Opening the black box of decision support tools in marine spatial planning: shedding light into reserve site selection algorithms for a balanced empowerment of stakeholders

Adrien Brunel[a]*, Juliette Davret[b], Brice Trouillet[b], Nicolas Bez[a], Julie Salvetat[a], Antoine Gicquel[a] and Sophie Lanco Bertrand[a]

[a]IRD, MARBEC (Univ. Montpellier, CNRS, IFREMER, IRD)
[b]Université de Nantes, CNRS, UMR LETG

Abstract

Marine spatial planning (MSP) is positioning itself as a rational decision-making process regulating uses of marine spaces and resources in order to reduce tensions between exploitation and conservation as well as between ocean stakeholders. As global political agendas identified marine protected areas as a key answer to biodiversity erosion, systematic reserve site selection became a critical component of MSP. Establishing an ocean zoning involves the analysis of large quantities of heterogeneous, multi-sources and spatially explicit data. This often leads to problems too complex to be solved by human intuition only, thus calling for optimisation tools to support the decisions. In that context, our work aims at informing practitioners about stakes, possibilities and limitations of MSP approach through reserve site selection tools. We first clarify the reserve site selection framework, especially the underlying mathematics - the problem formulation and the solving method. Then, we highlight potential pitfalls due to input data feeding the reserve-based planning approach. Finally, and more practically, we show to what extent parameters used in reserve selection tools shape the reserve outcome. These elements are explored and illustrated on a real case study, namely the Fernando de Noronha archipelago in the Brazilian tropical Atlantic. This work provides a brief overview of informational challenges brought by decision support tools in marine spatial planning negotiations.

Keywords: marine spatial planning; decision support tools; reserve site selection; optimisation; marxan; informational challenges;

*Corresponding author. Email: adrien.brunel.pro@protonmail.com; Postal address: UMR MARBEC (IRD, Univ. Montpellier, Ifremer, CNRS), Avenue Jean Monnet, 34203 Sète Cedex, France

1 Introduction

Marine environments are frequently seen as tomorrow's territories for «blue growth» (*The EU Blue Economy Report 2019*; *The EU Blue Economy Report 2020*; *WWF Briefing 2018: Principles for a Sustainable Blue Economy*). Yet these spaces are already being at the heart of multiple anthropogenic pressures: fishing, aquaculture, shipping routes, seabed exploitation, recreational activities, renewable energies, *etc.* 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 in order to reduce tensions between exploitation and conservation, as well as between ocean stakeholders. According to Ehler and Douvere (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". MSP broadly diffused in the last decades to eventually emerge as the favoured at-sea governance paradigm among management institutions seeking a sustainable development.

MSP strives to be a rational and evidence-based process (Pınarbaşı et al. 2017). In this framework, rooted in data analysis, decision support tools (DSTs) turned out to be indispensable 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 (Pınarbaşı et al. 2017).

The number and types of DSTs have grown continuously. Those that focus on systematic conservation planning and reserve site selection, such as Marxan or prioritizR, have particularly gained in popularity driven by different international agendas. Indeed, conservation institutions identified marine protected areas (MPAs) as an essential part of the solution to ensure biodiversity resilience. *De facto*, areas dedicated to conservation are proved to provide biotic communities global benefits (*e.g.* Stolton and Dudley 2010) especially for strict reserve (Claudet et al. 2020; Liu et al. 2017). Thus, the United Nations (UN) target for global ocean protection was established to 10% of the coastal and marine areas in MPA by 2020, as set forth by Aichi Target 11 under the Convention on Biological Diversity. The UN Sustainable Development Goal 14 reaffirms this commitment. Going further, many scientists emphasise that the 10% target is 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% for 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 2008[1]. More recently, the European Green Deal aims at 30% of MPAs within the European seas, among which 10% with a strict protection (*The European Green Deal, COM(2019) 640 final, Brussels; 11.12.2019,EU Biodiversity Strategy for 2030, Bringing nature into our lives, COM(2020) 380 finals,Brussels, 20.05.2020*). Consequently, tools for systematic selection of reserve sites are needed to delimit the ever-expanding areas devoted to conservation, and to avoid, as far as possible, *ad-hoc* and opaque reserve solutions (Pressey 1994; Pressey and Tully 1994). DSTs for reserve design have therefore rapidly become an issue for research and use at the global level, including for the management of MSP issues.

[1] `https://www.protectedplanet.net/marine`

Although early attempts were based on simple and intuitive ranking approaches (Kirkpatrick 1983; Margules et al. 1988) of areas based on a computed conservation value (Helliwell 1967; Tubbs and Blackwood 1971; Goldsmith 1975; Wright 1977), reserve site selection is now assessed as an optimisation problem involving an integer programming framework (Cocks and Baird 1989; Possingham et al. 1993; Possingham et al. 2000; Margules and Pressey 2000; Possingham et al. 2006). Practically, conservation-based planning tools aim at finding where to locate areas dedicated to conservation, i.e. nature reserve. This can be intended as a resource allocation optimisation problem. The purpose is to find the resource layout which minimises a given objective subject to a set of constraints. Mathematically, it can be modelled as a deterministic binary programming problem. Besides, recent emergence of efficient exact optimisation solvers have made possible to apply methods that used to be considered numerically unreasonable in the past (Church et al. 1996; Beyer et al. 2016; Schuster et al. 2020).

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 comes 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 boxes" 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.

Within that context, the purposes of this work are to (1) detail and question the mathematical functioning of these DSTs to end-users through graphical 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 DST 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 Material and methods

2.1 Study site

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 coasts[2] 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, a series of field research surveys were conducted since 2015, providing the spatially explicit data on fish and fisheries used hereafter.

2.2 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 $[32.65°\text{W}, 32.30°\text{W}]$ and $[3.95°\text{S}, 3.75°\text{S}]$ ranges. We chose a 0.01° resolution which results in considering N=36×21=756 planning units. Both our boundaries and resolution choices were justified to properly capture data feeding our case study (*cf.* Section 2.4). 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 2). 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.

2.3 Optimisation

2.3.1 General framework

Conservation-based planning tools practically aim at finding where to locate areas dedicated to conservation, *i.e.* nature reserves, in order to ensure a given level of biodiversity persistence and eventually ecosystem services provision. As it can be expressed as a resource allocation problem, the conservation science literature proposed and developed an optimisation framework

[2] 360 km from Natal, 545 km from Recife

to deal with such problems. The purpose is to find the reserve allocation layout which maximises/minimises a given objective subject to a set of constraints. In that context, reserve site selection tools provide optimisation solving methods suited for binary programming. Indeed, a reserve is mathematically represented with a vector $x \in \{0,1\}^N$. The row value x_j of the vector x is 1 if the planning unit $j \in \{1, \cdots, N\}$ is included within the reserve, 0 otherwise. Each planning unit is associated with a socio-economic cost[3] and with the amount of each considered conservation feature. Then, depending on the problem formulation, global conservation targets or total socio-economic budget are defined from the available ecological and socio-economic knowledge. Targets represent the least total amount of each conservation feature which must be included in the final reserve. Budget represents the maximum tolerated socio-economic cost. Explicit optimisation models are detailed in Section 2.3.2.

2.3.2 Formulation

Considering the reserve site selection question through an optimisation framework resulted from an encounter between operations research and conservation science. In this fruitful collaboration, various optimisation formulations emerged. The two main generic formulations, namely the *minimum set* and *maximum coverage* formulations, were detailed in Table 1.

Minimum set formulation	Maximum coverage formulation
$\begin{cases} \min\limits_{x\in\{0,1\}^N} & c^T x + \beta x^T B(1-x) \\ \text{s.t.} & Ax \geq t \end{cases} \quad (1)$	$\begin{cases} \max\limits_{x\in\{0,1\}^N} & \omega^T A x \\ \text{s.t.} & c^T x + \beta x^T B(1-x) \leq b \end{cases} \quad (2)$

Table 1 – Minimum set and maximum coverage formulations for conservation resource allocation 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 problem (*cf.* Equation (1)), the goal of systematic reserve site selection tools is to find which collection of planning units achieves *a priori* defined conservation targets at a minimum socio-economic cost. Alternatively, the maximum coverage problem (*cf.* Equation (2)) purpose is to find which planning unit collection covers the maximum amount of conservation features within the limits of a predefined socio-economic cost budget. Results of both approaches were detailed in Section 3.1.1. In both formulations, the compactness parameter β in Equation (1) and (2) allowed to include the reserve perimeter $x^T B(1-x)$ as a cost. The bigger the compactness parameter, the more spatially aggregated the computed reserve.

2.3.3 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

[3] assessed from a manager perspective

programming context, the quadratic element $x^T B(1-x)$ in Table 1 can be linearised (Billionnet 2007). Comparison between both solving methods were presented in Section 3.1.2. 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[4] project (Lougee-Heimer 2003) called through a dedicated code[5] 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 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).

2.4 Input data

2.4.1 Fish biomass

Recent acoustic surveys around Fernando de Noronha collected *in situ* data on fish distribution (Betrand 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 column. Figure 1 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 (see Section 4.2), 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 were presented in Panel B in Figure 2.

2.4.2 Habitats

Bathymetric data were collected from GEneral Bathymetric Chart of the Ocean (GEBCO) online platform[6]. 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 separated suitable habitats for benthic and demersal fish, 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

[4] Common Optimization INterface for Operations Research
[5] `https://github.com/AdrienBrunel/reserve-site-selection`
[6] `https://download.gebco.net/`

pixel to continental shelf. Results of this process were illustrated in Panel C and D in Figure 2. 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.

2.4.3 Fishery

Fishery data were composed of 69 GPS tracks from fishermen's 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 thought of 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 2.

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. 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 :

- $\mathbf{Cost_1 = 1}$
  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.

- $\mathbf{Cost_2 = 1 + FC}$
  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.

- $\mathbf{Cost_3 = 1 + \ln{(1 + FC)}}$
  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.

- **$\mathbf{Cost_4 = FC_{1\rightarrow 10}}$ 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.

- **$\mathbf{Cost_5 = FC_{1\rightarrow 100}}$ 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. Results of the associated sensitivity analysis were presented in Section 3.2.

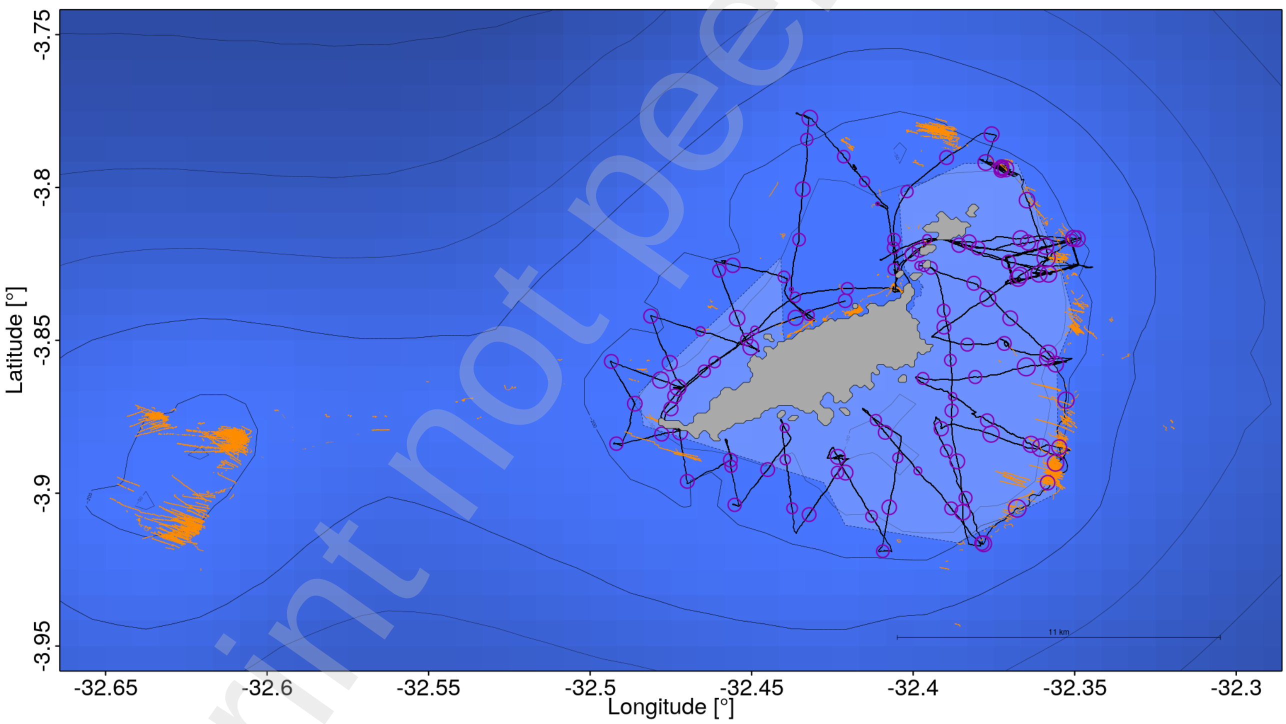


Figure 1 – Raw input data feeding the conservation problem. Bathymetric raw data (GEBCO 2014) is represented by a light to deep blue colour gradient with appended iso-depth solid thin black lines (50m, 200m, 1000m, 2000m, 3000m, 4000m). Fishermen's boats GPS points estimated in a fishing state are illustrated with orange dots. Acoustic raw data is 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.

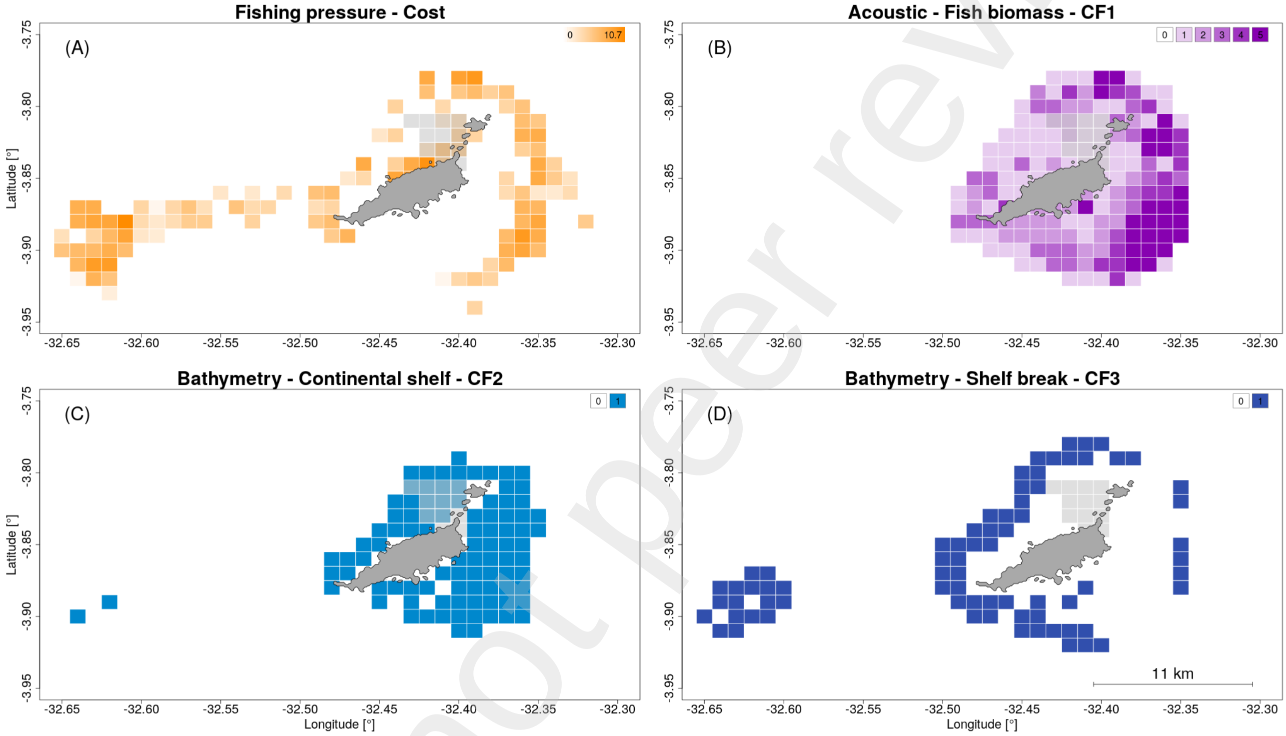


Figure 2 – Processed geographical layers feeding the conservation problem. (A) Fishery data representing fishing pressure is shown with an orange colour gradient indexed on the number of fishing points recorded in each planning unit. (B) Acoustic data interpreted as a fish biomass surrogate (conservation feature n°1) is depicted with a purple colour gradient indexed on interpolated class median value. (C) Continental shelf habitat surrogate (conservation feature n°2) is illustrated in light blue indexed on presence of depth between 0m and 50m. (D) Shelf break habitat surrogate (conservation feature n°3) is coloured in deep blue indexed on presence of depth between 50m and 200m. Transparent grey pixels are the locked-out planning units.

3 Results

3.1 Structural elements

3.1.1 Modelling 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 3. 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 2 : 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.

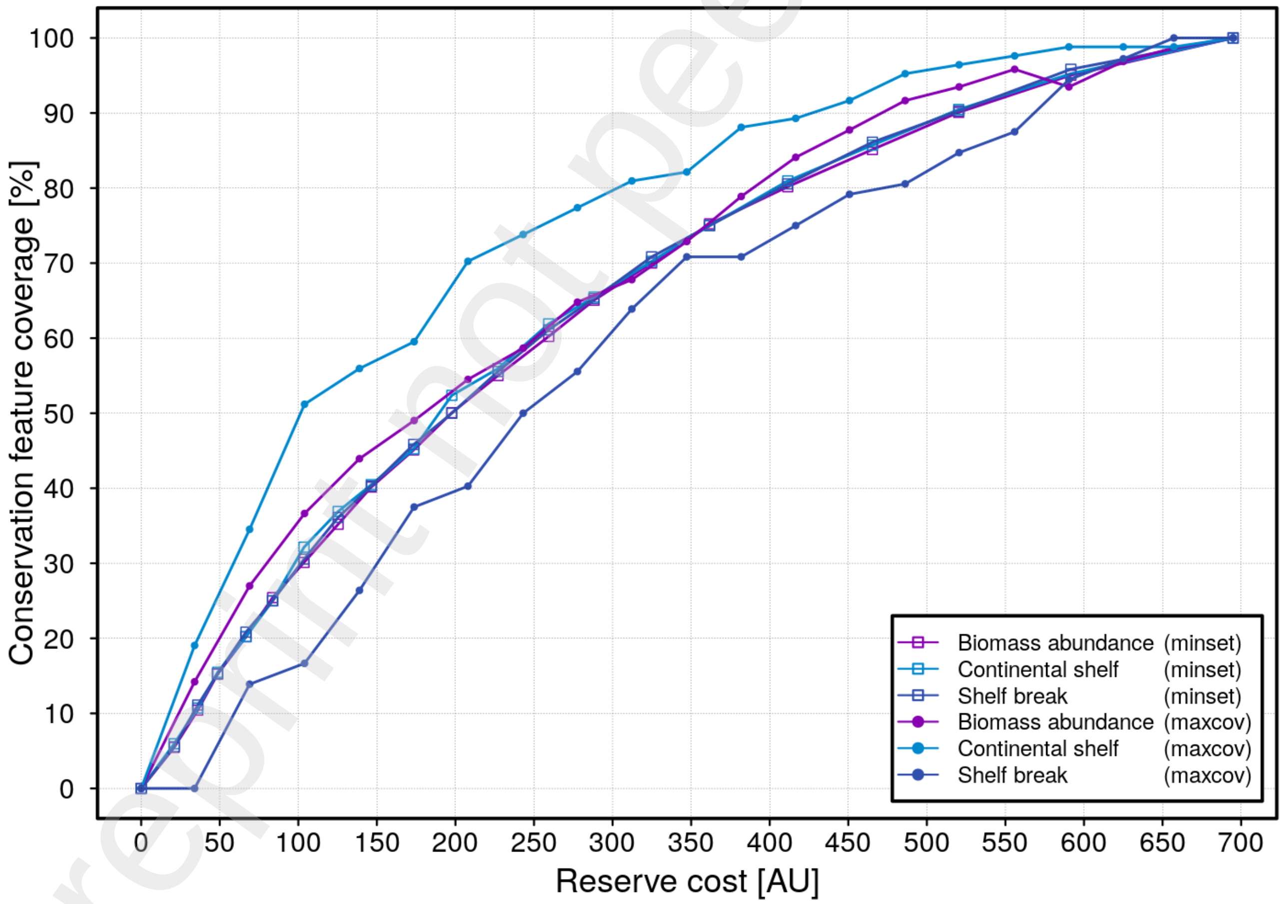


Figure 3 – Reserve cost in arbitrary units versus conservation feature coverage in % for both minimum set and maximum coverage formulations. The 3 conservation features coverage are shown in purple (biomass abundance), light blue (continental shelf) and deep blue (shelf break) while associated formulation 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.

3.1.2 Solving method

We here illustrated (*cf.* Figure 4) results provided by reserve site selection DSTs computed with both solving approaches (see Section 2.3.3). The metaheuristic results were represented by a green colour gradient representing the Marxan selection frequency. 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 2. The aggregated aspect was due to the active compactness penalty (see Section 3.3.2 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 2. Also, in this small size study case, a 5% gap was derived between optimal and mean metaheuristic solution scores (among the 100 Marxan runs).

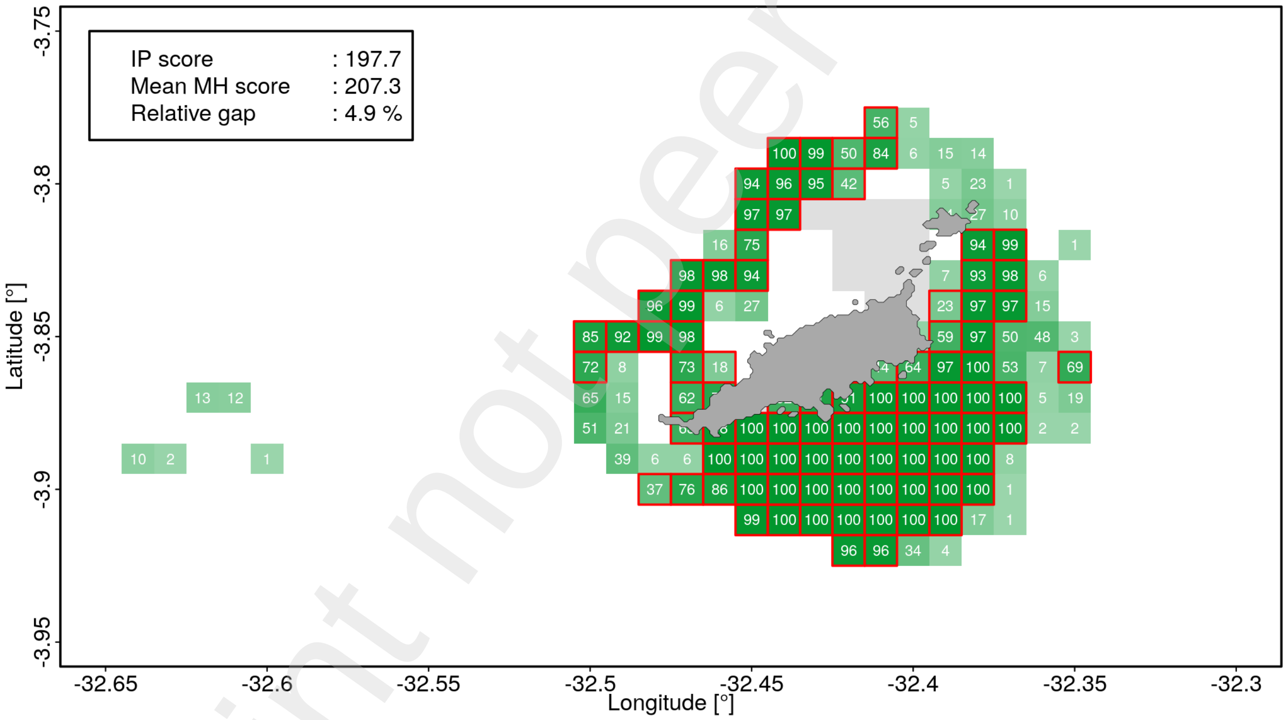


Figure 4 – Metaheuristic versus exact integer programming reserve solutions. Minimum set formulation, 3 conservation features each represented with a 50% protection target, considered cost layer worth $1 + \ln(1 + FC)$ and compactness parameter $\beta = 1$. Selected planning units within optimal reserve 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.

3.2 Influence of cost expression

Table 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 $\text{FC}_{1\rightarrow100}$ 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 correlation 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 implie 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 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		$1+\text{FC}$		$1+\ln(1+\text{FC})$		$\text{FC}_{1\to10}$ scale		$\text{FC}_{1\to100}$ scale	
	Cost	Solution	Cost	Solution	Cost	Solution	Cost	Solution	Cost	Solution
1	1		-	0.40	-	0.41	-	0.47	-	0.47
$1+\text{FC}$	⋆		1		0.55	0.93	0.85	0.79	1.00	0.82
$1+\ln(1+\text{FC})$	⋆		⋆		1		0.83	0.84	0.58	0.87
$\text{FC}_{1\to10}$ scale	⋆		⋆		⋆		1		0.87	0.89
$\text{FC}_{1\to100}$ scale	⋆		⋆		⋆		⋆		1	

Table 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 $\beta = 0$.

3.3 Parameters influence

3.3.1 Coverage targets

Figure 3 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.

3.3.2 Compactness parameter

Sensitivity analysis over the compactness parameter β is shown in Figure 5. 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 (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.

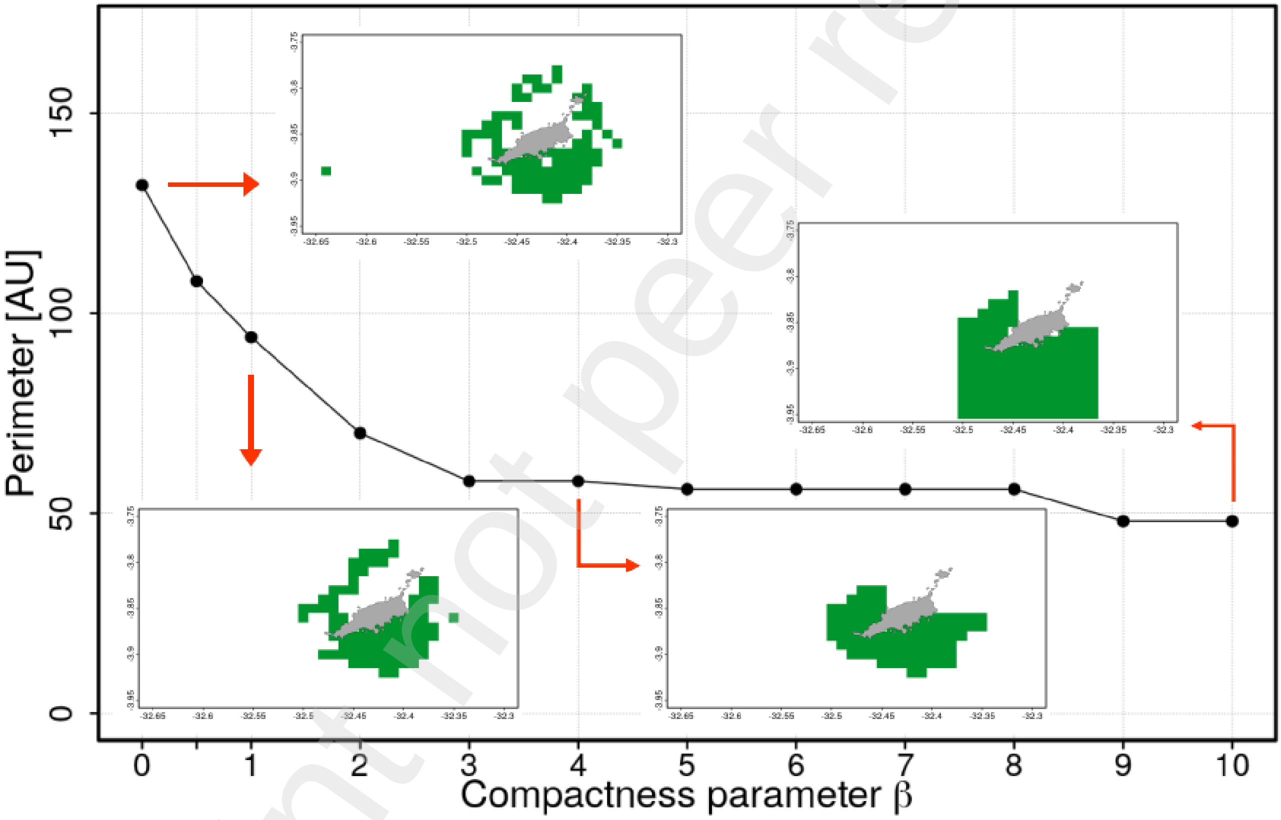


Figure 5 – Perimeter (in arbitrary unit) versus compactness parameter. Sensitivity analysis performed on the compactness parameter β with respect to the reserve solution perimeter. Minimum set formulation, 3 conservation features represented, considered cost layer worth $1 + \ln(1 + FC)$, exact solving thanks to Cbc solver.

4 Discussion

4.1 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.

4.1.1 Formulation

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 3 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.

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 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 2.3.3. 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 (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.

4.1.2 Solving method

Although metaheuristic were historically preferred due to Marxan developments, exact solvers should prevail in the future. In our case study (see Section 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 recursively 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.

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 2.4, 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.

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 3.2. 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.

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.

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.

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)

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.

Coverage targets If not ecologically guided, coverage targets can be used as tuning parameters. Indeed, as demonstrated in Section 3.1.1, these parameters were directly linked with the reserve cost (*cf.* Figure 3) 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.

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 A 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)). 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 3.3.2).

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 go 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, proxys 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.

References

Ardron, Jeff A., H. P. Possingham, and Carissa J. Klein (2010). *Marxan Good Practices Handbook*.

Ball, Ian R., Hugh P. Possingham, and Matthew 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.

Betrand, Arnaud (2019). "FAROFA 3 cruise, RV TUBARAO Tigre". In: Publisher: Sismer. DOI: 10.17600/18001381.

Beyer, Hawthorne 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.

Bez, Nicolas and Cheikh-Baye 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 Josef Michael Jech, pp. 709–718. DOI: 10.1139/cjfas-2013-0437.

Bezanson, Jeff et al. (2012). "Julia: A Fast Dynamic Language for Technical Computing". en. In: *arXiv:1209.5145 [cs]*. arXiv: 1209.5145.

Bezanson, Jeff et al. (2015). "Julia: A Fresh Approach to Numerical Computing". en. In: *arXiv:1411.1607 [cs]*. arXiv: 1411.1607.

Billionnet, Alain (2007). *Optimisation Discrète, de la modélisation à la résolution par des logiciels de programmation mathématique*. Dunod.

Carvalho, Sílvia 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.

Chiles, Jean-Paul and Pierre Delfiner (2012). *Geostatistics: modeling spatial uncertainty*. 2nd ed. Wiley series in probability and statistics. Hoboken, N.J: Wiley.

Church, Richard L., David M. Stoms, and Frank 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.

Claudet, Joachim 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.

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.

Commission, European (2020). *EU Biodiversity Strategy for 2030, Bringing nature into our lives, COM(2020) 380 finals,Brussels, 20.05.2020*. Tech. rep.

Commission, European (2019). *The European Green Deal, COM(2019) 640 final, Brussels; 11.12.2019*. Tech. rep.

Delavenne, Juliette 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.

Drira, Sabrine 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.

Dunning, Iain, Joey Huchette, and Miles 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.

Ehler, Charles N. and Fanny Douvere (2009). "Marine Spatial Planning, A Step-by-Step Approach toward Ecosystem-based Management". In: *Intergovernmental Oceanographic Com-*

mission and Man and the Biosphere Programme IOC Manual and Guides No. 53 ICAM Dossier No. 6. UNESCO, Paris, pp. 99.

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, Andreas and Bas Waterhout (2006). "Introducing Evidence-Based Planning". en. In: *disP - The Planning Review* 42.165, pp. 4–13. DOI: 10.1080/02513625.2006.10556950.

Fiorella, Kathryn 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.

Flannery, Wesley 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.

Forrest, John et al. (2018). *Coin-Or/Cbc: Version 2.9.9*. DOI: 10.5281/ZENODO.1317566.

Game, Edward T. and Hedley S. Grantham (2008). *Marxan User Manual For Marxan version 1.8.10*.

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.

Goodchild, Michael (2009). "NeoGeography and the nature of geographic expertise". en. In: *Journal of Location Based Services* 3.2, pp. 82–96. DOI: 10.1080/17489720902950374.

Goodchild, Michael F. (2010). "Towards Geodesign: Repurposing Cartography and GIS?" en. In: *Cartographic Perspectives* 66, pp. 7–22. DOI: 10.14714/CP66.93.

Hanson, Jeffrey O. et al. (2020). *prioritizr: Systematic Conservation Prioritization in R*.

Harter, R. et al. (2017). *Rsymphony: SYMPHONY in R*.

Helliwell, D. R. (1967). "The amenity value of trees and woodlands". en. In: *Arboricultural Association Journal* 1.5, pp. 128–131. DOI: 10.1080/00037931.1967.10590279.

IUCN (2016). "IUCN Congress 2016 Bulletin". In: 89.16, p. 43.

IUCN (2014). "IUCN World Parks Congress 2014 Bulletin". In: *International Institute for Sustainable Development* 89.16, p. 43.

Joo, Rocio et al. (2013). "Hidden Markov Models: The Best Models for Forager Movements?" en. In: *PLoS ONE* 8.8. Ed. by Gonzalo G. de Polavieja, e71246. DOI: 10.1371/journal.pone.0071246.

Kidd, Sue and Geraint 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.

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.

Liu, Peng 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.

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.

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.

Magris, Rafael 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 Maria Beger, pp. 198–215. DOI: 10.1111/ddi.13183.

Margules, C. and R. L. Pressey (2000). “Systematic conservation planning”. In: *Nature* 405, pp. 243–253.

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.

McClintock, Brett T. et al. (2020). “Uncovering ecological state dynamics with hidden Markov models”. en. In: *Ecology Letters* 23.12. Ed. by Tim Coulson, pp. 1878–1903. DOI: 10.1111/ele.13610.

Önal, Hayri (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.

Pınarbaşı, Kemal 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.

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 ares for conservation”. In: *National ASOR Conference*, pp. 536–545.

Possingham, Hugh, Ian Ball, and Sandy 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.

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. 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.

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.

Queffelec, Betty 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 Wesley Flannery, pp. 1–13. DOI: 10.1093/icesjms/fsab006.

Ralphs, Ted et al. (2019). *coin-or/SYMPHONY: Version 5.6.17*. DOI: 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, April E., Nathalie Butt, and Vanessa 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.

Schuster, Richard 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.

Simmonds, John and David MacLennan, eds. (2005). *Fisheries Acoustics*. en. Oxford, UK: Blackwell Publishing Ltd. DOI: 10.1002/9780470995303.

Stolton, Sue and Nigel Dudley (2010). *Arguments for Protected Areas: Multiple Benefits for Conservation Use*.

Trouillet, Brice (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.

Trouillet, Brice (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.

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.

Vanderkam, Robert P.D., Yolanda F. Wiersma, and Douglas 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.

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.

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.

WWF (2018). *WWF Briefing 2018: Principles for a Sustainable Blue Economy*. Tech. rep.

6 Appendix

A Compactness parameter correction

In 3.3.2, we mentioned an unwanted edge effect involved by the compactness parameter β. As a reminder, we observed in our work (*cf.* Figure 5) but also in other publications that planning units at the edge of the study area are more likely to be included in the reserve solution. It was simply explained by the fact they had less common frontiers with surrounding pixels due to their position at the border so they had 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 6). 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 missed 2 edges while another pixel at the border only missed 1. The fictive pixel was locked-out (*i.e.* never included in the reserve solution) thus leaving the rest of the optimisation problem undisturbed.

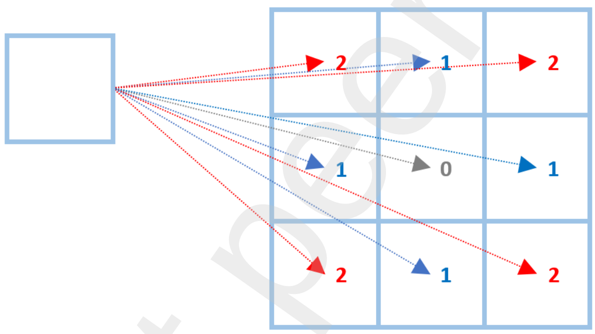


Figure 6 – Principle of the correction involving the addition of a fictive pixel connected to the external edge of the grid. The newly added fictive pixel is invisible in the original model (necessarily locked-out), but allows the outer edge of the study area to be taken into account within the computation 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 (1).

$$\boldsymbol{B}^* = \left(\begin{array}{ccc:c} & \boldsymbol{B} & & \begin{matrix} b^*_{1,N+1} \\ \vdots \\ b^*_{N,N+1} \end{matrix} \\ \hdashline b^*_{N+1,1} & \cdots & b^*_{N+1,N} & 0 \end{array}\right) \qquad \boldsymbol{x}^* = \begin{pmatrix} x^*_1 \\ \vdots \\ x^*_N \\ x^*_{N+1} \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \\ 0 \end{pmatrix} \tag{1}$$

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

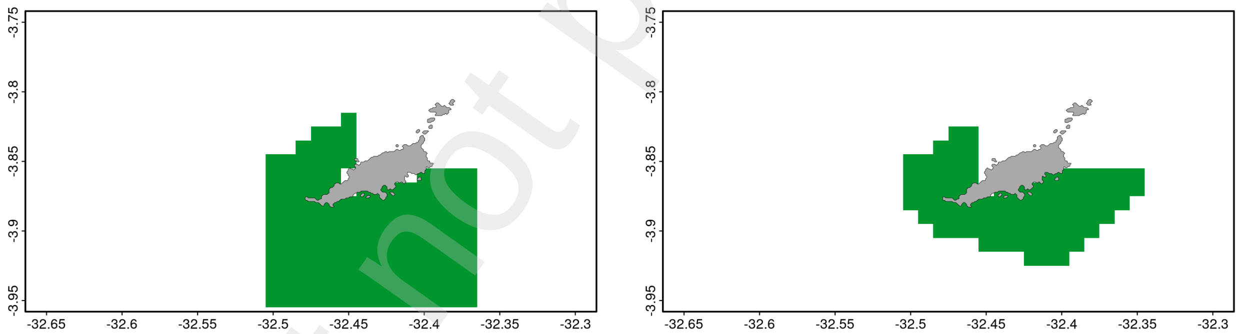
$\forall i \in \{1, ..., N\}$,

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

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 :

$$\begin{aligned} \boldsymbol{x^{*T} B^* (1 - x^*)} &= \sum_{i=1}^{N+1} \sum_{j=1}^{N+1} x^*_i b^*_{i,j} (1 - x^*_j) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i b_{i,j} (1 - x_j) + \sum_{i=1}^{N} x_i b^*_{i,N+1} \\ &= \boldsymbol{x^T B (1 - x) + x^T b^*} \end{aligned} \tag{2}$$

Denoting $b^* = (b^*_{1,N+1}, \cdots, b^*_{N,N+1})^T$ in Equation (2), we can see the new perimeter calculation was composed of two terms : the known quadratic term $x^T B(1-x)$ used previously to calculate the reserve perimeter but also $x^T b^*$ 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 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 B^* to remain a square and symmetric matrix eventually allowing to write the model in a compact form.



(a) Uncorrected reserve solution

(b) Corrected reserve solution

Figure 7 – 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 $1 + \ln(1 + \text{FC})$, compactness parameter $\beta = 10$, exact resolution provided by Cbc solver. Left panel reminds the original solution while the right panel shows the same solution but with correction implemented.

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 7 which shows side by side the solutions obtained with and without the proposed correction. It can be seen in Figure 7(b) that the selected reserve did not extend to the edge of the area like in the original model (Figure 7(a)) 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.