* excluded. N=nominal, COMP=compactness, CON=connectivity, GF=gap-free.

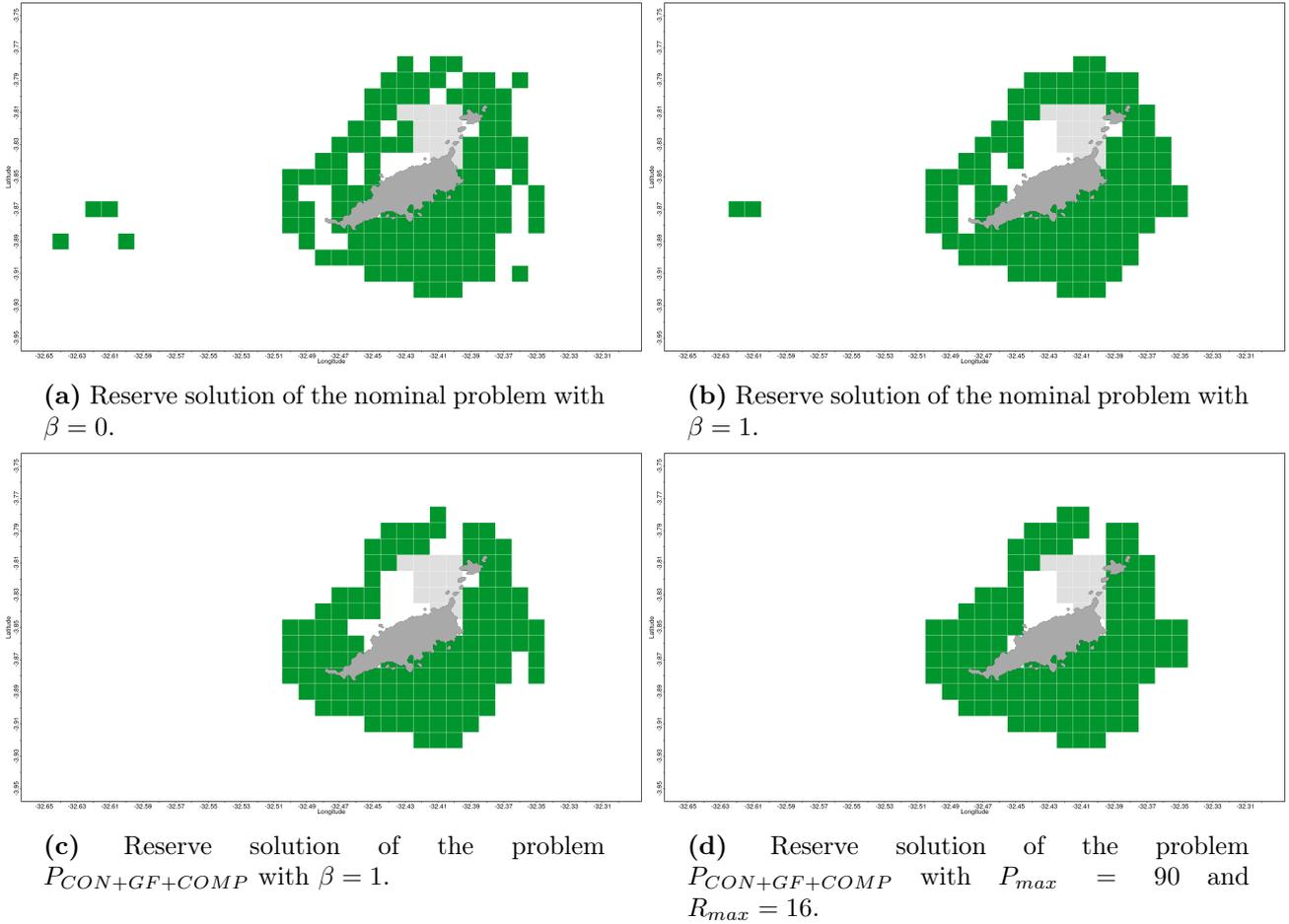


Figure 5.3: Reserve solutions of the real-world instance of Fernando de Noronha for several models. Conservation features targets are all set to 70%. Green planning units represent the reserve selection. Grey planning units are *a priori* excluded. N=nominal, COMP=compactness, CON=connectivity, GF=gap-free.

5.3.2 Feasibility assessment on generated instances

In this section, we tested our models for several generated instances to have a more accurate idea of the computation time needed and the extra-cost involved to obtain connected and gap-free reserve solutions. We used a systematic way of building scenarios with a user-defined complexity for our reserve site selection optimisation problems. More details can be found in Section 1.5.2. The generation of instances is different from what is done in (Billionnet 2016). We wanted more realistic instances, closer to a real dataset, and more likely to have gaps when targets are high. Table 5.2 and Table 5.3 provides the characteristics of the reserve solutions computed using the models described in Section 5.2 for generated instances of respectively 300 and 500 planning units.

In any case, the nominal problems P_N and P_{N+COMP} are solved very fast, mostly under 1 second and 3.4 seconds at worst. However, the reserve solutions with $\beta = 0$ are very scattered, with many gaps, for all instances: 23.6 connected components and 11.0 gaps in average for 300 planning units; 41.5 connected components and 21.2 gaps in average for 500 planning units. The reserve solutions with $\beta = 1$ are less scattered, but still have several connected components and gaps in general: 2.7 connected components and 2.2 gaps in average for 300 planning units; 5.2 connected components and 3.4 gaps in average for 500 planning units.

The reserve solutions using our complete model $P_{CON+GF+COMP}$ with $\beta = 1$ expectedly provides connected and gap-free reserves. For instances of 300 planning units, the mean computation time is 80.0 seconds with a standard deviation of 61.4 seconds, a minimum and maximum time of 15.8 and 229.6 seconds respectively. Obtaining a connected and gap-free reserve solution involves a mean relative extra cost of 2.7% (standard deviation of 1.8%, maximum of 5.7%) with respect to the state-of-the-art model with the same value of β . When we match the compactness demand, *i.e.* we constrain the perimeter to remain below the perimeter of the state-of-the-art reserve solution, the mean relative extra cost drops to 0.7% (standard deviation of 0.4%, maximum of 1.3%). The mean computation time for this model is 150.6 seconds. Once the reserve was compact, connected and gap-free, we evaluated the impact of an increase of compactness using the R_{max} constraints. To do so, we set R_{max} to the reserve radius obtained solving $P_{CON+GF+COMP}$ with $\beta = 1$ minus 1 so the constraints were activated. For instances of 300 planning units, the mean computation time is 182.3 seconds with a standard deviation of 280.4 seconds, a minimum and maximum time of 21.1 and 928.7 seconds. The inclusion of R_{max} constraints was sometimes associated with a decrease in computation time, sometimes with an increase, depending on the instances considered.

For instances of 500 planning units, the proof of optimality was not provided every time with a time limit set to 1000 seconds. Though, the incumbent solution returned is still compact, connected and gap-free. In the following, for instances reaching the time limit, we will consider a solving time of 1000 seconds and the characteristics of the incumbent suboptimal reserve solution. The mean computation time of the reserve solutions using our complete model $P_{CON+GF+COMP}$ with $\beta = 1$ is 480.0 seconds with a standard deviation of 382.8 seconds, a minimum and maximum time of 50.8 and 1000 seconds respectively. Obtaining a connected and gap-free reserve solution involves a mean relative extra cost of 3.4% (standard deviation of 1.7%, maximum of 6.9%) with respect to the state-of-the-art model with the same value of β . When we match the compactness demand, *i.e.* we constrain the perimeter to remain below the perimeter of the state-of-the-art reserve solution, the mean relative extra cost drops to 0.9% (standard deviation of 0.4%, maximum of 1.6%). For instances of 500 planning units, when a maximum radius was imposed to the reserve, it led to more computational difficulties. Unlike instances of 300 planning units, the addition of the R_{max} constraints systematically involves a

greater computation time or an incumbent solution further from optimality than without R_{max} . There are even 3 instances where no solutions are found within the time limit.

5.3.3 Assessment of the compactness models

In this section, we aim to assess the difference of computation time needed to obtain a compact solution whether using constraints associated to the use of β or P_{max} . To do so, we first solved $P_{CON+GF+COMP}$ with $\beta = 1$. The perimeter of the reserve solution obtained was then used for P_{max} when we solved $P_{CON+GF+COMP}$. We expectedly obtained the same solutions between the two models. For instances of 300 planning units, the mean computation time was 80.0 seconds for $P_{CON+GF+COMP}$ with β and 74.8 seconds with $P_{CON+GF+COMP}$ with P_{max} . Then, Table 5.4 does not show a systematic trend between the two models since it sometimes took more time, sometimes less time, depending on the instance. However, for instances of 500 planning units, Table 5.5 shows a clear trend: models with constraints associated to the use of β are solved faster than the models using the P_{max} constraint for every instance. When the time limit was reached, the model using the constraints associated with β provided a solution closer to optimality than the model with the P_{max} constraint.

Instance	Model	Parameters	Time	Perimeter	Cost	Radius	Components	Gaps
1	N	$\beta = 0$	0.1	272	460.1	-	21	10
1	N+COMP	$\beta = 1$	0.4	106	544.3	-	2	3
1	CON+GF+COMP	$\beta = 1$	15.8	76	575.5	12	1	0
1	CON+GF+COMP	$P_{max} = 106$	245.6	106	550.2	13	1	0
1	CON+GF+COMP	$\beta = 1, R_{max} = 11$	24.9	82	571.4	11	1	0
2	N	$\beta = 0$	0.2	286	480.9	-	25	14
2	N+COMP	$\beta = 1$	0.4	96	580.2	-	1	3
2	CON+GF+COMP	$\beta = 1$	80.1	86	592.5	13	1	0
2	CON+GF+COMP	$P_{max} = 96$	143.5	96	582.7	14	1	0
2	CON+GF+COMP	$\beta = 1, R_{max} = 12$	30.9	88	590.9	12	1	0
3	N	$\beta = 0$	0.1	296	466.4	-	26	9
3	N+COMP	$\beta = 1$	0.2	112	541.2	-	2	4
3	CON+GF+COMP	$\beta = 1$	124.0	102	554.9	12	1	0
3	CON+GF+COMP	$P_{max} = 112$	229.9	112	546.3	12	1	0
3	CON+GF+COMP	$\beta = 1, R_{max} = 11$	21.1	86	572.3	11	1	0
7	N	$\beta = 0$	0.1	276	454.0	-	19	4
7	N+COMP	$\beta = 1$	0.1	142	527.9	-	4	1
7	CON+GF+COMP	$\beta = 1$	40.2	132	540.4	15	1	0
7	CON+GF+COMP	$P_{max} = 142$	10.1	142	530.9	16	1	0
7	CON+GF+COMP	$\beta = 1, R_{max} = 14$	101.3	130	543.0	14	1	0
12	N	$\beta = 0$	0.1	284	491.9	-	23	14
12	N	$\beta = 1$	0.1	112	572.1	-	2	2
12	CON+GF+COMP	$\beta = 1$	19.8	110	575.7	17	1	0
12	CON+GF+COMP	$P_{max} = 112$	44.3	112	574.4	17	1	0
12	CON+GF+COMP	$\beta = 1, R_{max} = 16$	345.5	106	580.0	16	1	0
13	N	$\beta = 0$	0.4	268	464.9	-	21	10
13	N+COMP	$\beta = 1$	0.5	106	555.0	-	1	2
13	CON+GF+COMP	$\beta = 1$	69.6	106	555.9	12	1	0
13	CON+GF+COMP	$P_{max} = 106$	41.2	106	555.9	12	1	0
13	CON+GF+COMP	$\beta = 1, R_{max} = 11$	55.3	92	570.9	11	1	0
14	N	$\beta = 0$	0.1	286	452.7	-	30	15
14	N+COMP	$\beta = 1$	0.2	120	524.3	-	2	4
14	CON+GF+COMP	$\beta = 1$	65.9	102	548.2	12	1	0
14	CON+GF+COMP	$P_{max} = 120$	307.2	120	531.2	12	1	0
14	CON+GF+COMP	$\beta = 1, R_{max} = 11$	104.8	112	539.8	11	1	0
15	N	$\beta = 0$	0.2	302	490.4	-	25	14
15	N+COMP	$\beta = 1$	0.2	148	565.4	-	6	1
15	CON+GF+COMP	$\beta = 1$	76.5	136	582.4	18	1	0
15	CON+GF+COMP	$P_{max} = 148$	78.2	148	571.1	19	1	0
15	CON+GF+COMP	$\beta = 1, R_{max} = 17$	928.7	138	582.1	16	1	0
16	N	$\beta = 0$	0.1	280	472.3	-	21	5
16	N+COMP	$\beta = 1$	0.7	124	567.6	-	4	0
16	CON+GF+COMP	$\beta = 1$	229.6	120	573.8	17	1	0
16	CON+GF+COMP	$P_{max} = 124$	191.6	124	569.9	16	1	0
16	CON+GF+COMP	$\beta = 1, R_{max} = 16$	175.3	118	575.8	16	1	0
17	N	$\beta = 0$	0.0	274	418.2	-	25	15
17	N+COMP	$\beta = 1$	1.0	114	477.8	-	3	2
17	CON+GF+COMP	$\beta = 1$	78.7	94	499.4	12	1	0
17	CON+GF+COMP	$P_{max} = 114$	214.0	114	483.3	14	1	0
17	CON+GF+COMP	$\beta = 1, R_{max} = 11$	35.3	84	515.6	11	1	0

Table 5.2: Results for 10 generated instances of 20×15 planning units and 3 conservation features. A summary of the characteristics of the reserve solutions is provided: computation time in seconds, reserve perimeter, total cost, radius of the reserve graph, number of connected components, number of gaps. Conservation features targets are all set to 50%. N=nominal, COMP=compactness, CON=connectivity, GF=gap-free.

Instance	Model	Parameters	Time	Perimeter	Cost	Radius	Components	Gaps
4	N	$\beta = 0$	0.1	524	772.6	-	38	28
4	N+COMP	$\beta = 1$	0.4	248	925.0	-	3	7
4	CON+GF+COMP	$\beta = 1$	139.0	214	961.8	21	1	0
4	CON+GF+COMP	$P_{max} = 248$	993.1	248	930.8	22	1	0
4	CON+GF+COMP	$\beta = 1, R_{max} = 20$	284.3	204	973.4	20	1	0
5	N	$\beta = 0$	0.2	474	663.6	-	39	17
5	N+COMP	$\beta = 1$	0.8	224	778.9	-	7	4
5	CON+GF+COMP	$\beta = 1$	TL[0.5%]	222	792.8	21	1	0
5	CON+GF+COMP	$P_{max} = 224$	TL[0.7%]	224	791.6	21	1	0
5	CON+GF+COMP	$\beta = 1, R_{max} = 20$	TL[Infeasible]	-	-	-	-	-
6	N	$\beta = 0$	0.6	494	742.1	-	43	16
6	N+COMP	$\beta = 1$	3.4	248	882.8	-	6	2
6	CON+GF+COMP	$\beta = 1$	577.4	224	910.8	23	1	0
6	CON+GF+COMP	$P_{max} = 248$	969.8	248	889.1	23	1	0
6	CON+GF+COMP	$\beta = 1, R_{max} = 22$	TL[Infeasible]	-	-	-	1	0
8	N	$\beta = 0$	0.1	476	749.8	-	43	20
8	N+COMP	$\beta = 1$	0.2	196	880.8	-	5	3
8	CON+GF+COMP	$\beta = 1$	50.8	164	915.9	21	1	0
8	CON+GF+COMP	$P_{max} = 196$	105.1	196	886.4	21	1	0
8	CON+GF+COMP	$\beta = 1, R_{max} = 20$	179.7	166	914.1	20	1	0
9	N	$\beta = 0$	0.1	530	806.8	-	43	25
9	N+COMP	$\beta = 1$	1.4	228	957.5	-	3	4
9	CON+GF+COMP	$\beta = 1$	542.6	192	999.5	21	1	0
9	CON+GF+COMP	$P_{max} = 228$	717.2	228	965.1	24	1	0
9	CON+GF+COMP	$\beta = 1, R_{max} = 20$	TL[Infeasible]	-	-	-	1	0
10	N	$\beta = 0$	0.4	460	773.3	-	47	27
10	N+COMP	$\beta = 1$	2.6	194	912.6	-	3	3
10	CON+GF+COMP	$\beta = 1$	97.6	184	927.8	18	1	0
10	CON+GF+COMP	$P_{max} = 194$	TL[0.1%]	194	919.1	20	1	0
10	CON+GF+COMP	$\beta = 1, R_{max} = 17$	263.2	174	938.4	17	1	0
18	N	$\beta = 0$	0.2	494	771.8	-	41	17
18	N+COMP	$\beta = 1$	0.6	230	913.4	-	8	5
18	CON+GF+COMP	$\beta = 1$	TL[0.4%]	178	976.6	23	1	0
18	CON+GF+COMP	$P_{max} = 230$	TL[0.5%]	230	927.0	22	1	0
18	CON+GF+COMP	$\beta = 1, R_{max} = 22$	TL[4.1%]	222	976.8	21	1	0
19	N	$\beta = 0$	0.1	460	807.9	-	30	15
19	N+COMP	$\beta = 1$	0.5	210	940.9	-	6	2
19	CON+GF+COMP	$\beta = 1$	109.2	202	956.2	23	1	0
19	CON+GF+COMP	$P_{max} = 210$	TL[0.3%]	210	951.5	22	1	0
19	CON+GF+COMP	$\beta = 1, R_{max} = 22$	435.2	202	956.9	19	1	0
20	N	$\beta = 0$	0.1	506	800.6	-	50	26
20	N+COMP	$\beta = 1$	1.3	198	950.4	-	6	2
20	CON+GF+COMP	$\beta = 1$	895.1	184	969.5	20	1	0
20	CON+GF+COMP	$P_{max} = 198$	TL[0.3%]	198	957.9	24	1	0
20	CON+GF+COMP	$\beta = 1, R_{max} = 19$	TL[0.7%]	174	984.2	19	1	0
21	N	$\beta = 0$	0.3	470	736.0	-	41	21
21	N+COMP	$\beta = 1$	0.6	198	870.2	-	5	2
21	CON+GF+COMP	$\beta = 1$	388.3	166	905.8	18	1	0
21	CON+GF+COMP	$P_{max} = 198$	616.7	198	875.6	19	1	0
21	CON+GF+COMP	$\beta = 1, R_{max} = 17$	TL[0.2%]	162	911.4	17	1	0

Table 5.3: Results for 10 instances of 25×20 planning units and 3 conservation features. A summary of the characteristics of the reserve solutions is provided: computation time in seconds, reserve perimeter, total cost, radius of the reserve graph, number of connected components, number of gaps. Conservation features targets are all set to 50%. N=nominal, COMP=compactness, CON=connectivity, GF=gap-free. If the time limit is reached (TL=1000s), the optimality gap of the incumbent solution is given within brackets.

Instance	Model	Parameters	Time	Perimeter	Cost	Radius	Components	Gaps
1	CON+GF+COMP	$\beta = 1$	15.8	76	575.5	12	1	0
1	CON+GF+COMP	$P_{max} = 76$	12.0	76	575.5	12	1	0
2	CON+GF+COMP	$\beta = 1$	80.1	86	592.5	13	1	0
2	CON+GF+COMP	$P_{max} = 86$	83.6	86	592.5	13	1	0
3	CON+GF+COMP	$\beta = 1$	124.0	102	554.9	12	1	0
3	CON+GF+COMP	$P_{max} = 102$	47.4	102	554.9	12	1	0
7	CON+GF+COMP	$\beta = 1$	40.2	132	540.4	15	1	0
7	CON+GF+COMP	$P_{max} = 132$	35.8	132	540.4	15	1	0
12	CON+GF+COMP	$\beta = 1$	19.8	110	575.7	17	1	0
12	CON+GF+COMP	$P_{max} = 110$	118.0	110	575.7	17	1	0
13	CON+GF+COMP	$\beta = 1$	69.6	106	555.9	12	1	0
13	CON+GF+COMP	$P_{max} = 106$	35.9	106	555.9	12	1	0
14	CON+GF+COMP	$\beta = 1$	65.9	102	548.2	12	1	0
14	CON+GF+COMP	$P_{max} = 102$	108.6	102	548.2	12	1	0
15	CON+GF+COMP	$\beta = 1$	76.5	136	582.4	18	1	0
15	CON+GF+COMP	$P_{max} = 136$	171.6	136	582.4	18	1	0
16	CON+GF+COMP	$\beta = 1$	229.6	120	573.8	17	1	0
16	CON+GF+COMP	$P_{max} = 120$	98.5	120	573.8	17	1	0
17	CON+GF+COMP	$\beta = 1$	78.7	94	499.4	12	1	0
17	CON+GF+COMP	$P_{max} = 94$	36.3	94	499.4	12	1	0

Table 5.4: Assessment of the computation time needed to enforce compactness whether using β or P_{max} constraints for 10 instances of 20×15 planning units and 3 conservation features. A summary of the characteristics of the reserve solutions is provided: computation time in seconds, reserve perimeter, total cost, radius of the reserve graph, number of connected components, number of gaps. Conservation features targets are all set to 50%.

Instance	Model	Parameters	Time	Perimeter	Cost	Radius	Components	Gaps
4	CON+GF+COMP	$\beta = 1$	139.0	214	961.8	21	1	0
4	CON+GF+COMP	$P_{max} = 214$	438.9	214	961.8	21	1	0
5	CON+GF+COMP	$\beta = 1$	TL[0.5%]	222	792.8	21	1	0
5	CON+GF+COMP	$P_{max} = 222$	TL[1.0%]	222	794.1	21	1	0
6	CON+GF+COMP	$\beta = 1$	577.4	224	910.8	23	1	0
6	CON+GF+COMP	$P_{max} = 224$	630.1	224	910.8	23	1	0
8	CON+GF+COMP	$\beta = 1$	50.8	164	915.9	21	1	0
8	CON+GF+COMP	$P_{max} = 164$	109.3	164	915.9	21	1	0
9	CON+GF+COMP	$\beta = 1$	542.6	192	999.5	21	1	0
9	CON+GF+COMP	$P_{max} = 192$	717.2	192	999.5	21	1	0
10	CON+GF+COMP	$\beta = 1$	97.6	184	927.8	18	1	0
10	CON+GF+COMP	$P_{max} = 184$	305.8	184	927.8	18	1	0
18	CON+GF+COMP	$\beta = 1$	TL[0.4%]	178	976.6	23	1	0
18	CON+GF+COMP	$P_{max} = 178$	TL[0.1%]	178	974.8	23	1	0
19	CON+GF+COMP	$\beta = 1$	109.2	202	956.2	23	1	0
19	CON+GF+COMP	$P_{max} = 202$	130.7	202	956.2	23	1	0
20	CON+GF+COMP	$\beta = 1$	895.1	184	969.5	20	1	0
20	CON+GF+COMP	$P_{max} = 184$	930.6	184	969.5	20	1	0
21	CON+GF+COMP	$\beta = 1$	388.3	166	905.8	18	1	0
21	CON+GF+COMP	$P_{max} = 166$	727.4	166	905.8	18	1	0

Table 5.5: Assessment of the computation time needed to enforce compactness whether using β or P_{max} constraints for 10 instances of 25×20 planning units and 3 conservation features. A summary of the characteristics of the reserve solutions is provided: computation time in seconds, reserve perimeter, total cost, radius of the reserve graph, number of connected components, number of gaps. Conservation features targets are all set to 50%. If the time limit is reached (TL=1000s), the optimality gap of the incumbent solution is given within brackets.

5.4 Discussion

In this work, we proposed a global integer linear program that produces compact, connected and gap-free reserves. We used a multicommodity flow model to enforce the connectivity of both the reserve and the non-reserve areas, resulting in connected and gap-free reserve solutions. We chose a multicommodity flow model rather than a single flow model in order to express the flow constraints as lazy constraints. Therefore, using Julia callbacks functions, a lot of constraints were not activated and improved the solving time. Using hard constraints instead took much more time or even made it impossible to find solutions within the time limit of 1000s in most cases. We also provided a reduction of the problem, benefiting from the rectangular structure of the grid. Indeed, due to the shape of the graph inferred from the rectangular grid, we applied a "chessboard" to separate the nodes into two sets. According to their position on the chessboard, a node was labelled black or white and belonged to the corresponding set. Then, we applied the multicommodity flow model only to the black nodes to decrease the size of the problem without any loss. We illustrated on the real-world instance of Fernando de Noronha that our model could already be used in practice. We numerically assessed the generality of the proposed approaches on several generated instances by deriving average characteristics. Our numerical experiments performed for generated instances showed that our model can be useful for instances composed of 300 to 500 planning units. We compared the solutions of our model with state-of-the-art model both in computation time and solution quality with respect to spatial requirements. The code used for this work is free, open and available. The model we proposed is highly customizable regarding the way we enforce compactness. A good setting of the β multiplier, the maximum perimeter or the maximum radius allowed to finely shape the reserve spatial attributes.

Instead of removing reserve solutions with gaps from the search space (Billionnet 2016), we used lazy flow constraints to *a priori* build a connected non-reserve. We did in this work what was advocated in the discussion of (Billionnet 2016) by including *a priori* a model to prevent the formation of gaps within reserves. We also differ from (Billionnet 2016) because we used a multicommodity flow model and lazy constraints to enforce connectivity in the minimum set problem. Using a multicommodity flow in reserve site selection problems is not new and was already mentioned in (Billionnet 2021) but it is a different approach compared to (Billionnet 2012; Billionnet 2016; Williams 2006).

The main limitation of our work is the limited size of instances that can be solved with our model in a reasonable time. The optimisation problem we proposed is still challenging computationally. This was expected because solving models that consider spatial constraints tends to be more computationally demanding in general, especially as the problem size increases (Wang et al. 2018). Although the size of the instances considered is of the same order of magnitude as what is usually done in a similar literature, these are still small instances compared to applications that did not focus on meeting spatial requirements. On the other hand, this could be mitigated by the fact that we can potentially provide solutions without the optimality proof. Our results showed that obtaining a reserve that is compact, connected and gap-free only involved a small increase of the site selection cost with respect to state-of-the-art models used in decision support tools such as Marxan and PrioritizR. The only price to pay is a greater computation time. Therefore, obtaining spatially coherent reserves is a more computational challenge rather than a question of limited resources with respect to the considered cost. Since optimal solutions are not so expensive relatively to state-of-the-art models, a compact, connected and gap-free reserve solution that is slightly suboptimal could be interesting. In addition, taking a suboptimal reserve solution is often tolerated in the conservation literature, especially when

metaheuristics are used to solve the problem.

The wide spreading and "cobweb" aspect of reserves within the study area is a common feature in models enforcing connectivity of the reserve (Önal and Briers 2006; Billionnet 2012). We mitigated this pitfall inherent to connectivity models by constraining the reserve radius and perimeter to remain below a predefined threshold. Our results also showed that including compactness constraints with the β multiplier and the associated constraints or the P_{max} constraint help the model to find a solution faster. In order to keep a single objective approach, it may be better practice to constrain a maximum perimeter of the reserve instead of using a penalty in the objective, but our results (*cf.* Section 5.3.3) illustrated that it makes solving slower for 500 planning units instances.

In this work, we did not address the problem of representing conservation features in two distinct reserve components, which is sometimes a desirable characteristic to be robust to catastrophic events (*e.g.* epidemic, fire). Imposing a minimum distance between two selected sites containing the same conservation feature (Williams 2006) was the method to achieve such property. It could improve the model to include such constraint, which can occasion future developments. Another feature we did not account for in this work is the potential different nature of locked-out planning units: some can be crossed, while others cannot. For example, in marine reserve design, a planning unit made of land is locked-out and cannot be crossed. On the other hand, a locked-out planning unit located in the harbour can be crossed. The main modelling implication of this difference is that a path ensuring connectivity in flow models cannot cross every locked-out planning units. This is not included in our current models and could occasion future developments.

Chapter 6

Discussion

6.1 Synthesis of the developments

The thesis work addressed the properties of decision support tools commonly used in marine spatial planning negotiations, in particular those implemented to optimally identify the location of conservation areas (*e.g.* marine protected areas). We first (Chapter 1) provided an overview of existing tools, the reasons for their coexistence, their formalism and functioning, and their respective strengths and weaknesses. Having established that the approximate solving optimisation tools (metaheuristics) had a mainly an historical *raison d'être* (to overcome computational problems that have now disappeared), we chose to focus on the exact solving optimisation tools to develop a certain number of their functionalities that were lacking up to now. We first proposed algorithms that allow to produce a diverse set of reserve solutions. While metaheuristic algorithms provide by construction a random sample of the solution space, without any indication of the distance to the optimum, nor any control on the real diversity of these solutions (which implies to produce many of them), we have proposed (Chapter 2) algorithms allowing to produce alternative solutions to the optimum, while explicitly controlling their distance to the optimum and maximizing the difference between these solutions. Our algorithms therefore allow the construction of a parsimonious portfolio of solutions (few solutions are needed since they are constructed to be truly different), whose relaxation from what would constitute a mathematically optimal solution can be explicitly controlled. Another important area of development seemed to us to be the consideration of uncertainty in this type of model. In existing applications, observational data are taken for absolute truth, whereas it is known, for example, that each instrument has a measurement variance. Similarly, the effects of seasonal variation, for example, although often known, are rarely taken into account in these static models. We have therefore proposed (Chapter 3) algorithms that make it possible to explicitly introduce the existing knowledge on the different types of uncertainty affecting the input data and to construct reserve solutions that are robust to these variations around the observation data. Finally, we were able to identify that a mathematically optimal solution is often not necessarily practical to manage in its geometry: problems of discontinuity, non-compactness, enclaves can appear in the solutions proposed by the models, and are generally removed in the real world by an *a posteriori* simplification of the geometry of the reserve. In Chapter 4, we proposed efficient and explicit numerical solutions to avoid the appearance of these phenomena if this is the wish of the stakeholders, thus avoiding hazardous and *a posteriori* simplifications of the reserve's shape.

6.2 Findings and limitations

In order to evaluate the advances provided by this thesis work, we perform the following thought exercise: imagine a working group whose mission is to identify reserve areas. Let us assume that this group is composed of end users of reserve selection tools with limited numerical expertise. These users could be fishers, tourism companies, non-governmental organisations, citizens, social scientists, policymakers, *etc.* Without knowing the content of this thesis, this working group, after some research, would most probably have turned to the best known tool, *i.e.* Marxan. After having studied the voluminous documentation of this tool, they would probably have followed the paid training offered by the Marxan team in order to be able to really use the tool on their real case. After this training and after a few weeks, they would surely manage to launch Marxan. After an empirical adjustment of the important parameters of Marxan (namely the “Specie Penalty Factor”, the “Boundary Length Modifier” and the number of desired solutions), they would have finally obtained a set of reserve solutions for the considered scenario. The full parameterisation of the simulated annealing algorithm requires a deep understanding of the tool, and would probably be left to its default parameterisation. Once they have obtained

the set of reserve solutions, they would have difficulty extracting useful information from this solution set and would surely be overwhelmed by the statistical analyses to be conducted. In particular, they may not be aware that within this solution set, it is not guaranteed that all solutions satisfy the targets. Thus, in the best case scenario, they would have to set up sorting and selection procedures *a posteriori*. In any case, they would have great difficulty in grasping the numerical issues involved in using such methods. In order to have a better understanding of these issues, at least empirically, they would surely have had to conduct experiments similar to those in Chapter 2 to see the influence of certain parameters/choices on their case study. As a conclusion of this thought experiment, it is very likely that this working group would have relied on blind belief in the tool, which can lead to nonsense and misinterpretations. This is confirmed by published results containing errors and our experience in using this tool. We think in particular of the edge effect involved by the use of the penalty of the perimeter in the objective function. This undesirable numerical effect seems to be a common mistake even in published literature (Delavenne et al. 2012; Beyer et al. 2016; Magris et al. 2021).

The same working group, in the light of this work, would understand that the use of metaheuristic algorithms is not necessarily the only option and that this choice implies strong numerical issues. The use of exact optimisation methods instead is surely preferable. Although the problem to be solved is the same, the formalism of metaheuristics does not allow one to immediately grasp what the algorithm actually does. It is often too complex, and this is even more true for an end user with limited numerical expertise. Paradoxically, these tools are often implemented by users who do not have the technical expertise to understand numerical issues, as the survey in (Ardron et al. 2010) attests. On the other hand, understanding what an optimal solution is and what it corresponds to is extremely simple and straightforward: it is the least expensive solution that meets the predefined targets. For metaheuristics, it is much less clear: there are several solutions (the user sets the number of solutions) which may be cheap (the quality of the solutions, *i.e.* the difference from the minimum cost, is not known nor controlled) and which may meet the predefined targets (depending on the setting of the "Specie Penalty Factor" parameter, it is possible to have a large number of solutions which do not meet the targets). Exact optimisation methods are easier to learn, equally fast in computation time, and leave less room for empirical parameter settings whose meaning is not intuitive. Second, exact optimisation provides a single solution, which makes it easier to analyse and opens the door to multiple scenarios rather than multiple solutions to the same scenario.

And if one wishes to generate a set of distinct solutions, this is still possible in the light of Chapter 3. Moreover, in the case where the data contain uncertainty, we have provided methods in Chapter 4 which allow, without changing the theoretical framework, to explicitly specify this uncertainty and to compute reserve solutions which take it into account. Similarly, if we want to enforce our solutions to be spatially coherent, the models presented in Chapter 5 allow this. Figure 6.1 and Figure 6.2 provide a graphical summary of both the state-of-the-art models and the algorithmic developments achieved in this thesis. This thesis proposes tools using exact optimisation methods that today have no reason not to be used in comparison to metaheuristic algorithms as used in Marxan. A major limitation of the proposed methods compared to Marxan is that they are not yet available as an off-the-shelf tool. In order to overcome this limitation, it would be desirable to disseminate our developments by implementing them in existing tools such as PrioritizR. But before this can be possible, the numerical evaluation of the methods is necessarily missing. This numerical evaluation would consist of solving the model considered on many instances in order to draw the numerical limits of these models.

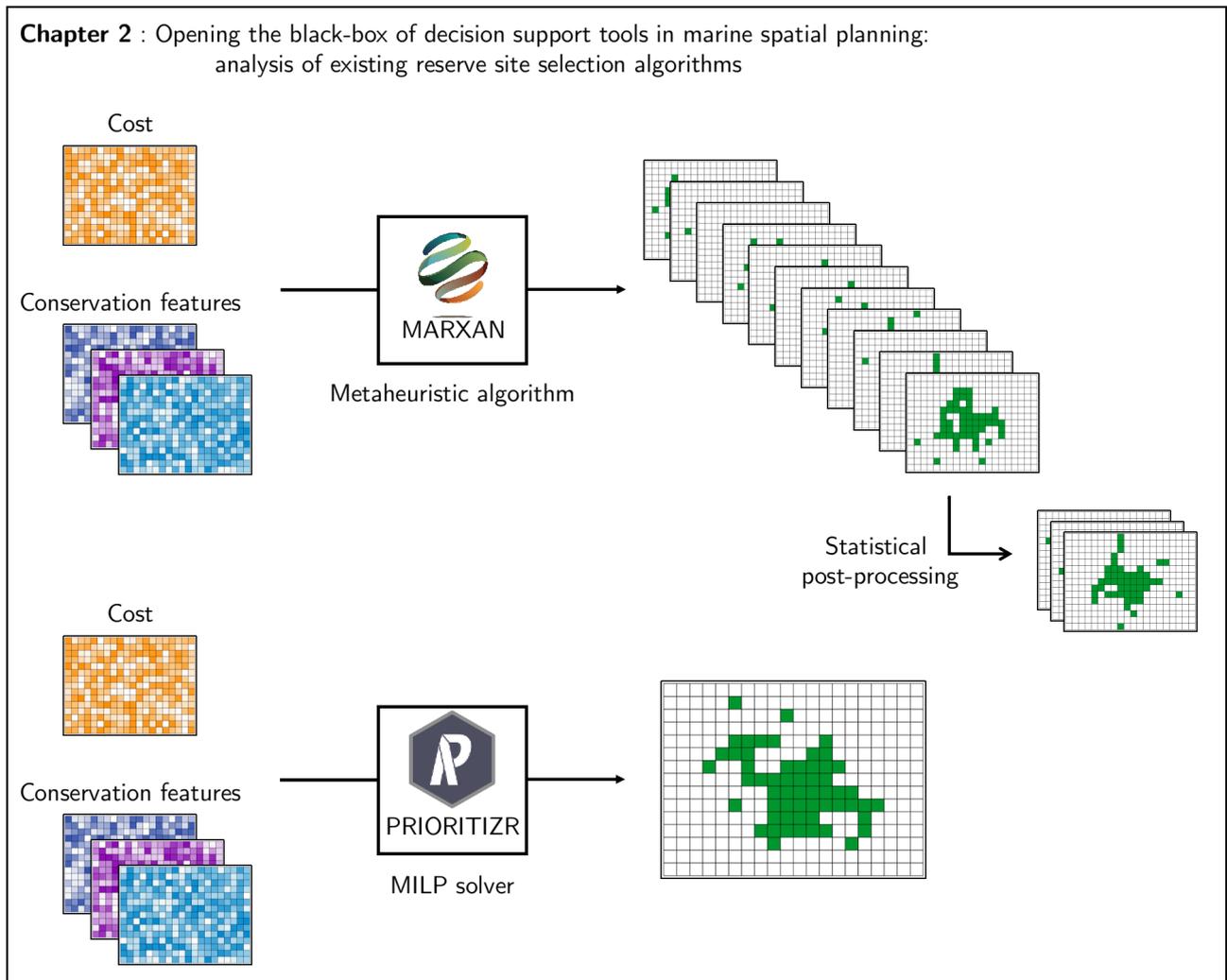


Figure 6.1: Graphical summary of the state-of-the-art tools used for reserve site selection. Marxan deploys a metaheuristic algorithm to build a user-defined number of reserve solutions that need a statistical post-processing. PrioritizR deploys exact optimisation methods to find the least-cost solution to the reserve site selection problem.

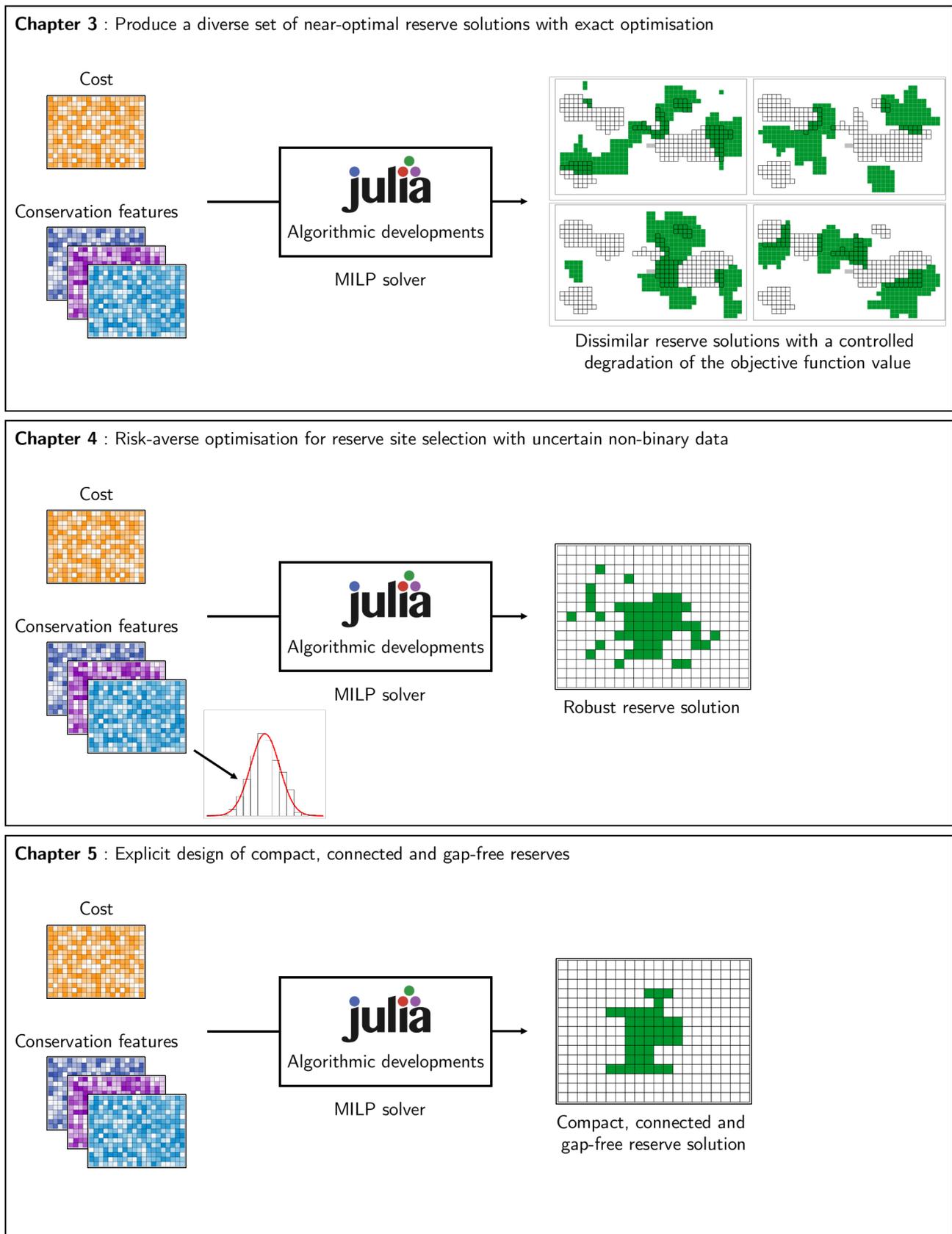


Figure 6.2: Graphical summary of the developments of the thesis. Chapter 3 provides exact optimisation methods to derive a diverse set of near-optimal solutions that are dissimilar between them and with a predefined degradation of the objective function value. Chapter 4 provides two approaches using exact optimisation to deal with an uncertainty that affects non-binary data of the reserve site selection problem. Chapter 5 provides exact optimisation methods that build compact, connected and gap-free reserves. The associated Julia code is open, free and available at <https://github.com/AdrienBrunel>. MILP = mixed integer linear programming.

6.3 Perspectives

In this thesis, we developed ideas and methods for the problem of reserve site selection. We could consider that solving the reserve site selection problem is the same as allocating the conservation human use. We could have done the same with another use and this is even the more general purpose of MSP. Indeed, the ultimate goal of marine spatial planning is to provide a spatial zoning to say which human activities can exist in which areas. For example, instead of selecting reserve sites, we can select locations for offshore wind turbines. Our methods can be a thinking basis for a more general spatial zoning, not only focused on conservation.

6.3.1 A different example of human use allocation: site selection of offshore wind farms

The multiplication of offshore wind turbines naturally raised the question of finding the most relevant areas for this activity at sea. Finding the optimal collection of planning units to deploy an offshore wind turbine is mathematically the same problem as finding the optimal collection of planning units to include within a reserve. In collaboration with Créocéan, we did the exercise to show how we can build an optimisation problem for this different need. Therefore, we proposed Créocéan to address the offshore wind turbine location problem through an integer programming framework. The only changing elements are the objective function and the constraints that must be adapted for the offshore wind turbine locations problem.

This problem is usually addressed through a multi-criteria decision-making (MCDM) formalism, since the feasibility and quality of sites are evaluated according to several criteria (Van Haaren and Fthenakis 2011; Kim et al. 2016; Rodríguez-Rodríguez et al. 2016). The choice of sites is mainly based on ad-hoc methods or heuristics. While few works exist (Mytilinou and Kolios 2017; Mytilinou et al. 2018), the optimisation formalism for the problem of wind farms site selection is not canonical yet. The fact that optimisation is not the main conceptual framework can be justified in the sense that many feasibility constraints does not leave that much choice for the location of wind farms. Optimisation models are sometimes used in this research area (Salcedo-Sanz et al. 2014; Serrano González et al. 2014), yet it is usually for finding the best wind turbine arrangement within the wind farm (known as the wind farm layout optimisation problem).

We worked with Créocéan to provide a preliminary example of wind farms site selection using the methods developed in this thesis. In this collaboration, the study area was discretised using 1.2 km square pixels. This distance represents the safety distance around an offshore wind turbine. The available data provided by Créocéan was levels of constraint represented by an integer in $\{0, 1, 2, 3\}$ and an average wind speed for every planning unit. The constraint were of various nature (Natura 2000 areas, military zone, ocean depth, fishing areas, maritime routes, *etc.*). The strongest level of constraint 3 indicated that the planning unit cannot host an offshore wind turbine. We removed every planning unit that had a level 3 for any constraint considered, which decrease the size of the problem. We imposed budgets t_{c_i} for every constraint c_i . We also imposed a maximum number of wind turbines n_e . The objective we chose to maximise was the total average wind speed v . Similarly, as in reserve site selection, we included the wind turbines selection perimeter in the objective to favour compact selection. We proposed the following optimisation model, inspired from reserve site selection models, and that can be adapted for other human use allocation:

$$\begin{cases} \max_{x \in \{0,1\}^N} & v^T x - \beta x^T B(1 - x) \\ \text{s.t.} & c_i^T x \leq t_{c_i} \\ & 1^T x \leq n_e \end{cases} \quad \forall i \in I \quad (6.1)$$

6.3.2 Toward a spatial zoning of human activities

The optimisation framework is by nature single objective, while the issues of MSP are inherently multi-objective. In order to overcome this limitation of our methods, it is advised to multiply the scenarios by changing the human activity considered to eventually confront the solutions between scenarios. However, the multi-objective optimisation literature could shed a new light on the reserve site selection problem (Basirati et al. 2021; Fox et al. 2019).

The literature for the spatial allocation of multiple uses is mainly based on the Marxan with Zones tool, implementing a metaheuristic approach (Watts et al. 2009). This tool has led to direct applications taking into account several human activities (fishing, transport, tourism, conservation, *etc.*) (White et al. 2012; Agostini et al. 2015; Markantonatou et al. 2021; Jumin et al. 2018). Taking into account ecological data and fisheries models allows an interesting application with the designation of "no take", "limited take" and "take" zones (Metcalfe et al. 2015). In line with the objectives of MSP, a conflict index can be explicitly minimised according to present or future conditions (Tuda et al. 2014). The outputs of metaheuristic approaches may require heavy statistical processing (Markantonatou et al. 2021). Also, the comparison between the outputs of the decision support tools and the decisions actually taken highlights the gap that still exists between the tools and reality (Agostini et al. 2015). The implementation of an exact solving for a spatial allocation problem is relatively rare in the literature. Indeed, a model solved by exact methods is developed in (Basirati et al. 2021) in order to allocate planning units to a human activity, the other existing human activities acting as constraints. The multi-objective optimisation model seeks to maximise the compactness of the allocation and an overall value of interest of the considered human activity.

We propose two development perspectives:

- A global approach: we directly seek for the best allocation of all considered uses. This approach can potentially be numerically difficult. A conceptual diagram is provided in Figure 6.3.
- An iterative approach: based on an explicit *a priori* preference (*e.g.* fishing > conservation > wind turbines), we successively allocate a use among the available planning units (*i.e.* planning units that were not already allocated). This approach should be more accessible numerically because it is a simplification of the global approach. A conceptual diagram is provided in Figure 6.4.

We think reasonable to assume that we can derive indexes of socio-ecological impacts and socio-economic gains such as:

- The socio-ecological impact $c_j^k \in \mathbb{R}$ of use $k \in K$ if allocated to the planning unit $j \in J$.
- The socio-economic gain $g_j^k \in \mathbb{R}_+$ of use $k \in K$ if allocated to the planning unit $j \in J$.

Let us consider M uses that seek to be deployed within a grid composed of N planning units. Let $K = \llbracket 1, M \rrbracket$ and $J = \llbracket 1, N \rrbracket$. Let $x^k \in \{0, 1\}^N$ the decision variable associated with the allocation of use $k \in K$. Use $k \in K$ is allocated to planning unit $j \in J$ if $x_j^k = 1$. We seek to find the "best" allocation $X = [x^1, x^2, \dots, x^M] \in \{0, 1\}^{N \times M}$.

We represent the set of admissible values \mathcal{X}^k for the decision variable x^k associated with the allocation of use $k \in K$. The compatibility between every pair of uses, *i.e.* if two uses can be allocated to the same planning unit, can be modelled by linear constraints. The constraint forbidding the allocation of more than one use within a planning unit can be expressed

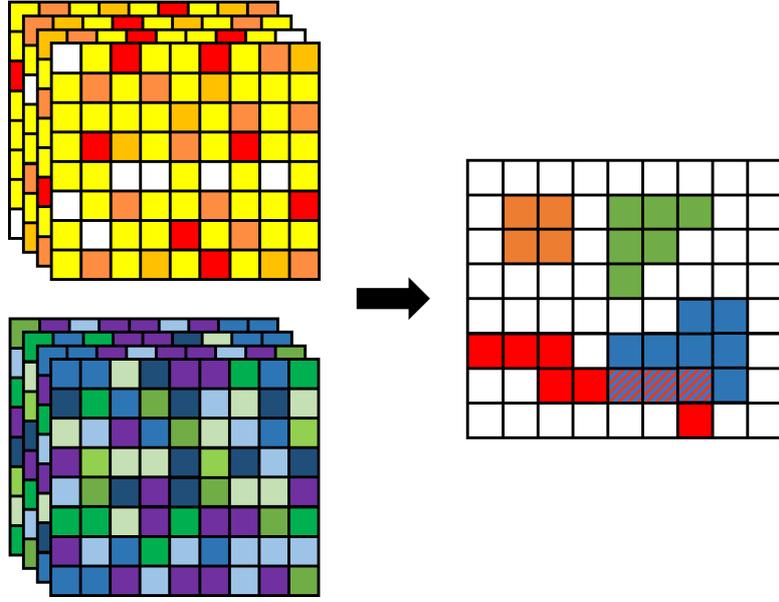


Figure 6.3: Conceptual diagram of the global approach. The allocation of uses are represented with orange, blue, red and green planning units. These planning units are allocated based on cost and gains layers. In this diagram, blue and red uses are considered compatible.

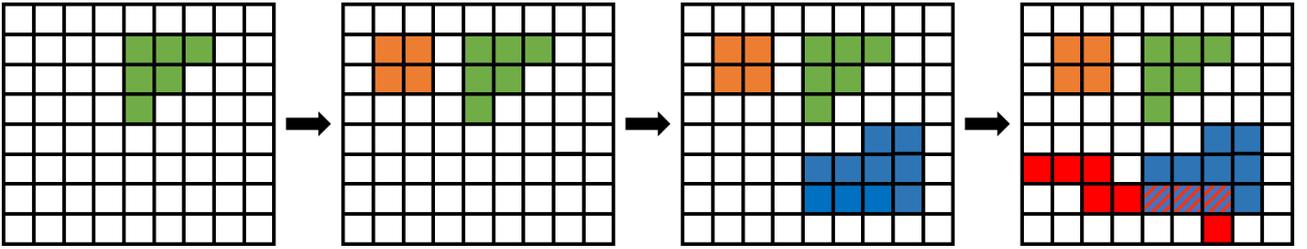


Figure 6.4: Conceptual diagram of the iterative approach. We assume that a preference is explicitly expressed between uses. In this case, we first allocate the green use, then the orange use, then the blue use and finally the red use. In this diagram, blue and red uses are considered compatible.

$\sum_{k \in K} x_j^k \leq 1$ and this $\forall j \in J$. Let us imagine that two activities k_1 et k_2 can be allocated in the same planning units, we can then write $\sum_{k \in K - \{k_1\}} x_j^k \leq 1$ et $\sum_{k \in K - \{k_2\}} x_j^k \leq 1$.

Regarding the models, we can express an objective that maximises the total gain of the allocation. The objective is then to maximise $\sum_{k \in K} \omega_k \sum_{j \in J} g_j^k x_j^k$. The corresponding model can be found in Equation (6.2). This formulation assume that the gains involved by the allocation of uses will contribute to a common total gain represented with a weighted sum. If we change our perspective, we could also seek to the allocation of uses that minimises the socio-ecological impacts. The corresponding model can be found in Equation (6.3). One can imagine putting negative c_j^k coefficients to represent the positive impact of a reserve, for instance.

A gain-oriented approach:

$$\left\{ \begin{array}{ll} \max_{x^1, \dots, x^M} & \sum_{k \in K} \omega_k \sum_{j \in J} g_j^k x_j^k \\ \text{s.t.} & x^k \in \mathcal{X}^k \quad \forall k \in K \\ & \sum_{k \in K} x_j^k \leq 1 \quad \forall j \in J \end{array} \right. \quad (6.2)$$

An impact-oriented approach:

$$\left\{ \begin{array}{l} \min_{x^1, \dots, x^M} \quad \sum_{k \in K} \sum_{j \in J} c_j^k x_j^k \\ \text{s.t.} \quad x^k \in \mathcal{X}^k \quad \forall k \in K \\ \sum_{k \in K} x_j^k \leq 1 \quad \forall j \in J \\ \sum_{j \in J} g_j^k x_j^k \geq t_k \quad \forall k \in K \end{array} \right. \quad (6.3)$$

In an iterative gain-oriented approach, and if we name $\hat{x}^1, \dots, \hat{x}^{k-1}$ the allocations of uses computed the iterations before, we have at iteration $k \in K$ the following optimisation problem:

$$\left\{ \begin{array}{l} \max_{x^k} \quad \sum_{j \in J} g_j^k x_j^k \\ \text{s.t.} \quad x^k \in \mathcal{X}^k \quad \forall k \in K \\ x_j^k \leq 1 - \hat{x}_j^l \quad \forall l \in \{1, 2, \dots, k-1\} \end{array} \right. \quad (6.4)$$

6.4 Ethical positioning

Considering marine space as an available and allocable resource is a strong ideological position in itself. The ideas that support sustainable development (and therefore MSP) are derived from the conservationist environmental ethics of Gifford Pinchot. However, we have seen that this position is at the extreme of a continuum of environmental ethics and that other paths are possible. Choosing MSP means accepting the idea of a possible sustainable development because there is no possibility in these tools to question human activities *a priori*. The "do nothing" scenario can be the most favourable scenario, especially when the knowledge around a study area is patchy. However, this scenario is not possible with the tools we are dealing with, since an activity is necessarily allocated to a space as soon as these tools are used. While the consequences on ecosystems are sometimes irreversible, the development of human activities continues to grow, as attested by the attractiveness of the blue economy. This is consistent with our economic models since they are based on a growth obligation. Nature reserves must not become an artefact to justify the constant and increasing exploitation of the environment. The advent of conservation policies should not be a signal that the solution to biodiversity loss has been identified but that the limits have been reached and that the only barrier to mitigate global change is to create areas without human pressure. The debate around the effectiveness of these zones can remain but it must be concomitant with that of the possibility of prohibiting the development of human activities for the sake of ecosystems.

All of this thesis work has been centred around the problem of reserve selection, but the final issue is indeed political and social. Although reserve selection tools only address a fraction of the MSP problem, we have been able to see the numerical issues that can exist. Thus, the proposed methods can be seen as providing technical improvements leading to better decision making. But this work as a whole allows us to increase transparency in the decision-making process, to clarify grey areas, to make choices more explicit, to simplify and make more intuitive the understanding of how these tools work. Indeed, more transparency means more equity in a discussion that brings together actors with heterogeneous levels of power in a MSP process. Overall, whether it is a question of constructing a set of truly different solutions, taking into account uncertainty or the spatial attributes of the reserves, we have been able to show that our needs can be explicitly taken into account in exact and therefore transparent integer optimisation models. Transparency leads to more fairness in negotiations and tends to reduce epistemic injustice, more specifically the hermeneutical injustice (Fricker 2007; McKinnon 2016). Knowledge, intentionally or unintentionally, is unevenly distributed and this leads

to powers imbalance: we speak of hermeneutical injustice. Those who lack resources to produce and access knowledge will be disadvantaged in a situation where this knowledge is involved, *e.g.* in MSP negotiations. Bringing more transparency is bringing knowledge closer to actors that can be victim of this epistemic injustice. It is therefore in the hope of rebalancing discussions, giving more control and understanding to end-users, regardless of their numerical expertise, and thus increasing the chances of practical implementation that this work makes sense.

The proposed approach is mathematical, whereas the issues are social and political. Mathematics do not make disappear power issues. They are crossed by these. Mathematical tools can even amplify the power imbalance between stakeholders (Queffelec et al. 2021). For example, it is reasonable to imagine that those who are able to understand or even produce these tools will have an important advantage in negotiations. The solution is not more "objective" and "rational" because it is derived from a mathematical tool. Despite our efforts to make the methods transparent and more equitable, these tools cannot be completely immune to these power relations. At best, we can only make these issues transparent, force decisions to be arguable, understandable by all so that the decisions taken tend to be more equitable.

Reserve selection tools could themselves be questioned, since these mathematical tools are often in the hands of users who rarely have the expertise to detect numerical problems (such as those induced by high "Boundary Length Modifier" values in Marxan, observed in the published literature). Therefore, there may be a temptation to place blind trust in tools that promise mathematical truth. But, an optimal solution is completely dependent on the data that shapes it as well as the formulation from which it is derived. As shown in Chapter 2, the solution of the "MaxCov" formulation is not the same as the solution of the "Minset" formulation. And we have only shown two formulations, but others exist.

References

- Abdelmaguid, T. (2018). “An Efficient Mixed Integer Linear Programming Model for the Minimum Spanning Tree Problem”. en. In: *Mathematics* 6.10, p. 183. DOI: [10.3390/math6100183](https://doi.org/10.3390/math6100183).
- Agostini, V. N. et al. (2015). “Marine zoning in St. Kitts and Nevis: A design for sustainable management in the Caribbean”. en. In: *Ocean & Coastal Management* 104, pp. 1–10. DOI: [10.1016/j.ocecoaman.2014.11.003](https://doi.org/10.1016/j.ocecoaman.2014.11.003).
- Ando, A. et al. (1998). “Species Distributions, Land Values, and Efficient Conservation”. In: *Science* 279.5359, pp. 2126–2128.
- Araújo, M. B. and P. H. Williams (2000). “Selecting areas for species persistence using occurrence data”. en. In: *Biological Conservation* 96.3, pp. 331–345. DOI: [10.1016/S0006-3207\(00\)00074-4](https://doi.org/10.1016/S0006-3207(00)00074-4).
- Ardron, J. A., H. P. Possingham, and Carissa J. Klein (2010). *Marxan Good Practices Handbook*.
- Arthur, J. L., M. Hachey, and K. Sahr (1997). “Finding all optimal solutions to the reserve site selection problem: formulation and computational analysis”. en. In: *Environmental and Ecological Statistics*, p. 13.
- Baer, J. G. (1968). “Aperçu historique de la protection de la nature”. fr. In: *Biological Conservation* 1.1, pp. 7–11. DOI: [10.1016/0006-3207\(68\)90004-9](https://doi.org/10.1016/0006-3207(68)90004-9).
- Baillie, J. and Y-P. Zhang (2018). “Space for nature”. en. In: *Science* 361.6407, pp. 1051–1051. DOI: [10.1126/science.aau1397](https://doi.org/10.1126/science.aau1397).
- Ball, I. R. and H. P. Possingham (2000). “Marine Reserve Design using Spatially Explicit Annealing”. en. In: p. 70.
- Ball, I. R., H. P. Possingham, and M. E. Watts (2009). “Marxan and relatives: software for spatial conservation prioritisation”. In: *Spatial Conservation Prioritisation: Quantitative Methods and Computational Tools*. Vol. 14, pp. 185–196.
- Basirati, M. et al. (2021). “Exact Zoning Optimization Model for Marine Spatial Planning (MSP)”. en. In: *Frontiers in Marine Science* 8, p. 726187. DOI: [10.3389/fmars.2021.726187](https://doi.org/10.3389/fmars.2021.726187).
- Bayon, R. and M. Jenkins (2010). “The business of biodiversity”. In: *Nature* 446, pp. 184–185.
- Beech, T. et al. (2008). “A stochastic approach to marine reserve design: Incorporating data uncertainty”. en. In: *Ecological Informatics* 3.4-5, pp. 321–333. DOI: [10.1016/j.ecoinf.2008.09.001](https://doi.org/10.1016/j.ecoinf.2008.09.001).
- Bell, D., T. Gray, and C. Haggett (2005). “The ‘Social Gap’ in Wind Farm Siting Decisions: Explanations and Policy Responses”. en. In: *Environmental Politics* 14.4, pp. 460–477. DOI: [10.1080/09644010500175833](https://doi.org/10.1080/09644010500175833).
- Beraldi, P. and A. Ruszczyński (2002). “The Probabilistic Set-Covering Problem”. en. In: *Operations Research* 50.6, pp. 956–967. DOI: [10.1287/opre.50.6.956.345](https://doi.org/10.1287/opre.50.6.956.345).
- Bertrand, A. (2019). “FAROFA 3 cruise, RV TUBARAO Tigre”. In: Publisher: Simer. DOI: [10.17600/18001381](https://doi.org/10.17600/18001381).
- Bertsimas, D. and M. Sim (2004). “The Price of Robustness”. en. In: *Operations Research* 52.1, pp. 35–53. DOI: [10.1287/opre.1030.0065](https://doi.org/10.1287/opre.1030.0065).
- Beyer, H. L. et al. (2016). “Solving conservation planning problems with integer linear programming”. en. In: *Ecological Modelling* 328, pp. 14–22. DOI: [10.1016/j.ecolmodel.2016.02.005](https://doi.org/10.1016/j.ecolmodel.2016.02.005).
- Bez, N. (2021). *Global estimation based on indicators factorization*. en. Tech. rep., p. 7.

- Bez, N. and C-B. Braham (2014). “Indicator variables for a robust estimation of an acoustic index of abundance”. en. In: *Canadian Journal of Fisheries and Aquatic Sciences* 71.5. Ed. by J. M. Jech, pp. 709–718. DOI: [10.1139/cjfas-2013-0437](https://doi.org/10.1139/cjfas-2013-0437).
- Bezanson, J. et al. (2012). “Julia: A Fast Dynamic Language for Technical Computing”. en. In: *arXiv:1209.5145 [cs]*. arXiv: 1209.5145.
- Bezanson, J. et al. (2015). “Julia: A Fresh Approach to Numerical Computing”. en. In: *arXiv:1411.1607 [cs]*. arXiv: 1411.1607.
- Billionnet, A. (2012). “Designing an optimal connected nature reserve”. en. In: *Applied Mathematical Modelling* 36.5, pp. 2213–2223. DOI: [10.1016/j.apm.2011.08.002](https://doi.org/10.1016/j.apm.2011.08.002).
- (2016). “Designing Connected and Compact Nature Reserves”. en. In: *Environmental Modelling & Assessment* 21.2, pp. 211–219. DOI: [10.1007/s10666-015-9465-3](https://doi.org/10.1007/s10666-015-9465-3).
- (2021). *Designing protected area networks: a mathematical programming approach*. en. OCLC: 9463862519.
- (2013). “Mathematical optimization ideas for biodiversity conservation”. en. In: *European Journal of Operational Research*, p. 21.
- (2007). *Optimisation Discrète, de la modélisation à la résolution par des logiciels de programmation mathématique*. Dunod.
- Bolderdijk, J. W. et al. (2012). “Comparing the effectiveness of monetary versus moral motives in environmental campaigning”. In: *Nature climate change* 3, pp. 413–416.
- Bookchin, M. (1971). *Post-scarcity anarchism*.
- (1987). “Social Ecology versus Deep Ecology: A Challenge for the Ecology Movement”. In: *Green Perspectives: Newsletter of the Green Program Project* 4-5.
- (1982). *The ecology of freedom*.
- Brennan, A. and Y-S. Lo (2002). “Environmental Ethics”. In: *The Stanford Encyclopedia of Philosophy*.
- Brill, E., S-Y. Chang, and L. D. Hopkins (1982). “Modeling to Generate Alternatives: The HSJ Approach and an Illustration Using a Problem in Land Use Planning”. en. In: *Management Science* 28.3, pp. 221–235. DOI: [10.1287/mnsc.28.3.221](https://doi.org/10.1287/mnsc.28.3.221).
- Brouwer, H. et al. (2016). *The MSP Guide: How to Design and Facilitate Multi-Stakeholder Partnerships*. en. Practical Action Publishing. DOI: [10.3362/9781780446691](https://doi.org/10.3362/9781780446691).
- Cabeza, M. et al. (2004). “Combining probabilities of occurrence with spatial reserve design”. en. In: *Journal of Applied Ecology* 41.2, pp. 252–262. DOI: [10.1111/j.0021-8901.2004.00905.x](https://doi.org/10.1111/j.0021-8901.2004.00905.x).
- Carson, R. (1962). *Silent Spring*.
- Carvalho, S. B. et al. (2011). “Conservation planning under climate change: Toward accounting for uncertainty in predicted species distributions to increase confidence in conservation investments in space and time”. en. In: *Biological Conservation* 144.7, pp. 2020–2030. DOI: [10.1016/j.biocon.2011.04.024](https://doi.org/10.1016/j.biocon.2011.04.024).
- Carvalho, S. B. et al. (2010). “Simulating the effects of using different types of species distribution data in reserve selection”. en. In: *Biological Conservation* 143.2, pp. 426–438. DOI: [10.1016/j.biocon.2009.11.010](https://doi.org/10.1016/j.biocon.2009.11.010).
- Chang, S-Y., E. D. Brill, and L. D. Hopkins (1982a). “Efficient random generation of feasible alternatives: a land use example”. en. In: *Journal of Regional Science* 22.3, pp. 303–314. DOI: [10.1111/j.1467-9787.1982.tb00754.x](https://doi.org/10.1111/j.1467-9787.1982.tb00754.x).
- (1982b). “Use of mathematical models to generate alternative solutions to water resources planning problems”. en. In: *Water Resources Research* 18.1, pp. 58–64. DOI: [10.1029/WR018i001p00058](https://doi.org/10.1029/WR018i001p00058).
- Chiles, J-P. and P. Delfiner (2012). *Geostatistics: modeling spatial uncertainty*. 2nd ed. Wiley series in probability and statistics. Hoboken, N.J: Wiley.

-
- Church, R. L., D. M. Stoms, and F. W. Davis (1996). “Reserve selection as a maximal covering location problem”. en. In: *Biological Conservation* 76.2, pp. 105–112. DOI: [10.1016/0006-3207\(95\)00102-6](https://doi.org/10.1016/0006-3207(95)00102-6).
- Claudet, J. et al. (2008). “Marine reserves: size and age do matter”. en. In: *Ecology Letters* 11.5, pp. 481–489. DOI: [10.1111/j.1461-0248.2008.01166.x](https://doi.org/10.1111/j.1461-0248.2008.01166.x).
- Claudet, J. et al. (2020). “Underprotected Marine Protected Areas in a Global Biodiversity Hotspot”. en. In: *One Earth* 2.4, pp. 380–384. DOI: [10.1016/j.oneear.2020.03.008](https://doi.org/10.1016/j.oneear.2020.03.008).
- Clemens, M. A., C. S. ReVelle, and J. C. Williams (1999). “Reserve design for species preservation”. en. In: *European Journal of Operational Research*, p. 11.
- Cocks, K.D. and I.A. Baird (1989). “Using mathematical programming to address the multiple reserve selection problem: An example from the Eyre Peninsula, South Australia”. en. In: *Biological Conservation* 49.2, pp. 113–130. DOI: [10.1016/0006-3207\(89\)90083-9](https://doi.org/10.1016/0006-3207(89)90083-9).
- Coleman, H. et al. (2011). *Decision Guide : Selecting Decision Support Tools for Marine Spatial Planning*. Tech. rep. Center for Ocean Solution.
- Commission, European (2020). *EU Biodiversity Strategy for 2030, Bringing nature into our lives, COM(2020) 380 final, Brussels, 20.05.2020*. Tech. rep.
- Csuti, B. et al. (1997). “A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon”. en. In: *Biological Conservation* 80.1, pp. 83–97. DOI: [10.1016/S0006-3207\(96\)00068-7](https://doi.org/10.1016/S0006-3207(96)00068-7).
- Delavenne, J. et al. (2012). “Systematic conservation planning in the eastern English Channel: comparing the Marxan and Zonation decision-support tools”. en. In: *ICES Journal of Marine Science* 69.1, pp. 75–83. DOI: [10.1093/icesjms/fsr180](https://doi.org/10.1093/icesjms/fsr180).
- Diamond, J. M. (1975). “The island dilemma: Lessons of modern biogeographic studies for the design of natural reserves”. en. In: *Biological Conservation* 7.2, pp. 129–146. DOI: [10.1016/0006-3207\(75\)90052-X](https://doi.org/10.1016/0006-3207(75)90052-X).
- Dinerstein, E. et al. (2019). “A Global Deal For Nature: Guiding principles, milestones, and targets”. en. In: *Science Advances* 5.4, eaaw2869. DOI: [10.1126/sciadv.aaw2869](https://doi.org/10.1126/sciadv.aaw2869).
- Doak, D. F. et al. (2014). “What is the future of conservation?” en. In: *Trends in Ecology & Evolution* 29.2, pp. 77–81. DOI: [10.1016/j.tree.2013.10.013](https://doi.org/10.1016/j.tree.2013.10.013).
- Douvere, F. (2008). “The importance of marine spatial planning in advancing ecosystem-based sea use management”. en. In: *Marine Policy* 32.5, pp. 762–771. DOI: [10.1016/j.marpol.2008.03.021](https://doi.org/10.1016/j.marpol.2008.03.021).
- Dira, S. et al. (2019). “Species-area uncertainties impact the setting of habitat conservation targets and propagate across conservation solutions”. en. In: *Biological Conservation* 235, pp. 279–289. DOI: [10.1016/j.biocon.2019.05.012](https://doi.org/10.1016/j.biocon.2019.05.012).
- Dunning, I., J. Huchette, and M. Lubin (2017). “JuMP: A Modeling Language for Mathematical Optimization”. en. In: *SIAM Review* 59.2. arXiv: 1508.01982, pp. 295–320. DOI: [10.1137/15M1020575](https://doi.org/10.1137/15M1020575).
- Ehler, C. N. and F. Douvere (2009). “Marine Spatial Planning, A Step-by-Step Approach toward Ecosystem-based Management”. In: *Intergovernmental Oceanographic Commission and Man and the Biosphere Programme IOC Manual and Guides No. 53 ICAM Dossier No. 6*. UNESCO, Paris, pp. 99.
- Ehrenfel, D. W. (1976). “The Conservation of Non-Resource”. In: *American Scientist* 64.6, pp. 648–656.
- European Commission and Directorate-General for Maritime Affairs and Fisheries (2019). *The EU Blue Economy Report 2019*. English. OCLC: 1111106476.
- European Commission. Directorate General for Maritime Affairs and Fisheries. (2020). *The EU Blue Economy Report 2020*. eng. LU: Publications Office.
- Faludi, A. and B. Waterhout (2006). “Introducing Evidence-Based Planning”. en. In: *disP - The Planning Review* 42.165, pp. 4–13. DOI: [10.1080/02513625.2006.10556950](https://doi.org/10.1080/02513625.2006.10556950).

- Fernandes, L. et al. (2005). “Establishing Representative No-Take Areas in the Great Barrier Reef: Large-Scale Implementation of Theory on Marine Protected Areas”. en. In: *Conservation Biology* 19.6, pp. 1733–1744. DOI: [10.1111/j.1523-1739.2005.00302.x](https://doi.org/10.1111/j.1523-1739.2005.00302.x).
- Ferretti, V., I. Pluchinotta, and A. Tsoukiàs (2019). “Studying the generation of alternatives in public policy making processes”. en. In: *European Journal of Operational Research* 273.1, pp. 353–363. DOI: [10.1016/j.ejor.2018.07.054](https://doi.org/10.1016/j.ejor.2018.07.054).
- Fiorella, K. et al. (2010). “Methodological considerations in reserve system selection: A case study of Malagasy lemurs”. en. In: *Biological Conservation* 143.4, pp. 963–973. DOI: [10.1016/j.biocon.2010.01.005](https://doi.org/10.1016/j.biocon.2010.01.005).
- Fischer, D. T. and R. L. Church (2005). “The SITES reserve selection system: A critical review”. en. In: *Environmental Modeling & Assessment* 10.3, pp. 215–228. DOI: [10.1007/s10666-005-9005-7](https://doi.org/10.1007/s10666-005-9005-7).
- Flannery, W. et al. (2020). “A critical turn in marine spatial planning”. en. In: *Maritime Studies* 19.3, pp. 223–228. DOI: [10.1007/s40152-020-00198-8](https://doi.org/10.1007/s40152-020-00198-8).
- Flower, J. et al. (2020). “Marine spatial planning on the Caribbean island of Montserrat: Lessons for data-limited small islands”. en. In: *Conservation Science and Practice* 2.4. DOI: [10.1111/csp2.158](https://doi.org/10.1111/csp2.158).
- Foley, M. M. et al. (2010). “Guiding ecological principles for marine spatial planning”. en. In: *Marine Policy* 34.5, pp. 955–966. DOI: [10.1016/j.marpol.2010.02.001](https://doi.org/10.1016/j.marpol.2010.02.001).
- Forrest, J. et al. (2018). *Coin-Or/Cbc: Version 2.9.9*. DOI: [10.5281/ZENODO.1317566](https://doi.org/10.5281/ZENODO.1317566).
- Fox, A. D. et al. (2019). “An Efficient Multi-Objective Optimization Method for Use in the Design of Marine Protected Area Networks”. en. In: *Frontiers in Marine Science* 6, p. 17. DOI: [10.3389/fmars.2019.00017](https://doi.org/10.3389/fmars.2019.00017).
- Fricker, M. (2007). *Epistemic Injustice*. Oxford University Press. DOI: [10.1093/acprof:oso/9780198237907.001.0001](https://doi.org/10.1093/acprof:oso/9780198237907.001.0001).
- Game, E. T. and H. S. Grantham (2008). *Marxan User Manual For Marxan version 1.8.10*.
- Game, E. T. et al. (2008). “Planning for persistence in marine reserves: a question of catastrophic importance”. en. In: *Ecological Applications* 18.3, pp. 670–680. DOI: [10.1890/07-1027.1](https://doi.org/10.1890/07-1027.1).
- Garey, M. R. and D. S. Johnson (1979). *Computers and intractability: a guide to the theory of NP-completeness*. en. 27. print. A series of books in the mathematical sciences. New York [u.a]: Freeman.
- Gehlbach, F. R. (1975). “Investigation, evaluation, and priority ranking of natural areas”. en. In: *Biological Conservation* 8.2, pp. 79–88. DOI: [10.1016/0006-3207\(75\)90033-6](https://doi.org/10.1016/0006-3207(75)90033-6).
- Goldsmith, F.B. (1975). “The evaluation of ecological resources in the countryside for conservation purposes”. en. In: *Biological Conservation* 8.2, pp. 89–96. DOI: [10.1016/0006-3207\(75\)90034-8](https://doi.org/10.1016/0006-3207(75)90034-8).
- Goodchild, M. (2009). “NeoGeography and the nature of geographic expertise”. en. In: *Journal of Location Based Services* 3.2, pp. 82–96. DOI: [10.1080/17489720902950374](https://doi.org/10.1080/17489720902950374).
- Goodchild, M. F. (2010). “Towards Geodesign: Repurposing Cartography and GIS?” en. In: *Cartographic Perspectives* 66, pp. 7–22. DOI: [10.14714/CP66.93](https://doi.org/10.14714/CP66.93).
- Haight, R. G., C. S. Revelle, and S. A. Snyder (2000). “An Integer Optimization Approach to a Probabilistic Reserve Site Selection Problem”. en. In: *Operations Research* 48.5, pp. 697–708. DOI: [10.1287/opre.48.5.697.12411](https://doi.org/10.1287/opre.48.5.697.12411).
- Hanson, J. O. et al. (2019). “Optimality in prioritizing conservation projects”. en. In: *Methods in Ecology and Evolution* 10.10. Ed. by D. Hodgson, pp. 1655–1663. DOI: [10.1111/2041-210X.13264](https://doi.org/10.1111/2041-210X.13264).
- Hanson, J. O. et al. (2020). *prioritizr: Systematic Conservation Prioritization in R*.
- Harris, L. R. et al. (2014). “Using multivariate statistics to explore trade-offs among spatial planning scenarios”. en. In: *Journal of Applied Ecology* 51.6. Ed. by P. Armsworth, pp. 1504–1514. DOI: [10.1111/1365-2664.12345](https://doi.org/10.1111/1365-2664.12345).

- Harroy, J-P. (1969). “L’Union Internationale pour la Conservation de la Nature et de ses Ressources: Origine et Constitution”. In: *Biological Conservation* 1.2, pp. 106–110.
- Harter, R. et al. (2017). *Rsymphony: SYMPHONY in R*.
- Helliwell, D. R. (1971). “A methodology for the assessment of priorities and value in nature conservation”. In: *Nature Conservancy Council*.
- (1967). “The amenity value of trees and woodlands”. en. In: *Arboricultural Association Journal* 1.5, pp. 128–131. DOI: [10.1080/00037931.1967.10590279](https://doi.org/10.1080/00037931.1967.10590279).
- (1969). “Valuation of wildlife resources”. en. In: *Regional Studies* 3.1, pp. 41–47. DOI: [10.1080/09595236900185051](https://doi.org/10.1080/09595236900185051).
- Holling, C.S. (1973). “Resilience and Stability of Ecological Systems”. en. In: *Annual Review of Ecology and Systematics* 4.1, pp. 1–23. DOI: [10.1146/annurev.es.04.110173.000245](https://doi.org/10.1146/annurev.es.04.110173.000245).
- IPBES (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. en. Tech. rep. Version Number: 1. Zenodo. DOI: [10.5281/ZENODO.3831673](https://doi.org/10.5281/ZENODO.3831673).
- IUCN (2016). “IUCN Congress 2016 Bulletin”. In: 89.16, p. 43.
- (2014). “IUCN World Parks Congress 2014 Bulletin”. In: *International Institute for Sustainable Development* 89.16, p. 43.
- Jafari, N. and J. Hearne (2013). “A new method to solve the fully connected Reserve Network Design Problem”. en. In: *European Journal of Operational Research* 231.1, pp. 202–209. DOI: [10.1016/j.ejor.2013.05.015](https://doi.org/10.1016/j.ejor.2013.05.015).
- Jenkins, R. E. and B. W. Bedford (1973). “The use of natural areas to establish environmental baselines”. en. In: *Biological Conservation* 5.3, pp. 168–174. DOI: [10.1016/0006-3207\(73\)90005-0](https://doi.org/10.1016/0006-3207(73)90005-0).
- Joo, R. et al. (2013). “Hidden Markov Models: The Best Models for Forager Movements?” en. In: *PLoS ONE* 8.8. Ed. by G. G. de Polavieja, e71246. DOI: [10.1371/journal.pone.0071246](https://doi.org/10.1371/journal.pone.0071246).
- Jumin, R. et al. (2018). “From Marxan to management: ocean zoning with stakeholders for Tun Mustapha Park in Sabah, Malaysia”. en. In: *Oryx* 52.4, pp. 775–786. DOI: [10.1017/S0030605316001514](https://doi.org/10.1017/S0030605316001514).
- Kareiva, P., R. Lalasz, and M. Marvier (2011). “Conservation in the Anthropocene: Beyond Solitude and Fragility”. In: *Breakthrough Journal* 2.
- Kareiva, P. and M. Marvier (2012). “What Is Conservation Science?” en. In: *BioScience* 62.11, pp. 962–969. DOI: [10.1525/bio.2012.62.11.5](https://doi.org/10.1525/bio.2012.62.11.5).
- Kidd, S. and G. Ellis (2012). “From the Land to Sea and Back Again? Using Terrestrial Planning to Understand the Process of Marine Spatial Planning”. en. In: *Journal of Environmental Policy & Planning* 14.1, pp. 49–66. DOI: [10.1080/1523908X.2012.662382](https://doi.org/10.1080/1523908X.2012.662382).
- Kim, T., J-I. Park, and J. Maeng (2016). “Offshore wind farm site selection study around Jeju Island, South Korea”. en. In: *Renewable Energy* 94, pp. 619–628. DOI: [10.1016/j.renene.2016.03.083](https://doi.org/10.1016/j.renene.2016.03.083).
- Kirkpatrick, J.B. (1983). “An Iterative Method for Establishing Priorities for the Selection of Nature Reserves: An Example From Tasmania”. In: *Biological Conservation* 25.2, pp. 127–134.
- Le Saout, S. et al. (2013). “Protected Areas and Effective Biodiversity Conservation”. en. In: *Science* 342.6160, pp. 803–805. DOI: [10.1126/science.1239268](https://doi.org/10.1126/science.1239268).
- Leopold, A. (1949). *A Sand County Almanac*.
- Linke, S. et al. (2011). “Using multivariate analysis to deliver conservation planning products that align with practitioner needs”. en. In: *Ecography* 34.2, pp. 203–207. DOI: [10.1111/j.1600-0587.2010.06351.x](https://doi.org/10.1111/j.1600-0587.2010.06351.x).
- Liu, P. et al. (2017). “What are the benefits of strictly protected nature reserves? Rapid assessment of ecosystem service values in Wanglang Nature Reserve, China”. en. In: *Ecosystem Services* 26, pp. 70–78. DOI: [10.1016/j.ecoser.2017.05.014](https://doi.org/10.1016/j.ecoser.2017.05.014).

-
- Lougee-Heimer, R. (2003). “The Common Optimization INterface for Operations Research: Promoting open-source software in the operations research community”. en. In: *IBM Journal of Research and Development* 47.1, pp. 57–66. DOI: [10.1147/rd.471.0057](https://doi.org/10.1147/rd.471.0057).
- Luedtke, J. and S. Ahmed (2008). “A Sample Approximation Approach for Optimization with Probabilistic Constraints”. en. In: *SIAM Journal on Optimization* 19.2, pp. 674–699. DOI: [10.1137/070702928](https://doi.org/10.1137/070702928).
- MacArthur, R. H. and Edward O. Wilson (1967). *The theory of island biogeography*. Princeton: Princeton University Press.
- Maclennan, D, P. G. Fernandes, and J. Dalen (2002). “A consistent approach to definitions and symbols in fisheries acoustics”. en. In: *ICES Journal of Marine Science* 59.2, pp. 365–369. DOI: [10.1006/jmsc.2001.1158](https://doi.org/10.1006/jmsc.2001.1158).
- Magris, R. A. et al. (2021). “A blueprint for securing Brazil’s marine biodiversity and supporting the achievement of global conservation goals”. en. In: *Diversity and Distributions* 27.2. Ed. by M. Beger, pp. 198–215. DOI: [10.1111/ddi.13183](https://doi.org/10.1111/ddi.13183).
- Makowski, D. et al. (2000). “A framework to study nearly optimal solutions of linear programming models developed for agricultural land use exploration”. en. In: *Ecological Modelling* 131.1, pp. 65–77. DOI: [10.1016/S0304-3800\(00\)00249-0](https://doi.org/10.1016/S0304-3800(00)00249-0).
- Margules, C. and R. L. Pressey (2000). “Systematic conservation planning”. In: *Nature* 405, pp. 243–253.
- Margules, C. and M.B. Usher (1981). “Criteria used in assessing wildlife conservation potential: A review”. en. In: *Biological Conservation* 21.2, pp. 79–109. DOI: [10.1016/0006-3207\(81\)90073-2](https://doi.org/10.1016/0006-3207(81)90073-2).
- Margules, C.R., A.O. Nicholls, and R.L. Pressey (1988). “Selecting networks of reserves to maximise biological diversity”. en. In: *Biological Conservation* 43.1, pp. 63–76. DOI: [10.1016/0006-3207\(88\)90078-X](https://doi.org/10.1016/0006-3207(88)90078-X).
- Maris, V. (2010). *Philosophie de la Biodiversité*.
- Markantonatou, V. et al. (2021). “Marine spatial plans focusing on biodiversity conservation: The case of the Aegean Sea”. en. In: *Aquatic Conservation: Marine and Freshwater Ecosystems* 31.8, pp. 2278–2292. DOI: [10.1002/aqc.3610](https://doi.org/10.1002/aqc.3610).
- May, R. M. (1975). “Islands biogeography and the design of wildlife preserves”. en. In: *Nature* 254.5497, pp. 177–178. DOI: [10.1038/254177a0](https://doi.org/10.1038/254177a0).
- McClintock, B. T. et al. (2020). “Uncovering ecological state dynamics with hidden Markov models”. en. In: *Ecology Letters* 23.12. Ed. by T. Coulson, pp. 1878–1903. DOI: [10.1111/ele.13610](https://doi.org/10.1111/ele.13610).
- McDonnell, M. D. et al. (2002). “Mathematical methods for spatially cohesive reserve design”. en. In: *Environmental Modeling and Assessment* 7.2, pp. 107–114. DOI: [10.1023/A:1015649716111](https://doi.org/10.1023/A:1015649716111).
- McKinnon, R. (2016). “Epistemic Injustice: Epistemic Injustice”. en. In: *Philosophy Compass* 11.8, pp. 437–446. DOI: [10.1111/phc3.12336](https://doi.org/10.1111/phc3.12336).
- Metcalf, K. et al. (2015). “Evaluating conservation and fisheries management strategies by linking spatial prioritization software and ecosystem and fisheries modelling tools”. en. In: *Journal of Applied Ecology* 52.3. Ed. by A. Punt, pp. 665–674. DOI: [10.1111/1365-2664.12404](https://doi.org/10.1111/1365-2664.12404).
- Moilanen, A., K. A. Wilson, and H. P. Possingham, eds. (2009). *Spatial conservation prioritization: quantitative methods and computational tools*. Oxford biology. OCLC: ocn277205285. Oxford ; New York: Oxford University Press.
- Moilanen, A. et al. (2006a). “Planning for robust reserve networks using uncertainty analysis”. en. In: *Ecological Modelling* 199.1, pp. 115–124. DOI: [10.1016/j.ecolmodel.2006.07.004](https://doi.org/10.1016/j.ecolmodel.2006.07.004).
- Moilanen, A. et al. (2006b). “Uncertainty Analysis for Regional-Scale Reserve Selection”. en. In: *Conservation Biology* 20.6, pp. 1688–1697. DOI: [10.1111/j.1523-1739.2006.00560.x](https://doi.org/10.1111/j.1523-1739.2006.00560.x).

-
- Muir, J. (1913). *The Writings of John Muir: The story of my boyhood and youth*.
- Mytilinou, V. and A. J. Kolios (2017). “A multi-objective optimisation approach applied to offshore wind farm location selection”. en. In: *Journal of Ocean Engineering and Marine Energy* 3.3, pp. 265–284. DOI: [10.1007/s40722-017-0092-8](https://doi.org/10.1007/s40722-017-0092-8).
- Mytilinou, V., E. Lozano-Minguez, and A. Kolios (2018). “A Framework for the Selection of Optimum Offshore Wind Farm Locations for Deployment”. en. In: *Energies* 11.7, p. 1855. DOI: [10.3390/en11071855](https://doi.org/10.3390/en11071855).
- Naess, A. (1973). *The Shallow and the Deep, Long-Range Ecology Movement*.
- Nalle, D. J., J. L. Arthur, and J. Sessions (2002). “Designing Compact and Contiguous Reserve Networks with a Hybrid Heuristic Algorithm”. en. In: *Forest Science* 48.1, p. 10. DOI: <https://doi.org/10.1093/forestscience/48.1.59>.
- Norton, B. G. (1994). “Economists’ Preferences and the Preferences of Economists”. en. In: *Environmental Values* 3.4, pp. 311–332.
- (1991). “Toward Unity Among Environmentalists”. In: *Environmental Values* 2.3, pp. 271–274.
- (1984). “Environmental Ethics and Weak Anthropocentrism”. In: *Environmental Ethics* 6, pp. 131–148.
- Odenbaugh, J. (2016). “Conservation Biology”. In: *The Stanford Encyclopedia of Philosophy*.
- Önal, H. (2004). “First-best, second-best, and heuristic solutions in conservation reserve site selection”. en. In: *Biological Conservation* 115.1, pp. 55–62. DOI: [10.1016/S0006-3207\(03\)00093-4](https://doi.org/10.1016/S0006-3207(03)00093-4).
- Önal, H. and R. A. Briers (2002). “Incorporating spatial criteria in optimum reserve network selection”. In: *Royal Society*.
- (2006). “Optimal Selection of a Connected Reserve Network”. en. In: *Operations Research* 54.2, pp. 379–388.
- Pinchot, G. (1910). *The fight for conservation*.
- Pınarbaşı, K. et al. (2017). “Decision support tools in marine spatial planning: Present applications, gaps and future perspectives”. en. In: *Marine Policy* 83, pp. 83–91. DOI: [10.1016/j.marpol.2017.05.031](https://doi.org/10.1016/j.marpol.2017.05.031).
- Polasky, S. et al. (2000). “Choosing reserve networks with incomplete species information”. en. In: *Biological Conservation* 94.1, pp. 1–10. DOI: [10.1016/S0006-3207\(99\)00171-8](https://doi.org/10.1016/S0006-3207(99)00171-8).
- Possingham, H. P., I. R. Ball, and S. Andelman (2000). “Mathematical Methods for Identifying Representative Reserve Networks”. en. In: *Quantitative Methods for Conservation Biology*. New York: Springer-Verlag, pp. 291–306. DOI: [10.1007/0-387-22648-6_17](https://doi.org/10.1007/0-387-22648-6_17).
- Possingham, H. P. and K. Shea (1999). “The business of biodiversity”. en. In: p. 3.
- Possingham, H. P. et al. (2006). “Protected areas: Goals, limitations, and design”. In: *Principles of Conservation Biology*. 3rd ed., pp. 507–549.
- Possingham, H. P. et al. (1993). “The mathematics of designing a network of protected areas for conservation”. In: *National ASOR Conference*, pp. 536–545.
- Pressey, R. L. (1994). “Ad Hoc Reservations: Forward or Backward Steps in Developing Representative Reserve Systems?” en. In: *Conservation Biology* 8.3, pp. 662–668.
- Pressey, R. L., H. P. Possingham, and J. R. Day (1997). “Effectiveness of alternative heuristic algorithms for identifying indicative minimum requirements for conservation reserves”. en. In: *Biological Conservation* 80.2, pp. 207–219. DOI: [10.1016/S0006-3207\(96\)00045-6](https://doi.org/10.1016/S0006-3207(96)00045-6).
- Pressey, R. L., H. P. Possingham, and C. R. Margules (1996). “Optimality in reserve selection algorithms: When does it matter and how much?” en. In: *Biological Conservation* 76.3, pp. 259–267. DOI: [10.1016/0006-3207\(95\)00120-4](https://doi.org/10.1016/0006-3207(95)00120-4).
- Pressey, R. L. and S. L. Tully (1994). “The cost of ad hoc reservation: A case study in western New South Wales”. en. In: *Austral Ecology* 19.4, pp. 375–384. DOI: [10.1111/j.1442-9993.1994.tb00503.x](https://doi.org/10.1111/j.1442-9993.1994.tb00503.x).

-
- Pressey, R. L. et al. (1993). “Beyond Opportunism: Key Principles for Systematic Reserve Selection”. en. In: 8.4, p. 5.
- Pressey, R.L. and A.O. Nicholls (1989). “Application of a numerical algorithm to the selection of reserves in semi-arid New South Wales”. en. In: *Biological Conservation* 50.1-4, pp. 263–278. DOI: [10.1016/0006-3207\(89\)90013-X](https://doi.org/10.1016/0006-3207(89)90013-X).
- Queffelec, B. et al. (2021). “Marine spatial planning and the risk of ocean grabbing in the tropical Atlantic”. en. In: *ICES Journal of Marine Science*. Ed. by W. Flannery, pp. 1–13. DOI: [10.1093/icesjms/fsab006](https://doi.org/10.1093/icesjms/fsab006).
- Rabe, F. W. and N. L. Savage (1979). “A methodology for the selection of aquatic natural areas”. en. In: *Biological Conservation* 15.4, pp. 291–300. DOI: [10.1016/0006-3207\(79\)90050-8](https://doi.org/10.1016/0006-3207(79)90050-8).
- Ralphs, T. et al. (2019). *coin-or/SYMPHONY: Version 5.6.17*. DOI: [10.5281/ZENODO.2656802](https://doi.org/10.5281/ZENODO.2656802).
- Regan, H. M., M. J. Ensbey, and M. Burgman (2009). “Conservation Prioritization and Uncertainty in Planning Inputs”. In: *Spatial Conservation Prioritisation: Quantitative Methods and Computational Tools*, pp. 145–157.
- Reside, A. E., N. Butt, and V. M. Adams (2018). “Adapting systematic conservation planning for climate change”. en. In: *Biodiversity and Conservation* 27.1, pp. 1–29. DOI: [10.1007/s10531-017-1442-5](https://doi.org/10.1007/s10531-017-1442-5).
- Rodrigues, A. S., J. Orestes Cerdeira, and K. J. Gaston (2000). “Flexibility, efficiency, and accountability: adapting reserve selection algorithms to more complex conservation problems”. en. In: *Ecography* 23.5, pp. 565–574. DOI: [10.1111/j.1600-0587.2000.tb00175.x](https://doi.org/10.1111/j.1600-0587.2000.tb00175.x).
- Rodrigues, A. S.L. and K. J. Gaston (2002). “Optimisation in reserve selection procedures—why not?” en. In: *Biological Conservation* 107.1, pp. 123–129. DOI: [10.1016/S0006-3207\(02\)00042-3](https://doi.org/10.1016/S0006-3207(02)00042-3).
- Rodríguez-Rodríguez, D. et al. (2016). “Achieving Blue Growth through maritime spatial planning: Offshore wind energy optimization and biodiversity conservation in Spain”. en. In: *Marine Policy* 73, pp. 8–14. DOI: [10.1016/j.marpol.2016.07.022](https://doi.org/10.1016/j.marpol.2016.07.022).
- Salcedo-Sanz, S. et al. (2014). “Offshore wind farm design with the Coral Reefs Optimization algorithm”. en. In: *Renewable Energy* 63, pp. 109–115. DOI: [10.1016/j.renene.2013.09.004](https://doi.org/10.1016/j.renene.2013.09.004).
- Salvetat, J. et al. (2022). “Comprehensive spatial distribution of tropical fish assemblages from multifrequency acoustics and video fulfils the island mass effect framework”. en. In: *Scientific Reports* 12.1, p. 8787. DOI: [10.1038/s41598-022-12409-9](https://doi.org/10.1038/s41598-022-12409-9).
- Sarkar, S. (2012). “Complementarity and the selection of nature reserves: algorithms and the origins of conservation planning, 1980–1995”. en. In: *Archive for History of Exact Sciences* 66.4, pp. 397–426. DOI: [10.1007/s00407-012-0097-6](https://doi.org/10.1007/s00407-012-0097-6).
- (2004). “Conservation Biology”. In: *The Stanford Encyclopedia of Philosophy*.
- (2002). “Defining "Biodiversity"; Assessing Biodiversity”. In: *The Monist* 85.1, pp. 131–155.
- Sarkar, S. and C. Margules (2002). “Operationalizing biodiversity for conservation planning”. en. In: *Journal of Biosciences* 27.4, pp. 299–308. DOI: [10.1007/BF02704961](https://doi.org/10.1007/BF02704961).
- Sarkar, S. et al. (2006). “Biodiversity Conservation Planning Tools: Present Status and Challenges for the Future”. en. In: *Annual Review of Environment and Resources* 31.1, pp. 123–159. DOI: [10.1146/annurev.energy.31.042606.085844](https://doi.org/10.1146/annurev.energy.31.042606.085844).
- Sarkar, S. et al. (2004). “Place prioritization for biodiversity conservation using probabilistic surrogate distribution data: Prioritization using distribution data”. en. In: *Diversity and Distributions* 10.2, pp. 125–133. DOI: [10.1111/j.1366-9516.2004.00060.x](https://doi.org/10.1111/j.1366-9516.2004.00060.x).
- Schuster, R. et al. (2020). “Exact integer linear programming solvers outperform simulated annealing for solving conservation planning problems”. en. In: *PeerJ* 8, e9258. DOI: [10.7717/peerj.9258](https://doi.org/10.7717/peerj.9258).
- Serra-Sogas, N. et al. (2020). *Marxan User Manual: For Marxan version 2.43 and above*.

- Serrano González, J. et al. (2014). “A review and recent developments in the optimal wind-turbine micro-siting problem”. en. In: *Renewable and Sustainable Energy Reviews* 30, pp. 133–144. DOI: [10.1016/j.rser.2013.09.027](https://doi.org/10.1016/j.rser.2013.09.027).
- Shirabe, T. (2005). “A Model of Contiguity for Spatial Unit Allocation”. en. In: *Geographical Analysis* 37.1, pp. 2–16. DOI: [10.1111/j.1538-4632.2005.00605.x](https://doi.org/10.1111/j.1538-4632.2005.00605.x).
- Simberloff, D. S. (1976). “Species Turnover and Equilibrium Island Biogeography”. In: *Science* 194.4265, pp. 572–578.
- Simberloff, D. S. and L. G. Abele (1976). “Island Biogeography Theory and Conservation Practice”. en. In: *Science* 191.4224, pp. 285–286. DOI: [10.1126/science.191.4224.285](https://doi.org/10.1126/science.191.4224.285).
- Simmonds, J. and D. MacLennan, eds. (2005). *Fisheries Acoustics*. en. Oxford, UK: Blackwell Publishing Ltd. DOI: [10.1002/9780470995303](https://doi.org/10.1002/9780470995303).
- Soule, M. E. (1985). “What Is Conservation Biology?” en. In: *BioScience* 35.11, p. 9.
- Soulé, M. (2013). “The “New Conservation””. en. In: *Conservation Biology* 27.5, pp. 895–897. DOI: [10.1111/cobi.12147](https://doi.org/10.1111/cobi.12147).
- Soulé, M. E. and D. S. Simberloff (1986). “What do genetics and ecology tell us about the design of nature reserves?” en. In: *Biological Conservation* 35.1, pp. 19–40. DOI: [10.1016/0006-3207\(86\)90025-X](https://doi.org/10.1016/0006-3207(86)90025-X).
- Stewart, R. R. and H. P. Possingham (2005). “Efficiency, costs and trade-offs in marine reserve system design”. en. In: *Environmental Modeling & Assessment* 10.3, pp. 203–213. DOI: [10.1007/s10666-005-9001-y](https://doi.org/10.1007/s10666-005-9001-y).
- Stolton, S. and N. Dudley (2010). *Arguments for Protected Areas: Multiple Benefits for Conservation Use*.
- Tans, W. (1974). “Priority ranking of biotic natural areas”. In: *Michigan Botanist*.
- Thoreau, H. D. (1854). *Walden, or Life in the Woods*.
- Ticco, P. C. (1995). “The use of marine protected areas to preserve and Enhance marine biological diversity: A case study approach”. en. In: *Coastal Management* 23.4, pp. 309–314. DOI: [10.1080/08920759509362274](https://doi.org/10.1080/08920759509362274).
- Tole, L. (2006). “Choosing reserve sites probabilistically: A Colombian Amazon case study”. en. In: *Ecological Modelling* 194.4, pp. 344–356. DOI: [10.1016/j.ecolmodel.2005.10.027](https://doi.org/10.1016/j.ecolmodel.2005.10.027).
- Trouillet, B. (2019). “Aligning with dominant interests: The role played by geo-technologies in the place given to fisheries in marine spatial planning”. en. In: *Geoforum* 107, pp. 54–65. DOI: [10.1016/j.geoforum.2019.10.012](https://doi.org/10.1016/j.geoforum.2019.10.012).
- (2020). “Reinventing marine spatial planning: a critical review of initiatives worldwide”. en. In: *Journal of Environmental Policy & Planning* 22.4, pp. 441–459. DOI: [10.1080/1523908X.2020.1751605](https://doi.org/10.1080/1523908X.2020.1751605).
- Tubbs, C.R. and J.W. Blackwood (1971). “Ecological evaluation of land for planning purposes”. en. In: *Biological Conservation* 3.3, pp. 169–172. DOI: [10.1016/0006-3207\(71\)90159-5](https://doi.org/10.1016/0006-3207(71)90159-5).
- Tuda, A. O., T. F. Stevens, and L. D. Rodwell (2014). “Resolving coastal conflicts using marine spatial planning”. en. In: *Journal of Environmental Management* 133, pp. 59–68. DOI: [10.1016/j.jenvman.2013.10.029](https://doi.org/10.1016/j.jenvman.2013.10.029).
- Tulloch, V. J. et al. (2013). “Incorporating uncertainty associated with habitat data in marine reserve design”. en. In: *Biological Conservation* 162, pp. 41–51. DOI: [10.1016/j.biocon.2013.03.003](https://doi.org/10.1016/j.biocon.2013.03.003).
- Tundi Agardy, M. (1994). “Advances in marine conservation: the role of marine protected areas”. en. In: *Trends in Ecology & Evolution* 9.7, pp. 267–270. DOI: [10.1016/0169-5347\(94\)90297-6](https://doi.org/10.1016/0169-5347(94)90297-6).
- Van Haaren, R. and V. Fthenakis (2011). “GIS-based wind farm site selection using spatial multi-criteria analysis (SMCA): Evaluating the case for New York State”. en. In: *Renewable and Sustainable Energy Reviews* 15.7, pp. 3332–3340. DOI: [10.1016/j.rser.2011.04.010](https://doi.org/10.1016/j.rser.2011.04.010).
- Vanderkam, R. P. D., Y. F. Wiersma, and D. J. King (2007). “Heuristic algorithms vs. linear programs for designing efficient conservation reserve networks: Evaluation of solution

-
- optimality and processing time”. en. In: *Biological Conservation* 137.3, pp. 349–358. DOI: [10.1016/j.biocon.2007.02.018](https://doi.org/10.1016/j.biocon.2007.02.018).
- Vaughan, D. and T. Agardy (2020). “Marine protected areas and marine spatial planning – allocation of resource use and environmental protection”. en. In: *Marine Protected Areas*. Elsevier, pp. 13–35. DOI: [10.1016/B978-0-08-102698-4.00002-2](https://doi.org/10.1016/B978-0-08-102698-4.00002-2).
- Virtanen, E.A. et al. (2022). “Balancing profitability of energy production, societal impacts and biodiversity in offshore wind farm design”. en. In: *Renewable and Sustainable Energy Reviews* 158, p. 112087. DOI: [10.1016/j.rser.2022.112087](https://doi.org/10.1016/j.rser.2022.112087).
- Visconti, P. et al. (2013). “Effects of Errors and Gaps in Spatial Data Sets on Assessment of Conservation Progress: Errors and Gaps in Spatial Data Sets”. en. In: *Conservation Biology* 27.5, pp. 1000–1010. DOI: [10.1111/cobi.12095](https://doi.org/10.1111/cobi.12095).
- Wang, Y-C. and H. Önal (2011). “Designing connected nature reserve networks using a graph theory approach”. en. In: *Acta Ecologica Sinica* 31.5, pp. 235–240. DOI: [10.1016/j.chnaes.2011.06.001](https://doi.org/10.1016/j.chnaes.2011.06.001).
- Wang, Y. and H. Önal (2013). “Designing a connected nature reserve using a network flow theory approach”. en. In: *Acta Ecologica Sinica* 33.5, pp. 253–259. DOI: [10.1016/j.chnaes.2013.07.004](https://doi.org/10.1016/j.chnaes.2013.07.004).
- Wang, Y., H. Önal, and Q. Fang (2018). “How large spatially-explicit optimal reserve design models can we solve now? An exploration of current models’ computational efficiency”. In: *Nature Conservation* 27, pp. 17–34. DOI: [10.3897/natureconservation.27.21642](https://doi.org/10.3897/natureconservation.27.21642).
- Watts, M. E. et al. (2009). “Marxan with Zones: Software for optimal conservation based land- and sea-use zoning”. en. In: *Environmental Modelling & Software* 24.12, pp. 1513–1521. DOI: [10.1016/j.envsoft.2009.06.005](https://doi.org/10.1016/j.envsoft.2009.06.005).
- Weise, H. et al. (2020). “Resilience trinity: safeguarding ecosystem functioning and services across three different time horizons and decision contexts”. en. In: *Oikos* 129.4, pp. 445–456. DOI: [10.1111/oik.07213](https://doi.org/10.1111/oik.07213).
- Westman, W. E. (1977). “How Much Are Nature’s Services Worth?” en. In: *Science* 197.4307, pp. 960–964. DOI: [10.1126/science.197.4307.960](https://doi.org/10.1126/science.197.4307.960).
- White, C., B. S. Halpern, and C. V. Kappel (2012). “Ecosystem service tradeoff analysis reveals the value of marine spatial planning for multiple ocean uses”. en. In: *Proceedings of the National Academy of Sciences* 109.12, pp. 4696–4701. DOI: [10.1073/pnas.1114215109](https://doi.org/10.1073/pnas.1114215109).
- Williams, J. C. (2003). “Convex Land Acquisition with Zero-One Programming”. en. In: *Environment and Planning B: Planning and Design* 30.2, pp. 255–270. DOI: [10.1068/b12925](https://doi.org/10.1068/b12925).
- (2006). “Optimal reserve site selection with distance requirements”. en. In: *Computers & Operations Research* 35.2, pp. 488–498. DOI: [10.1016/j.cor.2006.03.012](https://doi.org/10.1016/j.cor.2006.03.012).
- Williams, J. C. and C. S. ReVelle (1998). “Reserve assemblage of critical areas: A zero-one programming approach”. en. In: *European Journal of Operational Research* 104.3, pp. 497–509. DOI: [10.1016/S0377-2217\(97\)00017-9](https://doi.org/10.1016/S0377-2217(97)00017-9).
- Williams, J. C., C. S. ReVelle, and S. A. Levin (2005). “Spatial attributes and reserve design models: A review”. en. In: *Environmental Modeling and Assessment* 10, p. 19.
- (2004). “Using mathematical optimization models to design nature reserves”. en. In: *Frontiers in Ecology and the Environment* 2.2, pp. 98–105. DOI: [10.1890/1540-9295\(2004\)002\[0098:UMOMTD\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2004)002[0098:UMOMTD]2.0.CO;2).
- Williams, P. H. and M. B. Araújo (2000). “Using probability of persistence to identify important areas for biodiversity conservation”. en. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267.1456, pp. 1959–1966. DOI: [10.1098/rspb.2000.1236](https://doi.org/10.1098/rspb.2000.1236).
- Wright, D.F. (1977). “A site evaluation scheme for use in the assessment of potential nature reserves”. en. In: *Biological Conservation* 11.4, pp. 293–305. DOI: [10.1016/0006-3207\(77\)90042-8](https://doi.org/10.1016/0006-3207(77)90042-8).
- WWF (2018). *WWF Briefing 2018: Principles for a Sustainable Blue Economy*. Tech. rep.