BIOREGIONALIZATION OF THE SIOFA AREA BASED ON VME INDICATOR TAXA

BERTA RAMIRO SÁNCHEZ, SKIPTON WOOLLEY, BORIS LEROY

Project PAE2021-01

APRIL 2023

MUSEUM NATIONAL D'HISTOIRE NATURELLE
43 Rue Cuvier, 75005 Paris, France

CSIRO OCEANS & ATMOSPHERE
Australia

Funded by
the European Union

Suggested citation:

Ramiro-Sanchez, B., Woolley, S. and Leroy, B (2023). *Bioregionalization of the SIOFA area based on vulnerable marine ecosystem indicator taxa (Project PAE2021-01)*. Muséum d'Histoire Naturelle, France.

Table of Contents

Acknowledgements

This work was carried out in collaboration with Dr Piers Dunstan (CSIRO, Australia). We are thankful for his guidance in the development and interpretation of the extension of regions of common profile models.

Executive summary

SIOFA's project PAE2020-02 involved the development of bioregionalizations of the SIOFA area based on VME indicator taxa. Three predictive modelling approaches were implemented to produce maps depicting potential bioregions in the area. These three approaches are known as "group first, then predict", "predict first, then group", and "analyse simultaneously" (Ferrier & Guisan, 2006). At the end of project PAE2020-02, preliminary work was shown on the development of a bioregionalization based on the "analyse simultaneously" approach. We used a point process extension of Regions of Common Profile model (RCPs; Foster et al., 2013). RCPs simultaneously generate site-based classifications by analysing biological and environmental data in a single model. A benefit of this method is because the propagation of uncertainty from the observed species data through to the predicted bioregions. Most two-step approaches ("group first, then predict" & "predict first, then group") fail to transfer uncertain into predictions upon clustering. In this report we show the next iteration of the model and interpret the outcomes in relation to the "group first, then predict" and "predict first, then group" modelling approaches carried out previously in project PAE2020-02.

The point process extension of RCPs allowed us to generate bioregional predictions in a single step based on presence-only records for 134 VME indicator species and 31,535 unique sites. Using model-selection we found that the "best" model for these data suggesting seven bioregions at depths shallower than 2,500 m across the extent of the area of study. There were only three RCP groups with distinct distributions, whereas the remaining four RCPs displayed a similar, low pan-oceanic probabilistic distribution. This pan-oceanic distribution and low probability of presence might be related to the limited number of samples in Areas Beyond National Jurisdiction within the Indian Ocean. The distribution of the seven RCP groups appeared to reflect the ocean currents of the Indian Ocean, which was observed as latitudinal patterns and distinct spatial predictions in regions with unique environmental signatures. Some of the RCPs were also restricted to depth ranges.

A bridge between the approaches "group first, then predict" and "predict first, then group", the RCP bioregions had similarities to the bioregions predicted with both approaches, although there was no direct equivalence between predictions. The RCP bioregions are likely to provide an approximation of the expected number and distribution of bioregions that occur in the Southern Indian Ocean. Yet, the limited number of samples in Areas Beyond National Jurisdiction, the uncertainty in the model selection, and the low probabilities predicted for half of the RCP groups warrant caution in the interpretation of the model. The intermediate step in the "group first, then predict" approach, where the raw data is classified into groups, is an advantage over this approach, as it does not necessitate of model selection. Therefore, we suggest that conservation efforts in the SIOFA area should represent all the delineated bioregions and subregions in the "group first, then predict" approach.

1. Purpose of the report

This report aims to address Terms of Reference 1 "*Building on ongoing VME mapping activities being undertaken by SIOFA, to undertake classification of key biological, geological, and oceanographic data to provide a spatial framework for classifying the area's marine environment into bioregions and, where possible, smaller spatial entities within bioregions, that make sense ecologically and are at a scale useful for regional planning*" of project PAE2021-01 ("*Bioregionalization and Management of Vulnerable Marine Ecosystems (VMEs)").*

2. Introduction

The United Nations require states and regional fisheries management organisations (RFMOs) agreements to implement measures to prevent significant adverse impacts on vulnerable marine ecosystems (VMEs) (UNGA, 2007). Following UNGA resolutions, the Southern Indian Ocean Fisheries Agreement (SIOFA) requires scientific advice informed by maps of where VMEs are known to occur, or likely to occur, in the Agreement Area. Biogeographical classifications, or bioregionalizations (Woolley et al., 2020), are the building blocks for the planning and implementation process of management measures and are highly connected to the development of marine protected areas (Rice et al., 2011). This is because biogeographical classifications enable the identification of units that should be represented in a network, ensuring the protection of biogeographically unique areas and the development of a network that considers representativity, connectivity and replication of sites (Rice et al., 2011).

There are many predictive approaches to map bioregions, which can be generally classified into three categories (Ferrier & Guisan, 2006). The first one consists in grouping biological features into bioregions first, and then spatially predict these bioregions ("group first, then predict"). The second one consists in spatially predicting all biological features individually first, and then group them with a clustering approach ("predict first, then group"). The last one consists in grouping and predicting bioregions in a single modelling approach ("analyse simultaneously"). Each method has its set of advantages and shortcomings that make them complementary in the information they provide. Ultimately, however, the kind of data available will often limit the choice of method (Woolley et al., 2020).

Previous work for SIOFA (project PAE2021-01) developed bioregionalization schemes based on the three predictive approaches mentioned above (Ramiro-Sánchez & Leroy, 2022). One of the methods was however modified to follow state-of-the-art recommendations. Specifically, a "group first, then predict" modelling approach known as Generalized Dissimilarity Models was initially planned; however, it has been recently criticized for their inadequacy in identifying relevant predictors of turnover in species composition between regions (Woolley et al., 2017). Consequently, the methods of the consultancy (Generalized Dissimilarity Models or other appropriate techniques; Deliverable 3.3) were modified to include Regions of Common Profile (Foster et al., 2013), which corresponds to the approach "analyse simultaneously". Attempting to develop bioregions based on RCPs is important for three reasons. First, they offer repeatability because the mathematical model defines formally bioregions and the relationship with the environment (Hill et al., 2020; Warton et al.,

2015). Second, currently, they are the only approach able to quantify uncertainty given the probabilistic nature of the models (i.e., a site has some chance of belonging to more than one bioregion), calculating 95% confidence intervals of the predictions that allow to have a more conservative or a more optimistic scenario of the predictions. These appropriate measures of uncertainty are essential for spatial management applications to understand the risks associated with applying or not a certain measure in a location (Hill et al., 2020). Third, the simultaneous grouping and prediction for taxa has the potential of integrating dispersal limitation in the prediction. Thus, in this report, we provide the next iteration of the preliminary results from RCP model exploration and compare it with the other two bioregionalization approaches.

3. Bioregionalization of the Southern Indian Ocean

3.1 An extension of the Regions of Common Profile (RCPs; Foster et al., 2013)

As a third approach to generate biologically data-driven bioregions, we used a Poisson Point process extension of the Regions of Common Profile model (RCPs; Foster et al., 2013) . The Regions of Common Profile (RCPs; Foster et al., 2013) approach is a multivariate adaptation of a finite mixture-of-experts model (Jacobs et al., 1991) that jointly considers multi-species and environmental data. The RCP approach allows mapping of bioregions within which species share the same probability of being sampled at any site within that bioregion and predict these groups at unsampled sites through the relationship with the environmental data in a single analysis (Foster et al., 2013, 2017). This is advantageous compared to two-step classifications, since they do not propagate uncertainty from statistical models to the grouping step (clustering) (Warton et al., 2015). Moreover, the RCP approach allows to incorporate sampling artefacts in the analysis that otherwise may have the effect of conflating patterns in sampling artefacts with ecological patterns, leading to incorrect inferences (Foster et al., 2017).

The RCP framework described above considers scientifically collected data for which there is information on where species are present and where species were not observed (absences, e.g., Warton et al., 2013). However, at large scales like the one considered here, data are a compilation of records from scientific surveys and ad-hoc collections, the latter being typically kept in natural history museum collections and in on-line repositories. These ad-hoc data are missing in information of where species were not observed (absence) and, as such, it is not possible to estimate the probability of occupancy for these presence-only data (Guillera-Arroita et al., 2015). Instead, these presence-only data require special consideration, which has not been formulated into Foster et al.'s (2013) approach (Woolley et al., 2020).

A promising solution is to extend Foster et al.'s (2013) method to presence-only data cases by implementing a Poisson point process and accounting for sampling imbalance in the models (Woolley et al., 2020). Here the RCP model was extended to be an inhomogeneous Poisson point process (Cressie, 1993; Warton & Shepherd, 2010), where species occurrences (i.e., presence locations) are assumed to follow a Poisson distribution. The resultant model generates a probability of sighting a species within a bioregion (a density, based on the

presence points) rather than a probability of occurrence of a species within a bioregion. To account for the source of variation in the data collection, we estimated sampling bias for each grid cell using the target-group approach, as it has been shown to perform the best at emulating sampling effort (Barber et al., 2022). Specifically, we used all records of VMEs indicators in our dataset (at any taxonomic level) and used a 2D kernel density estimation to convert single points into a continuous probability surface (see Barber et al. (2022) for methods and code). By including it in the model, it is possible to estimate the distribution of bioregions with the confounding of sampling bias much reduced (Foster et al., 2017). Further details of RCP models can be found in Foster et al. (2013) and Foster et al. (2017).

3.2 Data analysis

We selected environmental variables from the World Ocean Atlas Data (WOA 2018) available at 1° spatial resolution: apparent oxygen utilization, percentage of oxygen, dissolved oxygen (µmol/kg), density, nitrate, phosphate, silicate, salinity (unitless), temperature (°C). We included mean net primary production (g C $m^{-2}year^{-1}$, NPP) generated from a vertically generalized production model (VGPM) across the years 2003 to 2010 (http://www.science.oregonstate.edu/ocean.productivity/). NPP is a function of satellite-derived chlorophyll (SeaWiFS and Modis). Depth was included from the GEBCO bathymetric model (GEBCO, 2008) originally available at a spatial resolution of 1 minute. We also included latitude and longitude as variables. All environmental variables were reprojected to Albers equal area projection and to same spatial resolution of the bathymetric model.

We checked for multicollinearity in the environmental variables using Spearman rank correlation. Strongly correlated variables ($|\rho|>0.7$) were removed from analyses to avoid issues with co-linearity of model coefficients (Dormann et al., 2013). For tied variables, we selected the variable the most biologically meaningful for VME indicator taxa. Our final set of variables was composed of seven variables: depth, dissolved oxygen, temperature, and salinity at seafloor, mean net primary production, and latitude and longitude. Specifically, we fitted a model with linear and quadratic polynomials of the environmental variables, with latitude and longitude as an interaction term, for 134 species that had more than 30 occurrences in the dataset. We restricted the model to depths shallower than 2,500 m because of the large number of grid cells for prediction.

When fitting an RCP model, the number of groups (i.e., RCPs) needs to be specified. The optimal number of RCPs is rarely known *a priori* but can be inferred by fitting models with varying numbers of groups and using the Bayesian Information Criteria (BIC) to choose the "best" number of groups (Foster et al., 2013; Hill et al., 2017). Here we ran models with 2 to 12 RCPs with a meaningful number of environmental variables. Specifically, we described the probability of sighting a species within a bioregion as a function of linear and quadratic polynomials of the explanatory variables for each of the 2-12 RCPs, with 30 random starts. Random starts are necessary to avoid getting stuck in a local likelihood maximum (Foster et al., 2013).

Model uncertainty was estimated via 20 Bayesian bootstrap replications, providing 95% confidence intervals (CI). We spatially predicted the preliminary optimal RCPs with the

selected number of groups. Finally, we calculated the species composition of each RCP multiplying the probability that a species occurs in a RCP by the probability of a site belonging to that RCP, and summed across all sites ($\Sigma_{RCPn}\alpha_j \times Posterior\ probability_j$).

We used the R "ecomix" package (https://github.com/skiptoniam) and "RCPmod" R package (https://github.com/cran/RCPmod) to fit and predict the RCP models. All analyses were conducted in the R computing environment v4.1.2 (R Development Core Team, 2021) and models run in the High-Performance Computing and Algorithmic Platform cluster of the French Natural History Museum (Plateforme de Calcul Intensif et Algorithmique PCIA, Muséum National d'Histoire Naturelle, Centre national de la recherche scientifique, UAR 2700 2AD, CP 26, 57 rue Cuvier, F-75231 Paris Cedex 05, France).

4. Bioregionalization results

4.1 Model selection

Out of 360 models fitted (30 runs per each number of RCPs), 326 were successfully modelled. Models fitted with 8 to 12 number of RCPs were considered misfits as the BIC spanned a large range of negative and positive values (Figure 1A), and the log-likelihoods were greater than zero (Figure 1B).

The lowest BIC (excluding RCPs ≥ 8) suggested 4 or 7 RCPs as the optimal number of groups in the data (Figure 1A). Inspection of the RCP model with 4 RCPs revealed a model misfit as there was only one RCP group predicted and the remaining RCPs were zero groups. This was likely a result of a spike in the log-likelihood, which can be observed in Figure 1B as a solitary negative value. Although model selection in mixture-of-experts models (i.e., RCPs) is performed using BIC values, BIC is not entirely reliable because it sometimes can suggest one RCP for a single site (i.e., grid cell) (McLachlan & Peel, 2000), as it was the case here. Thus, several models need to be inspected to understand whether they are a misfit. Consequently, we focused on the next lowest BIC, which corresponded to 7 RCPs. Inspection of fitted models with 7 RCPs revealed BIC values similar to BIC values from some models fitted with 6 and 8 RCPs. As a result, we selected ten models for further inspection (one model with 6 RCPs, four models with 7 RCPs and five models with 8 RCPs) and performed 20 Bayesian bootstrap samples to calculate uncertainty in the predictions resulting from the models. Next, we spatially predicted the ten models. Visual evaluation of the ten models indicated one model fitted with 7 RCPs as a potential good fit.

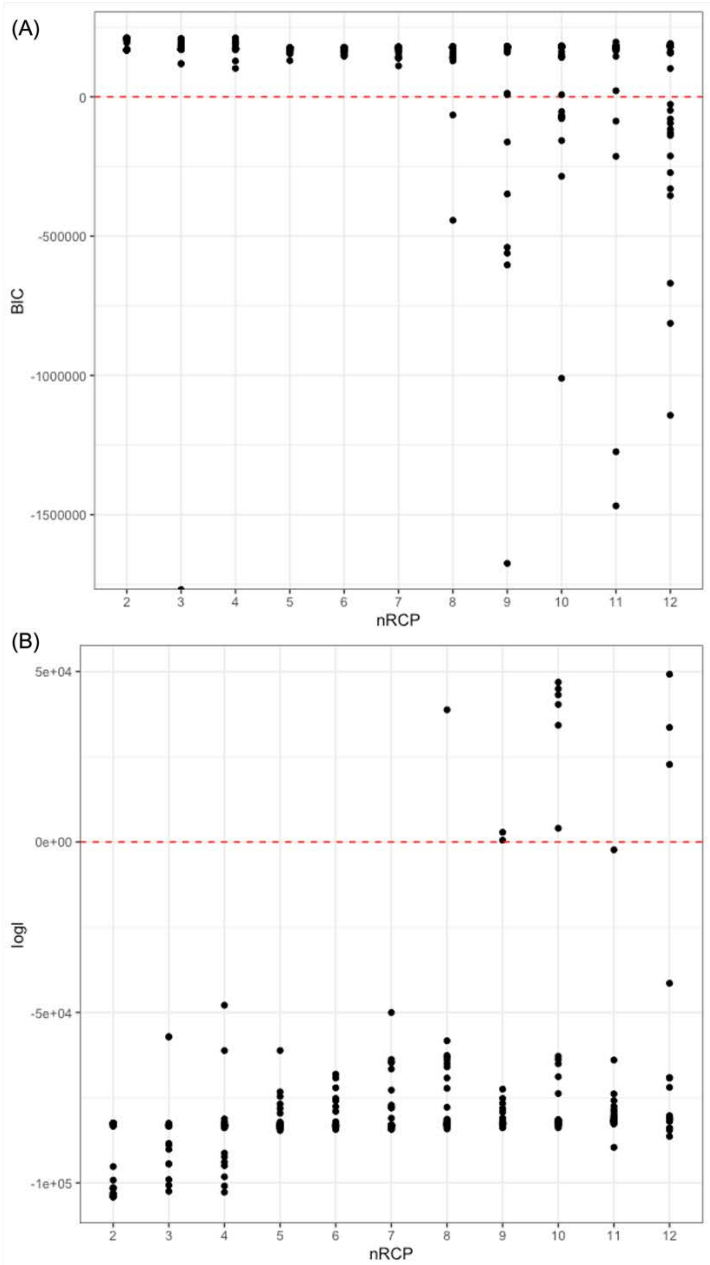

Figure 1. Selection of "best" number of RCPs based on (A) the Bayesian Information Criteria (BIC) and (B) maximum log likelihood. We truncated the y-axis in (A) to display better the BIC values and the characteristic exponential decay curve they should follow. We provide as an inset the full spectrum of models, where RCPs ≥ 10 are misfits given their large spanning positive and negative BIC values. Similarly, models for RCPs ≥ 10 exhibit log-likelihoods greater than zero also indicating a misfit of the models (B).

4.2 Model predictions

The final model comprised 7 RCP groups. The predicted spatial distribution of the RCPs across the Indian Ocean is shown in Figure 2. The probabilistic model output has been summarized by assigning each cell a hard class based on its most likely RCP (Figure 2A). The probability of the hard classification for each cell is also shown (Figure 2B). Maps of the average probability of occurrence (and 95% CI's) of each RCP are shown separately in Appendix A (Figure A1-Figure A7). The species composition for each RCP can also be found in Appendix A, Table A1.

RCP1 was predicted across all latitudes and longitudes of the study area with the lowest probabilities (< 0.4) but absent from the shallowest depths at Walter's Shoal, the Mozambique Ridge, Del Cano and Crozet Plateau (Figure A1). Net primary productivity appeared to be a key driver of RCP1's distribution (Figure 3).

RCP2 was distributed latitudes south of ~15°S, specifically south of Madagascar, Walters Shoal, the shallowest parts of the Southwest Indian Ridge and Del Cano and Crozet Plateau, Mozambique Ridge, Broken Ridge and east of Broken Ridge with the junction of Southeast Indian Ridge (Figure A2). RCP2 spanned all depths, although there was a distinct decrease in probability at depths > 2,000m (Figure 3). In addition, this group was characterised by high levels of dissolved oxygen, lower temperatures, and low net primary productivity (Figure 3).

RCP3 had the same pan-oceanic weak predicted probability of presence as RCP1, that is, it was excluded from the shallowest areas within SIOFA, except at Saya de Malha (Figure A3). This RCP revealed highest probability of detection at shallower areas (< 500 m) and great variability in its response to the remaining environmental covariates (Figure 3).

RCP4 was predicted across all the Indian Ocean very weakly (Figure A4) and across several depths (Figure 3). The stronger prediction was on Ninety-east Ridge and Broken Ridge. Although not within SIOFA's Convention Area, the RCP was not predicted along the east coast of Africa (Figure A4).

RCP5 followed the same distribution and similar response to the environment as RCP4 (Figure 3), but it was also predicted along eastern Africa (Figure A5). The probability of RCP membership was low (< 0.2, Figure A5).

RCP6 had a northern distribution, with high probability of detection at latitudes north of the Equator, particularly in the Carlsberg Ridge, Chagos-Laccadive Ridge, and northern Mascarene plateau (Seychelles) within SIOFA (Figure A6). The response to the environmental covariates appears to resemble the geographical location of its distribution (Figure 3).

Finally, RCP7 had an anti-meridian prediction, with strong probability of presence along the African coast, reaching the Agulhas Plateau, northwest of Madagascar, Saya de Malha and northwest Australia (Figure A7). The strongest predicted latitudes were approximately from 5°N to 20°S. The distribution of RCP7 seemed to be driven mainly by dissolved oxygen, salinity, and net primary productivity (Figure 3).

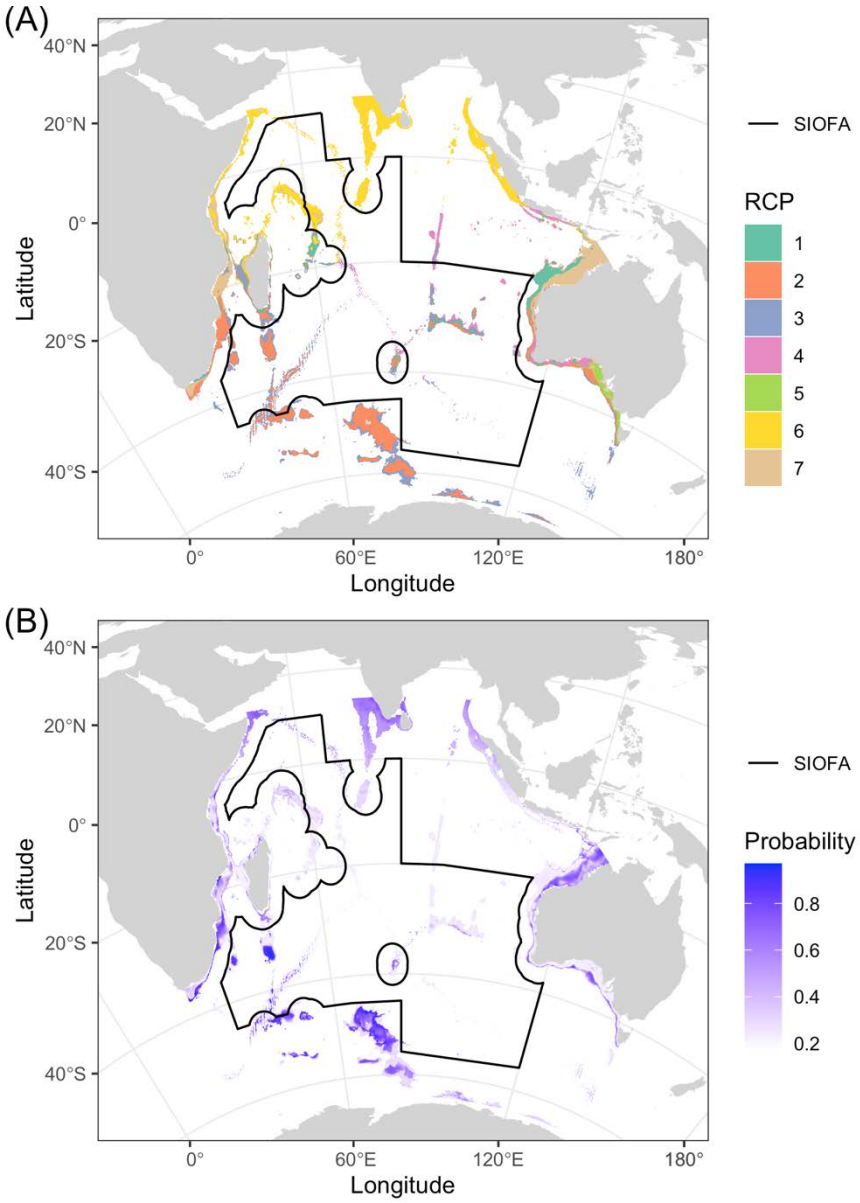

Figure 2. (A) Hard classed predictions of the spatial locations of each of the seven RCP groups and (B) the probability of occurrence associated with the hard classification. Each cell was assigned its most probable RCP group (a) and the probability of the group with the highest probability of presence at each site . White areas are not included in the model (depths > 2,500 m).

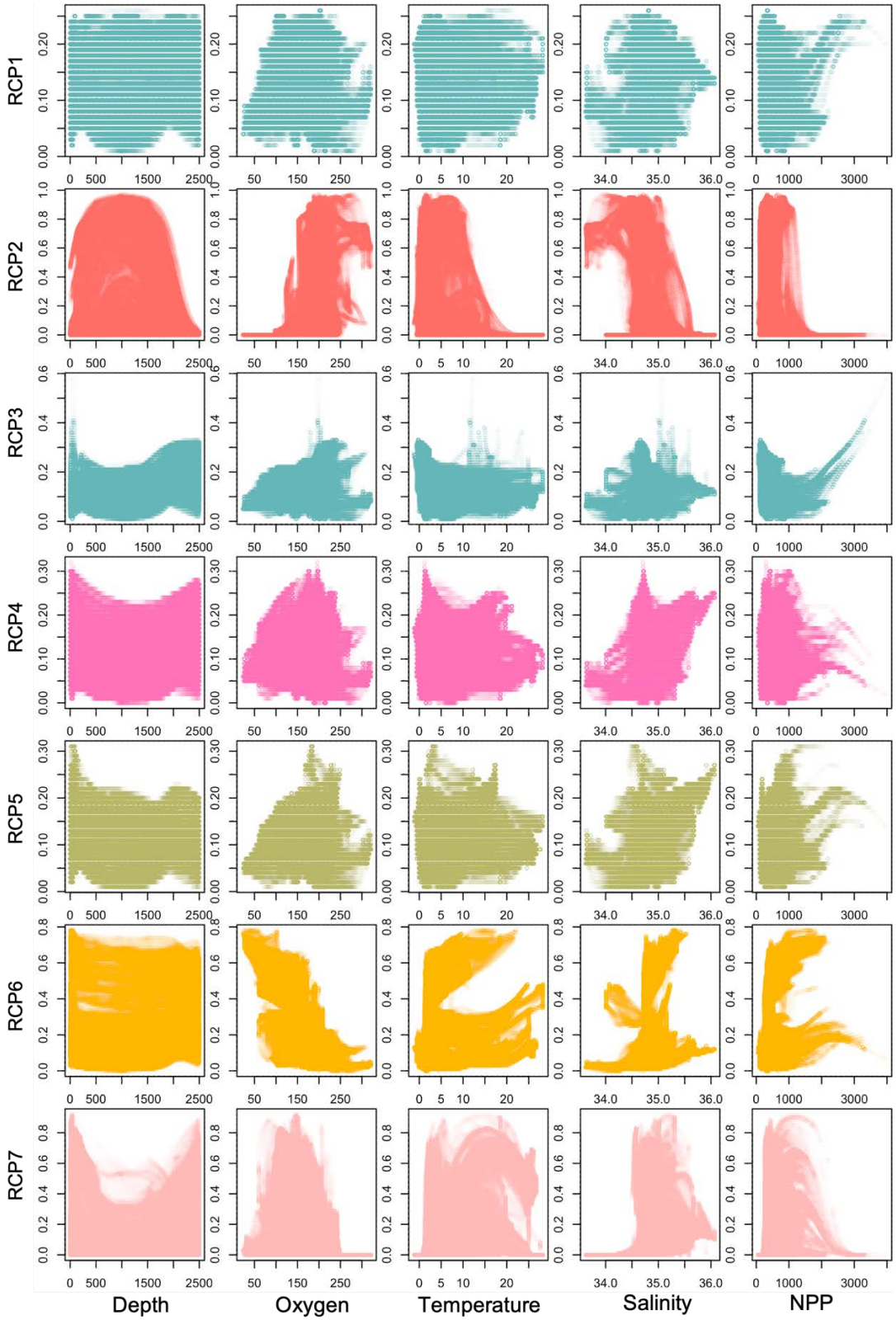

Figure 3. Response of each RCP to depth, and yearly mean of bottom dissolved oxygen, bottom temperature, bottom salinity, and surface net primary production. Plots were generated by predicting RCP membership for each trawl site based only on its environmental covariates.

5. Discussion

SIOFA's project PAE2020-02 investigated three predictive bioregionalization modelling approaches to classify biological, environmental, and oceanographic data into bioregions (Ramiro-Sánchez & Leroy, 2022). Results from the "analyse simultaneously" approach were preliminary at the time due to a late incorporation of the methodology. The present report completes the work and discusses the results in the context of the three predictive bioregionalization schemes (Ferrier & Guisan, 2006): "group first, then predict", "predict first, then group" and "analyse simultaneously". For the "analyse simultaneously" approach, we implemented an extension of the Regions of Common Profile (RCPs; Foster et al., 2013), which analyses and groups simultaneously biological and environmental information for presence-only data. This method allowed to group and predict, in a single step, bioregions based on 134 VME indicator species, suggesting seven groups at depths shallower than 2,500 m across the extent of the area of study. The hard-class classification revealed both a bathymetric and latitudinal distribution of the seven RCP groups. This classification has similarities with the bioregions predicted with the "group first, then predict" and the "predict first, then group" approach, but presents clear differences. Due to the uncertainties surrounding the model selection step, the present RCP classification remains to be validated. Nonetheless, it constitutes an important step towards the development of model-based bioregions using state-of-the-art methodologies for broad-scale presence-only data. The collection of new data in Areas Beyond National Jurisdiction will enable its refinement. We therefore recommend that conservation efforts should represent all the delineated bioregions and subregions detected in the "group first, then predict" approach.

Bioregions based on an extension of the "Regions of Common Profile"

There were only three RCP groups with distinct distributions (Figure A2, Figure A6 and Figure A7), whereas the remaining four RCPs displayed a similar pan-oceanic probabilistic distribution (Figure A1, Figure A3 – Figure A5). The low probability of presence for these four RCPs was observed in the homogeneous poor response to the environmental covariates used for their prediction (Figure 3). Unlike the previous preliminary model iteration (PAE2020-02; Ramiro-Sánchez & Leroy, 2022), where site specificities were not considered in the model (i.e., latitude and longitude were not predictors), the addition of the interaction term might have intensified the uncertainty in the spatial predictions. In addition, data shortfalls might have contributed to the low probabilities, wherein those well characterized sites likely formed the basis of the distinctly bounded RCPs (RCP2, RCP6 and RCP7). In fact, the pan-oceanic distribution and low probability of presence of RCP3, RCP4 and RCP5 might be related to the limited number of samples in Areas Beyond National Jurisdiction for our study area.

The RCP approach uses a single model to group and predict sites based on the species observed at a site and should be preferred upon methods that do this using two separate models, as per the "group first, then predict" and "predict first, then group" methods. The bioregions based on RCPs shared similarities with outputs from both the "group first, then predict" bioregions (Figure 4) and the "predict first, then group" bioregions (Figure 5), although there was no direct equivalence between predictions. For instance, on one hand, RCP2 and RCP3 resembled subregion 1.2, and RCP6 resembled subregion 1.1 of "group first,

then predict", whereas RCP1-RCP3 also resembled cluster 2, cluster 3 and cluster 5 of the "predict first, then group" approach.

The distribution of the seven RCP groups appeared to reflect the ocean currents of the Indian Ocean, which was observed as latitudinal patterns and distinct spatial predictions in regions with unique environmental signatures. This explains the weak environmental response curves (Figure 3), as the properties of currents and water masses are displayed but not their unique geographic location resulting from the influence of the seafloor topography and other factors, such as the presence of the Indian monsoon and the influx of the Antarctic Circumpolar Current from the south. Depending on taxa and depth of topographic features, these may act as potential barriers or as stepping-stones to dispersal by restricting the flow of currents and/or facilitating habitat for dispersing larvae across vast distances (McClain & Hardy, 2010). The signature of currents was evident for RCP6, RCP7 and RCP1 (Figure 2). RCP6 followed the westward flow of the South Equatorial Current and the eastward flow of the East Madagascar Current (Talley et al., 2011). RCP7 appeared to follow the North East Madagascar Current that further splits into a portion that flows into the Mozambique Channel, which eventually joins the Agulhas Current, and another portion that continues north joining the East African Coastal Current (Talley et al., 2011). Finally, RCP1, also appeared to follow the South Equatorial Current path, but at different depths.

The RCPs not only occupied latitudinal bands but were restricted to depth ranges, although the limitations of using global bathymetric models were evident (panel "Depth", Figure 3). The role of bathymetry in the distribution of benthic marine species is well known, where biodiversity at local and regional scales is often associated with depth bounds (Carney, 2005). Nevertheless, the depth at which species turnover occurs varies among taxa and latitude (e.g., O'Hara et al., 2011; Williams et al., 2001; Woolley et al., 2013). In our study, RCP2 and RCP3 showed the greatest patterns in depth bounds, whereas the patterns of the remaining RCPs reflected less nuanced differences, with a combination of other factors driving their distributional patterns.

The previous iteration of the model suggested seven RCPs for depths shallower than 3,500 m (Ramiro-Sánchez & Leroy, 2022). The first-iteration model was calibrated with the same environmental predictors as the present iteration, although the environmental data was sourced from the Bio-Oracle global oceanographic dataset (Assis et al., 2018; Tyberghein et al., 2012). Besides the environmental data sources, the main difference between the previous and the present iteration stems from the inclusion in the latter of the interaction term between latitude and longitude. The previous iteration revealed bimodal patterns within the same RCP group. For instance, there was one RCP predicted in the northern Indian Ocean and in the Southern Ocean. This prediction probably resulted from certain environmental variables presenting similar characteristics such as a minimum in dissolved oxygen at both latitudinal limits of the Indian Ocean. Thus, by including latitude and longitude, we attempted to ensure that specificities of sites were taken into account in the model.

Mixture-of-experts models are difficult to implement and are computationally intensive. Furthermore, as illustrated in the results, model selection is not straightforward. Whilst model selection can be informed using the BIC values, in mixture-of-experts different BIC values might be obtained for the same fitted model (Foster et al., 2017). Ideally, the model

run with the lowest BIC would be chosen, but because of the spikes in the log-likelihood some of these values are not reliable. As a result, a range of models need to be investigated. In our case, we explored ten models and chose one optimal model out of the investigated models (with a different number of RCP groups). However, for the same number of RCP groups, another model could have been selected (based on the BIC) that would have displayed a slightly different spatial configuration. Thus, we conclude that the present model remains tentative for the data used in the analysis.

Bioregions of the Southern Indian Ocean based on VME indicator taxa: where do we stand?

Project PAE2020-02 aimed to develop three predictive bioregionalization approaches to generate bioregions based on VME indicator taxa for the SIOFA area. Despite the scarcity of data, we managed to identify, predict, and map several biogeographical regions, although with varying levels of confidence across the three techniques. The ultimate purpose of these different classifications will be to support management decisions towards the protection of benthic areas. As we have discussed extensively (see Ramiro-Sánchez & Leroy, 2022 and this report), the results differ between the different approaches, and especially between bioregions of the "group first, then predict" approach and the bioregions of the "predict first, then group" approach, whereas the "analyse simultaneously" bioregions are a mixture of the two other classifications.

The "group first, then predict" approach suggested several bioregions and subregions whose distribution could be explained by depth and water masses in the Indian Ocean (Ramiro-Sánchez & Leroy, 2022). Because these bioregions were based on the known patterns of species co-occurrence in the area, they accounted for the known limits of species dispersal. The resulting bioregions can be interpreted as known bioregions to occur in the area based on observed samples and, as such, they provide an approximation of the expected number of bioregions to occur in the Southern Indian Ocean. In contrast, the "predict first, then group" approach does not account for actual distribution patterns of VME indicator taxa resulting from dispersal limitations and biotic interactions. Rather, it provides a classification of the environment into distinct groups at a fine resolution. Thus, this second set of methods can be interpreted as an illustration of the diversity of habitats for VME taxa, based on the available environmental data in the study area. In that sense, it is complementary to the "group first, then predict" approach. Finally, the bioregions based on RCPs, with similarities to both approaches, integrate species' environmental tolerance and dispersal limitation in the prediction. These bioregions are likely to also provide an approximation of the expected number and distribution of the bioregions to occur in the Southern Indian Ocean. Yet, as with the other two approaches, the limited number of samples in Areas Beyond National Jurisdiction, the uncertainty in the model selection, and the low probabilities predicted for half of the RCP groups also warrant caution in the interpretation of the model with the current data.

In summary, the bioregionalization maps suggest that the SIOFA area has a great diversity of bioregions, which is insightful for the ultimate objective of the identification of key areas for conservation. The three methodologies have offered complementary insights into the distribution of VME taxa in the Southern Indian Ocean. With the present data, bioregions

detected with both the “group first, then predict” and “analyse simultaneously” approaches appear to reflect actual patterns of geographical distinction in the distribution of VME indicator taxa. Although both methods warrant caution in their interpretation, we would suggest that conservation efforts in the SIOFA area should represent all the delineated bioregions and subregions in the “group first, then predict” approach, based on the available data. Although it is a two-step approach, the grouping step does not necessitate of model selection, reducing uncertainty to some extent. Nonetheless, there were two main limits to these results. First, the resolution at which we delineated bioregions and predicted them spatially is coarse (1° latitude-longitude). In this respect, the RCP classification offers more nuance in the identification of depth zones in certain places, such for RCP1 and RCP7 in the northwest of Australia (Figure 2). Second, there were large areas of uncertainty in our predictions, i.e., areas where none of our models were able to predict the suitability for any of the bioregions. Here too, the consideration of the RCP classification may inform future steps on how best to capitalise on the results and circumvent these limits to inform conservation plans for SIOFA VME taxa. Finally, we insist on the very limited availability of data in the SIOFA area, which implies that interpretation must be exerted with caution, and we urge for the necessity of acquiring more data on deep-sea benthic taxa in the SIOFA area. As new data become available, bioregionalization models will be updated and provide more confidence in their interpretation.

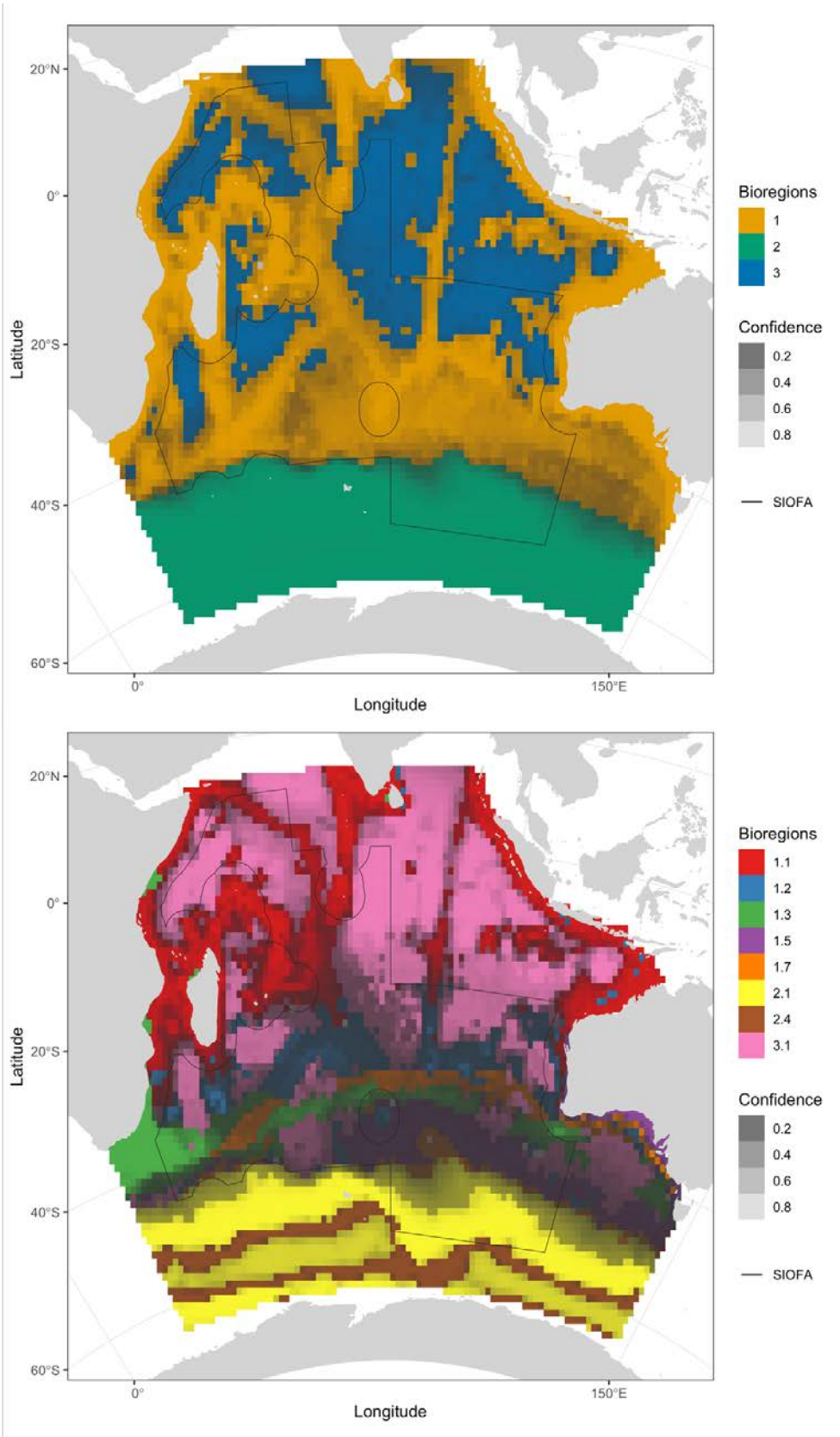

Figure 4 . Hard classed predictions of the spatial locations of the first-level (top panel) and nested second-level (bottom panel) bioregions predicted using the "group first, then predict" bioregionalization approach. See Ramiro-Sanchez and Leroy (2022) for methods.

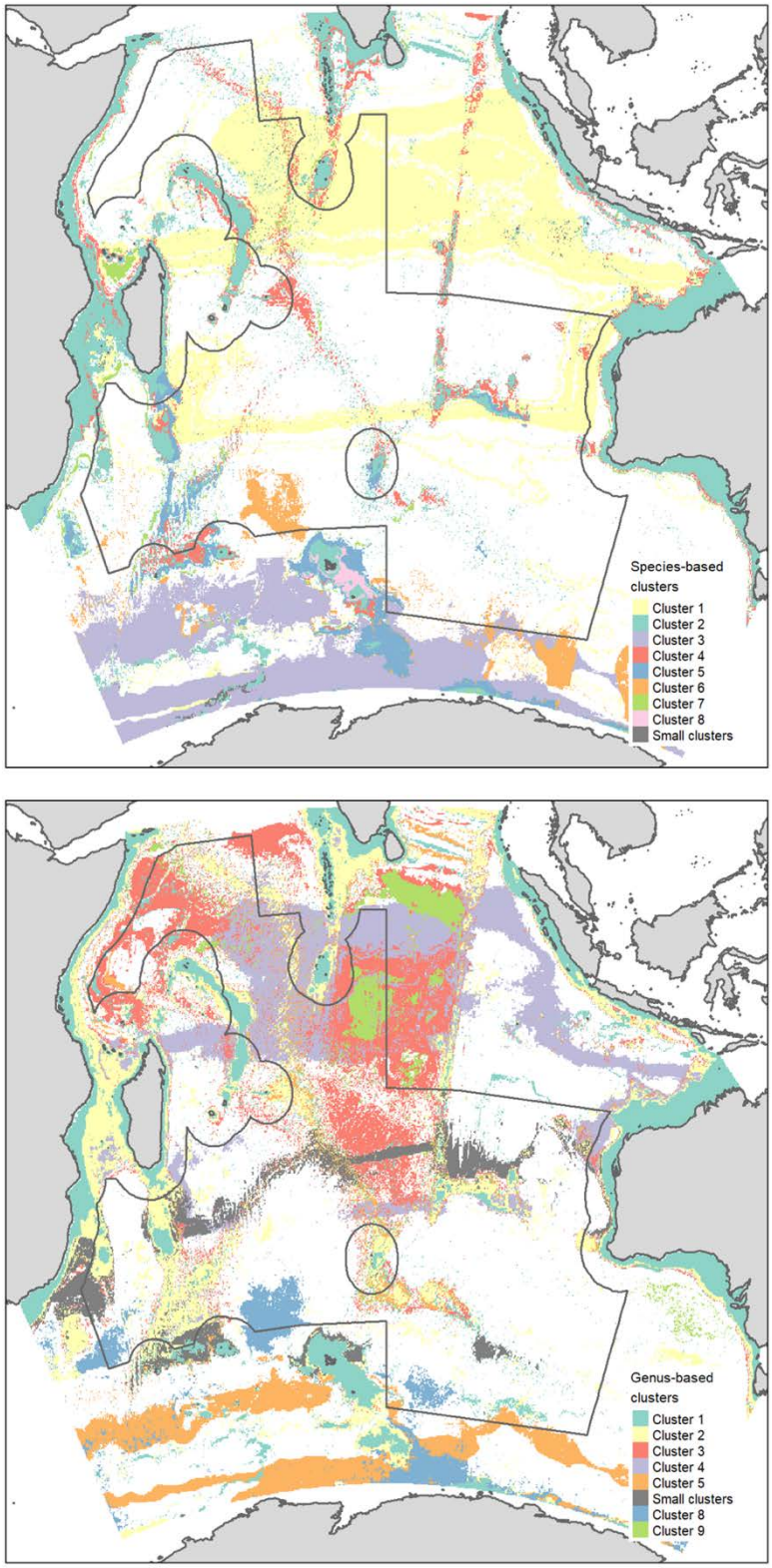

Figure 5 . Bioregions at species (top panel) and genus (bottom panel) taxonomic level predicted using the "predict first, then group" bioregionalization approach. See Ramiro-Sanchez and Leroy (2022) for methods.

6. Recommendations

Following on the work carried out we recommend the Scientific Committee to:

- Note that the work carried out for this Terms or Reference goes in conjunction with a previous consultancy (PAE2020-02).
- Consider the bioregionalization map produced by the "Group first, then predict" modelling approach, as model selection for the RCPs remains tentative.
- Note that, despite the model's overprediction, the bioregionalization map produced by the "Predict first, then group" approach will be more complete than any environmental classification.

References

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & de Clerck, O. (2018). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, *27*(3), 277–284. https://doi.org/10.1111/GEB.12693

Barber, R. A., Ball, S. G., Morris, R. K. A., & Gilbert, F. (2022). Target-group backgrounds prove effective at correcting sampling bias in Maxent models. *Diversity and Distributions*, *28*(1), 128–141. https://doi.org/10.1111/DDI.13442

Carney, R. S. (2005). Zonation of Deep Biota on Continental Margins. In *Oceanography and Marine Biology* (1st ed., pp. 221–288). CRC Press. https://doi.org/10.1201/9781420037449-8

Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., Mcclean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/J.1600-0587.2012.07348.X

Ferrier, S., & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, *43*(3), 393–404. https://doi.org/10.1111/J.1365-2664.2006.01149.X

Foster, S. D., Givens, G. H., Dornan, G. J., Dunstan, P. K., & Darnell, R. (2013). Modelling biological regions from multi-species and environmental data. *Environmetrics*, *24*(7), 489–499. https://doi.org/10.1002/ENV.2245

Foster, S. D., Hill, N. A., & Lyons, M. (2017). Ecological grouping of survey sites when sampling artefacts are present. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *66*(5), 1031–1047. https://doi.org/10.1111/RSSC.12211

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., Mccarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, *24*(3), 276–292. https://doi.org/10.1111/GEB.12268/SUPPINFO

Hill, N. A., Foster, S. D., Duhamel, G., Welsford, D., Koubbi, P., & Johnson, C. R. (2017). Model-based mapping of assemblages for ecology and conservation management: A

case study of demersal fish on the Kerguelen Plateau. *Diversity and Distributions*, *23*(10), 1216–1230. https://doi.org/10.1111/DDI.12613

Hill, N. A., Woolley, S. N. C., Foster, S. D., Dunstan, P. K., McKinlay, J., Ovaskainen, O., & Johnson, C. (2020). Determining marine bioregions: A comparison of quantitative approaches. *Methods in Ecology and Evolution*, *11*(10), 1258–1272. https://doi.org/10.1111/2041-210X.13447

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive Mixtures of Local Experts. *Neural Computation*, *3*(1), 79–87. https://doi.org/10.1162/NECO.1991.3.1.79

McClain, C. R., & Hardy, S. M. (2010). The dynamics of biogeographic ranges in the deep sea. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1700), 3533–3546. https://doi.org/10.1098/RSPB.2010.1057

McLachlan, G. J., & Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons.

O'Hara, T. D., Rowden, A. A., & Bax, N. J. (2011). A Southern Hemisphere Bathyal Fauna Is Distributed in Latitudinal Bands. *Current Biology*, *21*(3), 226–230. https://doi.org/10.1016/J.CUB.2011.01.002

Ramiro-Sánchez, B., & Leroy, B. (2022). *SIOFA bioregionalization and VME project*. https://siofa.org/sites/default/files/files/VMEMapping_FullReport.pdf

Rice, J., Gjerde, K. M., Ardron, J., Arico, S., Cresswell, I., Escobar, E., Grant, S., & Vierros, M. (2011). Policy relevance of biogeographic classification for conservation and management of marine biodiversity beyond national jurisdiction, and the GOODS biogeographic classification. *Ocean & Coastal Management*, *54*(2), 110–122. https://doi.org/10.1016/J.OCECOAMAN.2010.10.010

Talley, L. D., Pickard, G. L., Emery, W. J., & Swift, J. H. (2011). Indian Ocean. In *Descriptive Physical Oceanography: An Introduction* (6th ed., pp. 1–555). Elsevier Ltd. https://doi.org/10.1016/C2009-0-24322-4

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & de Clerck, O. (2012). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, *21*(2), 272–281. https://doi.org/10.1111/J.1466-8238.2011.00656.X

Resolution 61/105 Sustainable fisheries, including through the 1995 agreement for the Implementation of the Provisions of the United nations Convention on the law of the Sea of 10 December 1982 relating to the Conservation and management of Straddling fish Stocks and highly migratory fish Stocks, and related instruments, UNGA A/RES/61/105 21 (2007).

Warton, D. I., Foster, S. D., De'ath, G., Stoklosa, J., & Dunstan, P. K. (2015). Model-based thinking for community ecology. *Plant Ecology*, *216*(5), 669–682. https://doi.org/10.1007/S11258-014-0366-3/FIGURES/3

Warton, D. I., Renner, I. W., & Ramp, D. (2013). Model-Based Control of Observer Bias for the Analysis of Presence-Only Data in Ecology. *PLOS ONE*, *8*(11), e79168. https://doi.org/10.1371/JOURNAL.PONE.0079168

Warton, D. I., & Shepherd, L. C. (2010). Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *Https://Doi.Org/10.1214/10-AOAS331*, *4*(3), 1383–1402. https://doi.org/10.1214/10-AOAS331

Williams, A., Koslow, J. A., & Last, P. R. (2001). Diversity, density and community structure of the demersal fish fauna of the continental slope off western Australia (20 to 35°S). *Marine Ecology Progress Series*, *212*, 247–263. https://doi.org/10.3354/MEPS212247

Woolley, S. N. C., Foster, S. D., Bax, N. J., Currie, J. C., Dunn, D. C., Hansen, C., Hill, N. A., O'Hara, T. D., Ovaskainen, O., Sayre, R., Vanhatalo, J. P., & Dunstan, P. K. (2020). Bioregions in Marine Environments: Combining Biological and Environmental Data for Management and Scientific Understanding. *BioScience*, *70*(1), 48–59. https://doi.org/10.1093/BIOSCI/BIZ133

Woolley, S. N. C., Foster, S. D., O'Hara, T. D., Wintle, B. A., & Dunstan, P. K. (2017). Characterising uncertainty in generalised dissimilarity models. *Methods in Ecology and Evolution*, *8*(8), 985–995. https://doi.org/10.1111/2041-210X.12710

Woolley, S. N. C., Mccallum, A. W., Wilson, R., O'Hara, T. D., & Dunstan, P. K. (2013). Fathom out: biogeographical subdivision across the Western Australian continental margin – a multispecies modelling approach. *Diversity and Distributions*, *19*(12), 1506–1517. https://doi.org/10.1111/DDI.12119

Appendix A

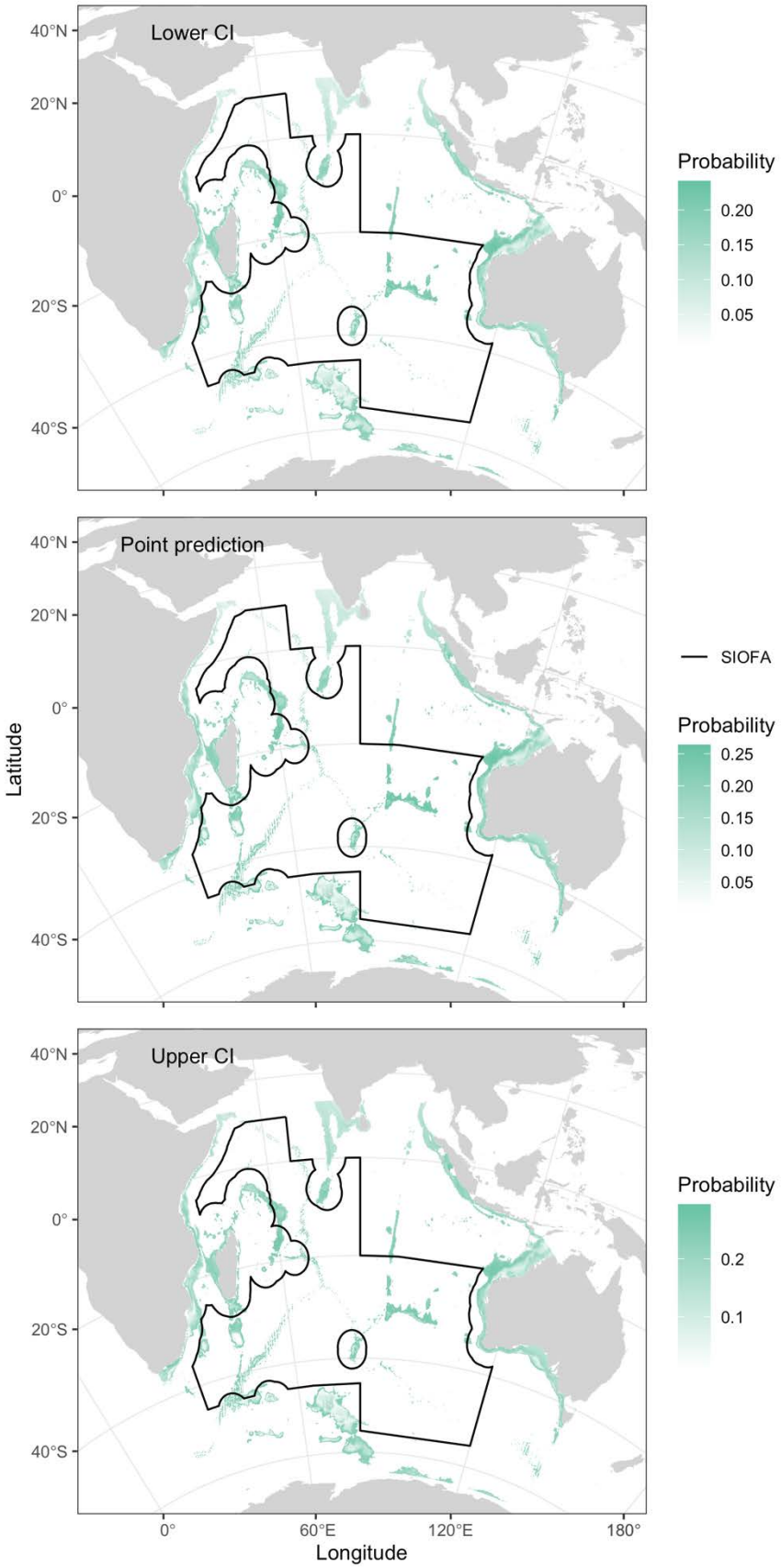

Figure A1. Probability distribution of RCP1 (Point prediction). Lower and upper confidence intervals (CI) (95%) are also provided.

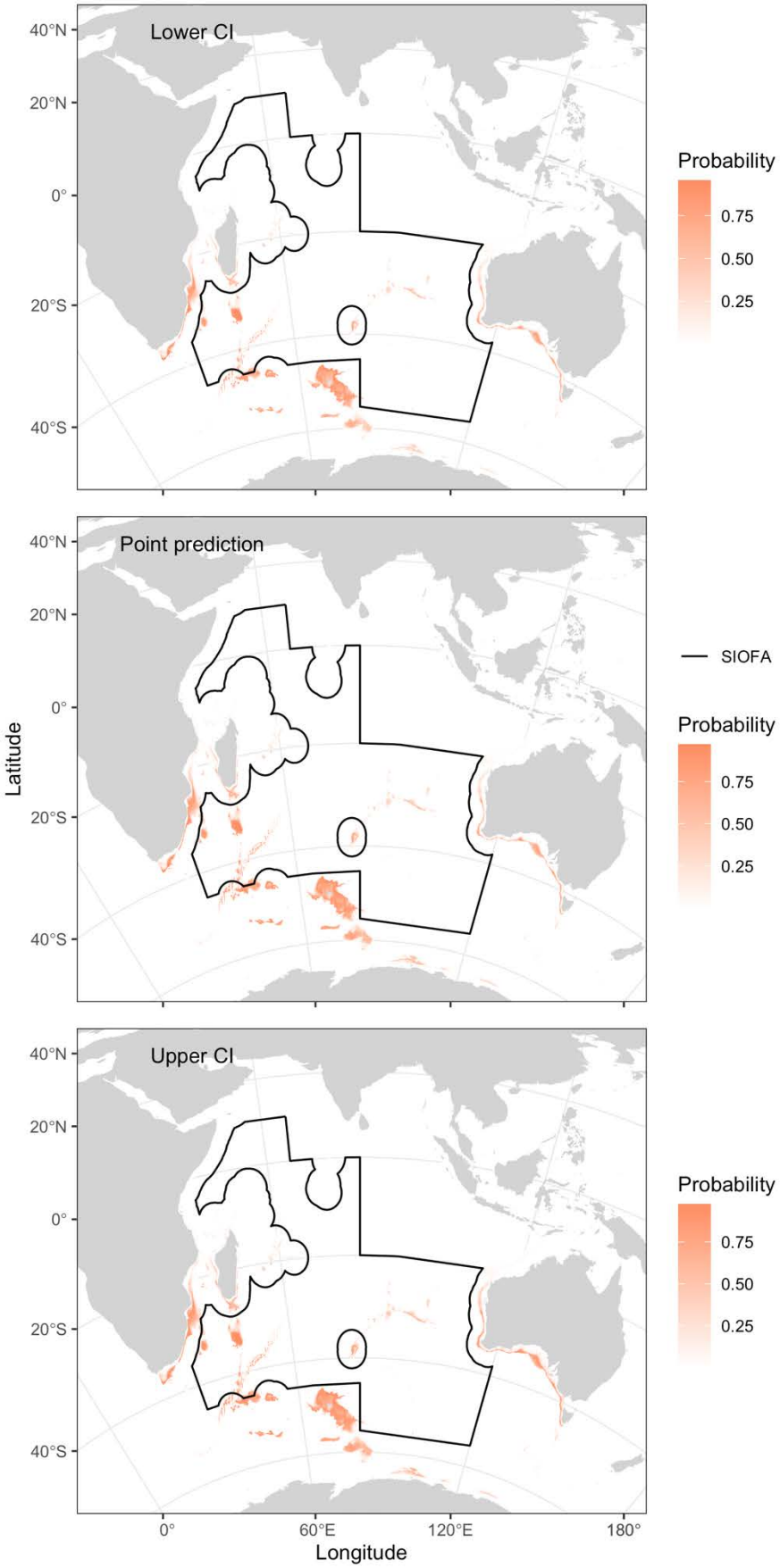

Figure A2. Probability distribution of RCP2 (Point prediction). Lower and upper confidence intervals (CI) (95%) are also provided.

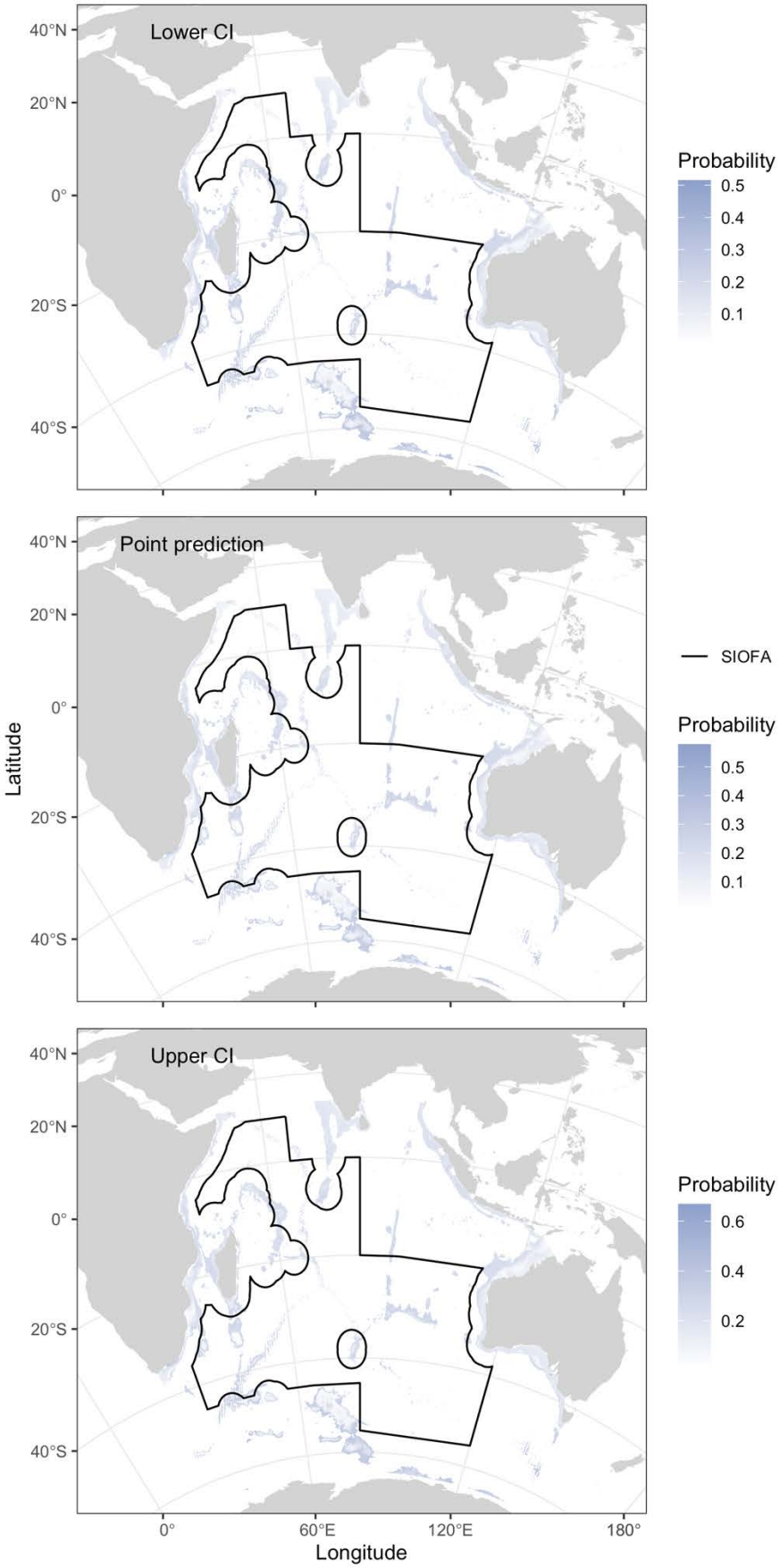

Figure A3. Probability distribution of RCP3 (Point prediction). Lower and upper confidence intervals (CI) (95%) are also provided.

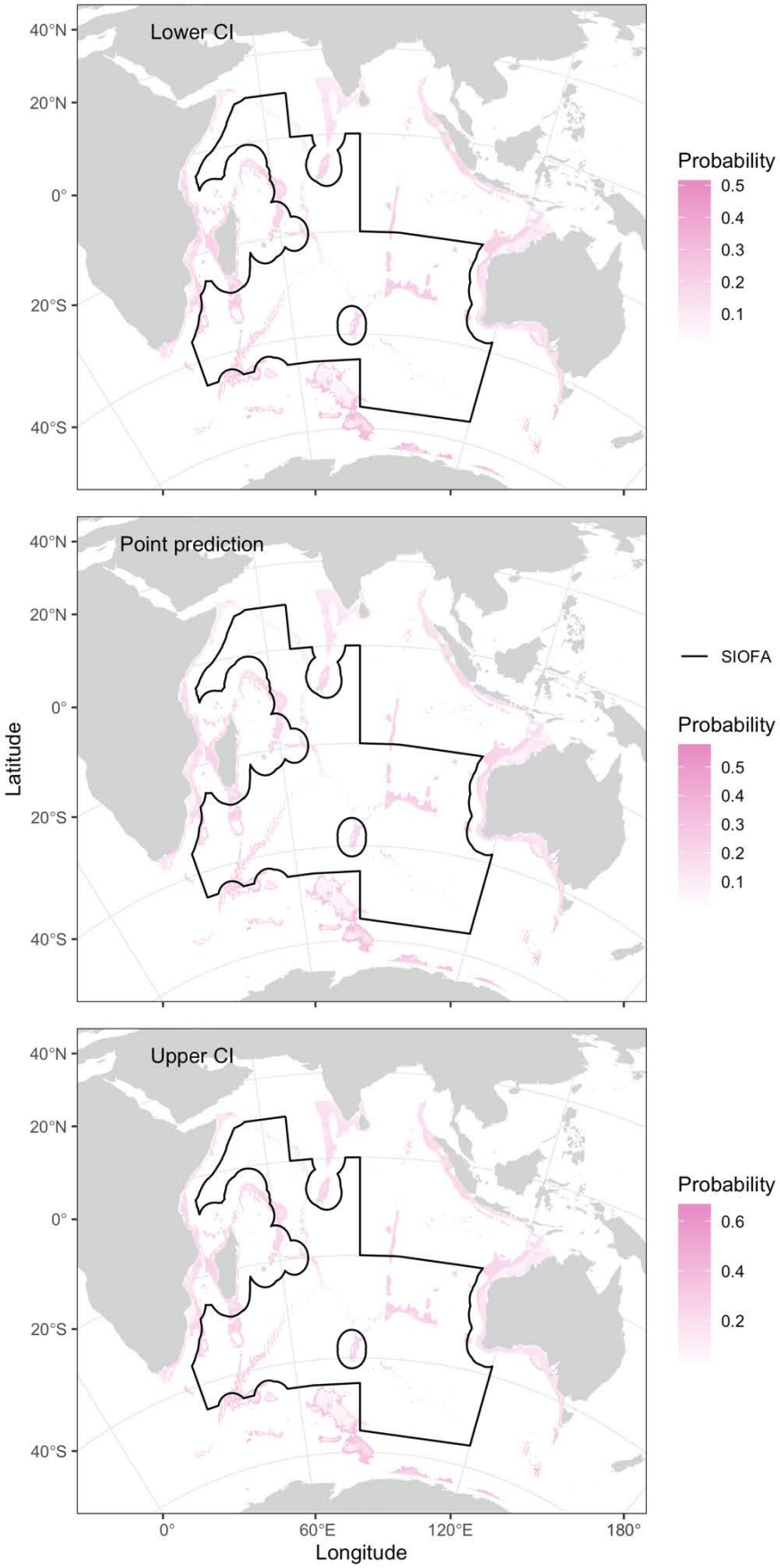

Figure A4. Probability distribution of RCP4 (Point prediction). Lower and upper confidence intervals (CI) (95%) are also provided.

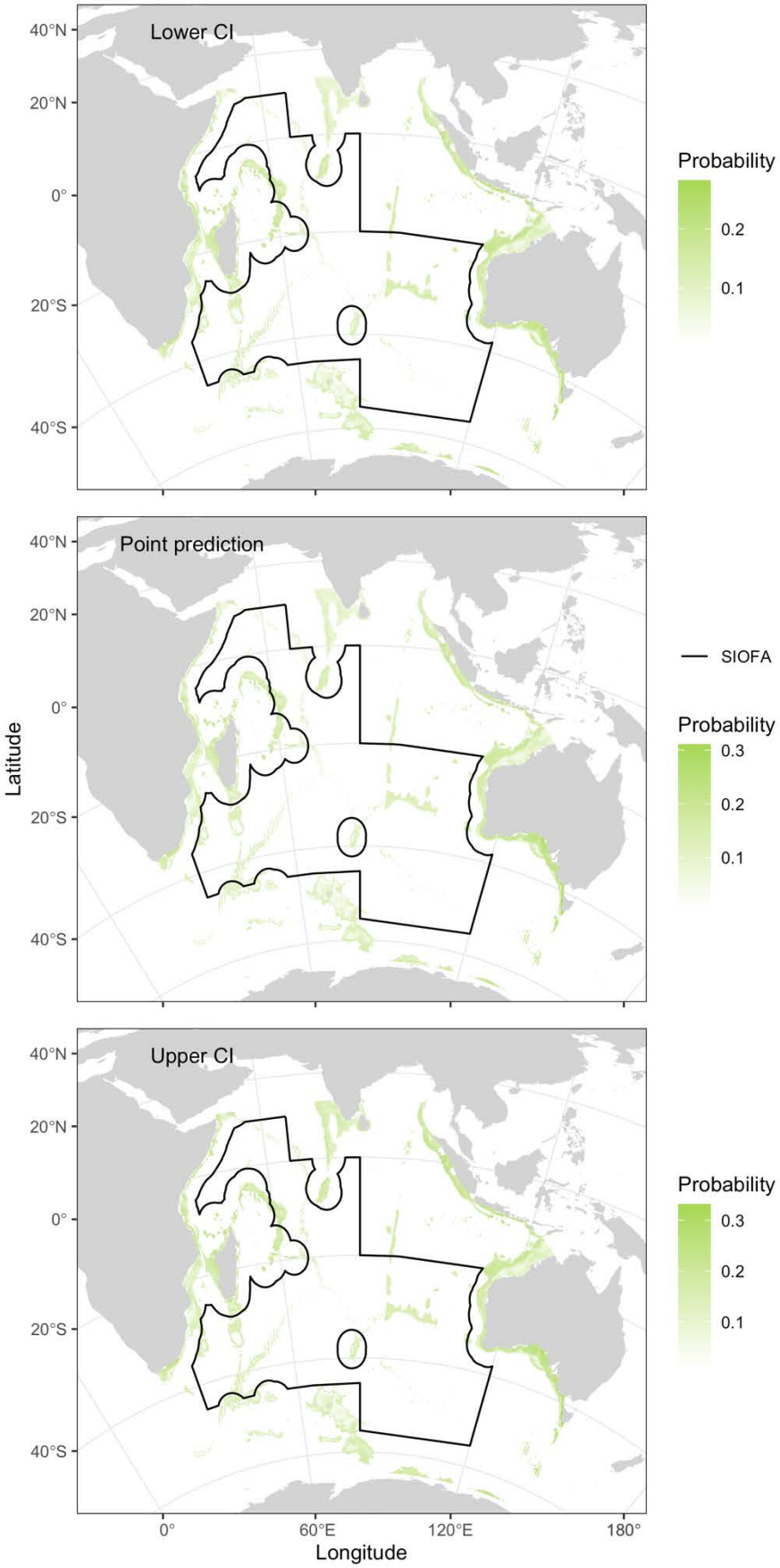

Figure A5. Probability distribution of RCP5 (Point prediction). Lower and upper confidence intervals (CI) (95%) are also provided.

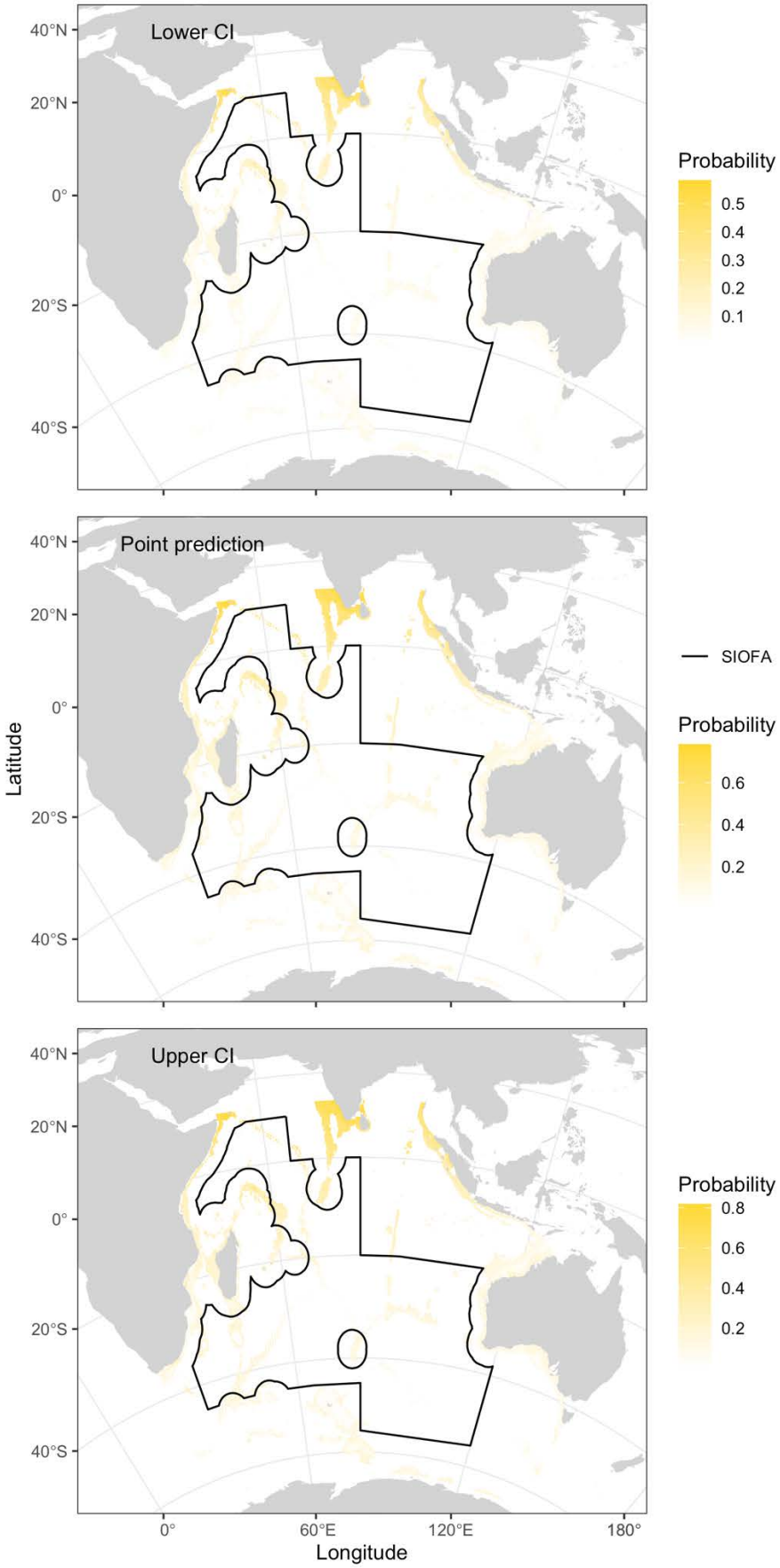

Figure A6. Probability distribution of RCP6 (Point prediction). Lower and upper confidence intervals (CI) (95%) are also provided.

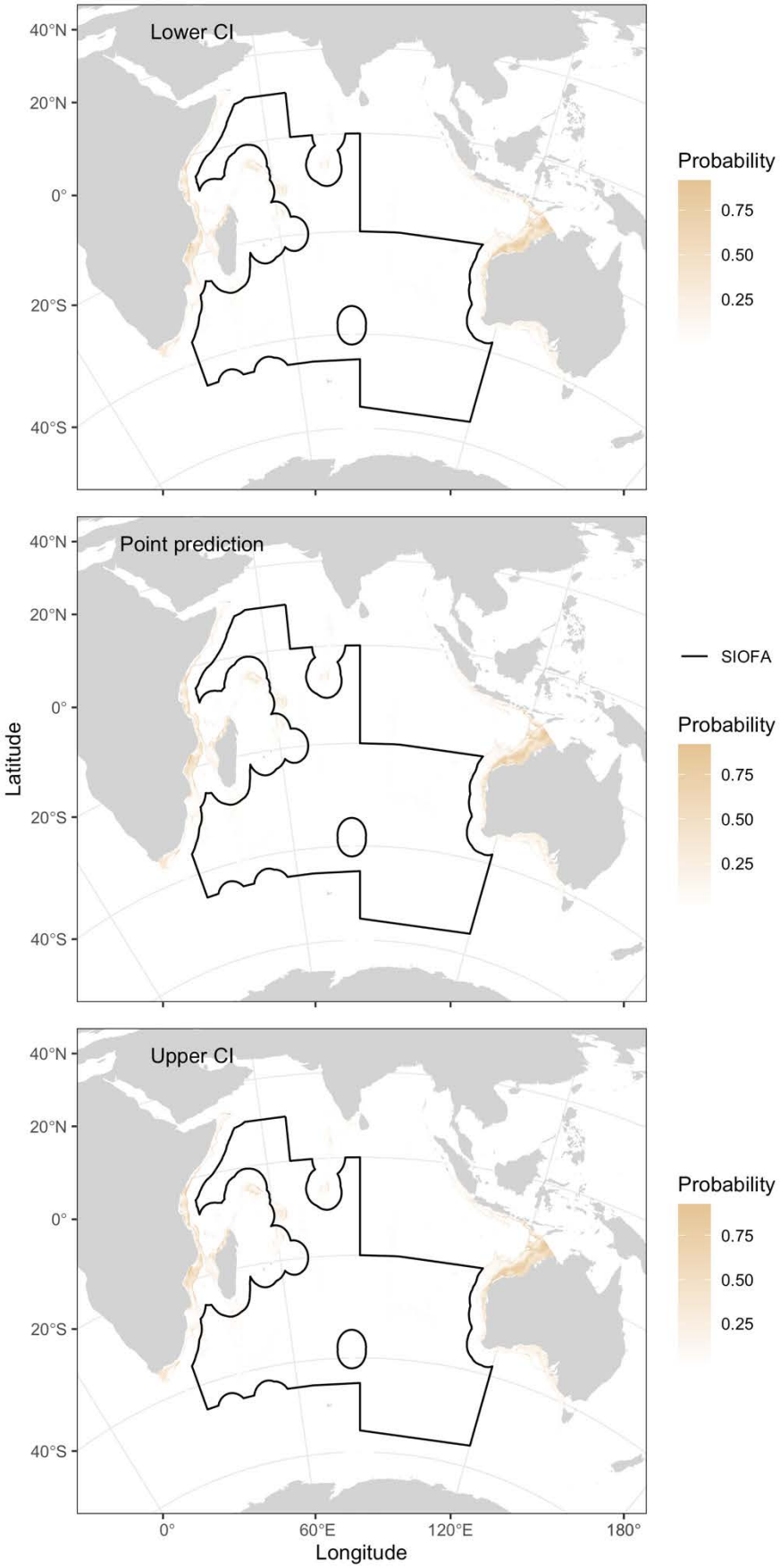

Figure A7. Probability distribution of RCP7 (Point prediction). Lower and upper confidence intervals (CI) (95%) are also provided.

SYSTEMATIC CONSERVATION PLANNING IN SIOFA

BERTA RAMIRO SÁNCHEZ, RODOLPHE DEVILLERS, BORIS LEROY

Project 2021-01

APRIL 2023

MUSEUM NATIONAL D'HISTOIRE NATURELLE
France

INSTITUT DE RECHERCHE POUR LE DÉVELOPPEMENT (IRD)
UMR ESPACE-DEV
France

Funded by the European Union

Suggested citation:

Ramiro-Sánchez, B., Devillers, R. and Leroy, B. (2023). *Systematic conservation planning in the SIOFA area (Project PAE2021-01)*. Muséum d'Histoire Naturelle, France.

Table of Contents

Acknowledgements

We are thankful to Dr Skipton Woolley (CSIRO, Australia) for fruitful discussions on the assessment of Indian Ocean conservation features.

Executive summary

International obligations require the implementation of conservation measures for the management of deep-sea bottom fisheries, due to the potential impacts they may have on seafloor biodiversity, especially on vulnerable marine ecosystems. Prioritising areas can help in the conservation of biodiversity from pressures imposed by human activities. Systematic conservation planning (SCP) has been promoted as an approach to ensure the long-term persistence of biodiversity using an explicit, objective, transparent, repeatable, and efficient methodology Thus, SCP, also referred to as the process of making informed decisions about the identification and preservation of areas of conservation value, can lead future work towards the prioritisation of areas in SIOFA.

Marxan is the most widely used SCP software to identify conservation priorities and help inform the creation of marine and terrestrial protected area systems. Marxan solves a "minimum-set strategy" optimisation problem, wherein the objective of the "minimum-set strategy" is to achieve a certain amount of every biodiversity feature for the smallest possible cost.

Here, we investigated the use of Marxan as a decision-support tool in the identification of biodiversity areas in SIOFA whilst minimising impacts to existing fisheries. We investigated the inclusion of three conservation features for protection: Ecologically or Biologically Significant Areas (EBSAs), bioregions based on VME indicator taxa, and geomorphic seafloor features. We set a low (5%), medium (10%), and high (15%) ecological conservation target for each feature and included fishing effort as tonnes of catch as a cost layer.

Following calibration of Marxan parameters, including the analysis of an appropriate level of reserve compactness, the three ecological target scenarios achieved approximately 4%, 7% and 10% protection of the overall area, respectively, without impacting the fishing footprint. These results should be interpreted relative to the selected conservation features for inclusion and the level of compactness of the reserve.

The investigation pointed to several aspects that SIOFA will need to take into consideration if a formal SCP is carried out for the area. First, convene over what conservation features to include and what ecological conservation targets to set. Second, understanding that there is no current approach to integrate certain information into the spatial analysis. Third, convene over the spatial resolution of the planning exercise.

This investigation intended to highlight which details could be retained or dropped in future work carried out by SIOFA. These Marxan outputs could be treated as a starting point for discussions about what to include in, but also what a draft network could look like. It is important to remember that Marxan is a decision-support tool in spatial decision-making process. That means that Marxan outputs should not be interpreted as final products but to guide managers towards areas that are important to protect. Decisions made about where to designate areas will need to factor in all possible available information that is not integrated into a spatial layer, such as scientific expertise and traditional ecological knowledge about the environment in question, practical and political constraints, and human conflicts.

1. Purpose of the report

This report aims to address the Terms of Reference 2 "*Investigate and advise on the identification of representative protected areas within the SIOFA management area, based on bioregionalization work*" of project PAE2021-01.

The analyses presented in this report aimed to develop solutions that help to identify priority areas for conservation while minimizing the economic impacts of conservation zones on the fisheries sector.

2. Introduction

International obligations require the implementation of conservation measures for the management of deep-sea bottom fisheries, due to the potential impacts they may have on seafloor biodiversity, especially on vulnerable marine ecosystems (UNGA, 2007). In this regard, SIOFA has initiated management efforts to protect VMEs by identifying a list of VME indicator taxa for Indian Ocean waters (CMM 2018/01) and has designated five interim Benthic Protected Areas (BPAs) where bottom-fishing trawling is not permitted (CMM 2019/01). Although these areas contain vulnerable habitats, the importance of selecting these areas has not yet been studied. In addition, SIOFA has mandated to map the potential biogeographical regions of VME indicator taxa using predictive modelling. Such work produced three bioregionalization schemes based on different algorithms (Ramiro-Sánchez & Leroy, 2022), which represents significant progress for the Indian Ocean, one of the lesser explored world's oceans.

In continuation of its advancements towards compliance with international obligations for the protection of vulnerable habitats, SIOFA wishes now to progress the work by investigating potential networks of reserves in their Convention Area based on the bioregionalization work. Despite the uncertainties of modelling, these models can help explore potential networks and prioritise areas for protection, together with other layers of information (e.g., *bona fide* records of VMEs), and SIOFA members' expert knowledge of the area.

Prioritising areas can help meet conservation targets set to protect biodiversity, including ecosystems, biological assemblages, species, and populations, from pressures imposed by human activities (Margules & Pressey, 2000). If different approaches can be used to prioritize areas for conservation, systematic conservation planning (SCP) has been promoted as an approach to ensure the long-term persistence of biodiversity using an explicit, objective, transparent, repeatable, and efficient methodology (Pressey et al., 1993). Thus, SCP, also referred to as the process of making informed decisions about the identification and preservation of areas of conservation value (Gaston et al., 2002; Moilanen et al., 2009), can lead future work towards the prioritisation of areas in SIOFA.

SCP involves several stages (Kukkala & Moilanen, 2013; Margules et al., 2002; Margules & Pressey, 2000). First, it requires making choices about the features to protect (e.g., species, habitats). Second, upon these features, explicit goals must be set that translate into operational and quantitative targets (e.g., X% of a given feature should be protected). Third, it identifies actions for achieving those goals. Forth, it considers existing actions that

complement new ones in achieving those goals (e.g., existing protected areas). Fifth, it applies explicit criteria to meet goals in the context of operational limitations (e.g., budget). Finally, it implements actions for maintaining the conditions within reserves that are required to foster the persistence of key natural features, together with monitoring of those features and adaptive management as required.

In SCP, the value of conservation features gets quantified within each spatial unit ("planning units" – PUs) that subdivide the study area. Those PUs can be arbitrary (e.g., a regular grid) or based on existing boundaries (e.g., administrative, natural watersheds). The value of conservation features can be associated to PUs in different ways, such as estimating the occupancy (presence or absence) of a population or species, an area of a species' habitat, an abundance, or a combination of measures that integrate occupancy and habitat condition. The conservation value of the PU must be quantified together with the cost of implementing an action. This is achieved by explicitly setting targets to achieve conservation goals. These targets are often subjective, set through expert opinion (e.g., Levin et al., 2013), community consensus (e.g., Game et al., 2011), or are informed by legislation or policy (e.g., Maina et al., 2020).

Resources for managing or protecting all valuable PUs are usually limited. As a result, there is a need for an approach to select a subset of PUs for conservation. This process can be translated into a mathematical optimization problem where the goal is to achieve all targets for the least cost (Beyer et al., 2016; Kukkala & Moilanen, 2013). In other words, management plans must be efficient so that maximising resources allows surplus turn over to be allocated to other problems. Otherwise, inefficient management plans may be too large and expensive to implement and less likely to succeed (Possingham et al., 2006).

Marxan as a decision-support tool

Marxan is the most widely used SCP software, allowing to solve the optimisation problem and helping to inform the creation of marine and terrestrial protected area systems (Ball et al., 2009). Marxan's optimization algorithm uses the "minimum-set strategy" to identify conservation priorities. The objective of the "minimum-set strategy" is to achieve a certain amount of every biodiversity feature for the smallest possible cost. The *objective function* is the mathematical formulation of the minimum-set problem. In protected area design, the problem we are trying to solve is to identify the protected area system that achieves our targets and spatial requirements for the least cost. Thus, a protected area configuration is given an objective function score to measure how well it performs and, in comparing alternative solutions, those with lower scores are better. Therefore, the objective function is a score that we want to minimize and is calculated as follows:

$$Cost = \sum_{PU} Cost + BLM \times \sum_{PU} Boundary\ Length + \sum_{Features} SPF\ for\ missing\ features \times Penalty$$

Where the cost of creating a network of protected areas is a function of the individual costs of PU included in the network, but also a function of other factors (i.e., Boundary length and SPF -Species penalty factor). A cost is assigned to each PU prior to planning, reflecting a virtual

cost of including this specific PU in the network (e.g., some PU could have more economic activities than others). Commonly, these are socio-economic costs, such as fishing effort or an index of cost. The total cost of a solution (a proposed network) is the sum of costs for all PU selected to form a protected area network. *Boundary length* of the protected area system refers to the spatial configuration of the reserve network (i.e., a single large system or several small systems). The reserve network boundary length is measured as the sum of the planning units that share a boundary with planning units outside the reserve network. Thus, fragmented reserve networks will have a large boundary length. A fragmented network will likely be difficult and costly to manage, as well as with increased edge effects and reduced connectivity. The objective function deals with this by using the *boundary length modifier* (BLM), which can be used to increase the importance of having a more compact reserve network. The BLM is user-defined and allows for control of the level of clumping in the solution. The larger the value of BLM, the more the system will be clumped and the higher the cost will be. For each alternative solution, Marxan calculates whether the target for each conservation feature is met or not. If a target is not met, then a user-defined penalty cost called the *species penalty factor (SPF)* can be applied. Making the SPF user-defined allows different weightings to be given to different feature targets. The higher the SPF, the higher the penalty when a conservation feature target is not met. *Penalty* refers to the penalty incurred for every feature that fails to meet its target**.** An appropriately high SPF will result in more costly protected areas with more targets met.

Marxan produces a number of solutions that each presents a proposed network of protected areas that could meet the targets set by the user. The two most important outputs from Marxan for any given scenario are the “best solution” and the “sum solution”, the latter also referred to as “selection frequency”. The “best solution” is the solution with the lowest overall cost (score). Marxan’s objective function states that this best or most efficient solution corresponds to selected sites at the lowest total cost. The “sum solution” or “selection frequency” reports on how often each PU is selected in a solution. For instance, the Marxan algorithm is run 100 times, the “selection frequency” reports on how many times each PU is selected by a given solution out of a hundred solutions. Marxan’s good practice is to present maps showing the “best solution” (or a number of solutions with low scores) and the “selection frequency”. This is because the selection frequency is useful to explore the irreplaceability of planning units, that is, how irreplaceable is a planning unit in creating efficient solutions.

The interpretation of Marxan results requires careful consideration of several aspects (see Ardron et al., 2010). First, the “selection frequency” map is built with an equal contribution of all solutions, even if some solutions have not met conservation objectives (this depends on SPF settings and other constraints). Therefore, it is important to investigate how many solutions are meeting or almost meeting all the conservation targets. Second, Marxan’s algorithm returns a near-optimal mathematical solution to the problem set (i.e., meeting objectives at minimum total cost) (Ball et al., 2009) within a number of given runs. However, best solution does not necessarily mean the “best spatial network configuration”. Indeed, within a given scenario there may be other good solutions (i.e., that meet all the conservation targets) that might form the basis of a network that is more practical to implement. As such, it is more useful for decision-makers to identify a range of near-optimal solutions that provide diverse options for a decision-maker, rather than a single optimal solution (Kirkpatrick et al.,

1983). Finally, it is important to note that, while Marxan can help find efficient solutions to spatial prioritization problems, the tool helps to inform decisions but does not make them. The software is designed to be a decision support tool and cannot replace a spatial planning process – no single Marxan solution will ever deliver a finished reserved network plan. Rather, Marxan solutions should be used in an iterative planning process within a larger decision-making process involving stakeholders and managers, where spatial configurations are reviewed, modified, and refined in multiple planning rounds (where not each necessarily has to involve the use of Marxan).

Aims of the study

In this study, we aimed to investigate SCP to create scenarios of a reserve network within SIOFA using Marxan as a decision-support tool. A standard Marxan cost layer based on fishing effort was used in an ecological Marxan analysis. Analyses were conducted for three different ecological targets (low, medium, and high).

3. Methods

In the following sections we describe the input layers required for the conservation planning, namely the PUs, the conservation features, and the cost layer. We then present the ecological scenarios run with Marxan.

3.1 Study area

The study area encompasses SIOFA's Convention Area (Figure 1), with a total size of 27,180,300 km^2.

SIOFA is subdivided into nine management subareas for fisheries. The main fisheries operating in the SIOFA area are for Patagonian toothfish (SIOFA subareas 3b, 7); orange roughy (topographic features in subareas 1, 2, 3a and 3b); alfonsino topographic features in subareas 1, 2, 3a and 3b); saurida and scads (subarea 8, mainly Saya de Malha); shallow-water snappers, emperors, and groupers (subarea 8, mainly Saya de Malha); deep-water snappers, lutjanids and Hapuka; and oilfish (southwest Indian Ocean) (report Summary Statistics SIOFA).

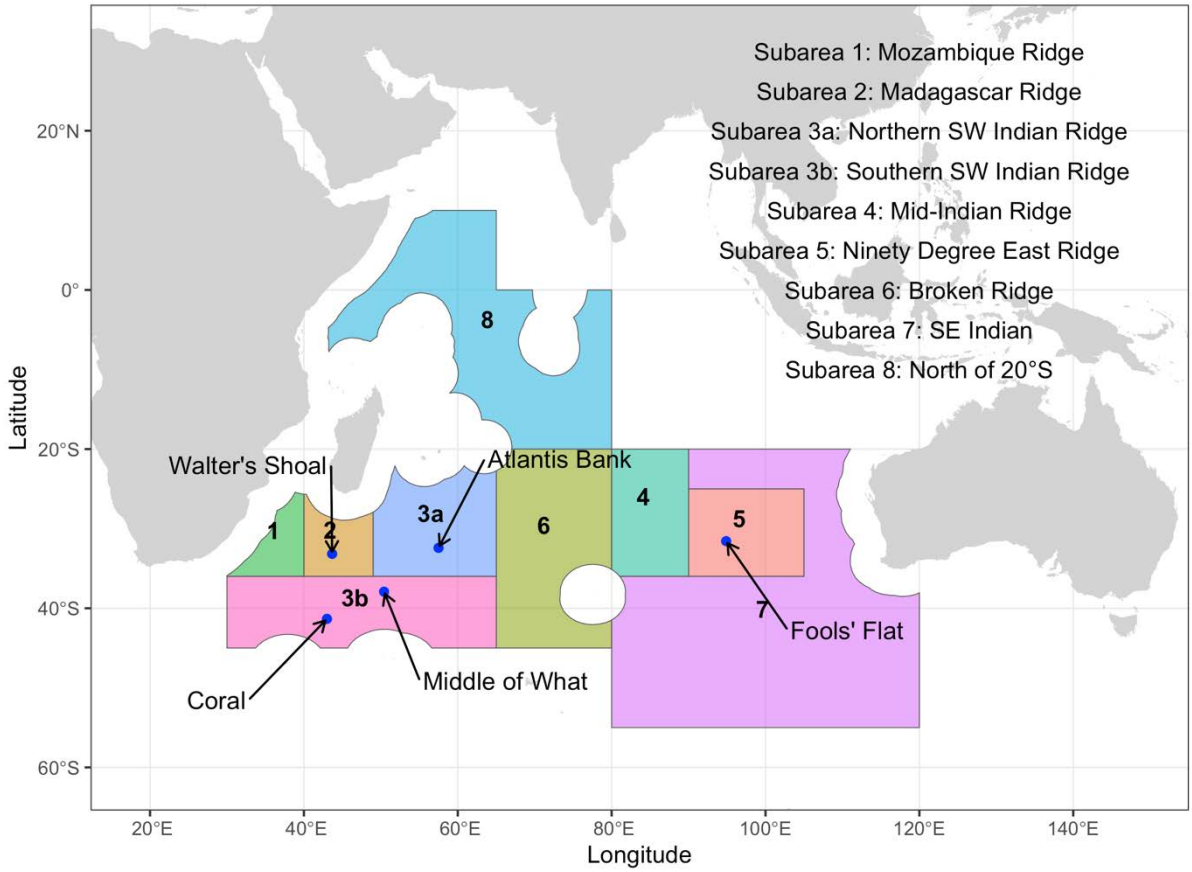

Figure 1. SIOFA's Convention Area divided into nine management subareas, and location of SIOFA's five interim protected areas (blue circles).

There are currently five interim Protected Areas (IPAs) in SIOFA (Figure 1) where restrictions to bottom fisheries should be applied by Cooperating and Contracting Parties (CCPs) until the adoption of a dedicated research and management plan. Current restrictions to fisheries in IPAs include a prohibition for CCPs to engage in bottom fishing, excluding line and trap methods, and an obligation to have a scientific observer onboard at all times while fishing inside those areas.

A geographic information system layer subdividing the study area in grid cells at 20 arc minutes and 30 arc minutes were provided by SIOFA Secretariat. We clipped those grid cells falling within SIOFA perimeter. At 20 arc minutes, this resulted in a total of 24,453 grid cells, hereafter "planning units". Each planning unit (PU) is based on an equal-area grid square of 1,369 km^2. We chose this resolution to reflect the regional scale of the analysis and to scale the fishing information proportionally. We ensured that our spatial grids aligned with SIOFA's spatial layers by using original SIOFA's spatial layers, which form the basis for their own work.

3.2 Cost

Cost was included as fishing effort, the most common way to capture cost in Marxan when applied to marine environments. Aggregated data for fishing effort were requested to SIOFA and received as number of fishing events and tonnes of catch per gear type: trawl, gillnets, line, and traps. Data were provided at a resolution of 10 arc min. Cost was included in the Marxan ecological analysis as catch weight in metric tonnes, aggregated for all gears.

Data, originally provided in vector format, were rasterised to an equal area grid of the same resolution to that of the planning unit, by summing all values of surrounding cells. The aggregation reduced the number of individual information from 1,154 individual data points at 10 arc minute resolution to 403 data points at 20 arc minute resolution, i.e., a total of 403 PU out of 24,453 had fishing effort as cost.

Exploration of the number of fishing events and catch data prior to the analyses indicated a potential outlier in the data that skewed the distribution. The catch value was one order of magnitude larger than the remainder of the dataset (Figure 2). The record was kept as the extreme value was buffered when aggregating effort data at 20 arc minute resolution. The data point was located in SIOFA subdivision 8.

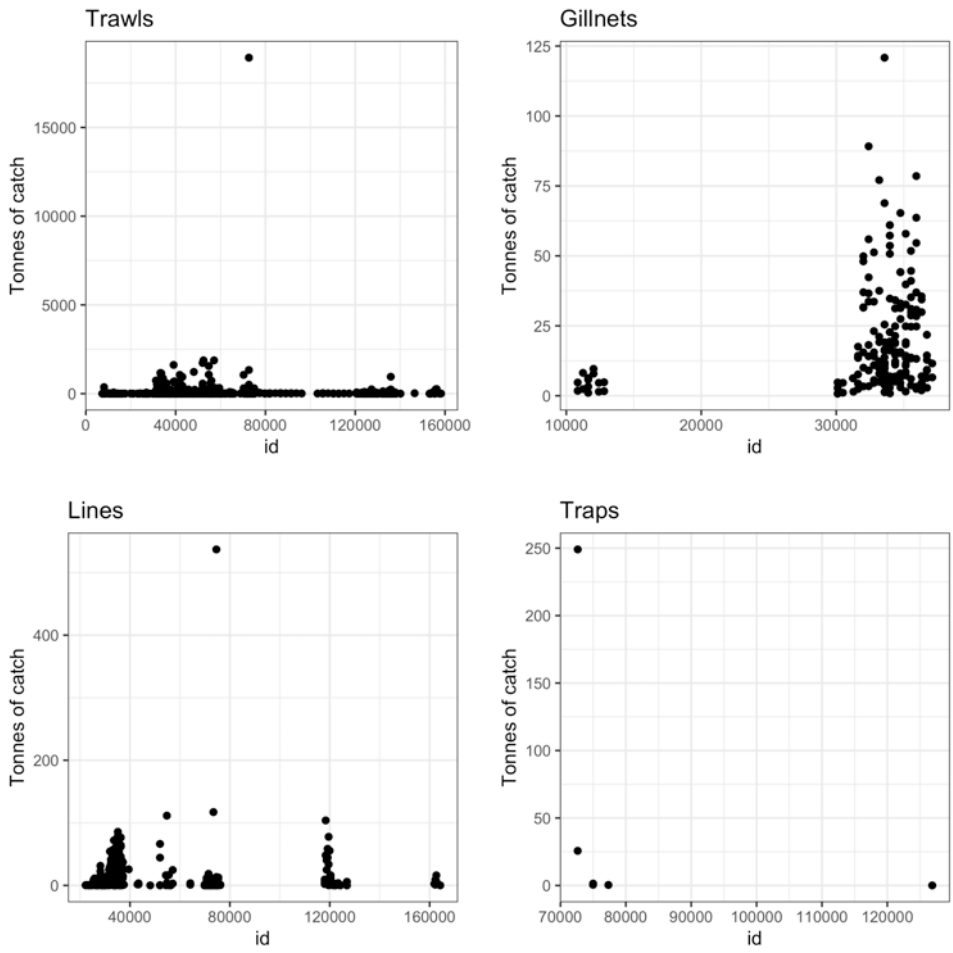

Figure 2. Fish catch (in tonnes) by gear type.

3.3 Conservation features

We identified three key data groups that should be considered as conservation features in the analyses: Ecologically or Biologically Significant Areas (EBSAs), geomorphic seafloor features, and bioregions of VME indicator taxa. In the following sections we elaborate on each of these features.

3.3.1 EBSAs

Ecologically or Biologically Significant Areas (EBSAs) are marine areas of potential importance to protect. These areas are thought to support the healthy functioning of the oceans and the services that it provides. Areas are identified based on scientific criteria that is similar to the criteria for the identification of Vulnerable Marine Ecosystems (VMEs (Annex I, Decision IX/20; CBD, 2008)):

1. Uniqueness or Rarity
2. Special importance for life history stages of species
3. Importance for threatened, endangered or declining species and/or habitats
4. Vulnerability, Fragility, Sensitivity, or Slow recovery
5. Biological Productivity
6. Biological Diversity
7. Naturalness

Selection of areas is carried out by States and intergovernmental organizations through regional workshops. The application of the EBSA criteria is an evolving process, where information is updated as new scientific and technical information becomes available. COP10 encouraged Parties, other Governments, and competent intergovernmental organizations to cooperate, to identify and adopt appropriate measures for conservation and sustainable use in relation to EBSAs, including by establishing representative networks of marine protected areas in accordance with international law, including the United Nations Convention on the Law of the Sea, and based on the best scientific information available.

Presently, there are 39 EBSAs in the Southern Indian Ocean (https://www.cbd.int/ebsa/), 11 of which fall fully or partially within SIOFA's Convention Area: Agulhas Front; Walters Shoal; Prince Edwards Islands, Del Caño Rise and Crozet Islands; Saya de Malha Bank; Rusky; East Broken Guyot; Mozambique Channel; Coral seamount and fracture zone feature; Atlantis Seamount; Central Indian Basin; Fools Flat (Figure 3). We downloaded the spatial layers for these 11 EBSAs as shapefiles (source: https://www.cbd.int/ebsa/) and merged all EBSAs into one single polygon, without distinguishing individual EBSAs. We used the layer as one single conservation layer since not all EBSAs fall in their totality within SIOFA.

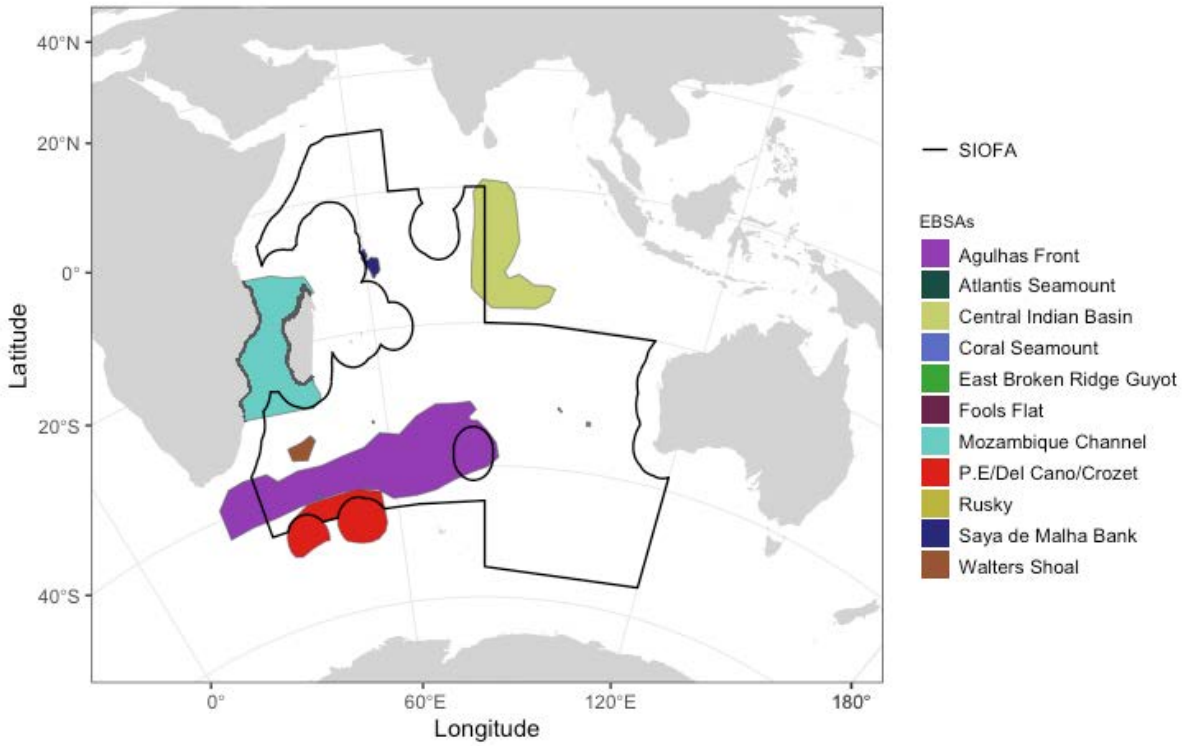

Figure 3. Ecologically or Biologically Significant Areas (EBSAs) falling partially or totally in SIOFA's Competence Area.

3.3.2 Seafloor geomorphology

Seabed geomorphology is a useful surrogate for biodiversity where ecological processes are known or suspected to link seabed structure and the distribution of benthic organisms (McArthur et al., 2010). Seabed geomorphology encompasses the shape and hardness of the seabed at a range of spatial scales and describes the processes behind these occurrences. At coarse scale, geomorphological features are categorical descriptors of the shape of the seabed (e.g., reefs, ridges, seamounts) describing structures that range in scale from thousands of km^2 (e.g., basins) to a few hundred m^2 (Harris et al., 2008). At a finer scale, substrate type has been shown to significantly differ in species composition, where hard substrate acts as a habitat to a larger proportion of sessile suspension feeders and soft sediments act as a habitat to a larger proportion of motile organisms (Harris et al., 2008).

Surrogates are particularly useful for assessing biodiversity based on minimal biological information. In the Southern Indian Ocean, there is very scarce biological information available for deep-sea habitats. As a result, we included Harris et al. (2014)'s seafloor geomorphic features map of the global ocean as a conservation feature (Figure 4). The map includes 29 geomorphic feature categories, where the seafloor base layers shelf, slope, abyss and hadal zones are overlain by classification layers and discrete feature layers (Harris et al., 2014). Two of the base layers (shelf and abyssal) are subdivided into classification layers based on roughness: low, medium, and high relief shelves; and abyssal plains, abyssal hills, and

abyssal mountains. Whereas the four base layers are mutually exclusive, the classification layers and feature layers may overlay each other (Harris et al., 2014).

We plotted the map for SIOFA's area and combined some of the 29 classes to reduce the risks of fragmentation of the proposed network in Marxan. Moreover, some of Harris et al. (2014)'s features overlay other features and some others are out of scope of the analysis (e.g., hadal depths). As a result, we reduced the classification to eleven features (Table 1), of which six were considered for this investigation due to their potential use as a surrogate for benthic biodiversity: Plateau, Rise, Shelf, Slope, Terrace, and Ridge. A detailed description of each feature type is presented in Harris et al. (2014). Seamounts were also considered, but we used the latest layer produced by Yesson et al. (2021), based on an updated bathymetric model. We describe this spatial layer in the next subsection.

Table 1. Reclassification of Harris et al. (2014)'s geomorphic seafloor features for use in the conservation planning exercise in SIOFA's Convention Area.

Harris et al. (2014)'s original classification	Modified or reclassified features from the original classification
Abyss	**Abyss**
Basin	**Basin**
Canyon	**Canyon**
Guyot	**Guyot**
Plateau	**Plateau**
Ridge	**Ridge**
Rise	**Rise**
Seamount	**Seamount**
Shelf	**Shelf**
Slope	**Slope**
Terrace	**Terrace**
Bridge	Not included (overlay other features)
Escarpment	Not included (overlay other features)
Hadal	Not included (depths at out of scope)
Trenches	Not included (depths at out of scope)
Sills	Not included (not equally mapped according to Harris *et al*. (2014))
Fan	**Rise** (reclassified, as fans overlay rises)
Rift valley	**Ridge** (reclassified; singular feature)
Spreading Ridge	**Ridge** (reclassified; singular feature)
Trough	**Abyss** (reclassified; at abyssal depths)

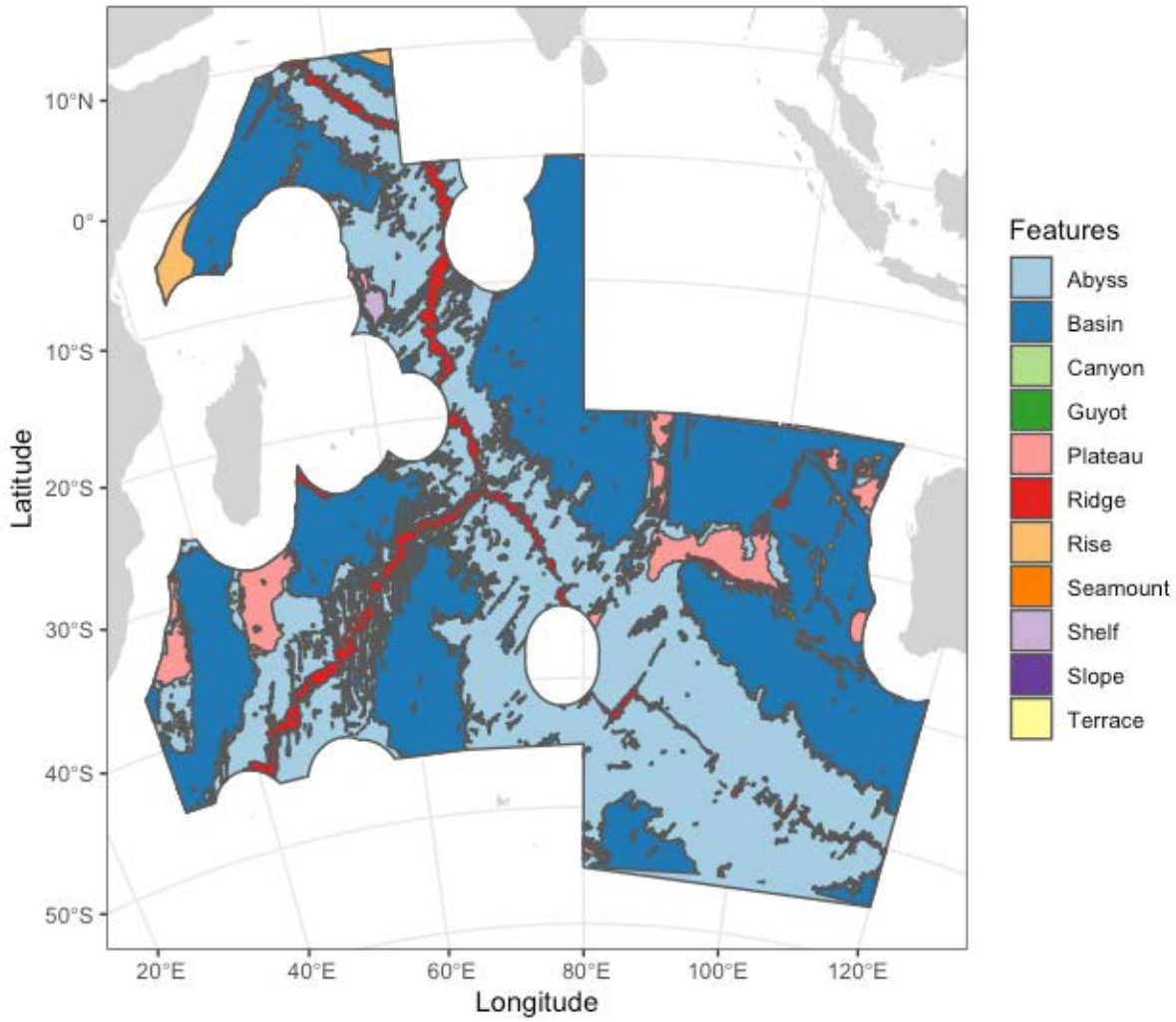

Figure 4. Seafloor geomorphology map modified from Harris et al. (2014) with combined geomorphic categories considered for this study.

Seamounts

Regardless of the methodology used, it is important to keep prediction datasets up to date with the latest bathymetry grids. We therefore used the updated list of seamounts by Yesson et al. (2021) who, using the same methodology as Yesson et al. (2011), updated their seamount predictions using the latest available bathymetry. Yesson et al. (2021) remark that "the revised predictions are higher than other predictions that post-date Yesson et al. (2011) such as 24,643 seamounts in the Kim & Wessel (2011) dataset and 10,234 of Harris et al. (2014), but are still lower than some other predictions, for example, 68,669 of Costello et al. (2010)". However, they note that each of these studies uses different ways of detecting seamounts where, for example, Harris et al. (2014) have a stricter definition of seamount that excludes features along ridges, while the methodology used in Yesson et al. (2021) (and Yesson et al. 2011) employs a distance-based filtering of adjacent features.

We downloaded their global seamount dataset from PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.757564) available in vector format and in two formats: as the centroids of the seamounts and as the bases of the predicted seamounts. We used the latter to calculate the feature contribution in each PU (Figure 5).

Thus, in total there were seven geomorphic features used as surrogates of biodiversity.

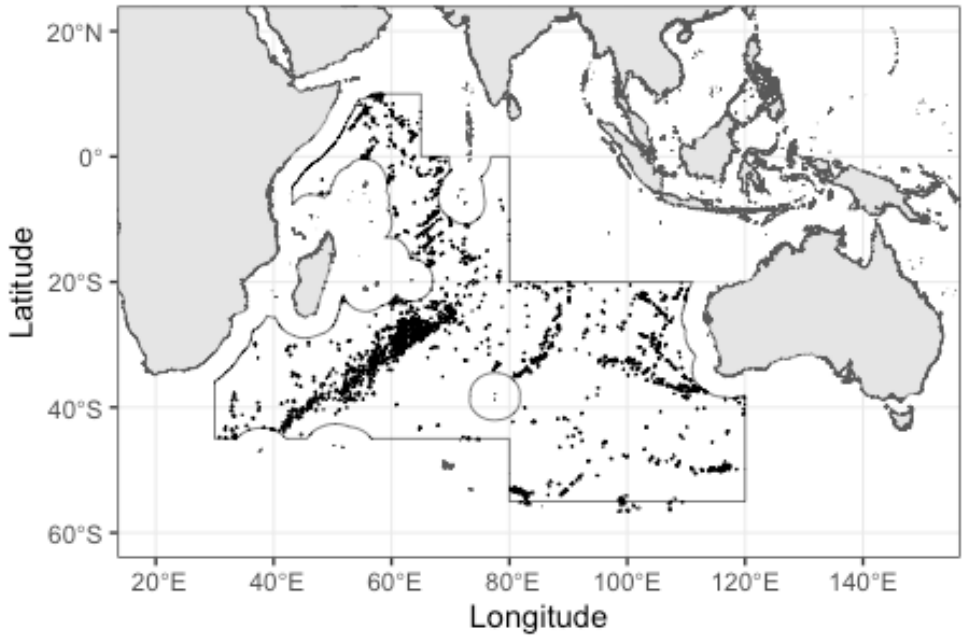

Figure 5. Seamount bases distribution in SIOFA area obtained from Yesson et al. (2021).

3.3.3 Bioregionalization

Biogeographical classifications, or bioregionalizations (Woolley et al., 2020), are the building blocks for the planning and implementation process of management measures and are highly connected to the development of marine protected areas (Rice et al., 2011). This is because bioregionalizations enable the identification of units that should be represented in a network, ensuring the protection of biogeographically unique areas and the development of a network that considers representativity, connectivity and replication of sites (Rice et al., 2011). Previous predictive bioregionalization work was carried out for SIOFA under project PAE2020-02 using three approaches commonly referred to as "group first, then predict", "predict first, then group" and "analyse simultaneously" (Ramiro-Sánchez & Leroy, 2022). An extensive discussion on the advantages and limitations of each resulting bioregionalization scheme was presented in the corresponding project report. Despite some of the limitations from these approaches, they still represent an improvement over other existing classifications (e.g., the Global Open Oceans and Deep Seabed classification). Out of the three approaches, the "group first, then predict" yielded most robust results. Consequently, we selected this biogeographical classification as an input layer in the conservation planning. Specifically, we

selected the nested subregions, as these provide finer detail of potential biogeographic patterns in the deep sea of the Southern Indian Ocean (Figure 6).

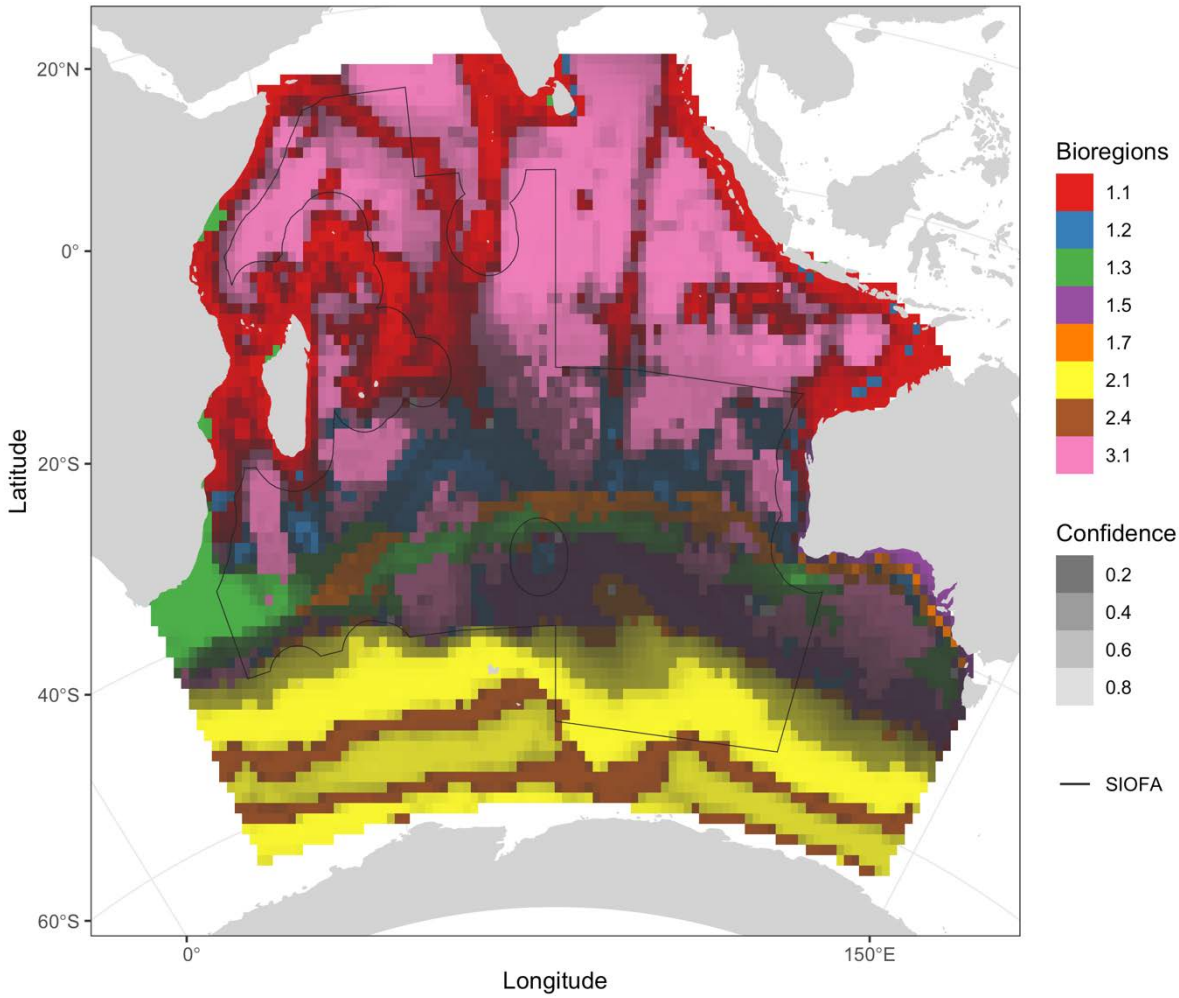

Figure 6. Predicted biogeographical regions of VME indicator taxa in the Southern Indian Ocean at the second level of the hierarchy. Areas with low confidence in the prediction are shown in darker shades of grey. Note that, because of the low number of points, we could not reliably evaluate these predictions.

3.4 Conservation planning with Marxan

To select priority areas for marine conservation, we initially defined three broad conservation goals as follows:

- to protect 10% of each of the 7 seafloor geomorphic habitats, reflecting CBD Aichi Target 11 protection targets of 10% (CBD, 2010);
- to protect 10% of each of the 8 bioregions in order to promote the long-term population viability of marine populations by maintaining natural habitats for the different communities of the Indian Ocean;

- to protect 10% of EBSA, helping to align national and regional priorities with globally recognized areas (Maina et al., 2020).

Thus, we considered the ecological conservation target of 10% as a baseline target for all conservation features, to which variations could be applied (i.e., lower or increase the target).

We included a cost layer based on the aggregated fishing effort values (as tonnes of catch) provided by SIOFA Secretariat, because including spatially explicit costs within Marxan allows meeting conservation targets while minimizing costs to fisheries, reflecting spatial planning best practice (Ban & Klein, 2009; Watts et al., 2017). Planning units were given a minimum cost value of "1" to allow for the Marxan algorithm to run (a standard practice in Marxan analyses).

Based on the initially defined baseline scenario, we defined two additional scenarios with a lower and higher ecological conservation target for each of the conservation features: 5% and 15% conservation targets. Thus, we had three scenarios that we rename henceforth as "low-target" (5% target), "medium-target" (10% target) and "high-target" (15% target) scenarios. In the three scenarios, cost was treated the same. An exploratory analysis suggested to lower the target for Bioregion 3.1 (the largest; Figure 6). On the premise that it will be less impacted by certain human activities because it characterises the abyssal zone, we lowered its target to 2.5% for all three scenarios. In addition, we set three additional scenarios with the same ecological targets and settings, except for cost. For these other three scenarios, there was no cost associated to the planning units (cost was set to 1 for all planning units to allow the Marxan algorithm to run), as the exploration of scenarios with no cost allows for the potential identification of significant biodiversity areas without any constraint. Therefore, in total, we explored six scenarios: three ecological target levels (low, medium, and high) for two cost scenarios (tonnes of catch or no cost) (Table 2).

Marxan provides the option to systematically include (lock-in) areas that should be part of the reserve network (e.g., existing protected areas that already contribute to the ecological targets) or systematically exclude (lock-out) areas that should not be part of the network (e.g., areas where protected areas should not be established). Marxan will tend to build onto locked-in areas (if there are ecological attributes nearby), whereas it will avoid attaching sites to locked-out areas (unless there are few other options). Those areas should normally be no-take zones as Marxan will assume that all the species and habitats in these zones are fully protected. Similarly, areas can be identified to be locked out if there is a reason those areas should not be part of any protected area network. Here, none of the scenarios had PU locked-in or locked-out.

For each scenario, we calibrated the number of iterations to ensure that the algorithm adequately searched the sampling space and looked for improvements; the number of runs (or solutions) to ensure that the algorithm did not get stuck in local minima; the Boundary Length Modifier (BLM) using the calibration method of Stewart & Possingham (2005) to plan for spatially clumped reserves, which minimizes the trade-off in reduced boundaries and increased costs; and the Species Penalty Factor (SPF) to meet all ecological targets. Details on the calibration of all parameters are found in Appendix A for scenarios with cost and in Appendix B for scenarios with no cost.

For each scenario with cost, we provide the "selection frequency" map and the "best run" (a near optimal configuration of zones and one with the lowest overall objective function score across runs) maps. For each scenario with no cost, we provide the "selection frequency" map and evaluate whether these maps encompass SIOFA's existing IPAs, by overlapping their spatial layers.

We used the QMarxan Toolbox v2.0.1 plugin (Apropos Information Systems Inc.) for QGIS (QGIS Development Team, 2022) to prepare Marxan's input files and calibrate the scenarios with the available calibration functions. For Marxan, we used Marxan v2.4.3. All manipulation of geographic layers, input calculations (e.g., area calculations) and results preparation were performed using the R Statistical Software (R Core Team, 2021).

Table 2. Summary of the scenarios with varying parameters: the boundary Length Modifier (BLM), cost layer, status of the planning units (locked-in or not in the analysis), and different conservation targets for the conservation features.

Scenario	Target	Conservation features	Cost
Low target	5 %	Bioregions, seafloor geomorphology, and EBSA	Catch (in metric tonnes)
Medium target	10 %	Bioregions, seafloor geomorphology, and EBSA	Catch (in metric tonnes)
High target	15 %	Bioregions, seafloor geomorphology, and EBSA	Catch (in metric tonnes)
Low target – no cost	5 %	Bioregions, seafloor geomorphology, and EBSA	Equal cost for all PUs
Medium target – no cost	10 %	Bioregions, seafloor geomorphology, and EBSA	Equal cost for all PUs
High target – no cost	15 %	Bioregions, seafloor geomorphology, and EBSA	Equal cost for all PUs

4. Results

In total, there were 16 ecological features to consider for conservation in the Marxan analyses. Table 3 lists the ecological features for which targets were set, grouped by theme. Individual PUs can include 1 to 7 ecological features, with only 26 PUs that do not have any ecological feature, found mainly in the eastern part of SIOFA (at 40°S, 110°E) and at coordinates 25°S, 70°E (Figure 7). Although we expect to have one ecological feature at least per PU given the probabilistic nature of the bioregionalization layer, the 26 PUs with no assigned bioregion corresponded to pixels that had equal maximum suitability for two bioregions at the same time. Since the final bioregionalization map only shows one subregion per pixel (i.e., bioregions are mutually exclusive), we decided to leave these pixels unclassified. On the other hand, areas with the highest concentration of features are found in Saya de Malha, the western side of Walters Shoal, the southwest Indian Ridge, the southern slope of Broken Ridge, and the ridges in general. The lowest numbers of ecological features

considered in this analysis are found at abyssal depths (> 3,500 m water depth) of SIOFA's Convention Area, since the selected geomorphic features did not include features at those depths.

Marxan calibrations were performed for all three ecological scenarios following the guidelines of the Marxan Good Practices Handbook (Ardron et al., 2010) and the QMarxan Toolbox plugin (Apropos Information Systems Inc.)(Apropos Information Systems, v2.0.2). Description of the results and process followed can be found in Appendix A and Appendix B.

In the next sections, we provide a detailed description of the Marxan results for each of the scenarios with the three ecological conservation targets, including selection frequency and best solution (i.e., best score) maps.

Table 3. Group of ecological features included in the ecological Marxan analyses, the number of planning units (PUs) and percentage of total PUs they occupy.

Group name	Conservation features	No. of features per group	Total no. of PUs present	% of Total Area
EBSA	**EBSA**	1	4,381	17.92
Bioregionalization	**Bio_1 (bioregion 1.1)**		2,114	8.65
	Bio_2 (bioregion 1.2)		3,944	16.12
	Bio_3 (bioregion 1.3)		3,029	12.39
	Bio_4 (bioregion 1.5)		2,760	11.29
	Bio_5 (bioregion 1.7)		2,514	10.28
	Bio_6 (bioregion 2.1)		3,805	15.56
	Bio_7 (bioregion 2.4)		90	0.36
	Bio_8 (bioregion 3.1)		9,644	39.44
	Total	8	27,900	114.09*
Seafloor geomorphology	**Ridge**		2,950	12.06
	Rise		203	0.83
	Plateau		1,374	5.62
	Canyon		148	0.61
	Slope		75	0.31
	Shelf		69	0.28
	Seamount		3,540	14.47
	Total	7	8,359	34.18

*This value amounts to more than 100% because features within the group overlap with each other.

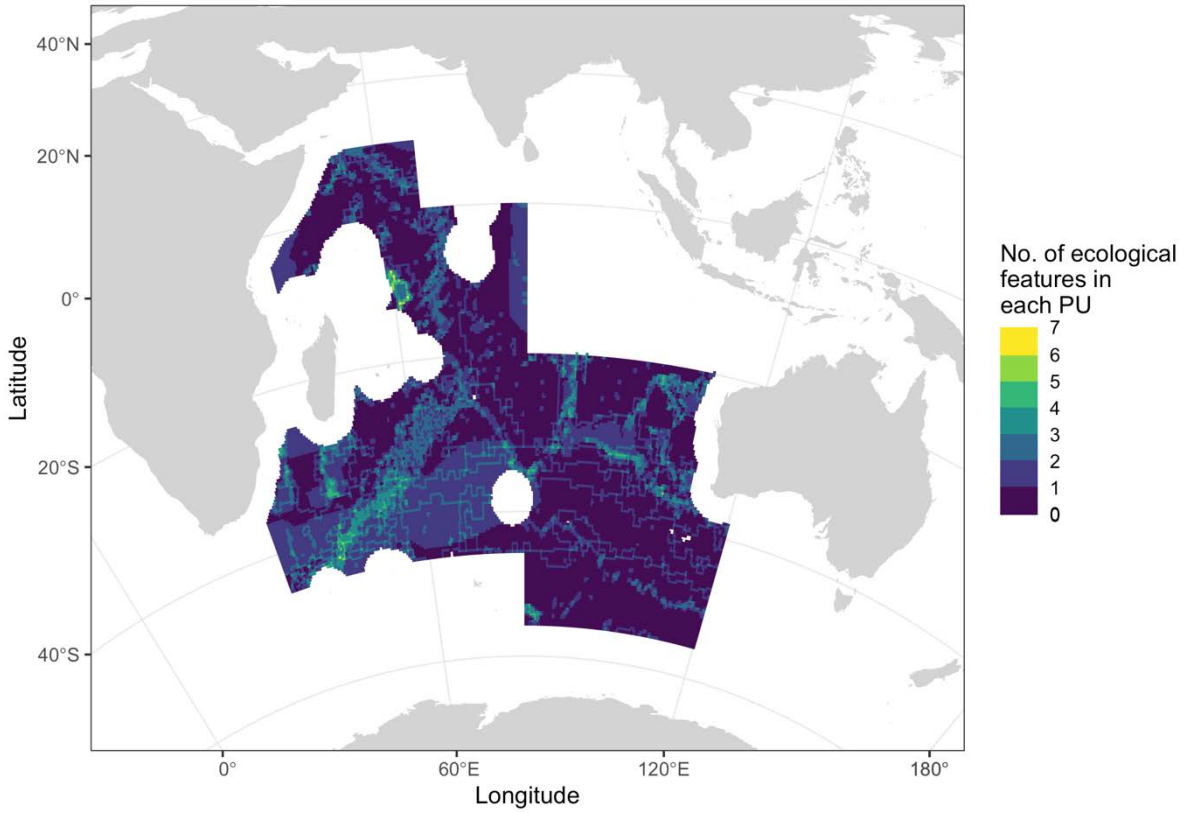

Figure 7. Total number of ecological features in each PU.

4.1 Scenarios with cost

4.1.1 Low target scenario

Marxan analyses used 80 million iterations, 100 solutions, a variable SPF (from 1 to 1.4) and a BLM of 0.01. The selection frequency map is presented together with three individual solutions, including the best solution (Figure 8).

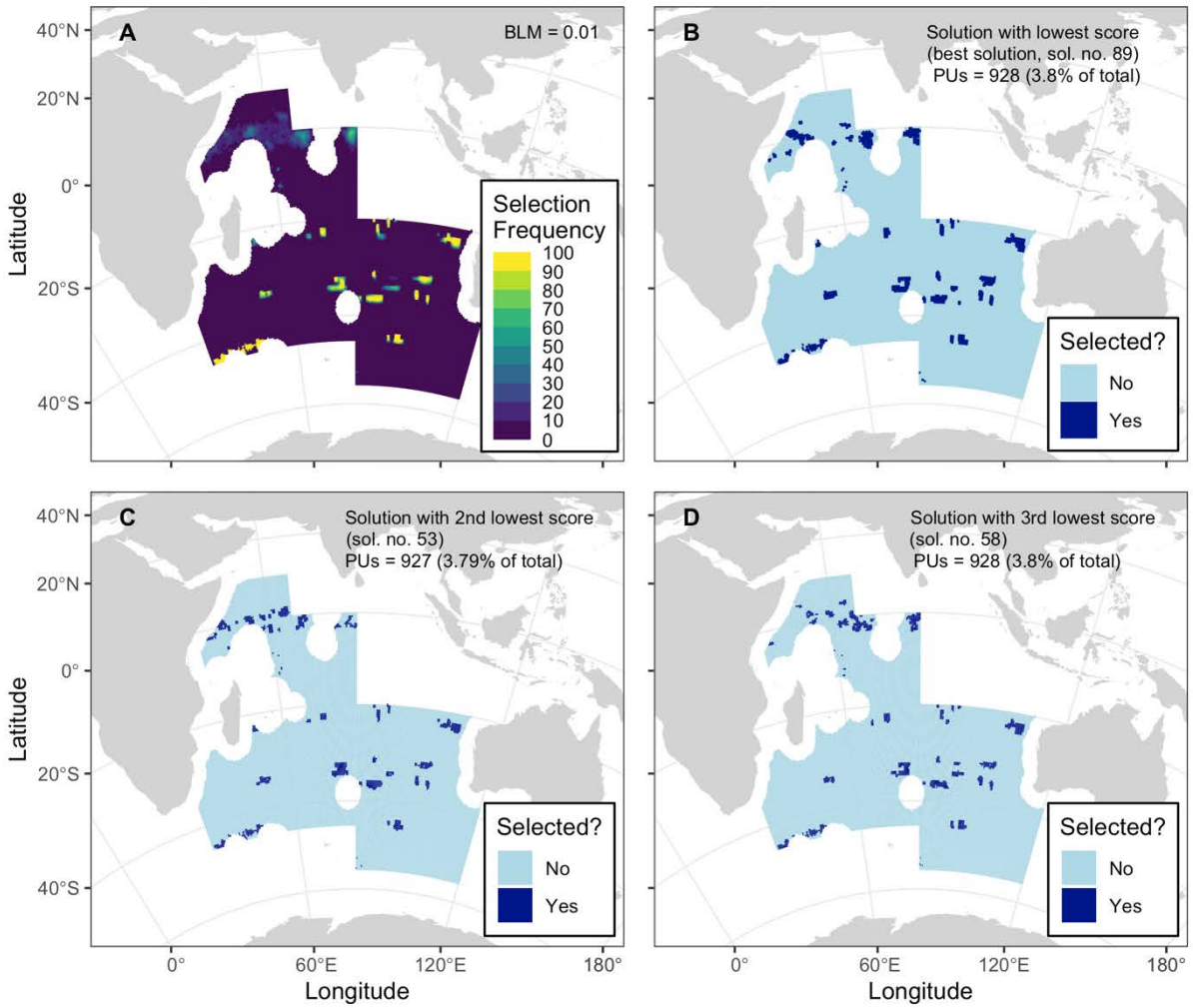

Figure 8. Ecological Marxan analysis results for the low-target scenario (SPF = 1-1.4; BLM = 0.01, iterations = 80 million, solutions = 100): selection frequency (Map A), which indicates the number of times a PU was selected in the analysis out of 100 possible solutions; best solution, or solution with the lowest Marxan score (Map B); second best solution (Map C); and third best solution (Map D).

The area needed to meet at least 99% of all targets based on the average values of 100 Marxan solutions is 3.8 % of the total study area. The three solutions with the lowest scores (including the best solution) are similar in terms of area selected, cost and boundary length (Table 4). For the best solution, the most contiguous set of PUs have similar areas. These are found in the northeast of Seychelles, SIOFA border at Del Caño Rise, an eastward area of the

SW Indian Ocean Ridge at the latitude of Walters Shoal, the northwest and east of Saint Paul and Amsterdam Islands (over the Mid-Indian Ridge), southeast Indian Ridge, south of Broken Ridge, Ninety East Ridge, and the Wallaby and Zenith Plateaus.

Table 4. Summary information on solutions presented from the low-target conservation scenario.

	100 Solution Avg.	Best solution (lowest score)	2nd Best	3rd Best
Score[a]	938.9	937.93	938.26	938.29
Cost (Tonnes)[b]	927.75	928	927	928
Cost (% from total)[c]	2.34	2.34	2.34	2.34
Area selected[d]	1,270,090	1,270,432	1,269,063	1,270,432
% of total area	3.79%	3.80%	3.79%	3.80%
Boundary Length[e]	1006.88	932	968	982
Shortfall[f]	1,190,313,150	6,915,49,000	1,688,810,000	482,255,000
Missing targets[g]	0.08	0	0	0
MPM	0.988	0.998	0.995	0.999

[a] Total score of the solution, which is the sum of the PU cost, the boundary cost, and the penalty (see Eq. 1)
[b] Cost of the solution, which in this case corresponds to tonnes of fish catch
[c] Total cost = 39,657.38 (with 1s); 15,607.38 (without 1s)
[d] Area measured in km^2. Size of PU 37 km x 37 km. Total area = 1,369 km^2 x 24,453 PUs = 33,476,157 km^2
[e] Total external perimeter of all the PUs forming a solution (set to 1; see Methods) = 49,338
[f] Combined shortfall for missing features (in m^2)
[g] Number of features missing targets

When exploring the distribution of the selection frequency values (Figure 8A), 13% (449 out of 3,409 PUs) of the PUs were selected in >80 solutions. Notably, below 20°S the percentage of PUs selected showed a consistent pattern with 34% of PUs (449 out of 1,326) selected in at least 80 solutions, whereas at latitudes north of 20°S there were no PUs appearing in at least 80 solutions (0 out of 2,083).

The amount of each ecological feature represented in the best solution is presented in Table 5. All ecological features met their specific targets in this scenario. Many features exceeded their target; this indicates that ecological features overlap in multiple different combinations in multiple locations with some of them present in most overlaps. For instance, the target set for the feature “Plateau” and “Ridge” is met at 143% and 142%, respectively. Having features that exceed their targets cannot be avoided in Marxan analysis and trying to reduce those targets is often impossible without reducing other targets below their 100% goal. It is almost impossible to select a combination of PUs that would meet every single target without overshooting at least some of the targets significantly. These “overshoots” do not mean that more area is included in the solutions than is necessary to meet all targets, i.e., they are not indicative of inefficient spatial configurations. Rather, they highlight that some features are driving the analysis much more than others.

Table 5. Detailed information on the amount held inside the best solution in the low-target scenario for all ecological features, the area that occupies inside the solution (no. of occurrences held by PU area size), the percentage occupied from the total area of the bioregion and if targets were met and by how much (in percentage).

Feature Name	Target	Amount Held	Occurrences Held	Area in solution (km^2)	% of total area	Target Met	Proportion met**
EBSA	2.2375E+11	2.4296E+11	222	303918	0.91	yes	108.6
Seamount	8.0733E+10	9.1257E+10	193	264217	0.79	yes	113.0
Shelf	2680618416	3395118970	3	4107	0.01	yes	126.7
Slope	1368180499	1552415529	3	4107	0.01	yes	113.5
Canyon	1496620585	1658627894	4	5476	0.02	yes	110.8
Plateau	5.1481E+10	7.3682E+10	111	151959	0.45	yes	143.1
Rise	9119262750	9914347185	10	13690	0.04	yes	108.7
Ridge	8.2223E+10	1.1694E+11	213	291597	0.87	yes	142.2
bio_7	2582694486	2592280050	4	5476	0.02	yes	100.4
bio_6	1.6086E+11	1.6084E+11	165	225885	0.67	yes	100.0
bio_3	1.3997E+11	1.3989E+11	140	191660	0.57	yes	99.9
bio_4	1.2236E+11	1.2207E+11	121	165649	0.49	yes	99.8
bio_5	1.0594E+11	1.0585E+11	126	172494	0.52	yes	99.9
bio_2	1.9315E+11	1.9312E+11	167	228623	0.68	yes	100.0
bio_8	2.7051E+11	2.7052E+11	221	302549	0.90	yes	100.0
bio_1	1.1835E+11	1.1817E+11	109	149221	0.45	yes	99.8

**This is calculated by Amount Held * 100/ Target Amount

4.1.2 Medium target scenario

Marxan analyses used 80 million iterations, 100 solutions, a variable SPF (from 1 to 1.3) and a BLM of 0.01. The selection frequency map is presented together with three individual solutions, including the best solution (Figure 9).

The area needed to meet at least 99% of all targets based on the average values of 100 Marxan solutions is 9.25 % of the total study area. The three solutions with the lowest scores (including the best solution) are similar in terms of area selected, cost and boundary length (Table 6). The solution is also similar to the low-target scenario in terms of the distribution of the reserves, where PUs have been selected and built around the same areas.

When exploring the distribution of the selection frequency values (Figure 9A), about 27 % (1,144 out of 4,193 PUs) of the PUs were selected in >80 solutions. Notably, below 20°S the percentage of PUs selected showed a consistent pattern with 48% of PUs (1,032 out of 2,164) were selected in at least 80 solutions, whereas at latitudes north of 20°S the percentage of PUs appearing in at least 80 solutions was 16% (325 out of 2,029).

The amount of each ecological feature represented in the best solution is presented in Table 7. All ecological features met their specific targets in this scenario. Three features exceeded their target too. In this case, the feature "Rise" met its target at 165%.

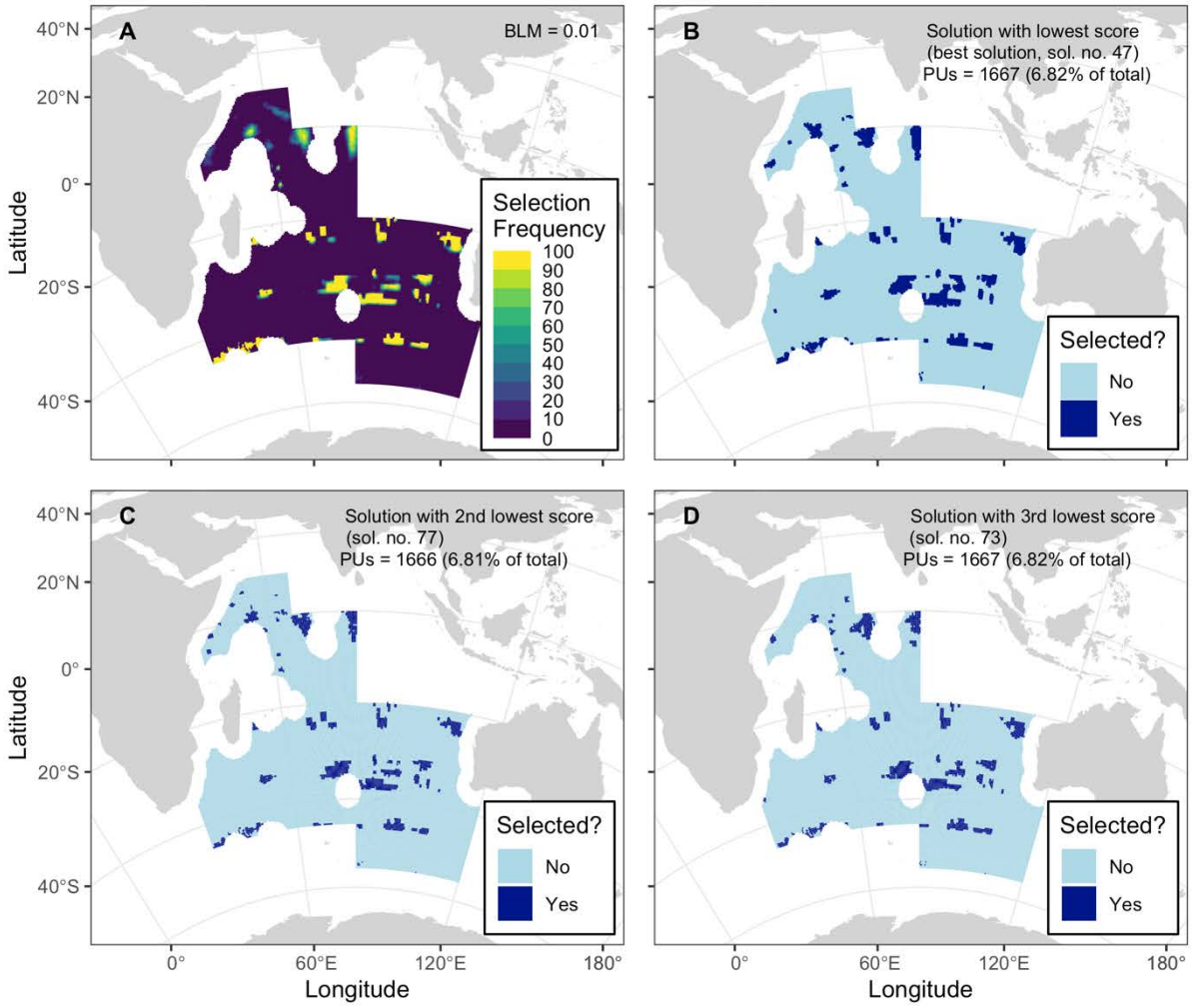

Figure 9. Ecological Marxan analysis results for the medium-target scenario (SPF = 1-1.3; BLM = 0.01, iterations = 80 million, solutions = 100): selection frequency (Map A), which indicates the number of times a PU was selected in the analysis out of 100 possible solutions; best solution, or solution with the lowest Marxan score (Map B); second best solution (Map C); and third best solution (Map D).

Table 6. Summary information on solutions presented from the medium-target conservation scenario.

	100 Solution Avg.	Best solution	2nd Best	3rd Best
Score[a]	1682.26	1681.14	1681.5	1681.54
Cost (Tonnes)[b]	1666.96	1667	1666	1667
Cost (% from total)[c]	4.20	4.20	4.20	4.20
Area selected[d]	2,282,068	2,282,123	2,280,754	2,282,123
% of total area	6.82	6.82	6.81	6.82
Boundary Length[e]	1409.15	1340	1384	1332
Shortfall[f]	1,339,261,210	7.47E+08	1.78E+09	1.44E+09
Missing targets[g]	0.01	0	0	0
MPM	0.992	0.999	0.999	0.999

[a] Total score of the solution, which is the sum of the PU cost, the boundary cost and the penalty (see Eq. 1)
[b] Cost of the solution, which in this case corresponds to tonnes of fish catch
[c] Total cost = 39,657.38 (with 1s); 15,607.38 (without 1s)
[d] Area measured in km^2. Size of PU 37 km x 37 km. Total area = 1,369 km^2 x 24,453 PUs = 33,476,157 km^2
[e] Total external perimeter of all the PUs forming a solution (set to 1; see Methods) = 49,338
[f] Combined shortfall for missing features (in m^2)
[g] Number of features missing targets

Table 7. Detailed information on the amount held inside the best solution in the medium-target scenario for all ecological features, the area that occupies inside the solution (no. of occurrences held by PU area size), the percentage occupied from the total area of the bioregion and if targets were met and by how much (in percentage).

Feature Name	Target	Amount Held	Occurrences Held	Area in solution (km^2)	% of total area	Target Met	Proportion met**
EBSA	4.475E+11	4.4757E+11	406	555814	1.66	yes	100.01
Seamount	1.6147E+11	1.6363E+11	326	446294	1.33	yes	101.34
Shelf	5361236831	6651421380	8	10952	0.03	yes	124.07
Slope	2736360997	2811823449	5	6845	0.02	yes	102.76
Canyon	2993241171	3011457597	10	13690	0.04	yes	100.61
Plateau	1.0296E+11	1.1732E+11	178	243682	0.73	yes	113.95
Rise	1.8239E+10	3.0195E+10	28	38332	0.11	yes	165.56
Ridge	1.6445E+11	1.665E+11	318	435342	1.30	yes	101.25
bio_7	5165388973	5158395590	7	9583	0.03	yes	99.86
bio_6	3.2173E+11	3.2137E+11	335	458615	1.37	yes	99.89
bio_3	2.7994E+11	2.7977E+11	289	395641	1.18	yes	99.94
bio_4	2.4472E+11	2.4461E+11	248	339512	1.01	yes	99.96
bio_5	2.1188E+11	2.1184E+11	248	339512	1.01	yes	99.98
bio_2	3.8631E+11	3.8626E+11	330	451770	1.35	yes	99.99
bio_8	2.7051E+11	2.705E+11	260	355940	1.06	yes	100.00
bio_1	2.367E+11	2.3667E+11	207	283383	0.85	yes	99.99

**This is calculated by Amount Held * 100/ Target Amount

4.1.3 High target scenario

Marxan analyses used 80 million iterations, 100 solutions, a variable SPF (from 1 to 1.1) and a BLM of 0.01. The selection frequency map is presented together with three individual solutions, including the best solution (Figure 10).

The area needed to meet at least 99% of all targets based on the average values of 100 Marxan solutions is 9.86% of the total study area. The three solutions with the lowest scores (including the best solution) are similar in terms of area selected, cost and boundary length (Table 8). Again, the solution is also similar to the low and medium-target scenarios in terms of the distribution of the reserves, where PUs have been selected and built around the same areas.

When exploring the distribution of the selection frequency values (Figure 10A), about 44 % (1,903 out of 4,344 PUs) of the PUs were selected in >80 solutions. Notably, below 20°S the percentage of PUs selected showed a consistent pattern with 58% of PUs (1,669 out of 2,902) were selected in at least 80 solutions, and at latitudes north of 20°S the percentage of PUs appearing in at least 80 solutions was 33% (481 out of 1,442).

The amount of each ecological feature represented in the best solution is presented in Table 9. All ecological features met their specific targets in this scenario. Some features exceeded their target by 15%.

Table 8. Summary information on solutions presented from the high-target conservation scenario.

	100 Solution Avg.	Best solution (lowest score)	2nd Best	3rd Best
Score[a]	2431.29	2430.19	2430.67	2430.68
Cost (Tonnes)[b]	2412.52	2412	2412	2412
Cost (% from total)[c]	6.08	6.08	6.08	6.08
Area selected[d]	3,302,740	3,302,028	3,302,028	3,302,028
% of total area	9.87	9.86	9.86	9.86
Boundary Length[e]	1763.26	1695	1725	1743
Shortfall[f]	1,252,885,980	1.35E+09	1.68E+09	1.24E+09
Missing targets[g]	0.02	0	0	0
MPM	0.993	0.998	0.995	0.986

[a] Total score of the solution, which is the sum of the PU cost, the boundary cost, and the penalty (see Eq. 1)
[b] Cost of the solution, which in this case corresponds to tonnes of fish catch
[c] Total cost = 39,657.38 (with 1s); 15,607.38 (without 1s)
[d] Area measured in km^2. Size of PU 37 km x 37 km. Total area = 1,369 km^2 x 24,453 PUs = 33,476,157 km^2
[e] Total external perimeter of all the PUs forming a solution (set to 1; see Methods) = 49,338
[f] Combined shortfall for missing features (in m^2)
[g] Number of features missing targets

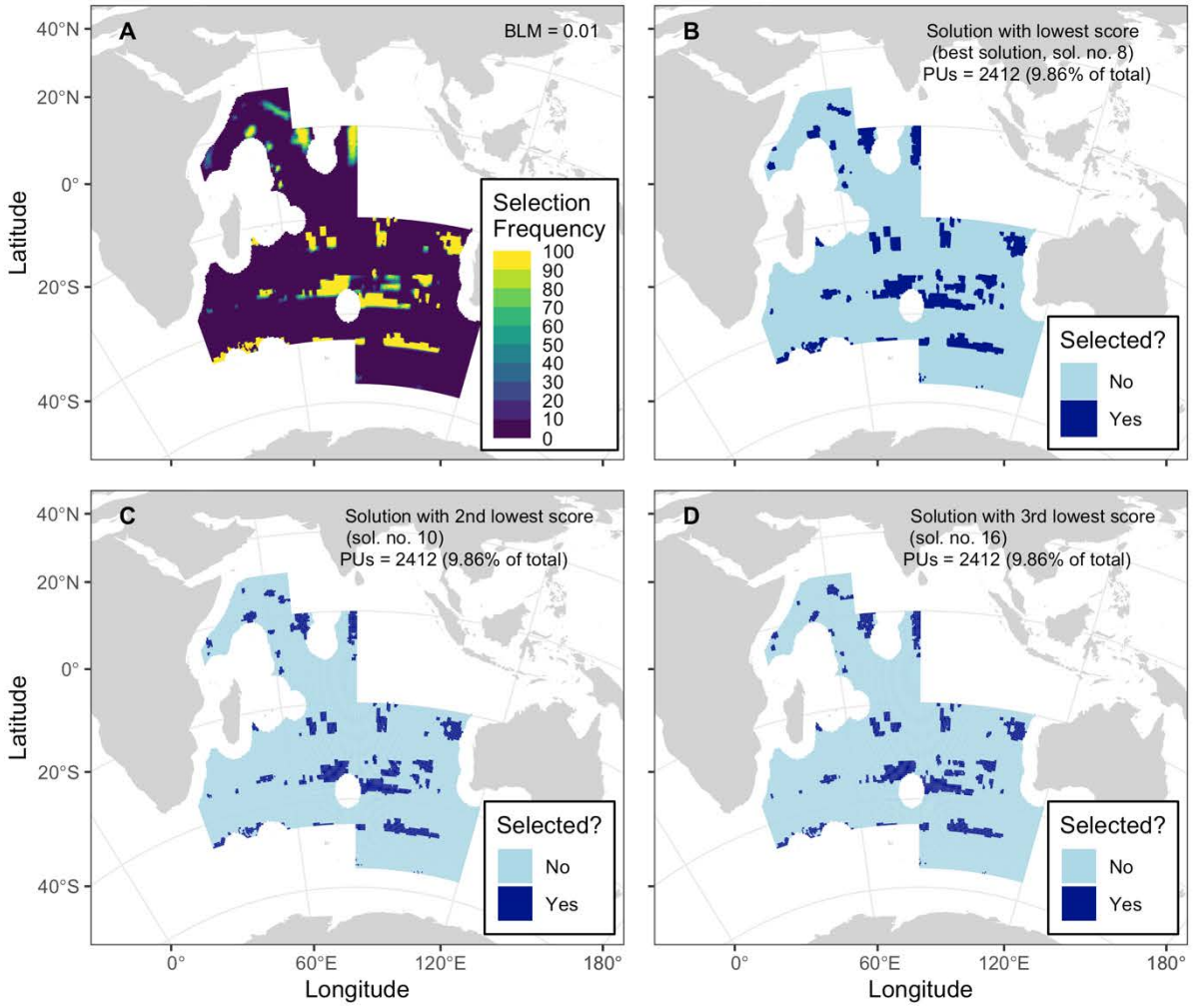

Figure 10. Ecological Marxan analysis results for the high-target scenario (SPF = 1 -1.1; BLM = 0.01, iterations = 80 million, solutions = 100): selection frequency (Map A), which indicates the number of times a PU was selected in the analysis out of 100 possible solutions; best solution, or solution with the lowest Marxan score (Map B); second best solution (Map C); and third best solution (Map D).

Table 9. Detailed information on the amount held inside the best solution in the high-target scenario for all ecological features, the area that occupies inside the solution (no. of occurrences held by PU area size), the percentage occupied from the total area of the bioregion and if targets were met and by how much (in percentage).

Feature Name	Target	Amount Held	Occurrences Held	Area in solution (km^2)	% of total area	Target Met	Proportion met**
EBSA	6.7126E+11	6.7137E+11	612	837828	2.50	yes	100.02
Seamount	2.422E+11	2.4323E+11	497	680393	2.03	yes	100.43
Shelf	8041855247	9286416921	13	17797	0.05	yes	115.48
Slope	4104541496	4239923171	9	12321	0.04	yes	103.30
Canyon	4489861756	4502194408	16	21904	0.07	yes	100.27
Plateau	1.5444E+11	1.5643E+11	229	313501	0.94	yes	101.28
Rise	2.7358E+10	3.1791E+10	31	42439	0.13	yes	116.21
Ridge	2.4667E+11	2.6498E+11	457	625633	1.87	yes	107.42
bio_7	7748083459	7736242524	10	13690	0.04	yes	99.85
bio_6	4.8259E+11	4.8232E+11	507	694083	2.07	yes	99.94
bio_3	4.1992E+11	4.1966E+11	431	590039	1.76	yes	99.94
bio_4	3.6707E+11	3.6688E+11	373	510637	1.53	yes	99.95
bio_5	3.1781E+11	3.1767E+11	362	495578	1.48	yes	99.95
bio_2	5.7946E+11	5.793E+11	499	683131	2.04	yes	99.97
bio_8	2.7051E+11	2.7051E+11	277	379213	1.13	yes	100.00
bio_1	3.5504E+11	3.5474E+11	299	409331	1.22	yes	99.92

**This is calculated by Amount Held * 100/ Target Amount

Best solutions for scenarios low, medium, and high target were compared with each other in terms of the area lost for fishing (or total area selected as part of the reserve network) if those areas were protected. This analysis was performed by combining the polygons defined by the best solution with the layer containing the aggregated fishing catch. None of the three scenarios overlapped with the PUs containing fishing information.

4.2 Scenarios with no cost

The scenarios with no cost were also calibrated based on the calibration of scenarios with cost. We present here only the selection frequency maps, as the aim of these no cost scenarios is to highlight potential biodiversity areas. We present the selection frequency maps when there is no constraint in level of compactness (BLM = 0) and at the same level of compactness (BLM = 0.01) calibrated for the cost scenarios. Calibration details for these scenarios can be found in Appendix B.

4.2.1 Low target, no cost

Marxan analyses used 80 million iterations, 100 solutions, and a variable SPF from 1 to 1.6 for a BLM set to 0. For a BLM of 0.01, the SPF varied from 1 to 1.7. The selection frequency map of each scenario configuration is presented in Figure 11.

Assessment of the presence of any of the five interim BPAs in the selection frequency maps revealed the inclusion of part of "Coral" one time when there was no constraint for the level of compactness (i.e., BLM = 0). When the constraint of compactness increased (i.e., BLM = 0.01), no PUs that included IPAs were selected.

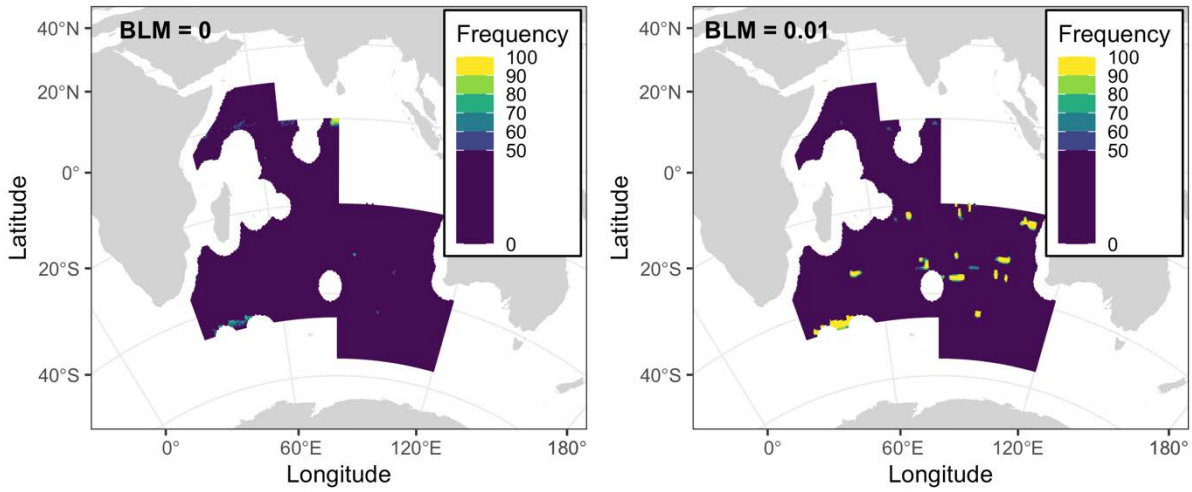

Figure 11. Selection frequency maps for the low target, no cost, scenario for two BLM values.

4.2.2 Medium target, no cost

Marxan analyses used 80 million iterations, 100 solutions, and a variable SPF from 1 to 1.3 for a BLM set to 0. For a BLM of 0.01, the SPF varied from 1 to 1.4. The selection frequency map of each scenario configuration is presented in Figure 12.

When there was no constraint for the level of compactness (i.e., BLM = 0), the assessment of the presence of any of the five IPAs in the selection frequency maps revealed the inclusion of "Coral" in six PUs with a selection frequency of 1 to 7 times. It also revealed the inclusion of "Fool's Flat" in three PUs with a selection frequency of 1 to 2 times. When the constraint of compactness increased (i.e., for BLM = 0.01), no PUs that included IPAs were selected.

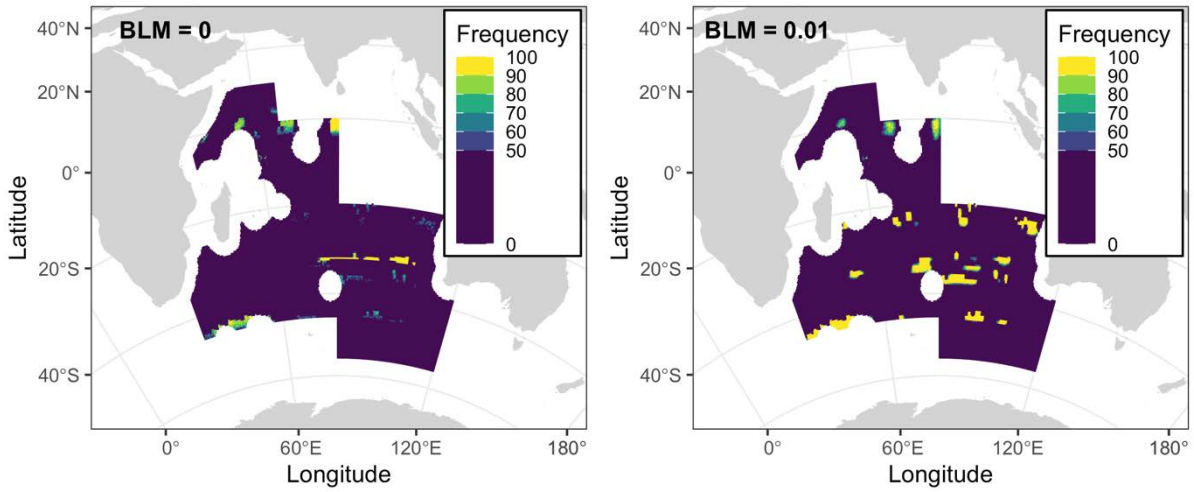

Figure 12. Selection frequency maps for the medium target, no cost, scenario for two BLM values.

4.2.3 High target, no cost

Marxan analyses used 80 million iterations, 100 solutions, and a variable SPF from 1 to 1.2 for a BLM set to 0. For a BLM of 0.01, the SPF varied from 1 to 1.1. The selection frequency map of each scenario configuration is presented in Figure 13.

When there was no constraint for the level of compactness (i.e., BLM = 0), the assessment of the presence of any of the five IPAs in the selection frequency maps revealed the inclusion of four IPAs: "Fools Flat" in one PU (selected three times), "Walters Shoal" in five PUs (selected ranging from one to five times), "Coral" in ten PUs (with a selection frequency of 1 to 13 times), and "Middle of What" in one PU (selected five times). When the constraint in compactness increased (i.e., for BLM = 0.01), no PUs that included IPAs were selected.

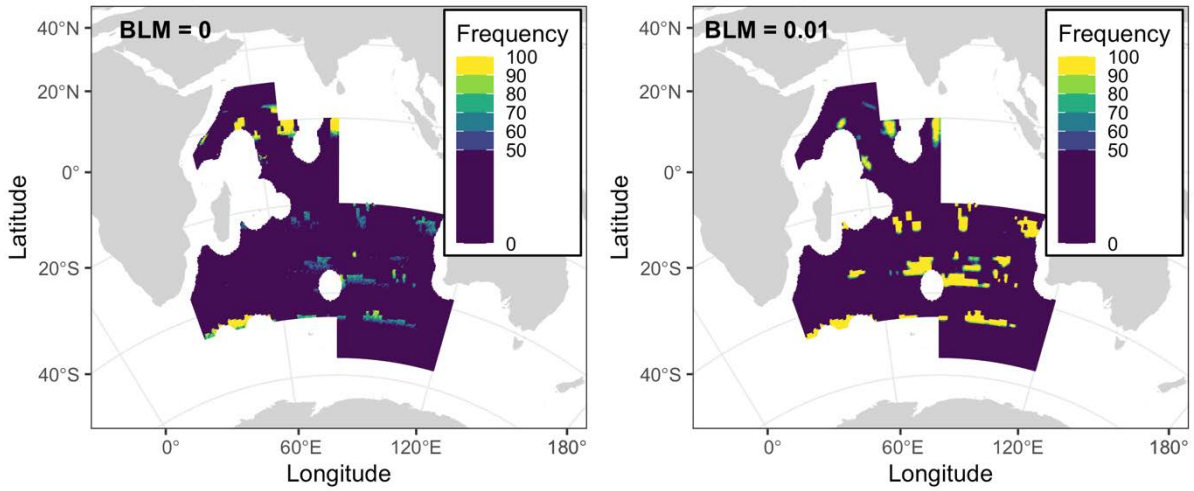

Figure 13. Selection frequency maps for the high target, no cost, scenario for two BLM values.

5. Discussion

This study aimed to investigate systematic conservation planning for SIOFA using the decision-support tool Marxan. We described the procedure to perform systematic conservation planning, the software Marxan, the parameters involved in the analysis, the rationale for the choice of conservation features and the calibration needed to meet targets. For each of the three cost scenarios run, we showed the selection frequencies and the best solutions in terms of lowest score according to the objective function that Marxan solves. These configurations should not be interpreted as best solutions, rather, they are a very good solution within a continuum of options (Ardron et al., 2010).

The scenarios presented met the ecological targets set for the planning, and for some even exceeded them. As previously mentioned, having features that exceed their targets cannot be avoided in Marxan analysis and trying to reduce those targets is often impossible without reducing other targets below their 100% goal. The total area needed to meet targets varied according to the cost function and the magnitude of the ecological targets. The low-ecological target scenario required 3.8% of the overall area, the medium target 6.82% and the high target 9.83% of the overall area. These proportions are relatively low due to the fact that the number of conservation features was low and, hence, there were less constraints on Marxan to find efficient solutions.

We carried out a conventional Marxan analysis where Marxan attempts to balance network design achieving ecological targets while taking into account the spatial distribution of associated economic costs of different places – in this case, based on aggregated catch. The presented solutions did not overlap with the aggregated layer of fishing catch, as Marxan sought to avoid losing potentially valuable areas for fishing. Indeed, fishing areas were completely avoided. In this case, even if equity across fisheries was not considered, despite different gears having different impact on the seafloor, none of the individual fishing

footprints (by gear type) would be impacted. However, these results should be interpreted relative to the BLM chosen, that is, the level of compactness and the number of conservation features included in the investigation. Other level of compactness and inclusion of conservation features could come with actual loss of fishing areas.

The strong role of the spatial distribution of cost in the results can be understood by two aspects. First, there is the sheer aspect of space: SIOFA encompasses a very large area where bottom fishing is concentrated in particular areas. Thus, there is "enough" remaining area for Marxan's algorithm to reach efficient solutions without impacting important fishing grounds. Second, all models are dependent on the information which they are fed and Marxan's objective function is no different. In other words, results are dependent on the ecological input layers that one deems appropriate to explore. Here, as a first exploration, our choice of ecological targets (i.e., the percentage for conservation) was based on the literature and international agreements. However, ecologically, this may result in fragmented or inadequate or incoherent levels of protection. For instance, one of the conservation features in this investigation was "Canyons", as canyons are an important habitat for benthic organisms (Pearman et al., 2020). Setting aside 10% of "canyon" habitat for protection makes little sense ecologically. Rather, an alternative would be to consider canyons as a whole and aim for the protection of the number of canyons (e.g., 10% of canyons, or any other ecological target), but not the actual area. The same rationale could be applied for seamounts (see Watling & Auster, 2021 for a discussion on the protection of seamounts). In contrast, for layers such as that of the bioregions, it makes sense to deal with the total area or coverage rather than with individual occurrences. Thus, although in this investigation we only considered coverage, SIOFA should consider the nature of the ecological features included in future analyses and set appropriate targets that make ecological sense. This implementation is possible because Marxan allows to include conservation features in three ways: as occurrences, as proportion (as we did here), or as a set amount.

In addition to considering how to integrate each layer of information, SIOFA should convene on appropriate target levels for the protection of each feature. On the one hand, there is the issue of uncertainty associated with some of the conservation layers. Some features, such as the bioregionalization, are work in progress but considerably influence the Marxan analysis. In order to reduce the associated uncertainty to this layer, higher target levels could be applied to other "more certain" features, such as the seamounts, or the individual EBSAs. The Southern Indian Ocean, especially the areas beyond national jurisdiction, are poorly known; therefore, how to include the selected features requires careful evaluation.

On the other hand, there is the issue of varying sizes between conservation features. For instance, in a first exploratory analysis, we included the conservation feature "bioregion 3.1" (here as "bio_8") with the same percentage of protection as the other features. Bioregion 3.1 was previously characterised as the prevalent bioregion at abyssal depths (>3,500m water depth). This feature occupies the greatest number of PUs in the study area (Table 3) which, when setting its protection percentage to the same as other bioregions, also resulted in large areas in the reserve network. Therefore, different protection levels (i.e., ecological targets) can be set for the conservation features. We exemplified it by lowering the ecological target for Bioregion 3.1 on the premise that it will be less impacted by certain human activities (e.g., here fishing because the depths are out of technological capacity). Yet, caution must be

exerted because analysis of fishing activity in ToR5 indicated that fishing occurs within this bioregion, therefore this bioregion cannot be excluded from the Marxan analyses. The reason behind this apparent discrepancy is due to the spatial resolution of the bioregionalization models (at 1° latitude-longitude). A coarse spatial resolution will aggregate into one single value many different depths, which becomes especially critical when these cells are on the edge of topographic complex areas, such as ridges, where depth usually varies more rapidly along vertical than horizontal scales. Thus, our potential solution was to set a lower conservation target for this feature to reduce the protection area (0.025% for all target scenarios). Other options could include the weighting of some areas according to the level of confidence of each pixel (based on the bioregionalization results).

Future work should assess the effect of varying the size of the planning units to 1° latitude-longitude. For this investigation we selected planning units of dimension 37 km x 37 km on the basis of a compromise between manageable units and informative layers. This resolution was however not shared by any of the two most important layers: fishing effort (the 'cost') was available at 10 arc minutes and the bioregionalization model at 1° latitude-longitude (approximately 111 km x 111 km at the equator). In practice this meant that we aggregated the fishing data and disaggregated raster cells for the bioregions to align resolutions at 37 x 37 km – but note that this alignment of resolutions does not improve the resolution of bioregions. In fact, the consequence of the difference in the resolutions was evidenced in Figure 7 (sum of the number of features per PU), where there are some lines that illustrate the effect of the coarser resolution of bioregions. Indeed, when Figure 6 is compared against the maps of the individual features, it becomes clear that the lines stem from the bioregionalization map: the original grid cells (at 1° latitude-longitude) become sparse when disaggregated into smaller units. This is the reason why the discrepancy between the original spatial resolution of the input layers, namely the bioregions and the fishing footprint, should be further investigated using a grid with PUs of 1° latitude-longitude. Although not appropriate for management because of such coarse resolution, it will still be valuable to examine the potential differences in the spatial configuration of the reserves and changes in cost.

We did not consider the inclusion of SIOFA's five IPAs in the analyses as Marxan considers that such areas should be entirely closed to all human activities that can impact species. Nonetheless, the ecological Marxan analysis provided some insights about the adequacy of the location of these IPAs. To assess if these areas were selected in any of the scenarios or highlighted as important areas (relative to the conservation features included in this investigation), we overlapped the resulting selection frequency maps and reserve networks of each ecological scenario with cost with the IPAs. Whilst none of the maps integrated these areas, the selection frequency maps of the scenarios with no cost selected PUs that included, with a low frequency, some of the IPAs. This does not mean that those areas (the IPAs) are not biologically important. Rather, it means that we don't have sufficient data to be able to detect the importance of these areas. As evidenced in Ramiro-Sánchez & Leroy (2022) where we showed data shortfalls, the current knowledge on the distribution of biodiversity and VMEs in the SIOFA area is extremely deficient. The suggested areas identified by Marxan in this report are based on the analysis of all the available knowledge in the SIOFA area. Alternatively, we suggest considering additional useful strategies to cope with data shortfalls,

which could include locations where the move-on rule has been triggered or lock-in areas with significant biological evidence.

We also insist on the necessity to acquire more data on the distribution of VME indicators in the SIOFA area, which remains deficient. Indeed, data shortfalls made it impossible to map the observed VME diversity, because almost the entire SIOFA area was empty from samples, or to predict the spatial distribution of VME diversity, because the models were too uncertain to be useful (see report on ToR 3). Therefore, the VME diversity could not be included as a layer in the Marxan exercise, which is an important limit, because the selected areas did not account for the most diverse areas where VME indicators are likely to persist. The only way to address this issue will be to selectively acquire more data on the distribution of VMEs in the SIOFA area, targeting specifically the most deficient areas visible in Ramiro-Sánchez & Leroy (2022) (report on VME occurrence data / completeness). Nonetheless, we encourage the inclusion of the predicted hotspots calculated using stacked-taxa distribution modelling in TOR3 (of the present project PAE2021-01) in future SCP assessments in SIOFA, provided that the uncertainty associated with the modelling methods is understood.

Finally, it is important to remember that Marxan is a decision-support tool in spatial decision-making process. That means that Marxan outputs should not be interpreted as final products but to guide managers towards areas important to protect. As evidenced above with the IPAs, these areas are biologically significant (there is evidence of a VME habitat) but were not selected in the reserve. Therefore, decisions made about where to designate areas will need to factor in all possible available information that is not integrated into a spatial layer, such as scientific expertise and traditional ecological knowledge about the environment in question, practical and political constraints, and human conflicts. This is why Marxan analyses are placed into consultations that involve different stakeholders.

This investigation intended to highlight which details could be retained or dropped in future work carried out by SIOFA. Notably, the future agreement on ecological targets for individual features and the selection of conservation features. Thus, these Marxan outputs could be treated as a starting point for discussions about what to include in, but also what a draft network could look like. The human and ecological complexity involved in spatial planning makes planning iterative, with continuous feedback and guidance to shape and improve plans at each iteration. As such, it does not matter that a ”scenario” is not perfect (e.g., whether there are gaps in ecological data, missing costs, or whether the BLM setting could be other). Rather, these outputs could be used in a workshop with scientific experts and stakeholders to react against and come up with an improved network configuration, in a process of negotiations and trade-offs.

6. Conclusions

This analysis has explored the effects of cost on a hypothetical reserve network planning in the SIOFA area using three groups of conservation features and the decision-support tool Marxan. The integration of all fisheries did not result in a loss for the sector, as the fishing footprint was completely avoided by the planning software. Moreover, with the ecological

targets set, approximately only 4%, 7% and 10% of the overall area was set aside for protection in the low, medium, and high target cost scenarios, respectively. The key selected areas consistently selected across all three scenarios with cost included the northeast of Seychelles, SIOFA border at Del Caño Rise, an area of the SW Indian Ridge at the latitude of Walters Shoal, the northwest and east of Saint Paul and Amsterdam Islands (mid-Indian Ridge), south of Broken Ridge, Ninety East Ridge, and the Wallaby and Zenith Plateaus. Results should however be interpreted with caution and relative to the parameters and conservation features selected for the investigation. Indeed, this analysis focused on selected features (e.g., EBSAs) but could not consider local knowledge about specific areas of importance (that may not have been captured by EBSA), such as SIOFA's existing interim protected areas. Hence, it should be regarded as a starting point for a process, not the end. Part of this process will entail detailed evaluation and consultation of the ecological targets and the working spatial resolution of the analysis, all placed in a consultation context.

7. Recommendations

We recommend the Scientific Committee to:

- Note that the results of this investigation should be interpreted relative to the conservation features and ecological targets used for the analysis.
- Note that "best solutions" are solutions within a larger set of other nearly as good solutions.
- Note that not all available information for the area can be integrated into spatial layers due to large gaps in their spatial coverage.
- Note that decisions made about where to designate areas will need to factor in all possible available information that is not integrated into a spatial layer, and in a wider consultation context.
- Note that these results should serve as a starting to point to discuss what conservation features to include and what not to include.

References

Apropos Information Systems Inc. (n.d.). *QMarxan Toolbox* (v2.0.1). https://aproposinfosystems.com/.

Ardron, J. A., Possingham, H. P., & Klein, C. J. (2010). *Marxan Good Practices Handbook, Version 2*. Pacific Marine Analysis and Research Association.

Ball, I. R., Possingham, H. P., & Watts, M. E. (2009). Marxan and relatives: Software for spatial conservation prioritization. In A. Moilanen, K. A. Wilson, & H. P. Possingham (Eds.), *Spatial conservation prioritization. Quantitative methods & computational tools* (pp. 185–195). Oxford University Press.

Ban, N. C., & Klein, C. J. (2009). Spatial socioeconomic data as a cost in systematic marine conservation planning. *Conservation Letters*, *2*(5), 206–215. https://doi.org/10.1111/J.1755-263X.2009.00071.X

Beyer, H. L., Dujardin, Y., Watts, M. E., & Possingham, H. P. (2016). Solving conservation planning problems with integer linear programming. *Ecological Modelling*, *328*, 14–22. https://doi.org/10.1016/J.ECOLMODEL.2016.02.005

CBD (Convention on Biological Diversity). (2008). *Annex I, COP 9 Decision IX/20. Marine and coastal biodiversity*.

CBD (Convention on Biological Diversity). (2010). *COP 10 decision X/2: Strategic plan for biodiversity 2011–2020. 10th Meeting of the Conference of the Parties to the Convention on Biological Diversity.* .

Costello, M. J., Cheung, A., & de Hauwere, N. (2010). Surface area and the seabed area, volume, depth, slope, and topographic variation for the world's seas, oceans, and countries. *Environmental Science and Technology*, *44*(23), 8821–8828. https://doi.org/10.1021/ES1012752

Game, E. T., Lipsett-Moore, G., Hamilton, R., Peterson, N., Kereseka, J., Atu, W., Watts, M., & Possingham, H. (2011). Informed opportunism for conservation planning in the Solomon Islands. *Conservation Letters*, *4*(1), 38–46. https://doi.org/10.1111/J.1755-263X.2010.00140.X

Gaston, K. J., Pressey, R. L., & Margules, C. R. (2002). Persistence and vulnerability: Retaining biodiversity in the landscape and in protected areas. *Journal of Biosciences*, *27*(4), 361–384. https://doi.org/10.1007/BF02704966

Harris, P. T., Heap, A. D., Whiteway, T., & Post, A. (2008). Application of biophysical information to support Australia's representative marine protected area program. *Ocean & Coastal Management*, *51*(10), 701–711. https://doi.org/10.1016/J.OCECOAMAN.2008.07.007

Harris, P. T., Macmillan-Lawler, M., Rupp, J., & Baker, E. K. (2014). Geomorphology of the oceans. *Marine Geology*, *352*, 4–24. https://doi.org/10.1016/J.MARGEO.2014.01.011

Kim, S. S., & Wessel, P. (2011). New global seamount census from altimetry-derived gravity data. *Geophysical Journal International*, *186*(2), 615–631. https://doi.org/10.1111/J.1365-246X.2011.05076.X/3/186-2-615-FIG015.JPEG

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, *220*(4598), 671–680. https://doi.org/10.1126/SCIENCE.220.4598.671

Kukkala, A. S., & Moilanen, A. (2013). Core concepts of spatial prioritisation in systematic conservation planning. *Biological Reviews*, *88*(2), 443–464. https://doi.org/10.1111/BRV.12008

Levin, N., Watson, J. E. M., Joseph, L. N., Grantham, H. S., Hadar, L., Apel, N., Perevolotsky, A., DeMalach, N., Possingham, H. P., & Kark, S. (2013). A framework for systematic conservation planning and management of Mediterranean landscapes. *Biological Conservation*, *158*, 371–383. https://doi.org/10.1016/J.BIOCON.2012.08.032

Maina, J. M., Gamoyo, M., Adams, V. M., D'agata, S., Bosire, J., Francis, J., & Waruinge, D. (2020). Aligning marine spatial conservation priorities with functional connectivity across maritime jurisdictions. *Conservation Science and Practice*, *2*(2), e156. https://doi.org/10.1111/CSP2.156

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature 2000 405:6783*, *405*(6783), 243–253. https://doi.org/10.1038/35012251

Margules, C. R., Pressey, R. L., & Williams, P. H. (2002). Representing biodiversity: Data and procedures for identifying priority areas for conservation. *Journal of Biosciences*, *27*(4), 309–326. https://doi.org/10.1007/BF02704962

McArthur, M. A., Brooke, B. P., Przeslawski, R., Ryan, D. A., Lucieer, V. L., Nichol, S., McCallum, A. W., Mellin, C., Cresswell, I. D., & Radke, L. C. (2010). On the use of abiotic surrogates to describe marine benthic biodiversity. *Estuarine, Coastal and Shelf Science*, *88*(1), 21–32. https://doi.org/10.1016/J.ECSS.2010.03.003

Moilanen, A., Arponen, A., Stokland, J. N., & Cabeza, M. (2009). Assessing replacement cost of conservation areas: How does habitat loss influence priorities? *Biological Conservation*, *142*(3), 575–585. https://doi.org/10.1016/J.BIOCON.2008.11.011

Pearman, T. R. R., Robert, K., Callaway, A., Hall, R., lo Iacono, C., & Huvenne, V. A. I. (2020). Improving the predictive capability of benthic species distribution models by incorporating oceanographic data – Towards holistic ecological modelling of a submarine canyon. *Progress in Oceanography*, *184*, 102338. https://doi.org/10.1016/J.POCEAN.2020.102338

Possingham, H., Wilson, K. A., Andelman, S. J., & Vynne, C. H. (2006). Protected areas: Goals, limitations, and design. In M. J. Groom, G. K. Meffe, & C. R. Carroll (Eds.), *Principles of Conservation Biology* (3rd ed., pp. 507–549). Sinauer Associates. http://espace.library.uq.edu.au/view/UQ:73195#.VjGIJbGnSgY.mendeley

Pressey, R. L., Humphries, C. J., Margules, C. R., Vane-Wright, R. I., & Williams, P. H. (1993). Beyond opportunism: Key principles for systematic reserve selection. *Trends in Ecology & Evolution*, *8*(4), 124–128. https://doi.org/10.1016/0169-5347(93)90023-I

QGIS Development Team. (2022). *QGIS Geographic Information System* (3.28 Firenze). QGIS Association.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (v4.1.2). R Foundation for Statistical Computing.

Ramiro-Sánchez, B., & Leroy, B. (2022). *SIOFA bioregionalization and VME project*.

Rice, J., Gjerde, K. M., Ardron, J., Arico, S., Cresswell, I., Escobar, E., Grant, S., & Vierros, M. (2011). Policy relevance of biogeographic classification for conservation and management of marine biodiversity beyond national jurisdiction, and the GOODS biogeographic classification. *Ocean & Coastal Management*, *54*(2), 110–122. https://doi.org/10.1016/J.OCECOAMAN.2010.10.010

Stewart, R. R., & Possingham, H. P. (2005). Efficiency, costs and trade-offs in marine reserve system design. *Environmental Modelling and Assessment*, *10*, 203–213.

Watling, L., & Auster, P. J. (2021). Vulnerable Marine Ecosystems, Communities, and Indicator Species: Confusing Concepts for Conservation of Seamounts. *Frontiers in Marine Science*, *8*, 572. https://doi.org/10.3389/FMARS.2021.622586/BIBTEX

Watts, M. E., Stewart, R. R., Martin, T. G., Klein, C. J., Carwardine, J., & Possingham, H. P. (2017). Systematic Conservation Planning with Marxan. In *Learning Landscape Ecology* (pp. 211–227). Springer, New York, NY. https://doi.org/10.1007/978-1-4939-6374-4_13

Woolley, S. N. C., Foster, S. D., Bax, N. J., Currie, J. C., Dunn, D. C., Hansen, C., Hill, N. A., O'Hara, T. D., Ovaskainen, O., Sayre, R., Vanhatalo, J. P., & Dunstan, P. K. (2020). Bioregions in Marine Environments: Combining Biological and Environmental Data for Management and Scientific Understanding. *BioScience*, *70*(1), 48–59. https://doi.org/10.1093/BIOSCI/BIZ133

Yesson, C., Clark, M. R., Taylor, M. L., & Rogers, A. D. (2011). The global distribution of seamounts based on 30 arc seconds bathymetry data. *Deep Sea Research Part I: Oceanographic Research Papers*, *58*(4), 442–453. https://doi.org/10.1016/J.DSR.2011.02.004

Yesson, C., Letessier, T. B., Nimmo-Smith, A., Hosegood, P., Brierley, A. S., Hardouin, M., & Proud, R. (2021). Improved bathymetry leads to >4000 new seamount predictions in the global ocean – but beware of phantom seamounts! *UCL Open Environment*, *4*. https://doi.org/10.14324/111.444/UCLOE.000030

Appendix A -Calibration of Marxan parameters

Standard Marxan calibrations were conducted to identify optimal number of iterations, number of solutions, SPF and BLM values. We describe in the next sections the different steps.

1. Number of iterations

With increasing number of iterations, Marxan succeeds more consistently in locating a global optimum, or at least better local optima. However, solution time increases linearly with the number of iterations, such that there are practical limits on the number of iterations that can be considered reasonable. To calibrate the number of iterations, the SPF was set to 1, the BLM value set to 0, and the number of solutions to 100. The number of iterations was increased from 1 million to 110 million, by 5 million steps from 1 to 10 million, and then by 10 million steps onwards. Trade-off curves comparing the average selected cost, score and boundary (out of 100 solutions) to the number of iterations was generated (Figure 1), resulting from cumulative distribution functions. After 30 million iterations, the cost, score, and area of the solutions differed only slightly, and 80 million has similar values to score and area than iterations higher than 30 million. It was then decided to use 30 million iterations to calibrate the SPF and BLM and use 80 million iterations to produce the final results.

Calibration of the number of iterations was based on the medium target scenario and not repeated for scenarios low and high. This is because the three scenarios (low, medium, and high) contain the same socio-economic feature and number of planning units, which are the parameters most likely to impact the number of iterations and the effects on the solutions cost (Ardron et al., 2010).

2. Number of runs

It is common practice to set Marxan to run 100 times, i.e., to provide 100 solutions, for each scenario. Nevertheless, it is good practice to explore the effect of increasing the number of solutions to assess if the characteristic set of solutions obtained changes or remains the same (Ardron et al., 2010). In certain circumstances, it may be more practical to assess the number of solutions than the number of iterations (given the linear increase in computation time), since unfeasible solutions can always be discarded (Ardron et al., 2010).

The number of runs, or number of solutions, was calibrated by increasing the number from 100 to 500 by 50 runs steps, keeping the SPF at 1, BLM at 0, and the number of iterations at 30 million. Again, this calibration was based on the medium target scenario. Figure 2 shows that the difference in the average cost of the solutions is minimal as we increase the number of solutions and the spatial distribution of the selection frequency remained the same, which indicated that 100 number of solutions were adequate to explore the solution space.

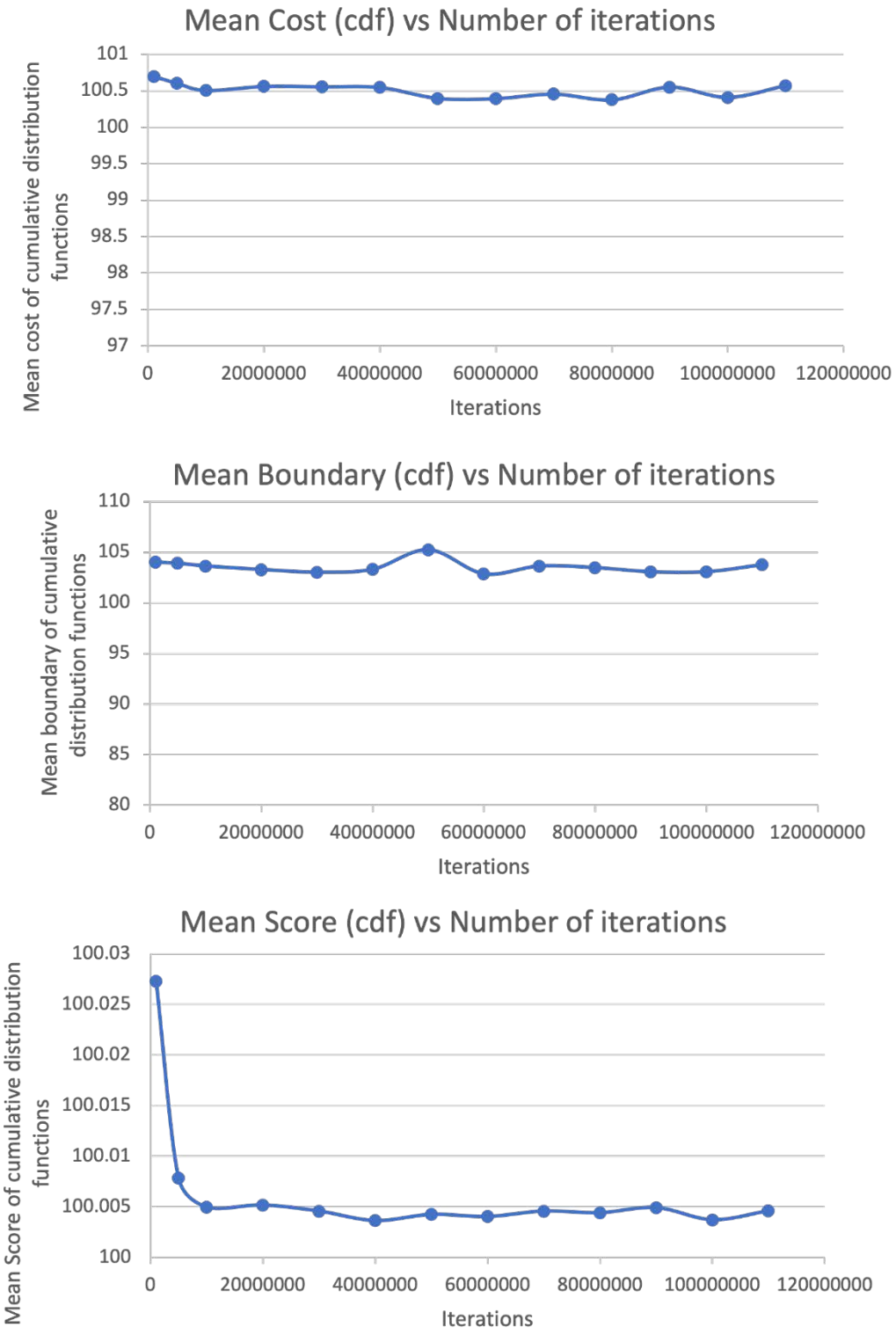

Figure 1. Calibration for the number of iterations for the medium target scenario resulting from averaging results of cumulative distribution functions (cdf) for cost, boundary, and score of solutions. Trade-off curves comparing mean cost (top row), mean boundary (middle row), and mean score (bottom row) against number of iterations.

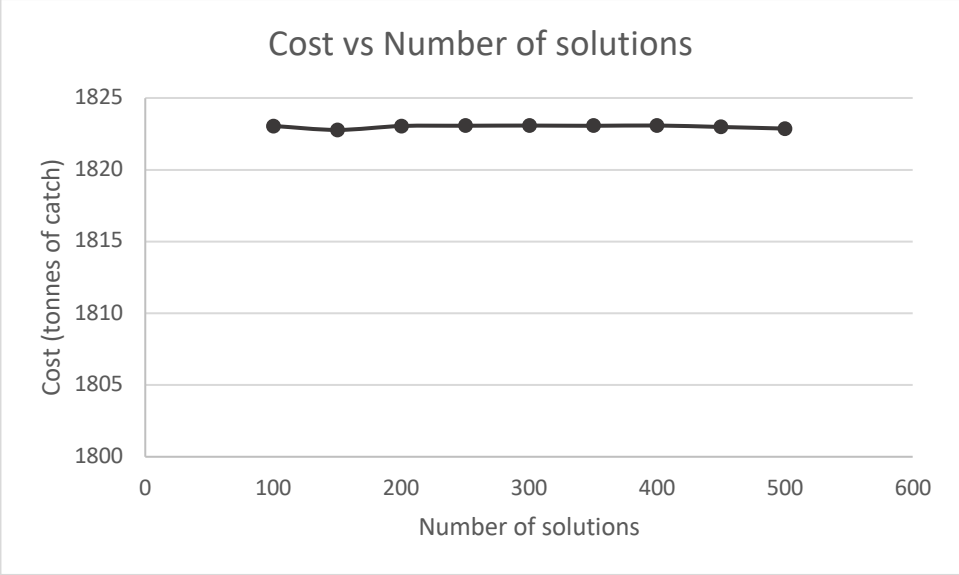

Figure 2. Calibration for the number of runs for the medium target scenario.

3. Boundary Length Modifier

The BLM determines the compactness of the solutions generated by Marxan. A BLM of 0 leads to spatially fragmented solutions consisting of many planning units individually selected. In contrast, a higher BLM leads to solutions consisting of fewer and larger spatial solutions. An exploratory analysis (Table 1) revealed that five out of the 16 features did not meet their targets when a SPF was set equal to 1 for all features. Following Marxan Good Practice Handbook, it was decided to calibrate first the BLM and, if after selecting an appropriate level of compactness, features did not still meet targets, the SPF value would be increased for those specific features alone.

Table 1. Summary statistics of baseline scenarios. Here "baseline" scenarios are those scenarios that did not go through the calibration process for BLM or SPF. In all scenarios the BLM is set to 0, and the SPF to 1.

	BLM	Score	Cost	Nb. Of Pus	Boundary Length	Shortfall	Nb. of missing features	MPM
Low	0	926.80	748.19	748	1356	1.9887E+11	5.79	0.63
Medium	0	1664.18	1259.55	1260	1824	4.3697E+11	5	0.44
High	0	2408.44	1859.94	1860	2372	5.9924E+11	5	0.55

3.1 Calibration of the low target scenario

The following method was used to calibrate the BLM values for low-target scenario:

1) The first calibration explored ten values for BLM ranging from 0 to 1000, using 100 solutions and 30 million iterations, and SPF set to 1 for all conservation features. Specifically, we tested scenarios for BLM values set to 0.0, 0.001, 0.01, 0.1, 1.0, 10, 100, and 1000.
2) A second calibration was set to explore eight more BLM values ranging from 0 to 10 to focus the search for an adequate BLM. We tested for 0, 0.001, 0.004, 0.005, 0.006, 0.01, 0.05, 0.1, 1.0, and 10.
3) Trade-off curves were constructed after each calibration to identify the effects of changing BLM values to the area selected and the boundary length.

3.2 Calibration of the medium target scenario

The following method was used to calibrate the BLM values for the medium-target scenario:

1) The first calibration explored ten values for BLM ranging from 0 to 0.001, using 100 solutions and 30 million iterations, and SPF set to 1 for all conservation features. Specifically, we tested scenarios for BLM values set to 0, 0.001, 0.01, 0.1, 1.0, 10, 100, and 1000.
2) A second calibration was set to explore nine more BLM values ranging from 0 to 10 to focus the search for an adequate BLM. We tested for 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1.0 and 10.
3) Trade-off curves were constructed after each calibration to identify the effects of changing BLM values to the area selected and the boundary length.

3.3 Calibration of the high target scenario

The following method was used to calibrate the BLM values for high-target scenario:

1) The first calibration explored 10 values for BLM ranging from 0 to 1000, using 100 solutions and 30 million iterations, and SPF set to 1 for all conservation features. Specifically, we tested scenarios for BLM values set to 0.0, 0.001, 0.01, 0.1, 1.0, 10, 100, and 1000.
2) A second calibration was set to explore six more BLM values ranging from 0.01 to 1 to focus the search for an adequate BLM. We tested for 0, 0.01, 0.05, 0.1, 0.5 and 1.0.
3) Trade-off curves were constructed after each calibration to identify the effects of changing BLM values to the area selected and the boundary length.

Figure 3, Figure 4, and Figure 5 show the results of the BLM calibration for the low, medium, and high target scenarios, respectively. The three figures show the expected response: an increase of BLM values increases the average area selected in the solutions, whilst decreasing the mean boundary length of the solutions. As mentioned earlier, this creates a more compact but more costly network, that is, solutions consisting of fewer, but larger, clusters of planning

units. Following the common method for selecting BLMs (Steward and Possingham, 2005), we selected the BLM value along the inflection point in the curve as the optimal BLM value for each target scenario. We also demonstrate in Figure 6, Figure 7, and Figure 8 the effect of using different BLM values in the design of the network for each of the target scenarios, respectively.

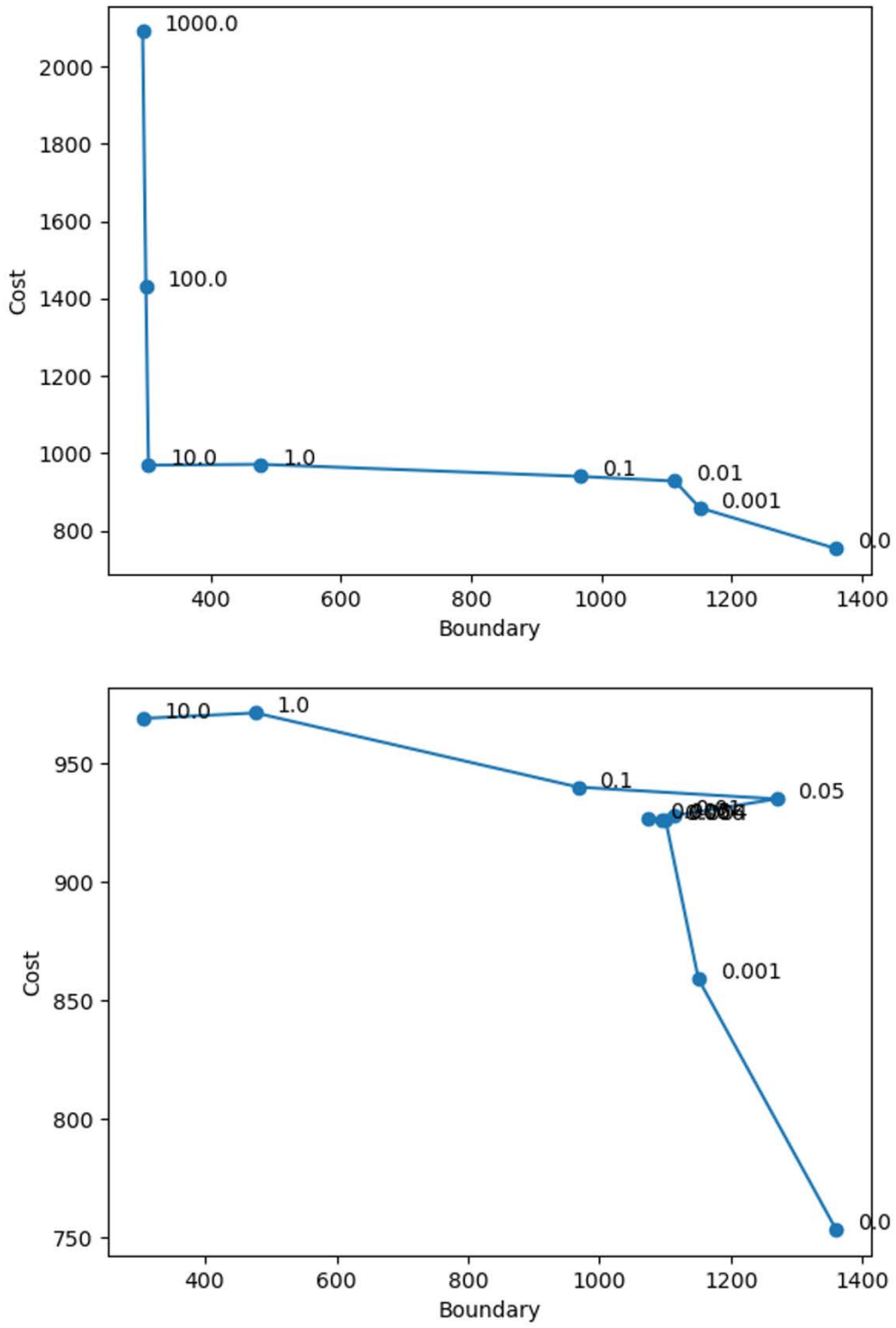

Figure 3. BLM calibration for the low target scenario. Effects of varying BLM values on the average area selected (cost) and average boundary length of the areas selected over 100 solutions. Here we look for the steep areas on the graph where the boundary length is quickly reduced for a small increase in cost. If that value does not produce the desired level of clumping, shift to a higher BLM (knowing that this includes a trade-off in terms of cost).

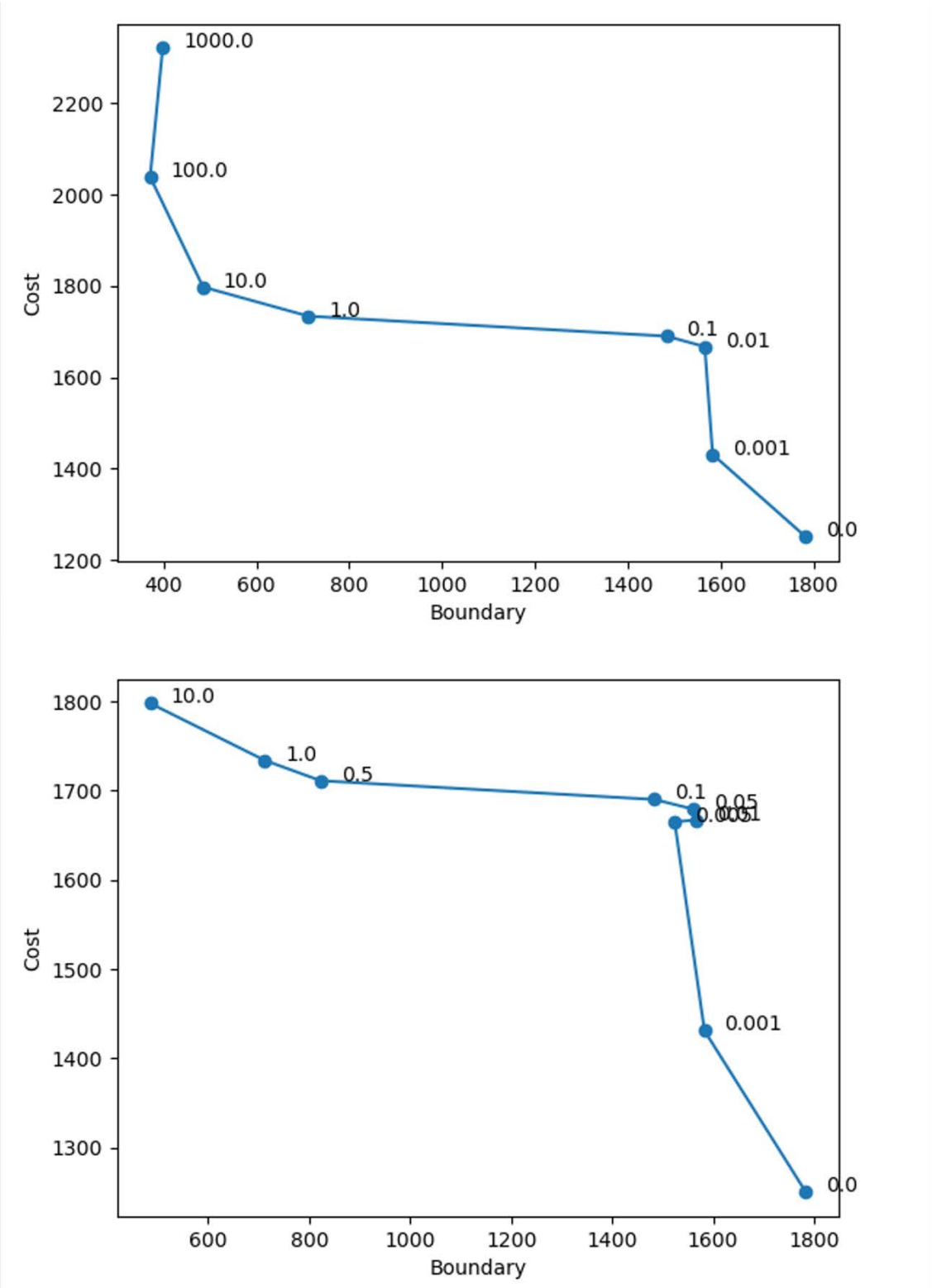

Figure 4. BLM calibration for the medium target scenario. Effects of varying BLM values on the average area selected (cost) and average boundary length of the areas selected over 100 solutions. Here we look for the steep areas on the graph where the boundary length is quickly reduced for a small increase in cost. If that value does not produce the desired level of clumping, shift to a higher BLM (knowing that this includes a trade-off in terms of cost).

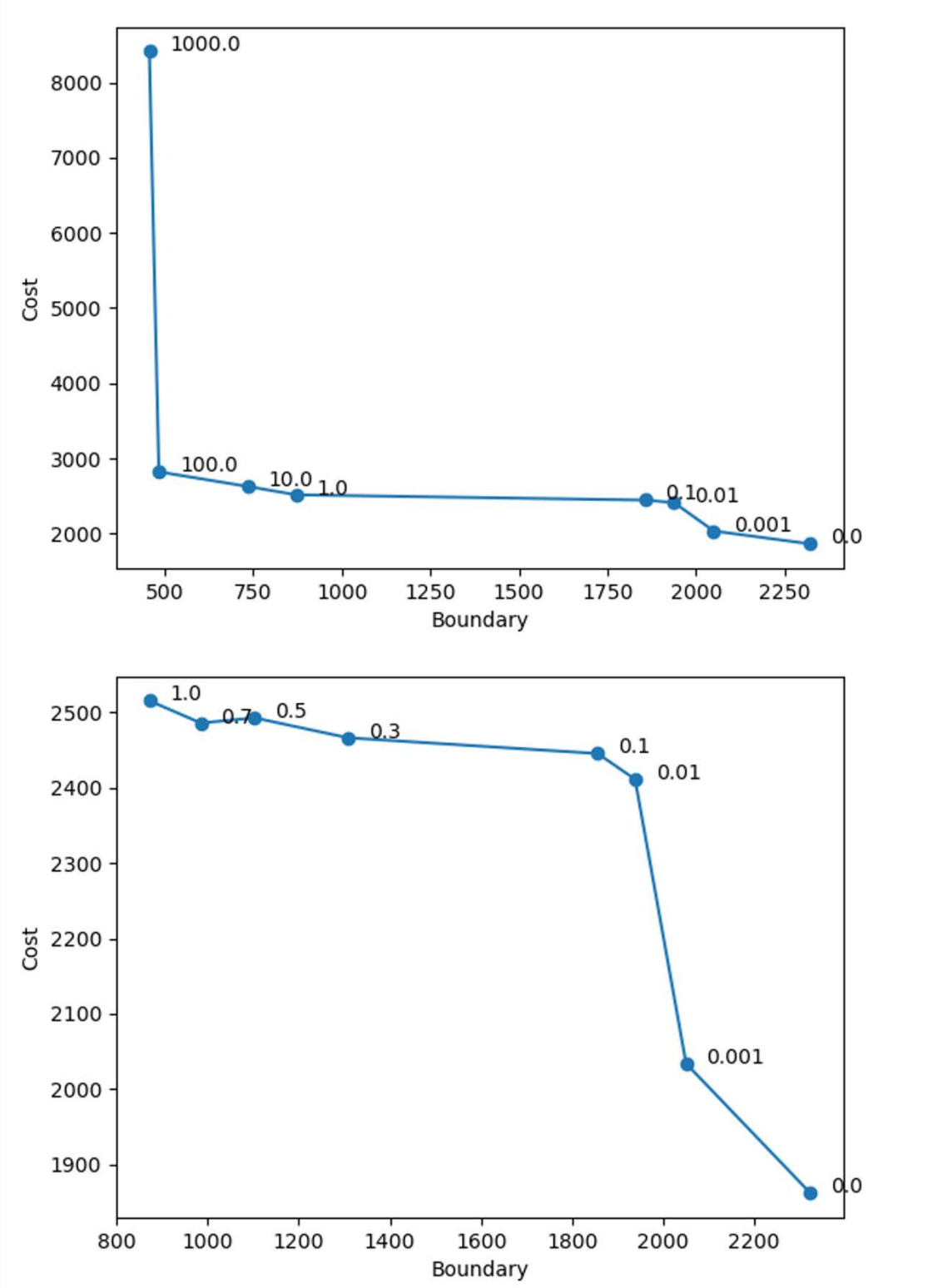

Figure 5. BLM calibration for the high target scenario. Effects of varying BLM values on the average area selected (cost) and average boundary length of the areas selected over 100 solutions. Here we look for the steep areas on the graph where the boundary length is quickly reduced for a small increase in cost. If that value does not produce the desired level of clumping, shift to a higher BLM (knowing that this includes a trade-off in terms of cost).

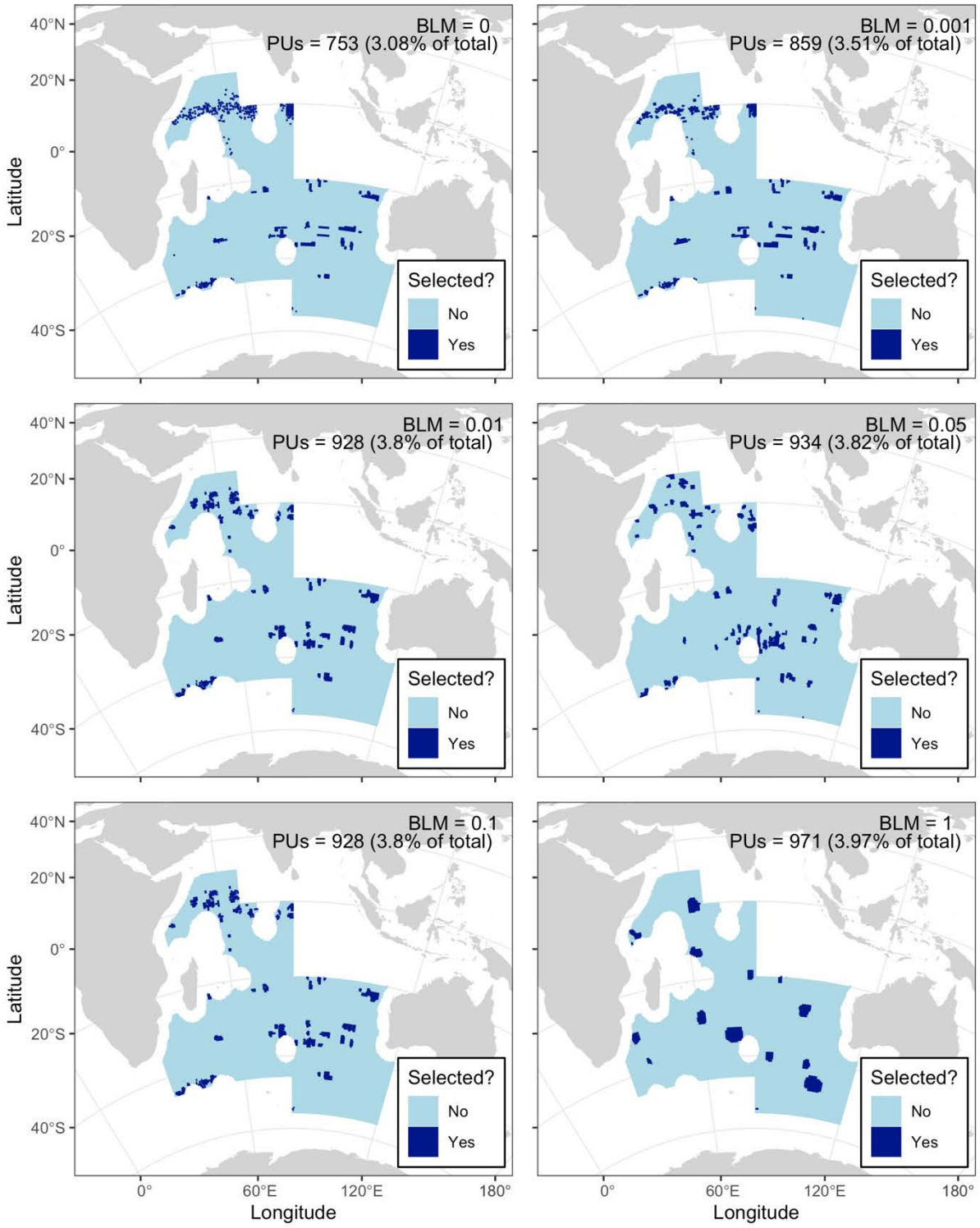

Figure 6. Effect of using different BLM values on the best solution calculated by Marxan for the low target scenario. The total amount of the selected area by the solution is presented.

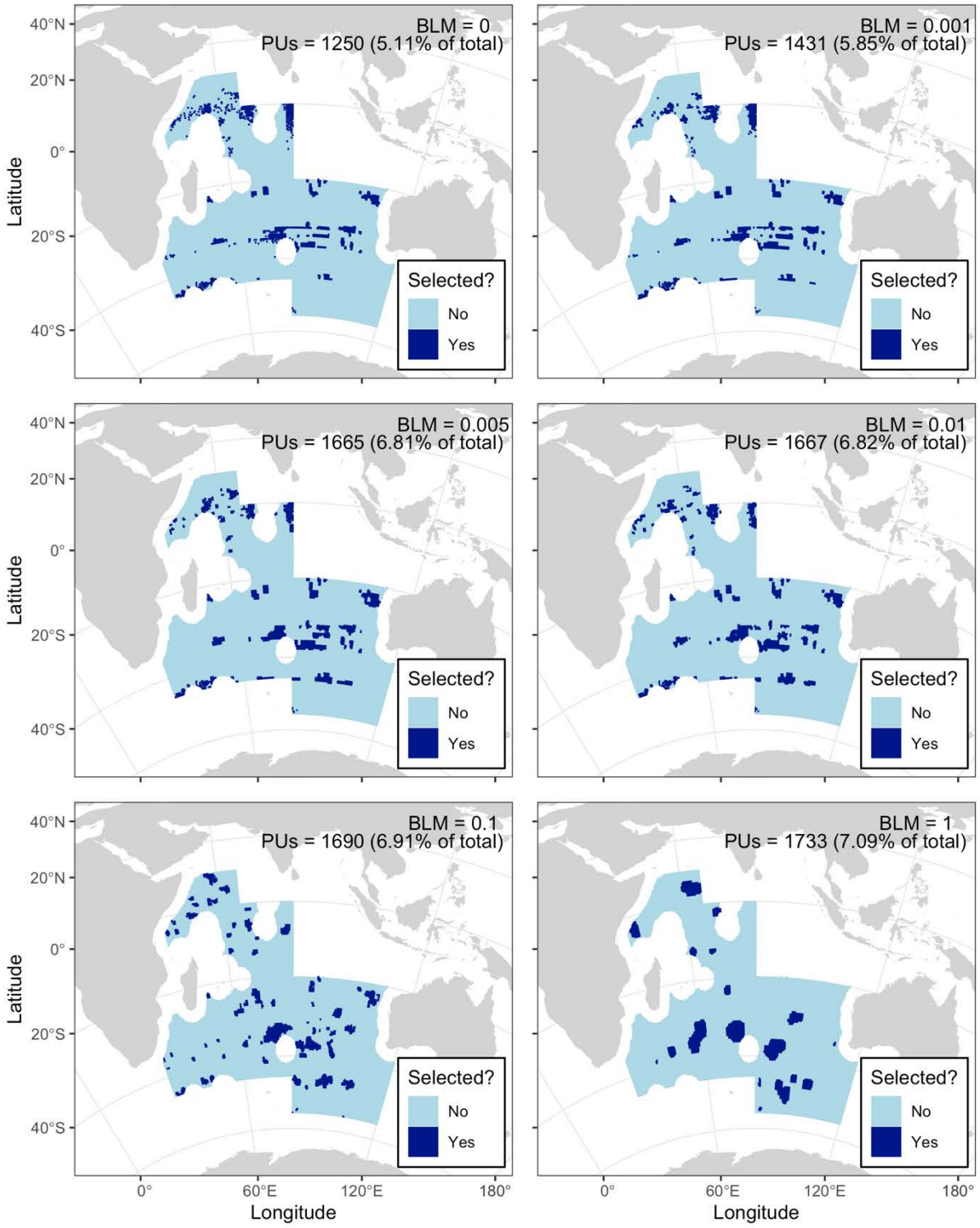

Figure 7. Effect of using different BLM values on the best solution calculated by Marxan for the medium target scenario. The total amount of the selected area by the solution is presented.

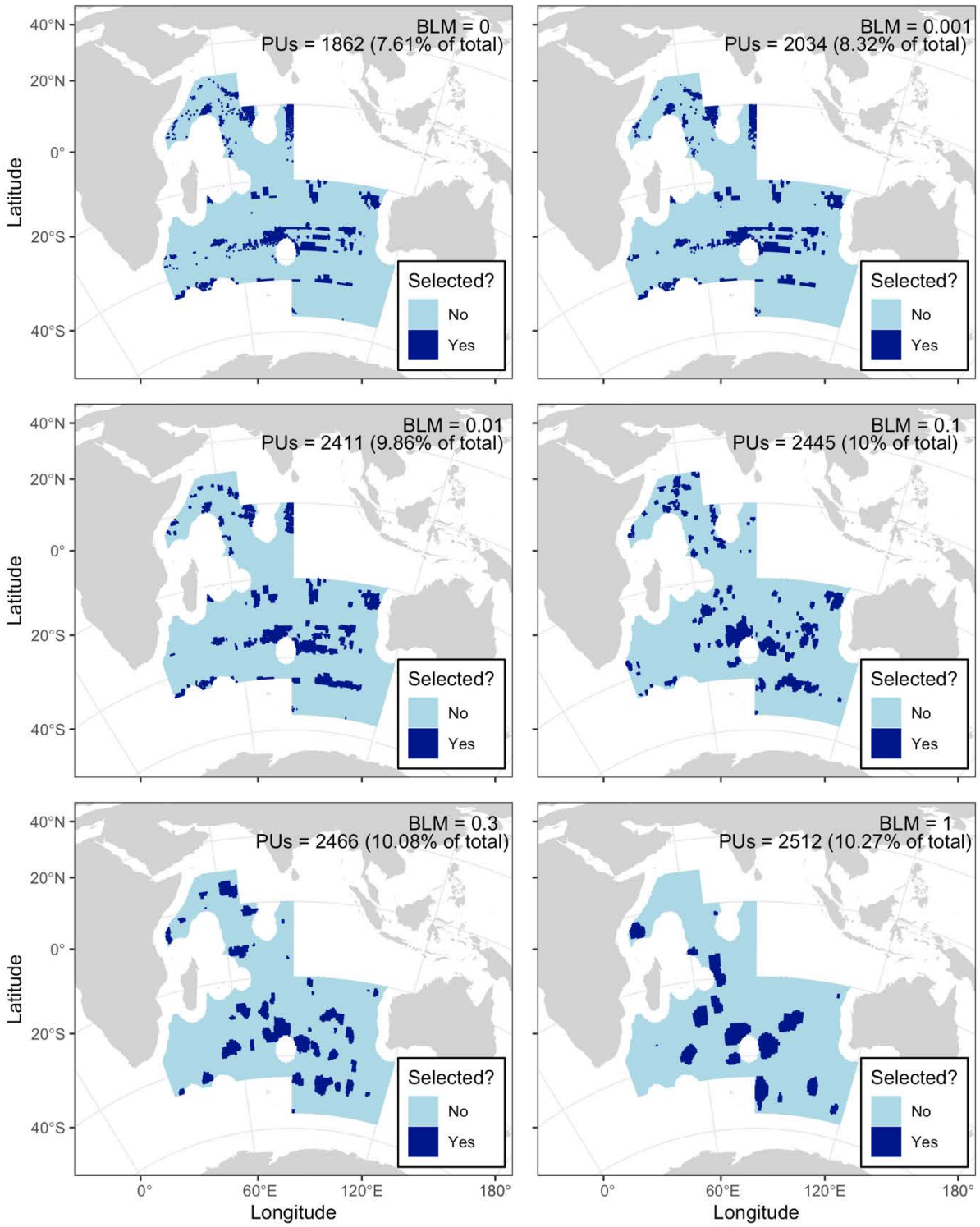

Figure 8. Effect of using different BLM values on the best solution calculated by Marxan for the high target scenario. The total amount of the selected area by the solution is presented.

4. Species Penalty Factor

An exploratory analysis (see Table 2) confirmed that on average one ecological feature at least did not meet its target when a SPF was set equal to 1 for all features at the selected level of compactness (BLM). It was then decided to calibrate the SPF with the application of a higher SPF for those specific features missing their targets.

For the SPF calibration we used the function available at QMarxan Toolbox. We set the SPF values using a step size increase of 0.1 and a percentage target success of 99 %. To adjust SPF values, we used the "As group" method where all features whose targets are not met have their SPF values increased together and other features whose targets are met, have their SPF values unaltered (Qmarxan Toolbox). For comparison, SPF values were also calibrated (with a BLM equal 0) to explore what SPF values are needed to achieve targets.

Table A2. Summary statistics (average) of baseline scenarios for 100 solutions with potential BLM values. Here "baseline" scenarios are those scenarios that did not go through the calibration process for SPF. In all scenarios the SPF was set to 1 and iterations to 30 million.

	BLM	Score	Cost	Nb. Of PUs	Boundary Length	Shortfall	Nb. of missing features	Minimum proportion met
Low	0.001	928.44	859.79	860	1176	16.8353E+10	1.91	0.66
	0.01	940.92	927.65	928	1181	1513074370	0.95	0.93
	0.1	1056.57	941.45	941	1137	1112413420	0.91	0.93
Medium	0.005	1674.92	1661.71	1662	1610	5053344030	0.98	0.93
	0.01	1685.07	1666.95	1667	1642	1705979630	0.99	0.93
	0.1	1860.24	1691.53	1691	1673	1099341783	0.63	0.95
High	0.001	2411.97	2035.13	2035	2089	4.0275E+11	4	0.63
	0.01	2434.47	2412.67	2413	2002	1748016110	0.98	0.94
	0.1	2654.91	2447.03	2447	2066	965123979	0.33	0.98

Differences in Marxan scores between the use of a consistent SPF of 1 for all features, and the application of a higher SPF for the features missing their targets, is described in Table 3. Analyses confirmed that more efficient solutions were obtained when using a customized SPF.

Table 3. Summary statistics of scenarios with a calibrated SPF for those features that did not meet targets.

	BLM	Score	Cost	Nb. Of Pus	Boundary Length	Shortfall	Nb. of missing features	Min. Proportion met
Low	0.01	941.06	927.99	928	1187	11338198910	0.01	0.99
Medium	0.005	1674.98	1662.45	1662	1602	4509657500	0	0.99
	0.01	1685.23	1667.57	1668	1645	1332639560	0	0.99
	0.1	1859.78	1691.3	1691	1674	895924990	0	0.99
	0.001	2435.96	2430.13	2430	4379	1685654270	0.01	0.99
High	0.01	2434.54	2413.36	2413	2002	1279665040	0	0.99

Table 4. SPF values applied to each ecological feature for each of the three scenarios after selecting an adequate BLM (i.e., level of compactness) and still not meeting targets.

Feature name	Low (BLM=0.01)	Medium (BLM=0.01)	High (BLM=0.01)
EBSA	1	1	1
Seamnt	1	1	1
Shelf	1	1	1
Slope	1	1	1
Canyon	1	1	1
Plateau	1	1	1
Rise	1	1	1
Ridge	1	1	1
bio_7 (bioregion 2.4)	1.4	1.3	1.1
bio_6 (bioregion 2.1)	1	1	1
bio_3 (bioregion 1.3)	1	1	1
bio_4 (bioregion 1.5)	1	1	1
bio_5 (bioregion 1.7)	1	1	1
bio_2 (bioregion 1.2)	1	1	1
bio_8 (bioregion 3.1)	1	1	1
bio_1 (bioregion 1.1)	1	1	1

Appendix B– Calibration of Marxan parameters for scenarios with no cost

Using the same calibration parameters as for the scenarios with cost (iterations = 30 million, and 100 runs), the SPF was calibrated for the low, medium, and high no cost scenarios. For this, we set the BLM to 0.

1. Species Penalty Factor

An exploratory analysis (see Table 1) confirmed that on average one ecological feature at least did not meet its target when a SPF was set equal to 1 for all features and BLM was set to 0. It was then decided to calibrate the SPF with the application of a higher SPF for those specific features missing their targets.

For the SPF calibration we used the function available at QMarxan Toolbox. We set the SPF values using a step size increase of 0.1 and a percentage target success of 99 %. To adjust SPF values, we used the "As group" method where all features whose targets are not met have their SPF values increased together and other features whose targets are met, have their SPF values unaltered (QMarxan Toolbox) (Table 2).

For comparison, SPF values were also calibrated after calibrating the BLM (i.e., level of compactness). The first calibration explored 10 values for BLM ranging from 0 to 1000, using 100 solutions and 30 million iterations, and SPF set to 1 for all conservation features. Specifically, we tested scenarios for BLM values set to 0.0, 0.001, 0.01, 0.1, 1.0, 10, 100, and 1000. The graphs suggested a BLM of 0.01 for all three, no cost scenarios (Figure 1, Figure 2, Figure 3). We then proceeded to calibrate the SPF values with that BLM (Table 3).

Table 1. Summary statistics of baseline no cost scenarios. Here "baseline" scenarios are those scenarios that did not go through the calibration process SPF. In all scenarios the BLM is set to 0, and the SPF to 1.

	Score	Cost	Nb. Of Pus	Shortfall	Nb. of missing features	Minimum Proportion met
Low, no cost	926.03	738.72	739	2.0801E+11	5.83	0.60
Medium, no cost	1662.53	1275.69	1276	4.2308E+11	4.86	0.54
High, no cost	2405.16	1822.38	1822	6.3823E+11	4.56	0.53

Table 2. SPF values applied to each ecological feature for each of the three no cost scenarios and a BLM (i.e., level of compactness) set to 0 to meet targets.

Feature name	Low (BLM=0)	Medium (BLM=0)	High (BLM=0)
EBSA	1	1	1
Seamount	1	1	1
Shelf	1	1	1
Slope	1	1	1
Canyon	1	1	1
Plateau	1	1	1
Rise	1	1	1
Ridge	1	1	1
bio_7 (bioregion 2.4)	1.6	1.3	1.2
bio_6 (bioregion 2.1)	1.1	1.1	1.1
bio_3 (bioregion 1.3)	1.1	1.1	1.1
bio_4 (bioregion 1.5)	1.1	1.1	1.1
bio_5 (bioregion 1.7)	1.1	1	1.1
bio_2 (bioregion 1.2)	1.1	1.1	1.1
bio_8 (bioregion 3.1)	1	1	1
bio_1 (bioregion 1.1)	1	1	1

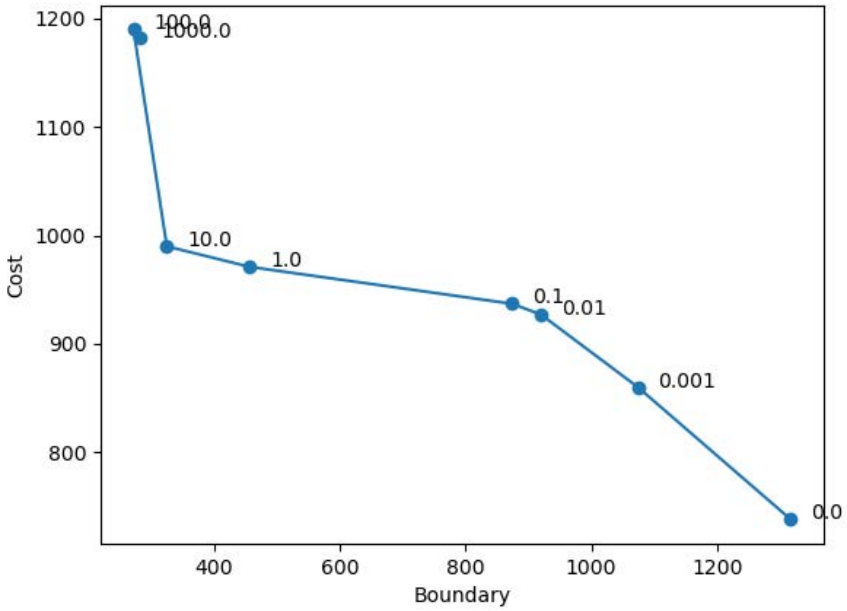

Figure 1. BLM calibration for the low target no cost scenario. Effects of varying BLM values on the average area selected (cost) and average boundary length of the areas selected over 100 solutions. Here we look for the steep areas on the graph where the boundary length is quickly reduced for a small increase in cost.

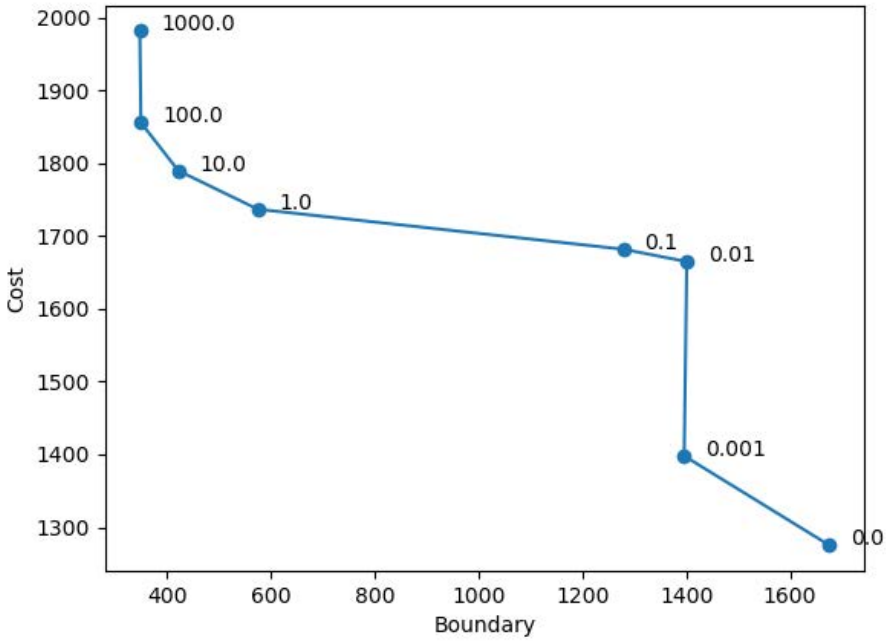

Figure 2. BLM calibration for the medium target no cost scenario. Effects of varying BLM values on the average area selected (cost) and average boundary length of the areas selected over 100 solutions. Here we look for the steep areas on the graph where the boundary length is quickly reduced for a small increase in cost.

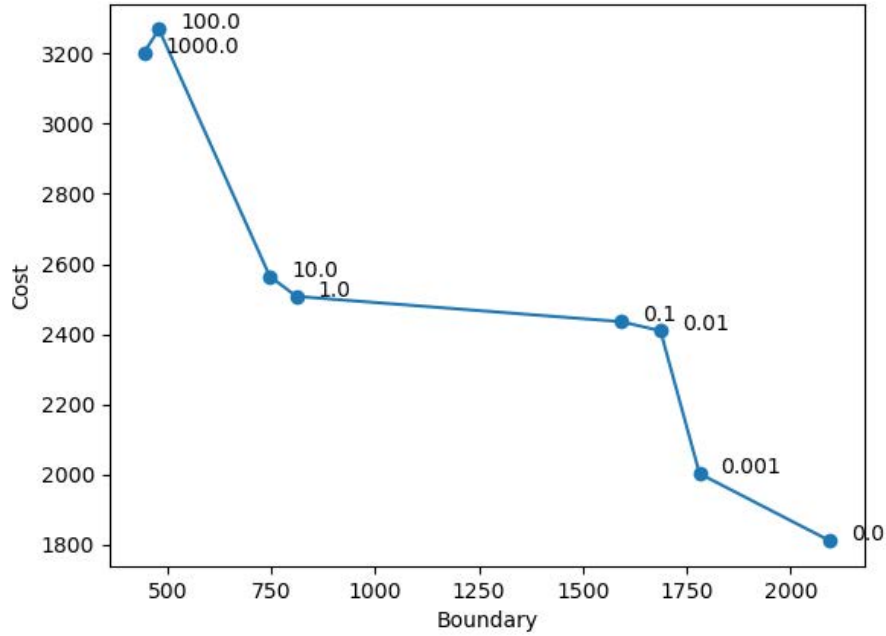

Figure 3. BLM calibration for the high target no cost scenario. Effects of varying BLM values on the average area selected (cost) and average boundary length of the areas selected over 100 solutions. Here we look for the steep areas on the graph where the boundary length is quickly reduced for a small increase in cost.

Table 3. SPF values applied to each ecological feature for each of the three no cost scenarios after selecting an adequate BLM (i.e., level of compactness) and still not meeting targets.

Feature name	Low (BLM=0.01)	Medium (BLM=0.01)	High (BLM=0.01)
EBSA	1	1	1
Seamount	1	1	1
Shelf	1	1	1
Slope	1	1	1
Canyon	1	1	1
Plateau	1	1	1
Rise	1	1	1
Ridge	1	1	1
bio_7 (bioregion 2.4)	1.7	1.4	1.1
bio_6 (bioregion 2.1)	1	1	1
bio_3 (bioregion 1.3)	1	1	1
bio_4 (bioregion 1.5)	1	1	1
bio_5 (bioregion 1.7)	1	1	1
bio_2 (bioregion 1.2)	1	1	1
bio_8 (bioregion 3.1)	1	1	1
bio_1 (bioregion 1.1)	1	1	1

BIODIVERSITY MODELS BASED ON VME INDICATOR TAXA IN THE SIOFA AREA

BERTA RAMIRO SÁNCHEZ, BORIS LEROY

Project PAE2021-01

APRIL 2023

MUSEUM NATIONAL D'HISTOIRE NATURELLE
43 Rue Cuvier, 75005 Paris, France

Funded by
the European Union

Suggested citation:

Ramiro-Sánchez, B. and Leroy, B. (2023). *Biodiversity models based on VME indicator taxa in the SIOFA area (Project PAE2021-01)*. Muséum d'Histoire Naturelle, France.

Table of Contents

Executive summary

Species richness is a widely used measure for biodiversity. Ideally, we would have representative sampling across taxonomic groups to calculate species richness; however, this information often has a limited spatial and temporal coverage. To overcome this limitation, biodiversity models, i.e., models that aim to predict species diversity (richness) and composition can be used.

There are two approaches to model species richness: stacking of individual species predictions and direct modelling of species numbers. The first approach, referred to as stacked species distribution modelling (S-SDMs) consists of first predicting the distribution of each species independently using species distribution models and then stacking them to provide species richness and assemblage composition. The second approach, referred to as macroecological modelling (MEM), statistically relates the number of species in a geographic unit to environmental variables that characterize the same unit. S-SDMs rely in our ability to model the distributions of individual species, wherein species with too little data (i.e., number of records) are usually excluded from further analysis for statistical reasons. MEMs, on the other hand, make use of all available data across all species, regardless of the number of records per species, which may help in modelling rare species.

In this study, we investigated the use of both approaches for predicting richness of VME indicators in the Southern Indian Ocean, and their use in highlighting potential biodiversity hotspots. For S-SDMs, we stacked individual habitat suitability predictions of VME indicators produced in previous work for SIOFA (project PAE2020-02) and applied three increasing thresholds to highlight the most diverse areas (i.e., areas with the top 2.5%, top 5% and top 10% taxa richness). For MEM, we attempted to model the observed richness of the area, accounting for sampling bias.

The S-SDMs highlighted Walters Shoal and the Mascarene Plateau as the areas with the highest index of richness. The application of the hotspot threshold highlighted the same areas, with a progressive emergence of the ocean ridges with a less conservative definition. In an extreme case of missing data and sampling bias, the MEM model failed to predict the observed richness, which prevents its current use.

Whilst the S-SDMs performed well in predicting richness in the area, it is based only on taxa that could be modelled (i.e., taxa that had a sufficient number of occurrences), thereby remaining incomplete in coverage because it does not include rare taxa. However, the index could be further used in the application of the definition of hotspot to highlight areas of greater biodiversity in SIOFA. Such information can be further utilised for conservation planning, for example.

1. Purpose of the report

This report aims to address Terms of Reference 3 "*Investigate and advise on the use of habitat suitability modelling in predicting benthic species diversity and distribution in the SIOFA management area, including assessing data availability for such modelling*" of project PAE2021-01 ("Bioregionalization and Management of Vulnerable Marine Ecosystems (VMEs)"*).*

2. Introduction

International obligations require the implementation of conservation measures for the management of deep-sea bottom fisheries, due to the potential impacts they may have on seafloor biodiversity, especially on vulnerable marine ecosystems (VMEs) (UNGA, 2007). As a basis to inform conservation measures for the protection of VMEs, prioritisation of areas can help in meeting conservation targets set to protect biodiversity (Margules & Pressey, 2000). One way to highlight important areas is to use species richness, a simple but an important widely used measure for biodiversity, which has also been identified as one of the Essential Biodiversity Variables (EBVs; Pereira et al., 2013). Therefore, investigating the distributional patterns of richness of VME indicators in the SIOFA area can also help in the prioritisation of sites for conservation.

Species richness is the number of taxa occurring in a defined geographic area. Ideally, we would have representative sampling across taxonomic groups to calculate species richness; however, this information often has a limited spatial and temporal coverage (Meyer et al., 2015). To overcome this limitation, biodiversity models, i.e., models that aim to predict species diversity (richness) and composition can be used, which can help to establish conservation strategies or to predict future patterns of biodiversity under global change (Ferrier & Guisan, 2006; Guisan & Rahbek, 2011).

There are two approaches to model species richness: stacking of individual species predictions and direct modelling of species numbers (Ferrier & Guisan, 2006). The first approach, referred to as stacked species distribution modelling (S-SDMs, Guisan & Rahbek, 2011), consists of first predicting the distribution of each species independently using species distribution models (SDM; Guisan & Thuiller, 2005) and then stacking them to predict species assemblages, providing species richness and assemblage composition. SDMs are correlative models that relate species occurrences to a set of environmental variables (Guisan & Zimmermann, 2000). In S-SDMs, assemblages will theoretically be composed of species with partially the same environmental requirements (Guisan & Rahbek, 2011). Although S-SDMs have been criticised for overpredicting richness in the past (Guisan & Rahbek, 2011), it has been shown to be related to thresholding individual SDMs to binary outputs prior to stacking (Calabrese et al., 2014). Instead, the preferred approach of stacking requires only the summation of per-site suitabilities from individual SDMs, which yields similar performance to macroecological models (Calabrese et al., 2014), especially when predicting current climate scenarios (Biber et al., 2020).

The second approach, referred to as macroecological modelling (MEM), statistically relates the number of species in a geographic unit to environmental variables that characterize the same unit (Gotelli et al., 2009). In the MEM, the number of species found in a geographic unit is expected to depend on available energy, environmental heterogeneity, disturbance, or history, with scale effects and some level of stochasticity (Gotelli et al., 2009).

Currently, the two approaches can be used separately to make the same prediction, although there are methodological and theoretical differences between the two approaches. S-SDMs rely in our ability to model the distributions of individual species. In this case, species with too little data (i.e., number of records) are usually excluded from further analysis for statistical reasons. With S-SDM we have information on the identity of species occurring at a given location. MEMs, on the other hand, make use of all available data across all species, regardless of the number of records per species, which may help in modelling rare species. However, MEMs do not provide information on the identity of species occurring at a given site. Despite these differences, both approaches can use a similar set of environmental predictors but with different underlying theoretical hypotheses. For instance, in S-SDMs, species ranges are constrained by environmental factors, while in MEM environmental factors limit the number of species that coexist in an area (Guisan & Rahbek, 2011). Thus, comparing the two approaches provides insights into our ability to predict patterns by analysing convergences and divergences between them.

In this study, we explored the use of both approaches for predicting diversity of VME indicators in the Southern Indian Ocean. The present study builds on previous work developing predictive bioregionalization schemes based on VME indicator taxa in the SIOFA area (Ramiro-Sánchez & Leroy, 2022). The three schemes addressed three methodological approaches: “group first, then predict”, “predict first, then group”, and “analyse simultaneously” (Woolley et al., 2020). For the exploration of S-SDMs, we made use of habitat suitability predictions developed during the “predict first, then group” approach, and explore the application of hotspot definition to identify areas of greater biodiversity within SIOFA. For the exploration of the MEM, we used the same biodiversity dataset collated for the bioregionalization work. We present the results and discuss the limitations from both approaches.

3. Methods

In the next section we introduce the biodiversity inventory serving as the basis for both approaches, an overview of the S-SDM, including an overview of the methods followed for the development of habitat suitability models, and the methods for the MEM development.

Biological inventory

The VME indicators list adopted by SIOFA (CMM 2018/01) at order, class, and phylum levels, includes the following categories of deep-sea (generally >200 m) benthic taxa: Cnidaria (Gorgonacea, Anthoathecatae, Stylasteridae, Scleractinia, Antipatharia, Zoantharia, Actiniaria, Alcyonacea, Pennatulacea), Porifera (Hexctanilleda, Demospongiae), Ascidiacea,

Bryozoa, Brachiopoda, Pterobranchia, Serpulidae, Xenophyophora, Bathylasmatidae, Crinoidea (stalked species only), Euryalida, Cidaroida. We downloaded all occurrence records under these categories from the public databases the Ocean Biodiversity Information System (OBIS, https://obis.org/) (accessed on 10/11/2020 and 02/04/2021), the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/) (accessed on 09/11/2020 and 14/04/2021; the full list of GBIF occurrence downloads is provided in Appendix A), NOAA's Deep-sea Corals Data Portal (https://deepseacoraldata.noaa.gov/) (accessed on 16/11/2020 and 09/04/2021), and Smithsonian Natural History Museum (https://collections.nmnh.si.edu/) (accessed on 16/11/2020). We also obtained occurrence records from SIOFA's observer programme, and research campaigns led by the Muséum National d'Histoire Naturelle. We obtained records for the whole Southern Indian Ocean to account for ecological continuity, that is, SIOFA's management area and all exclusive economic zones bordering it (latitudes 13°N – 65°S and longitudes 20°E – 147°E).

We applied verification procedures for taxonomic consistency, error detection, as well as evaluation of records in the environmental space. Specifically, we first checked species names against the most updated authority, the World Register of Marine Species (WoRMS, 2021) for synonyms and fossil records. Secondly, we applied automatic error and outlier detection using the function clean-coordinates from the R package CoordinateCleaner version 2.0-18 (Zizka et al., 2019). We tested for equal coordinates, coordinates over land using the Natural Earth data ocean shapefile version 4.1.0 (www.naturalearthdata.com, accessed November 2020), and zero coordinates. Finally, we used the catalogue number and geographical coordinates to filter out potential duplicates across GBIF and OBIS.

Our download strategy incorporated numerous shallow water species in the dataset, particularly zooxanthellate corals (i.e., corals with photosynthetic algae) – however, only deep-sea species fall under the definition of VME indicator taxa. Although very few zooxanthellate corals occur below 50 m (Cairns, 2007), using the typical definition of deep sea as waters below 200 m would exclude deep-water species that expand into shallower depths. Consequently, we aimed to integrate the ecology of taxa and the general definition of deep sea by applying several filters to exclude zooxanthellate corals and other strictly shallow-water taxa. First, we individually assessed the depth distribution of each species, and we kept a species if 90% of its records were below 200 m water depth. To assess the depth range of records, we relied on their original recorded depth as indicated in the sample record. In the cases where this information was not available, we used the General Bathymetric Chart of the Oceans (GEBCO, 2021) to assign depths. Second, we further filtered species known to be shallow water as we worked through with peer-reviewed deep-sea taxa lists (Cairns, 2017; Kocsis et al., 2018). Our filtered dataset of deep-sea VME indicator taxa comprised 1,991 species (of which 1,312 had an observation date). The list can be found in Ramiro-Sánchez & Leroy (2022).

3.1 Stacked-Taxa Distribution Models

Habitat suitability predictions of species form the basis of Stacked-Species Distribution Models (S-SDMs). During SIOFA's project PAE2020-02, habitat suitability models were built for VME indicators at family, genus, and species level (Ramiro-Sánchez & Leroy, 2022). These

models formed the basis for a subsequent bioregionalization using network analysis, following the approach known as "predict first, then group" (Ferrier & Guisan, 2006). For consistency we provide an overview of the habitat suitability modelling step, although for a full description of the methods we refer the reader to Ramiro-Sánchez & Leroy (2022).

All taxa (species, genus, family) with more than five records were modelled with down-sampled presence-only random forests, using a set of 50,000 randomly sampled background points. We chose down-sampled random forests because these techniques are one of the best performing techniques for presence-only distribution models, especially in situations with low occurrence numbers (i.e., between 5 and 30 – see Valavi, Elith, et al., 2021; Valavi, Guillera-Arroita, et al., 2021). In addition, random forests are relatively fast single models, which is an important aspect here given the large number of taxa to model. Because most VME indicator taxa had only a limited number of occurrences, we adapted our modelling process depending on the number of occurrences, with a decreasingly complex procedure (Ramiro-Sánchez & Leroy, 2022).

As discussed in Ramiro-Sánchez & Leroy (2022), the resulting predicted distributions were likely overpredicted given that the environmental niche was not well characterised. In particular, we observed that our modelling at species and genus level was less precise than at family level, because there were less data for species and genus than for family; thus, large areas were predicted at the species and genus level that were not predicted at family level. As a result, we estimated that the genus-level was the best compromise between a fine taxonomic resolution and a reduced (albeit still very present) overprediction of ranges due to the lack of data. Based on this rationale, for the present study we chose to work at the genus and family taxonomic level, therefore collating the individual habitat suitability models at genus (429 genera, list name is provided in Table B1 of Appendix B) and family (210 families, list name is provided in Table B2 of Appendix B) levels.

In order to proceed with the stacking, for each taxonomic level we calculated an index of richness with the cumulative sum of suitability scores per grid cell, which we refer hereafter as stacked-taxa distribution models (S-TDMs). Finally, to identify hotspots, we first spatially restricted the index of richness to only encompass the SIOFA area and then used threshold quantiles to represent the highest proportion of biodiversity in the area. Specifically, we calculated the top 10%, 5% and 2.5% quantiles for the index of richness.

3.2 Macroecological Models

Biological data

Occurrence data comprised 755 genera for inclusion in the macro-ecological model.

Environmental data

We selected environmental variables from the World Ocean Atlas Data (WOA 2018) available at 1° spatial resolution: apparent oxygen utilization, percentage of oxygen, density, nitrate, phosphate, silicate, salinity, temperature. We included mean net primary production (g C $m^{-2}year^{-1}$, NPP) generated from a vertically generalized production model (VGPM) across the years 2003 to 2010 (http://www.science.oregonstate.edu/ocean.productivity/). NPP is a

function of satellite-derived chlorophyll (SeaWiFS and Modis). We also included the latest compilation for sediment thickness data (GlobSed , Straume et al., 2019) as a further determining variable, as sedimentation is a crucial indicator for ecosystem types and biodiversity (e.g., Snelgrove, 1999; Zeppilli et al., 2016). Finally, we used the global bathymetric model GEBCO2020 originally available at a spatial resolution of 0.004° and aggregated it to 1° resolution using the maximum value from surrounding cells.

We checked for multicollinearity in the environmental variables using Spearman rank correlation. Strongly correlated variables ($|\rho|>0.7$) were removed from analyses to avoid issues with co-linearity of model coefficients (Figure A1) (Dormann et al., 2013). For tied variables, we selected the variable the most biologically meaningful for VME indicator taxa. Our final set of variables was composed of seven variables: mean salinity, dissolved oxygen, sediment thickness at seafloor, minimum depth, and mean net primary production at surface.

We aimed to account for the source of variation in the data collection. We estimated sampling bias for each grid cell using a target-group approach, as it has been shown to perform the best at emulating sampling effort (Barber et al., 2022). Specifically, we used all records of VMEs indicators in our dataset (at any taxonomic level) and used a 2D kernel density estimation to convert single points into a continuous probability surface (see Barber et al. (2022) for methods and code).

Statistical analysis

Following preliminary analyses, we modelled diversity of genera by using three algorithms: generalised linear mixed model, generalised additive mixed model, and random forest. We fitted the generalised linear mixed model with a negative binomial distribution and log link function, with the MASS R package (Venables & Ripley, 2002). We fitted linear, and second-order polynomial functions for each predictor variable given that a number of studies have emphasized the importance of unimodal relationships with temperature (Witman et al., 2004) and POC flux (Tittensor et al., 2010). The generalised additive mixed model was fit with a negative binomial distribution, random effects (individual grid cells), and a maximum of 4 knots per predictor variable, with the mgcv R package (Wood, 2011). Finally, the random forest model was fit with 1000 trees, with the randomforest R package (Liaw & Wiener, 2002). The three models included the estimated sampling effort for each grid cell as an offset.

We calibrated the models with the non-collinear environmental variables and evaluated them with a 4-fold cross-validation procedure, computing the residual mean square error (RMSE) between the predicted and observed richness. Models with a small RMSE were selected for inclusion in an ensemble model. For the ensemble model, we selected the median of the models. We rescaled the ensemble prediction from 0 to 100 to obtain a diversity index, in a further attempt to correct for the notorious undersampling in the area and the low trust in absolute numbers.

4. Results

4.1 Stacked-Species Distribution Models

At genus level, the index based on cumulative probabilities of occurrence suggested high richness across the coasts and shallower depths of the Southern Indian Ocean (Figure 1). Specifically, within SIOFA, the areas predicted with highest index of richness were Saya de Malha, the bathyal depth zone of the Southwestern Indian Ridge and Broken Ridge, Mozambique Ridge and Walter's Shoal. The quantiles showing the highest proportion of biodiversity at increasing levels of conservation, showed the same patterns but within a more constrained area and with a progressive emergence of the ocean ridges with less conservative thresholds (Figure 2). The top 10% starting at an index of richness of 52, the top 5% at 69 and the top 2.5% at 89 (Table A1).

At family level, the index of richness (Figure 3) and hotspots of biodiversity (Figure 4) showed similar trends to those at genus level . The starting index values of richness were 28 for the top10%, 38 for the top 5%, and 49 for the top 2.5% (Table A1).

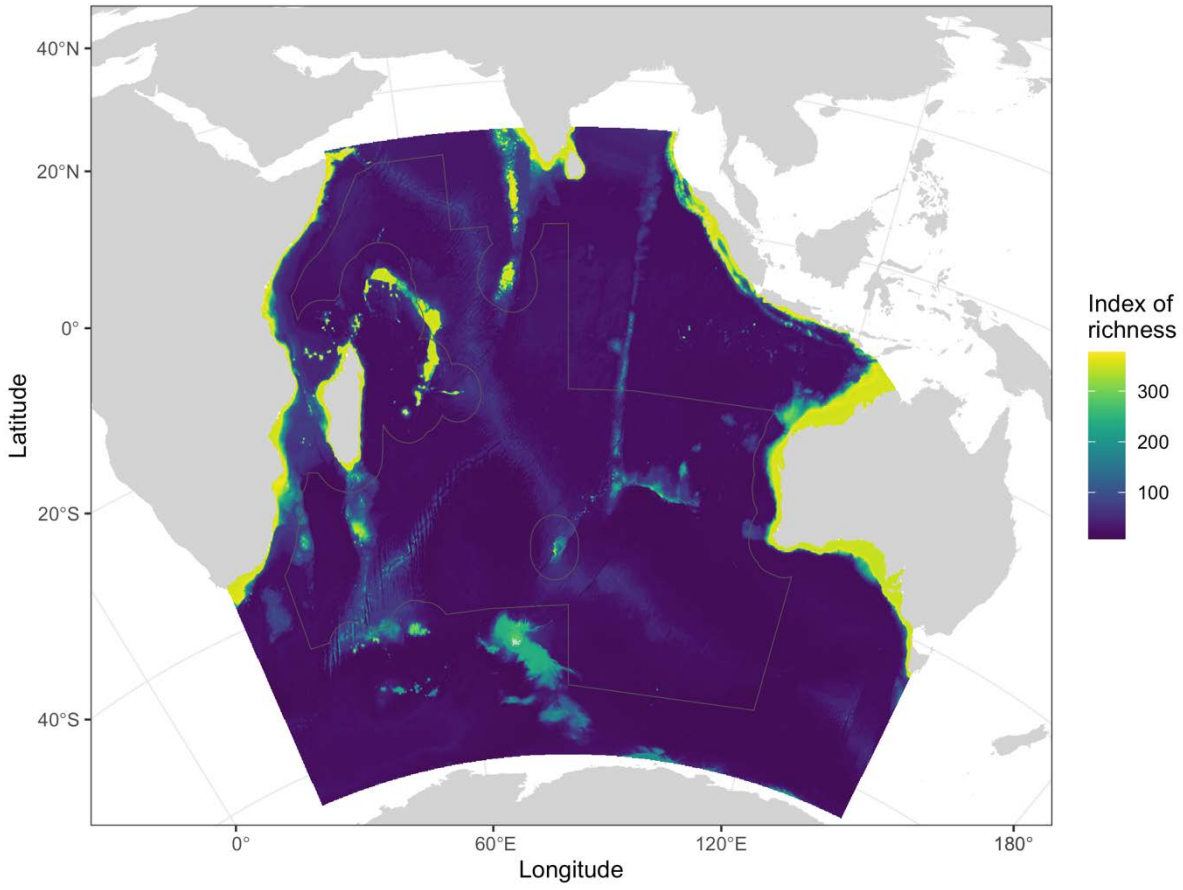

Figure 1. Index of richness at genus taxonomic level. The map shows the cumulative sum of occurrence probabilities per grid cell (resolution at 0.08333° latitude-longitude) based on habitat suitability models for genera of VME indicators.

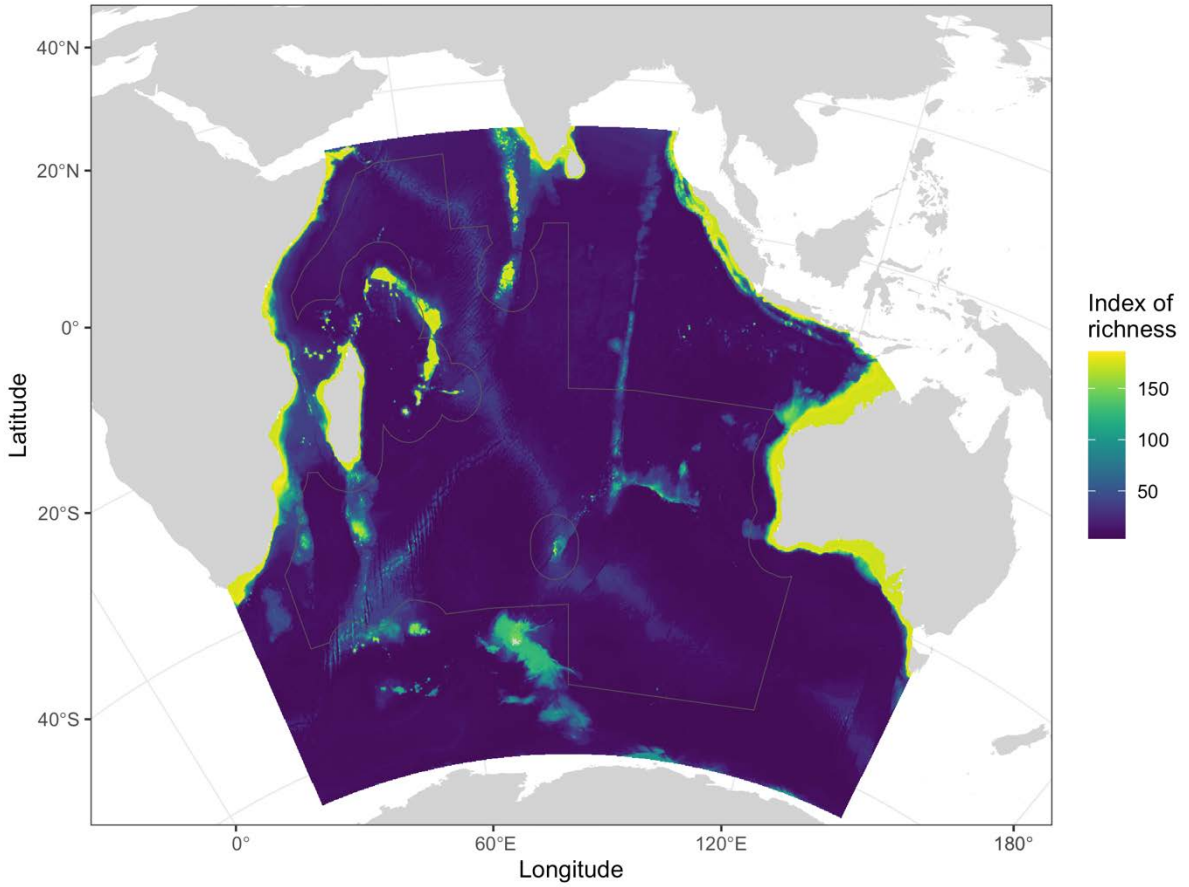

Figure 2. Index of richness at family taxonomic level. The map shows the cumulative sum of occurrence probabilities per grid cell (resolution at 0.08333° latitude-longitude) based on habitat suitability models for families of VME indicators.

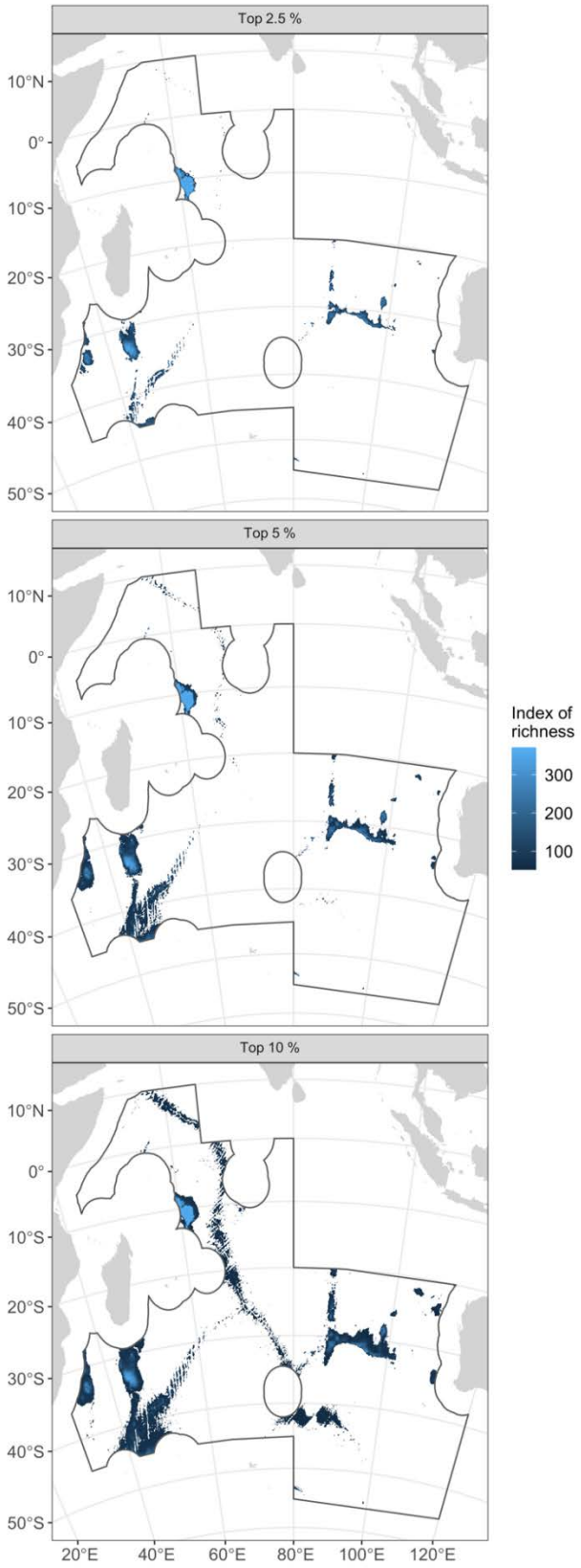

Figure 3. Biodiversity hotspots based on stacked-species distribution models of genera of VME indicators. The panels show an index of richness (sum of stacked suitability values in a cell) defining hotspot as grid cells with values in the top 2.5%, top 5% and top 10% richness (from more to less conservative). Note the progressive emergence of the ocean ridges with less restrictive biodiversity thresholds.

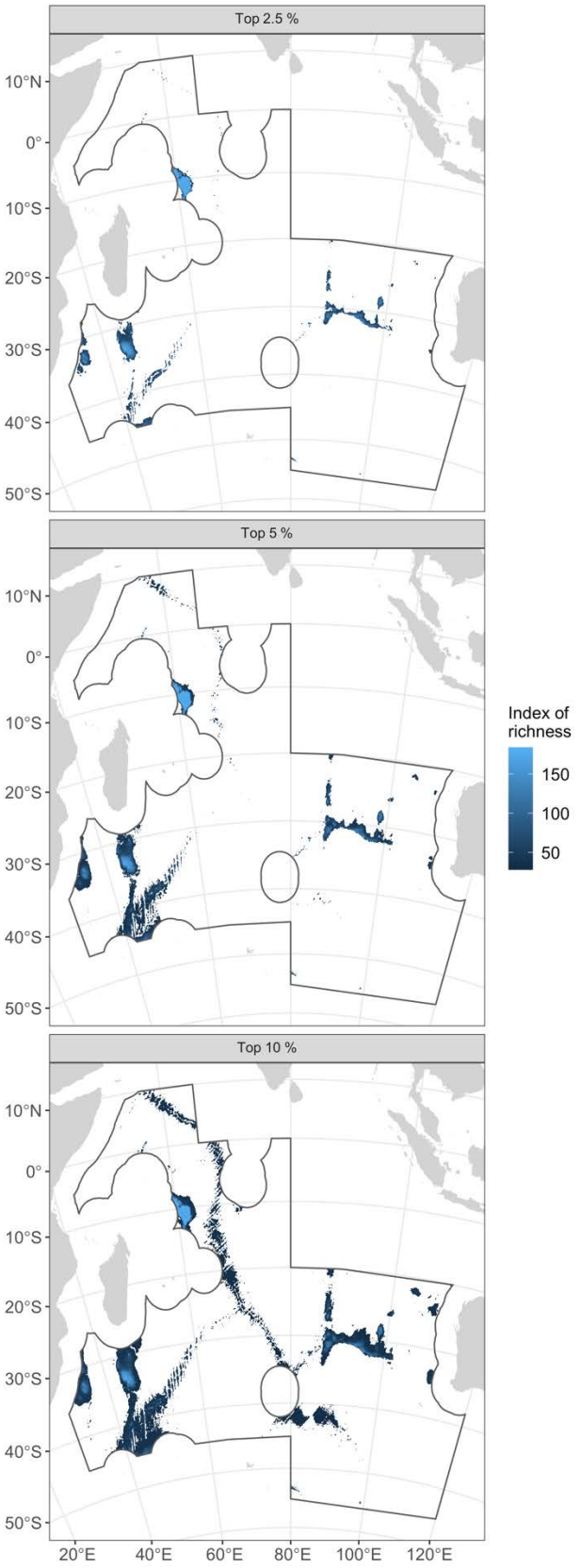

Figure 4. Biodiversity hotspots based on stacked-species distribution models of families of VME indicators. The panels show an index of richness (sum of stacked suitability values in a cell) defining hotspot as grid cells with values in the top 2.5%, top 5% and top 10% richness (from more to less conservative). Note the progressive emergence of the ocean ridges with less restrictive biodiversity thresholds.

4.2 Macroecological Models

We included in the calibration of the full model oxygen, salinity, mean net primary production, global sediment layer and maximum depth as predictor variables. We excluded undersampled cells, i.e., cells that had strictly less than three genera sampled. Model check of the posterior predictive probabilities for the generalised linear mixed model and the generalised additive mixed model indicated that the models were not adequate to predict richness at low numbers when compared with the observed data. Furthermore, model assumptions were not met: there was under and over dispersion in data and great heterogeneity of variance in the residuals. The generalised linear model explained only 25% of the deviance at the genus level, also suggesting an inadequate model fit.

The ensemble model included nine models in total (three per algorithm). Although we cannot trust the ensemble model as individual model assumptions were not met, we present the model results for consistency (Figure 5). In terms of spatial patterns, the highest prediction of species richness was in the Saya de Malha, northwest of Madagascar and Chagos Laccadive Ridge. The bathyal depth zone (300 – 3500 m) showed the next strongest trends. In particular, this corresponded with the topographic complex areas of the Southwestern Indian Ridge, Broken Ridge and the Kerguelen Plateau. The ocean basins showed higher predicted biodiversity than the S-TDM predictions.

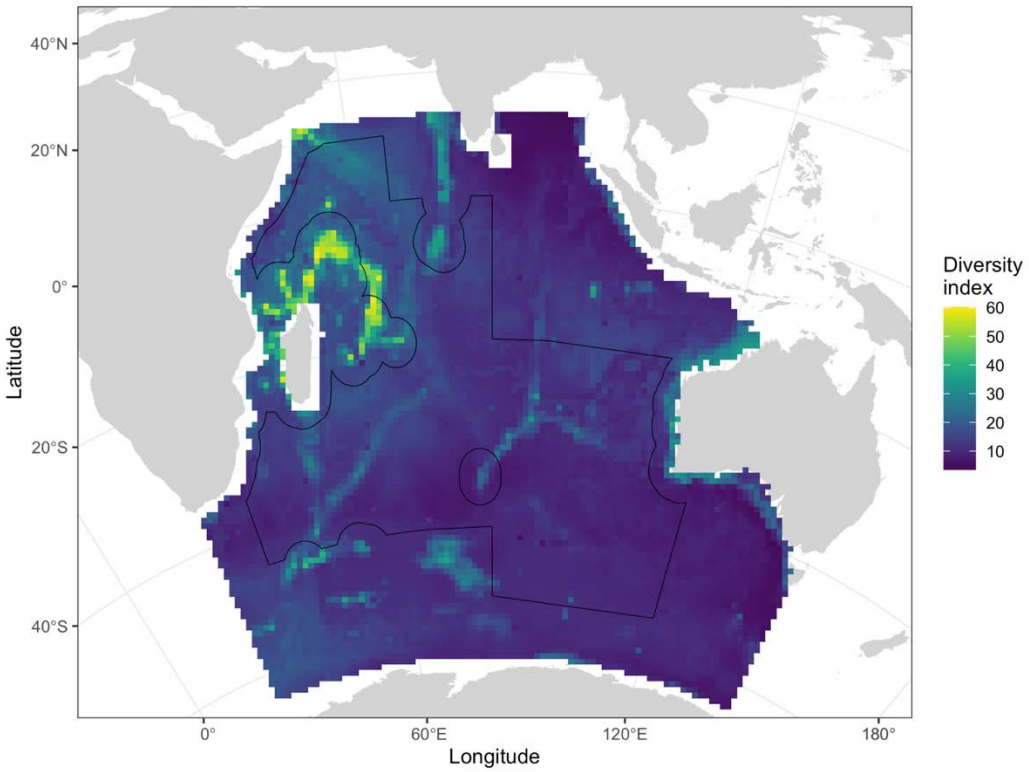

Figure 5. Macroecological model predicting genera richness in the Southern Indian Ocean based on observed richness of VME indicator genera.

5. Discussion

This work aimed to investigate and advise on the use of habitat suitability modelling in predicting benthic species diversity and distribution in SIOFA's management area. Here, we aimed to compare two approaches used for the prediction of diversity: stacked-taxa distribution models (S-TDMs) and macroecological models (MEMs). S-TDMs predictions performed well in predicting richness, although these models do not include rare taxa. In contrast, MEMs models failed to predict richness: VME indicator data shortfalls do not allow for direct modelling of observed richness. Yet, in terms of spatial patterns, the two approaches highlighted similar areas where more biodiversity is expected; however, we could not compare both indices of diversity in terms of absolute numbers. More complex methods estimating richness based on an expected sampling coverage might be useful to predict richness but will still necessitate of the inclusion of additional benthic data.

The S-TDMs performed well in predicting richness in the SIOFA area. The index of diversity was based on models that were statistically robust and that captured the areas where taxa occurred, even if these models showed overprediction in some areas (see Ramiro-Sánchez & Leroy, 2022 for an in-depth explanation of the models' results). The index could be used as a basis to apply the definition of hotspot to highlight areas of greater biodiversity in SIOFA. Such information can be further considered in conservation planning, for example. However, a disadvantage of the S-TDM approach is that it is based only on taxa that could be modelled, i.e., taxa that had a sufficient number of occurrences. The index therefore remains incomplete in coverage because it does not include rare taxa. In contrast, the MEM model failed to predict the observed richness in the area: the models did not meet model assumptions in an extreme case of missing data, which prevents their implementation. Complimentary to the S-TDM approach in that it includes rare taxa, the development of the MEM approach will require the acquisition of additional data.

In terms of spatial patterns, the two modelling approaches highlighted the mid-ocean ridges and plateaus as areas of higher diversity and, particularly within SIOFA, Saya de Malha, Walter's Shoal and Broken Ridge. Effectively, it is the bathyal depths that were predicted with more diversity. Although this is consistent with the radiation hypothesis (Holt, 1985) that predicts that deep-water diversity is maintained by immigration from bathyal sources (Rex et al., 2005), the insufficient samples prevent us from drawing strong conclusions to differentiate across bathyal geographic zones that potentially could vary in diversity. As a result, the S-TDM index of richness could be preliminary used in qualitative terms to highlight areas of lower, medium, or higher expected diversity. However, other studies predicting global benthic diversity show similar spatial patterns of diversity in the Indian Ocean. For ophiuroids, Saya de Malha and Chagos-Laccadive Ridge are also areas of high diversity at depths shallower than 2000 m, whereas at the same depths at higher latitudes the diversity decreases (Woolley et al., 2016). In contrast, at depths greater than 2000 m, diversity is predicted similarly, but it is enhanced in the western Indian Ocean, specifically in the Southwestern Indian Ridge, the Mozambique and the Agulhas Plateau (Woolley et al., 2016).

The low number of samples, the prevailing sampling bias, and high variability among sampled grid cells (regardless of the spatial resolution used) had consequences for the present study

investigating biodiversity models. Project PAE2020-02 ("VME Bioregionalization") assessed data availability by estimating the completeness of the biodiversity inventory collated for the project (Ramiro-Sánchez & Leroy, 2022) and used in this study. The completeness index of the gathered distributional database, that is, the ratio between observed richness and estimated richness, indicated that 72% of the species known to occur in the area had been included in the inventory. However, this estimated richness constitutes an upper-bound estimate only (Chao et al., 2020) because of the data paucity in the SIOFA area – therefore, the true completeness is most certainly much lower than the measured completeness. Indeed, the spatial distribution of the completeness index suggested that the SIOFA area is critically under-sampled: few sites had a high completeness, which corresponded with areas of higher observed species richness and sampling intensity in coastal areas and therefore outside of SIOFA's management area. In other words, the sampling is currently too inadequate to properly model richness in the area, which explains the failure of macroecological richness models. More complex modelling approaches are also unlikely to solve the evident data shortfalls of the Southern Indian Ocean and, more specifically, in the SIOFA area.

6. Conclusions

- There is insufficient data to predict the observed richness of VME indicator taxa in SIOFA.
- Because of the low number of samples, sampling bias and high variability among sampled grid cells, we can only make inferences about spatial patterns of diversity rather than considering absolute numbers.
- We recommend only the consideration of the S-TDMs models to further analyse potential patterns.
- The application of the definition of hotspot (i.e., areas with top 5%, or top 10% of biodiversity) to the S-TDM index of richness might help in the evaluation of important areas for biodiversity, such as including that information in a systematic conservation planning.

7. Recommendations

We recommend the Scientific Committee to:

- Note to consider the report and work carried out for this Terms of Reference.
- Note to consider only the biodiversity model based on S-TDM.
- Note to consider the application of the definition of hotspot to the S-TDM index of richness to use in future conservation planning.

References

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & de Clerck, O. (2018). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, *27*(3), 277–284. https://doi.org/10.1111/GEB.12693

Barber, R. A., Ball, S. G., Morris, R. K. A., & Gilbert, F. (2022). Target-group backgrounds prove effective at correcting sampling bias in Maxent models. *Diversity and Distributions*, *28*(1), 128–141. https://doi.org/10.1111/DDI.13442

Biber, M. F., Voskamp, A., Niamir, A., Hickler, T., & Hof, C. (2020). A comparison of macroecological and stacked species distribution models to predict future global terrestrial vertebrate richness. *Journal of Biogeography*, *47*(1), 114–129. https://doi.org/10.1111/JBI.13696

Cairns, S. D. (n.d.). On line appendix: Phylogenetic list of the 711 valid Recent azooxanthellate scleractinian species with their junior synonyms and depth ranges. In *Cold-Water Corals: The Biology and Geology of Deep-Sea Coral Habitats.* (p. 28). Cambridge University Press. Retrieved April 4, 2022, from https://www.marinespecies.org/introduced/aphia.php?p=sourcedetails&id=142534

Cairns, S. D. (2007). Deep-water corals: An overview with special reference to diversity and distribution of deep-water scleractinian corals. *Bulletin of Marine Science*, *81*(3), 311–322.

Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, *23*(1), 99–112. https://doi.org/10.1111/GEB.12102/SUPPINFO

Chao, A., Kubota, Y., Zelený, D., Chiu, C. H., Li, C. F., Kusumoto, B., Yasuhara, M., Thorn, S., Wei, C. L., Costello, M. J., & Colwell, R. K. (2020). Quantifying sample completeness and comparing diversities among assemblages. *Ecological Research*, *35*(2), 292–314. https://doi.org/10.1111/1440-1703.12102

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., Mcclean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/J.1600-0587.2012.07348.X

Ferrier, S., & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, *43*(3), 393–404. https://doi.org/10.1111/J.1365-2664.2006.01149.X

GEBCO. (2021). *GEBCO Gridded Bathymetry Data*. https://www.gebco.net

Gotelli, N. J., Anderson, M. J., Arita, H. T., Chao, A., Colwell, R. K., Connolly, S. R., Currie, D. J., Dunn, R. R., Graves, G. R., Green, J. L., Grytnes, J. A., Jiang, Y. H., Jetz, W., Kathleen Lyons, S., McCain, C. M., Magurran, A. E., Rahbek, C., Rangel, T. F. L. V. B., Soberón, J., … Willig, M. R. (2009). Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, *12*(9), 873–886. https://doi.org/10.1111/J.1461-0248.2009.01353.X

Guisan, A., & Rahbek, C. (2011). SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, *38*(8), 1433–1444. https://doi.org/10.1111/J.1365-2699.2011.02550.X

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, *8*(9), 993–1009. https://doi.org/10.1111/J.1461-0248.2005.00792.X

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*(2–3), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Holt, R. D. (1985). Population dynamics in two-patch environments: Some anomalous consequences of an optimal habitat distribution. *Theoretical Population Biology*, *28*(2), 181–208. https://doi.org/10.1016/0040-5809(85)90027-9

Kocsis, Á. T., Reddin, C. J., & Kiessling, W. (2018). The stability of coastal benthic biogeography over the last 10 million years. *Global Ecology and Biogeography*, *27*(9), 1106–1120. https://doi.org/10.1111/GEB.12771

Kull, M., & Flach, P. (2022). *prg: Software to create Precision-Recall-Gain curves and calculate area under the curve* (0.5.1). https://github.com/meeliskull/prg

Leroy, B., Delsol, R., Hugueny, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M., & Bellard, C. (2018). Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography*, *45*(9), 1994–2002. https://doi.org/10.1111/JBI.13402

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22. http://www.stat.berkeley.edu/

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature 2000 405:6783*, *405*(6783), 243–253. https://doi.org/10.1038/35012251

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, *6*(1), 1–8. https://doi.org/10.1038/ncomms9221

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Höft, R., Hurtt, G., Jetz, W., … Wegmann, M. (2013). Essential biodiversity variables. *Science*, *339*(6117), 277–278. https://doi.org/10.1126/SCIENCE.1229931/SUPPL_FILE/1229931.PEREIRA.SM.PDF

Ramiro-Sánchez, B., & Leroy, B. (2022). *SIOFA bioregionalization and VME project*. https://siofa.org/sites/default/files/files/VMEMapping_FullReport.pdf

Rex, M. A., McClain, C. R., Johnson, N. A., Etter, R. J., Allen, J. A., Bouchet, P., & Warén, A. (2005). A source-sink hypothesis for abyssal biodiversity. *The American Naturalist*, *165*(2), 163–178. https://doi.org/10.1086/427226/ASSET/IMAGES/LARGE/FG5.JPEG

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ECOG.02881

Smith, A. B. (2020). *enmSdm: Faster, better, smarter ecological niche modeling* (5.3.0). https://github.com/adamlilith/enmSdm

Snelgrove, P. V. R. (1999). Getting to the bottom of marine biodiversity: Sedimentary habitats: Ocean bottoms are the most widespread habitat on earth and support high biodiversity and key ecosystem services. *BioScience*, *49*(2), 129–130. https://doi.org/10.2307/1313538/2/BISI.1999.49.2.129-FU01.JPEG

Straume, E. O., Gaina, C., Medvedev, S., Hochmuth, K., Gohl, K., Whittaker, J. M., Abdul Fattah, R., Doornenbal, J. C., & Hopper, J. R. (2019). GlobSed: Updated Total Sediment Thickness in the World's Oceans. *Geochemistry, Geophysics, Geosystems*, *20*(4), 1756–1772. https://doi.org/10.1029/2018GC008115

Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. vanden, & Worm, B. (2010). Global patterns and predictors of marine biodiversity across taxa. *Nature 2010 466:7310*, *466*(7310), 1098–1101. https://doi.org/10.1038/nature09329

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & de Clerck, O. (2012). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, *21*(2), 272–281. https://doi.org/10.1111/J.1466-8238.2011.00656.X

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, *10*(2), 225–232. https://doi.org/10.1111/2041-210X.13107

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. *Ecography*, *44*(12), 1731–1742. https://doi.org/10.1111/ECOG.05615

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2021). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs*, *92*(1), e01486. https://doi.org/10.1002/ECM.1486

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer.

Witman, J. D., Etter, R. J., & Smith, F. (2004). The relationship between regional and local species diversity in marine benthic communities: A global perspective. *Proceedings of the National Academy of Sciences*, *101*(44), 15664–15669. https://doi.org/10.1073/PNAS.0404300101

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), 3–36. https://doi.org/10.1111/J.1467-9868.2010.00749.X

Woolley, S. N. C., Foster, S. D., Bax, N. J., Currie, J. C., Dunn, D. C., Hansen, C., Hill, N. A., O'Hara, T. D., Ovaskainen, O., Sayre, R., Vanhatalo, J. P., & Dunstan, P. K. (2020). Bioregions in Marine Environments: Combining Biological and Environmental Data for Management and Scientific Understanding. *BioScience*, *70*(1), 48–59. https://doi.org/10.1093/BIOSCI/BIZ133

Woolley, S. N. C., Tittensor, D. P., Dunstan, P. K., Guillera-Arroita, G., Lahoz-Monfort, J. J., Wintle, B. A., Worm, B., & O'Hara, T. D. (2016). Deep-sea diversity patterns are shaped by energy availability. *Nature*, *533*(7603), 393–396. https://doi.org/10.1038/nature17937

WoRMS. (2021). *WoRMS - World Register of Marine Species*. https://www.marinespecies.org/

Zeppilli, D., Pusceddu, A., Trincardi, F., & Danovaro, R. (2016). Seafloor heterogeneity influences the biodiversity–ecosystem functioning relationships in the deep sea. *Scientific Reports 2016 6:1*, *6*(1), 1–12. https://doi.org/10.1038/srep26352

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Daniel Edler, |, Farooq, H., Herdean, A., Ariza, M., Ruud Scharn, |, Svantesson, S., Wengström, N., Zizka, V., & Antonelli, | Alexandre. (2019). CoordinateCleaner: Standardized cleaning of occurrence

records from biological collection databases. *Methods Ecol Evol*, 1–8. https://doi.org/10.1111/2041-210X.13152

Appendix A

List of all GBIF occurrence data downloads used for analysis.

GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.9gyvnr
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.5cnauy
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.tpkv6f
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.g7trjv
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.qmg7mm
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.532e4a
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.nfwf4e
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.ujfmwm
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.p3fdas
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.8x49vt
GBIF.org (14 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.s528ag
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.csn8ss
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.4sup3k
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.x3psvp
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.8utndr
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.9db6ad
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.8sjem7
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.p88rng
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.7jckj5
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.dun29p
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.5jmrg4
GBIF.org (07 April 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.psarr7
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.8kygyt
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.f6hkv7
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.29bzd2
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.qx93xv
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.amhv2z
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.jex2bf
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.kc6f2c
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.2eptnr
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.grub32
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.y5bpdj
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.dhv3xd
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.984ku6
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.xvk58n
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.jz8e93
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.rb2jgj
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.u96fx3
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.nhd8sd
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.ge6bqe
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.s82sx7
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.2nks2p
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.zesz8a
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.vhg2c5
GBIF.org (09 November 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.em7j8c

Appendix B

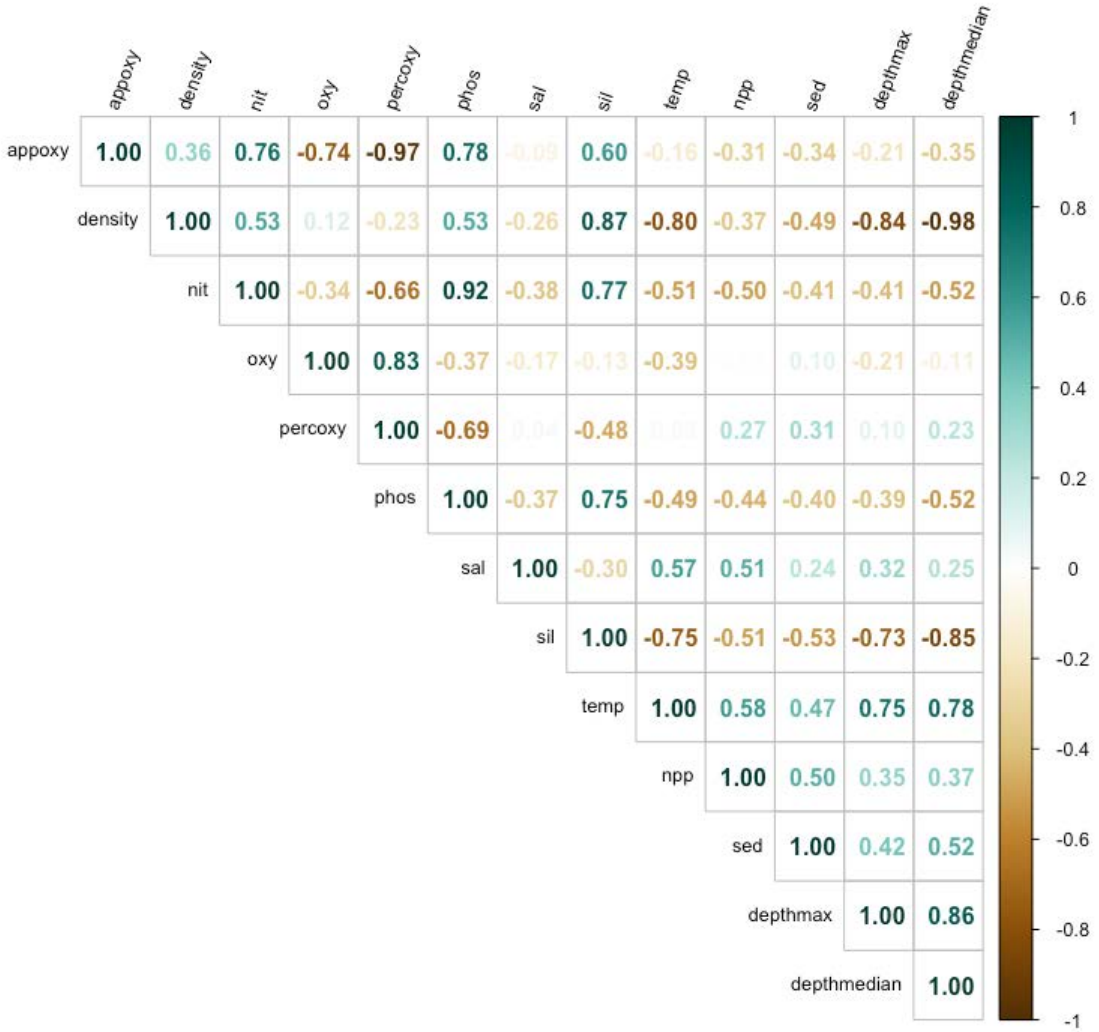

Figure B1. Spearman correlation plot between pairs of environmental variables selected for inclusion in the analysis of the macroecological model.

Table B1. List of families included in the Stacked-Taxa Distribution Model.

Acanthogorgiidae	Crambeidae	Isodictyidae	Romancheinidae
Acarnidae	Cribrilinidae	Isselicrinidae	Rossellidae
Actinernidae	Crisiidae	Kophobelemnidae	Schizopathidae
Actiniidae	Ctenocidaridae	Kraussinidae	Scleractinia.incertae.sedis
Adeonidae	Cupuladriidae	Lacernidae	Scleritodermidae
Agariciidae	Dallinidae	Latrunculiidae	Scleroptilidae
Agelasidae	Darwinellidae	Leiopathidae	Scopalinidae
Agneziidae	Deltocyathidae	Leiosalpingidae	Serpulidae
Alcyoniidae	Dendrophylliidae	Lepraliellidae	Siderastreidae
Amphianthidae	Desmacellidae	Leptastreidae	Siphonicytaridae
Ancorinidae	Diazonidae	Lichenoporidae	Smittinidae
Antedonidae	Dictyodendrillidae	Liponematidae	Solanderiidae
Anthemiphylliidae	Dictyonellidae	Lobophylliidae	Sphenopidae
Anthothelidae	Didemnidae	Mariametridae	Spirastrellidae
Antipathidae	Diploastreidae	Megathyrididae	Stannomidae
Aphrocallistidae	Discinidae	Membraniporidae	Steginoporellidae
Arachnopusiidae	Dyscoliidae	Micrabaciidae	Stelligeridae
Asterometridae	Dysideidae	Microporellidae	Stenocyathidae
Asteronychidae	Echinoptilidae	Microporidae	Styelidae
Astrocoeniidae	Electridae	Milleporidae	Stylasteridae
Aulocalycidae	Ellisellidae	Molgulidae	Stylocordylidae
Basiliolidae	Epizoanthidae	Monorhaphididae	Suberitidae
Bathylasmatidae	Escharinidae	Mycalidae	Tedaniidae
Beaniidae	Esperiopsidae	Myriopathidae	Terebratellidae
Bifaxariidae	Eudendriidae	Myxillidae	Terebratulidae
Bitectiporidae	Euphylliidae	Niphatidae	Terviidae
Bubaridae	Euplectellidae	Octacnemidae	Tessaradomidae
Bugulidae	Euretidae	Oculinidae	Tethyidae
Calescharidae	Euryalidae	Order_Cheilostomatida	Tetillidae
Callyspongiidae	Exochellidae	Pachastrellidae	Thalamoporellidae
Calwelliidae	Farciminariidae	Paragorgiidae	Thalassianthidae
Cancellothyrididae	Farreidae	Parazoanthidae	Thalassometridae
Candidae	Fenestrulinidae	Pennatulidae	Theneidae
Caryophylliidae	Flabellidae	Pentametrocrinidae	Theonellidae
Catenicellidae	Flustridae	Perophoridae	Thoosidae
Cellariidae	Foveolariidae	Petrosiidae	Timeidae
Celleporidae	Fungiacyathidae	Pheronematidae	Tubiporidae
Chalinidae	Fungiidae	Phloeodictyidae	Turbinoliidae
Chaperiidae	Funiculinidae	Phymaraphiniidae	Umbellulidae
Chlidonophoridae	Geodiidae	Platidiidae	Verticillitidae
Chondrillidae	Gorgonocephalidae	Plerogyridae	Vesiculariidae
Chondrosiidae	Guyniidae	Plesiastreidae	Virgulariidae

Chrysogorgiidae	Halichondriidae	Plexauridae	Vulcanellidae
Chunellidae	Halisarcidae	Pocilloporidae	Xeniidae
Cidaridae	Heliodomidae	Podospongiidae	Zancleidae
Cladopathidae	Hemithiridídae	Polymastiidae	
Cladorhizidae	Himerometridae	Primnoidae	
Clavulariidae	Hippothoidae	Protopolyclinidae	
Cleidochasmatidae	Histocidaridae	Psammettidae	
Coelogorgiidae	Hormathiidae	Psamminidae	
Coelosphaeridae	Hyalonematidae	Psammocoridae	
Colobometridae	Hymedesmiidae	Pyuridae	
Conescharellinidae	Ianthellidae	Quadricellariidae	
Corellidae	Irciniidae	Rhizangiidae	
Coscinaraeidae	Isididae	Rhizocrinidae	

Table A2. List of genera included in the Stacked-Taxa Distribution Model.

Aaptos	*Crisia*	*Ichthyaria*	*Polymastia*
Abyssascidia	*Crotalometra*	*Inflatella*	*Polyphyllia*
Abyssopathes	*Crypthelia*	*Iophon*	*Polysyncraton*
Acanthastrea	*Cryptodendrum*	*Ircinia*	*Porites*
Acanthella	*Ctenactis*	*Isodictya*	*Pourtalopsammia*
Acanthocidaris	*Ctenella*	*Isopora*	*Premocyathus*
Acanthogorgia	*Ctenocella*	*Isozoanthus*	*Primnoisis*
Actinernus	*Ctenocidaris*	*Jaffaia*	*Proagnesia*
Adagnesia	*Culicia*	*Jaspis*	*Promachocrinus*
Adeona	*Cyamon*	*Javania*	*Protosuberites*
Adeonella	*Cyathotrochus*	*Junceella*	*Psammetta*
Aerothyris	*Cycloseris*	*Kaliapsis*	*Psammochela*
Agelas	*Cynarina*	*Kinetoskias*	*Psammocinia*
Alcyonium	*Dactylostega*	*Kraussina*	*Psammocora*
Alertigorgia	*Damiria*	*Labioporella*	*Pseudodiazona*
Amastigia	*Danafungia*	*Labyrinthocyathus*	*Pseudosiderastrea*
Amorphinopsis	*Darwinella*	*Lamellodysidea*	*Pseudosuberites*
Amphilectus	*Decametra*	*Latrunculia*	*Pteroeides*
Amphimedon	*Deltocyathoides*	*Leiopathes*	*Raspailia*
Amphimetra	*Deltocyathus*	*Leiosalpinx*	*Regadrella*
Anacropora	*Democrinus*	*Lepidopora*	*Retiflustra*
Annella	*Dendrilla*	*Leptastrea*	*Rhodelinda*
Anomastraea	*Dendrophyllia*	*Leptopsammia*	*Rhombopsammia*
Anomocora	*Dercitus*	*Leptoria*	*Rhopalaea*
Antarctotetilla	*Desmophyllum*	*Leptoseris*	*Rhynchocidaris*
Anthemiphyllia	*Dichrometra*	*Letepsammia*	*Rhynchozoon*
Antho	*Diploastrea*	*Liosina*	*Rossella*
Antipathes	*Diplonotos*	*Liothyrella*	*Sandalolitha*
Aphrocallistes	*Discodermia*	*Liponema*	*Saracrinus*
Arachnopusia	*Discoporella*	*Lissodendoryx*	*Sarcophyton*
Argyrotheca	*Disporella*	*Lithoplocamia*	*Scapophyllia*
Artemisina	*Distichopora*	*Litophyton*	*Scleranthelia*
Asajirus	*Ditrupa*	*Lobactis*	*Scleroptilum*
Ascidia	*Domosclerus*	*Lobophyllia*	*Scuticella*
Asteronyx	*Dyscolia*	*Lobophytum*	*Scytalium*
Asteroporpa	*Dysidea*	*Lophelia*	*Semperella*
Asteropus	*Echinophyllia*	*Lophophysema*	*Seriatopora*
Astrea	*Echinopora*	*Madracis*	*Seriocarpa*
Astreopora	*Echinoptilum*	*Madrepora*	*Serpula*
Astroceras	*Ecionemia*	*Magellania*	*Setosellina*
Astrodia	*Ecteinascidia*	*Malakosaria*	*Siderastrea*
Astromuricea	*Ectyonopsis*	*Megaciella*	*Sigmosceptrella*
Astrosierra	*Eguchipsammia*	*Megerlia*	*Sinularia*

Astrothorax	*Electra*	*Megerlina*	*Sinuorota*
Astrotoma	*Eleutherobia*	*Menipea*	*Siphonicytara*
Atriolum	*Ellisella*	*Merulina*	*Siphonodictyon*
Auletta	*Emballotheca*	*Metacrinus*	*Smitticellaria*
Aulocyathus	*Enallopsammia*	*Metavermilia*	*Smittina*
Aulospongus	*Endectyon*	*Micromussa*	*Smittoidea*
Axinyssa	*Endopachys*	*Micropora*	*Solanderia*
Balanophyllia	*Epizoanthus*	*Millepora*	*Solenosmilia*
Basiliola	*Errina*	*Molgula*	*Sphenotrochus*
Bathylasma	*Erylus*	*Molguloides*	*Spirastrella*
Bathynanus	*Escharella*	*Monanchora*	*Spirobranchus*
Bathypathes	*Escharoides*	*Monorhaphis*	*Spongionella*
Bathystyeloides	*Eucalathis*	*Muricella*	*Stannophyllum*
Bathytelesto	*Eucidaris*	*Mycale*	*Steginocellaria*
Beania	*Eudendrium*	*Mycedium*	*Steginoporella*
Bernardpora	*Euginoma*	*Myxilla*	*Stelletta*
Bicrisia	*Eupaxia*	*Narella*	*Stellitethya*
Bifaxaria	*Euphyllia*	*Nellia*	*Stenocyathus*
Biflustra	*Euplectella*	*Neopetrosia*	*Stephanocyathus*
Blastomussa	*Eurypon*	*Niphates*	*Stephanophyllia*
Bolocera	*Farrea*	*Notocoryne*	*Stereocidaris*
Bubaris	*Fenestrulina*	*Notoplites*	*Sthenocephalus*
Bugula	*Figularia*	*Oceanapia*	*Stichopathes*
Caberea	*Fimbriaphyllia*	*Octacnemus*	*Styela*
Caleschara	*Flabellum*	*Ogivalia*	*Stylaraea*
Callogorgia	*Flustramorpha*	*Oligotrema*	*Stylaster*
Callyspongia	*Formosocellaria*	*Ophiocreas*	*Stylissa*
Calyx	*Foveolaria*	*Osthimosia*	*Stylobates*
Camptoplites	*Fungia*	*Oulophyllia*	*Stylocidaris*
Cancellothyris	*Fungiacyathus*	*Oxypora*	*Stylocoeniella*
Cantharellus	*Funiculina*	*Pachastrella*	*Stylocordyla*
Carbasea	*Galatheammina*	*Pachyseris*	*Suberites*
Carijoa	*Galaxea*	*Palythoa*	*Sycozoa*
Caryophyllia	*Galeopsis*	*Paraconotrochus*	*Synoicum*
Caulophacus	*Gardineroseris*	*Paracyathus*	*Tedania*
Cavernularia	*Gelliodes*	*Paragorgia*	*Telesto*
Cellaria	*Geodia*	*Paraminabea*	*Tentorium*
Celleporina	*Gephyrophora*	*Paramontastraea*	*Terebratulina*
Chalinula	*Goniastrea*	*Parantipathes*	*Tervia*
Chaperia	*Goniocidaris*	*Parastenella*	*Tessaradoma*
Chaperiopsis	*Goniocorella*	*Paratetilla*	*Tethya*
Characodoma	*Gorgonocephalus*	*Pavona*	*Tetilla*
Chlidonophora	*Guynia*	*Pectinia*	*Thenea*
Chondrilla	*Gyrosmilia*	*Pelagodiscus*	*Theonella*

Chondrocladia	*Halichondria*	*Pemphixina*	*Thesea*
Chondrosia	*Haliclona*	*Penares*	*Thouarella*
Cinachyra	*Halisarca*	*Pennatula*	*Thysanometra*
Ciocalypta	*Halomitra*	*Pentametrocrinus*	*Timea*
Cladiella	*Helicogorgia*	*Petrosia*	*Topsentia*
Cladorhiza	*Herdmania*	*Phakellia*	*Trochocyathus*
Clavularia	*Herpolitha*	*Phaneropora*	*Truncatoflabellum*
Cnemidocarpa	*Heteroxenia*	*Phonicosia*	*Tubastraea*
Coelocarteria	*Higginsia*	*Phyllospongia*	*Tubipora*
Coelogorgia	*Himantozoum*	*Physogyra*	*Umbellula*
Coeloseris	*Hippothoa*	*Pione*	*Vaceletia*
Coelosphaera	*Histocidaris*	*Platidia*	*Vermiliopsis*
Colobometra	*Histodermella*	*Platygyra*	*Virgularia*
Columnella	*Holascus*	*Plerogyra*	*Walteria*
Comatella	*Homaxinella*	*Plesiastrea*	*Xenobrochus*
Conotrochus	*Horastrea*	*Pleuractis*	*Xestospongia*
Cornucopina	*Hyalonema*	*Plicatellopsis*	*Zanclea*
Cornulella	*Hydnophora*	*Pocillopora*	*Zyzzya*
Coscinaraea	*Hymedesmia*	*Podabacia*	
Craniella	*Hyrtios*	*Poecillastra*	
Craterastrea	*Ianthella*	*Polycarpa*	

Table A3. Index of richness for genus and family habitat suitability models within SIOFA at quantiles showing the highest proportion of biodiversity in the top 10%, 5% and 2.5%.

	Index of richness	Top 10%	Top 5%	Top 2.5%
Genus	Minimum value	51.64	69.40	89.18
	Maximum value	371.10	371.10	371.10
Family	Minimum value	27.91	38.45	48.57
	Maximum value	183.86	183.86	183.86

IDENTIFYING SIGNIFICANT ADVERSE IMPACTS ON VULNERABLE MARINE ECOSYSTEMS: A REVIEW OF MEASURES AND APPLICABILITY TO SIOFA

BERTA RAMIRO SÁNCHEZ, BORIS LEROY

Project PAE2021-01

APRIL 2023

MUSEUM NATIONAL D'HISTOIRE NATURELLE
43 Rue Cuvier, 75005 Paris, France

Funded by the European Union

Suggested citation:

Ramiro-Sanchez, B. and Leroy, B. (2023). *Identifying significant adverse impacts on vulnerable marine ecosystems: a review of measures and applicability to SIOFA (Project PAE2021-01)*. Muséum d'Histoire Naturelle, France.

Table of Contents

Executive summary

To ensure the sustainability of the exploitation of fisheries resources, a series of United Nations General Assembly (UNGA) resolutions called upon States to assess and prevent significant adverse impacts (SAIs) on vulnerable marine ecosystems (VMEs). In line with international resolutions, the objectives of the Southern Indian Ocean Fisheries Agreement (SIOFA) are "to ensure the long-term conservation and sustainable use of the fishery resources in the Area through cooperation among the Contracting Parties, and to promote the sustainable development of fisheries in the Area, taking into account the needs of developing States bordering the Area that are Contracting Parties to this Agreement, and in particular the least developed among them and small-island developing States" (www.apsoi.org). The aim of this report is to provide an overview of the available measures for assessing and preventing Significant Adverse Impacts (SAIs) on VMEs in SIOFA's Convention Area.

A literature review of management measures to prevent SAIs is presented and these are discussed in relation to the information that is available to SIOFA. We first review the concept of VME developed by the FAO Guidelines and its practical application. Then, we continue to list management measures with their advantages and disadvantages, the potential outcomes of their combination, as well as examples of their implementation. We put particular focus on the so-called 'move-on rules', a form of spatial measure, which is currently the main mechanism to prevent SAIs in SIOFA. Finally, we assessed the use of distribution models and avenues of these predictive methodologies to overcome shortcomings of individual VME taxon predictions.

SIOFA's current Conservation Management Measures (CMMs) to prevent SAIs include the designation of five interim benthic protected areas where trawling is not permitted. In addition, SIOFA has set up encounter thresholds, which are presently the main instrument to prevent SAIs. SIOFA is increasingly building layers of information useful to inform the designation of more spatial measures, in particular of spatial closures, including predictive models of the extent of VME bioregions and individual taxa. Other available information includes the consideration of Ecologically and Biologically Significant Areas (EBSAs), connectivity models developed in the Western Indian Ocean, and current fishing regimes in adjacent territorial waters of the fishing footprint. As for now, VME bycatch data is of insufficient temporal resolution (2017 start) to provide reliable estimates of encounter thresholds.

The uncertainty in VME distributional data and predictive modelling outputs, renders fishing information as the most accurate and well documented source to derive results. This means that these layers of information should be precise and of high-resolution (within the limits of data confidentiality) in order to derive recommendations that protect benthic habitats and that are not too restrictive to fishing. The importance of high-resolution maps is key if the intention is to use a systematic planning procedure to design spatial closures.

1. Purpose of the report

This report aims to address Terms of Reference 4 "*Investigate and advise on a holistic framework for assessing and preventing Significant Adverse Impacts (SAIs) on vulnerable marine ecosystems (VMEs)*" of project PAE2021-01.

2. VME concepts: a key prelude to management

UNGA Resolution 61/105 called upon States to assess and prevent significant adverse impacts on vulnerable marine ecosystems (VMEs). The implementation of this call has been reinforced through a series of successive resolutions which:

- mandated the establishment of protocols, definitions of evidence of an encounter with a VME, thresholds levels and indicator species, and implementation of the FAO International Guidelines for the Management of Deep-sea Fisheries in the High Seas (FAO, 2009) (UNGA Resolution 64/72);
- called for the use of scientific research outputs such as habitat mapping and the implementation of Conservation and Management measures (UNGA Resolution 66/68);
- called upon RFMOs to use the full criteria in the FAO Deep-sea Fisheries Guidelines to identify , as well as the implementation of Sustainable Development Goals in accordance with the post -2015 agenda (UNGA Resolutions 71/123 and 72/72).

The International Guidelines for the management of deep-sea fisheries in the High Seas (the "Guidelines"; FAO, 2009) were developed to assist in the identification of VMEs, the definition of significant adverse impacts, their assessment and prevention. The Guidelines provide five criteria to identify VMEs:

Criterion 1. Uniqueness or rarity – an area or ecosystem that is unique or that contains rare species whose loss could not be compensated for by similar areas or ecosystems. These include:
• habitats that contain endemic species;
• habitats of rare, threatened, or endangered species that occur only in discrete areas; or
• nurseries or discrete feeding, breeding, or spawning areas.

Criterion 2. Functional significance of the habitat – discrete areas or habitats that are necessary for the survival, function, spawning/reproduction or recovery of fish stocks, particular life-history stages (e.g., nursery grounds or rearing areas), or of rare, threatened, or endangered marine species.

Criterion 3. Fragility – an ecosystem that is highly susceptible to degradation by anthropogenic activities.

Criterion 4. Life-history traits of component species that make recovery difficult – ecosystems that are characterized by populations or assemblages of species with one or more of the following characteristics:

- slow growth rates;
- late age of maturity;
- low or unpredictable recruitment; or
- long-lived.

Criterion 5. Structural complexity – an ecosystem that is characterized by complex physical structures created by significant concentrations of biotic and abiotic features. In these ecosystems, ecological processes are usually highly dependent on these structured systems. Further, such ecosystems often have high diversity, which is dependent on the structuring organisms.

The criteria described in the Guidelines to identify VMEs are purposedly broad and flexible to allow for their adaptation to different regional specificities. It also accounts for the limited knowledge of deep-sea biodiversity and poor sampling coverage in areas beyond national jurisdiction. Certainly, many deep-sea ecosystems and habitats have yet to be described in terms of occurring species and their rarity, resistance, resilience, vulnerability, and functionality, which renders difficult to use the criteria set out by the Guidelines. To address this difficulty, several taxa have been acknowledged as VME indicator taxa. VME indicators are taxa that most likely occur in habitats, or create biogenic habitats, meeting the criteria for VMEs (FAO, 2016). These comprise a number of listed habitats and habitat-forming taxa including cold-water corals reefs, coral gardens from soft-bottom and coral garden from hard-bottoms, sponge-dominated communities, communities that are composed of epifauna that provide a structural habitat for other associated species and bryozoan patches. These are all organisms attached to the seafloor, although seascapes such as seamounts, canyons, and knolls, are also considered VME indicators (FAO, 2016).

Nonetheless, the flexibility in the Guidelines has hindered progress on the implementation of measures for protection to VMEs because of the lack of explicit descriptions, measurements, and interchange of terms of VME and VME indicators. The loose semantic use of terms and clarification of concepts makes efforts to establish consistent systematic approaches to conservation difficult (Costello, 2009; Costello et al., 2020). Watling and Auster (2021) argue that the concept of indicator species has been conflated with the ecosystem itself so that the sparse distribution of an indicator species is interpreted as the presence and absence of a vulnerable marine ecosystem. The use of indicator species to describe sites or habitats is however a classic approach in ecology (e.g., Fitzpatrick et al., 2021). Indicator species are species that, due to their niche preferences, can be used as ecological indicators of community types, habitat conditions, or environmental changes (Carignan & Villard, 2002; McGeoch & Chown, 1998; Niemi & McDonald, 2004). They are normally used to represent all other species not sampled in the ecosystem; importantly, they do not define the composition of an associated community, species interactions in the community, or the flows of materials and energy that define the bounds of an ecosystem (Watling & Auster, 2021). An ecosystem, on the other hand, is the combination of species, their interactions, and the physical and chemical processes in their environment in a defined area (Costello et al., 2020). In that sense, ecosystems are composed of multiple habitats that support different types of communities. In contrast to habitats, ecosystems are usually considered over large areas and their distribution reflects current environmental and ecological conditions over large areas

(Costello et al., 2020). Likewise, species do not define ecosystems; rather, species composition defines a habitat (Costello et al., 2020). Thus, using indicator species does not account for the whole ecosystem and its components, such as, for example, spatially distinct communities within a seamount (Watling & Auster, 2021).

Although the aim of the Guidelines was to highlight the ecological importance of deep-sea species, habitats and ecosystems, the flexibility in terms has led to confusion in concepts, which, in turn, hinders the development of conservation measures. At the same time, it has allowed RFMO/As to adapt and interpret the term of VME and VME indicator accordingly to their region. Indeed, paragraph 43 of the Guidelines underlines that "These criteria should be adapted, and additional criteria should be developed as experience and knowledge accumulate, or to address particular local or regional needs". Furthermore, as exemplified in Costello et al. (2020), despite the conflation of ecological concepts, "the context always makes the meaning clear" and the Guidelines mention for example 'habitats or areas' in their criteria to identify VMEs (see Criterion 1, for instance). For instance, the International Council for the Exploration of the Sea (ICES) makes the difference (and provides context) between "VME habitats" as those records for which there is unequivocal evidence for a VME (e.g., ROV observations of a coral reef) and "VME indicators", which are records that suggest the presence of a VME with varying degrees of uncertainty. In order to facilitate the understanding of ecological terms, we present a glossary of ecological terminology as reviewed by Costello et al. (2020) (Table 1).

Yet, the scientific literature also reflects the sparingly semantic use of terms (Watling & Auster, 2021), perhaps to ascribe to the norm (i.e., the Guidelines), perhaps to keep things simple, perhaps due to professional bias of implicit appropriate interpretation. Watling and Auster (2021) exemplify the case with the work by Rowden et al. (2017) on the predicted distribution of the stony coral *Solenosmilia variabilis* in the South Pacific. They note that, although Rowden et al. (2017) used the term "*Solenosmilia* VME", in reality it corresponds to the *Solenosmilia* community, which is only one of the many communities supported on those seamounts.

Seamounts serve indeed as habitat for a large number of VME indicator species (Clark et al., 2010; Rogers, 2018). As such, a community dominated by a VME indicator species might extend along the whole or part of a seamount chain, or range over seamounts located in large sections of an ocean basin. Since deep-sea fisheries develop around aggregations of resident fish around seamounts and ridges (Clark et al., 2007; Kerry et al., 2022), these topographically complex areas are usually a target for conservation. As a result, , Watling & Auster (2021) make the following recommendations in order to take into account the relevance of seamount conservation and the potential large spatial extent of seamount ecosystems:

(i) use indicator species to identify individual seamount communities, but recognize that protecting part of a seamount identified only by the presence and distribution of an indicator species occurrence is not enough (Watling & Auster, 2017);
(ii) use the seamount classification system of Clark et al. (2011) or its equivalent to delimit groups of similar seamounts to focus conservation management efforts and to distinguish between rare and abundant seamount types;

(iii) examine the similarities among adjacent groups of seamounts to see whether they should be considered to be part of a larger ecosystem group;

(iv) evaluate the spatial extent of these larger units (larger ecosystem group) so that "significant adverse impacts" measures can be used to determine whether to allow some bottom fishing within a seamount ecosystem group.

In summary, the correct interpretation and understanding of ecological concepts are essential to the implementation of conservation management measures that are ecologically meaningful and that support sustainable fishing. Although the FAO Guidelines and provisions within may be imprecise, they also provide the opportunity to adapt to each regional case. In the following sections we continue to review available management actions; measures taken by other RFMO/As; the use of modelling for scenarios to inform management, and the applicability to SIOFA of available measures with the available information.

Table 1. Summary of the definitions and examples of ecological terminology. Taken and modified from Costello et al. (2020).

Concept	Definition	How is it measured?	Examples
Realm	Geographic area distinguished by a high proportion of endemic species arising from evolutionary history.	Geographic distribution of species.	Marine realms (Costello et al., 2017)
Biome	A large region characterised by similar habitat-forming plant growth forms enduring over years.		Kelp forest, mangrove forest, seagrass meadow, salt-marsh, shallow coral reefs.
Ecosystem	A large area of similar environmental conditions where it is considered that biological interactions and energy fluxes are greater within the ecosystem than across its boundaries to adjacent ecosystems.	Cluster analysis of environmental variables that would define a water mass and be related to ecological conditions, such as nutrients, salinity, light, and temperature.	The 3D "Ecological Marine Units" of (Sayre et al., 2017).
Seascape	A topographic or physiographic area defined by the shape of the coastline, bathymetry, and/or hydrography.	May be defined visually, by acoustic seabed mapping, aerial photography, spectrophotometric sensing (ocean colour), temperature, salinity.	Fjord, banks, canyon, trench, ridge, shelf, slope, seamount, basin, plain, depth zones.
Habitat	A physical environment inhabited by a named species or community over a timescale of years and replicated spatially.	Depends on the species or community of interest. The habitat for a marine mammal, seabird, and fish may be very different in size and physical environment.	*Desmophyllum pertusum* reefs, coral gardens, carbonate mounds, sea-pen and burrowing megafauna communities (OSPAR Agreement 2004-6).
Biogenic habitat	A three-dimensional habitat created by a species or group of species and occurring over a timescale of years and replicated spatially.		Biomes at a local scale where the species are named, e.g., *Zostera marina* meadow.
Assemblage	A group of species found together. Whether they interact or not is not known or assumed.		In practice conflated with community.

Concept	Definition	How is it measured?	Examples
Community	An assemblage of interacting species in a particular habitat occurring over a timescale of years and replicated spatially.	Visual observation, photography, samples of substrata and biota	E.g., *Solenosmilia*-dominated communities (Rowden et al., 2017)
Region (bio- eco-region, province, zone, section)	Expert opinion and/or based on ad hoc use of biogeographic, oceanographic or management criteria.	Variously defined depending on the context of the study. Bioregions may be defined by environmental conditions and selected species distributions.	Administrative environmental management areas for fisheries, conservation. Reviewed in Costello (2009) and (Zhao & Costello, 2020).
FAO Guidelines interpretation (ad-hoc addition to Costello et al. (2020) review)			
VME indicator	Taxa most likely to occur in habitats or that create biogenic habitats meeting the criteria for VMEs (FAO, 2016). ICES differentiates: 1) 'VME habitats' that are records for which there is unequivocal evidence for a VME (e.g., ROV observations of a coral reef). 2) 'VME indicators' are records that suggest the presence of a VME with varying degrees of uncertainty. A weighting system of vulnerability and uncertainty is used to aid interpretation.	Groups of species, communities, habitats, and seascapes such as seamounts.	Cold-water coral reefs, coral gardens from soft-bottoms (e.g., cup corals, cauliflower coral fields, soft-bottom gorgonians and black coral gardens) and coral gardens from hard-bottoms (hard-bottom gorgonians and black coral gardens, colonial scleractinians, non-reefal scleractinian aggregations), sponge dominated communities, (e.g., North-East Atlantic Fisheries Commission, FAO, 2016).
VME	Large areas containing habitats, communities, and populations whose life-histories, functional and structural significance characteristics render them susceptible to anthropogenic disturbance.	Area, habitat, ecosystem.	Cold-water coral reefs, hydrothermal vents, seeps.

3. Measures to prevent impacts on VMEs and VME indicator taxa

To prevent potential impacts on vulnerable species and habitats from bottom fishing, RMFO/As need to adopt conservation and management measures (CMMs), or close areas to bottom fishing until such CMMs are adopted (UNGA Resolution 66/68). In addition, they are required to assess the potential impacts of bottom fishing (UNGA Resolutions 71/123 and 72/72), for which the FAO Guidelines provide some criteria against which to identify what constitutes a significant impact. With regards to management measures, they can be in the form of spatial, technical or effort controls. These measures have varying levels of implementation success due to the uncertainty around VME biodiversity. For instance, a type of real-time spatial measure known as the "move on rule" requires a quantitative definition of what constitutes a VME. However, for other spatial measures, such as spatial closures or reserve design, predictions from species distribution models for VME indicator taxa can help to circumvent data shortfalls.

In this section, we first introduce the FAO definition of significant adverse impacts to provide context ("3.1. Significant adverse impacts"); we then elaborate on available management measures to address potential bottom-fishing impacts, with examples from other regions ("3.2 Management measures); we continue with the subject on the 'move-on rule' ("3.2.1 Move-on protocols"), and finish with a discussion on spatial measures and their dependency on distribution models for designating spatial closures or reserve design ("3.2.2 Dependency of spatial measures on distribution models").

3.1 Significant adverse impacts

The FAO Guidelines describe significant adverse impacts (SAIs) as those that affect the ecosystem structure and function so that they: (i) impair the ability of affected populations to replace themselves; (ii) degrade the long-term natural productivity of habitats; or (iii) cause, on more than a temporary basis, significant loss of species richness, habitat, or community types (FAO, 2009). When assessing an impact, the spatial and temporal (duration and frequency) extent of the impact on the species, habitat, or ecosystem, must be considered.

When determining the scale and significance of an impact, the following six factors should be considered:

i. the intensity or severity of the impact at the specific site being affected;
ii. the spatial extent of the impact relative to the availability of the habitat type affected;
iii. the sensitivity/vulnerability of the ecosystem to the impact;
iv. the ability of an ecosystem to recover from harm, and the rate of such recovery;
v. the extent to which ecosystem functions may be altered by the impact; and
vi. the timing and duration of the impact relative to the period in which a species needs the habitat during one or more of its life-history stages.

Temporary impacts are those that are limited in duration and that allow the particular ecosystem to recover over an acceptable time frame. Such time frames should be decided on a case-by-case basis and should be in the order of 5-20 years, taking into account the specific features of the populations and ecosystems (Paragraph 19).

In determining whether an impact is temporary, both the duration and the frequency at which an impact is repeated should be considered. If the interval between the expected disturbance of a habitat is shorter than the recovery time, the impact should be considered more than temporary. In circumstances of limited information, States and RFMO/As should apply the precautionary approach in their determinations regarding the nature and duration of impacts (Paragraph 20).

3.2 Management measures

The FAO Guidelines require States to have a 'functioning regulatory framework [that] should include an appropriate set of rules and regulations for the management of existing fisheries, as well as for the opening of new areas to exploratory fishing, consistent with these Guidelines and other relevant instruments. Such a framework should also include regulations to protect vulnerable populations, communities, and habitats. Management measures and voluntary industry actions (hereafter 'measures') that can be used to reduce and manage trawling effects on seabed habitats and biota can be grouped into four classes (McConnaughey et al., 2020):

1) Technical measures: refer to changes in gear design and operations;
2) Spatial controls: include gear-specific prohibitions, freezing the trawling footprint, nearshore restrictions and coastal zoning, prohibitions by habitat type including real time (i.e., 'move-one rules) and multipurpose habitat management (e.g., marine protected areas, MPAs),
3) Impact quotas: output controls that include invertebrate bycatch or habitat-impact quotas;
4) Effort controls: affect the overall distribution of trawling.

The four classes of measures can be evaluated using qualitative and quantitative performance metrics, although the preferred metrics will ultimately depend on the local, national, or regional context (McConnaughey et al., 2020). McConnaughey et al. (2020) propose four metrics for evaluation that include the positive and negative effects of trawling on (a) benthic biota; (b) sustainable fish populations and food productions, (c) ecosystems and ecosystem services and (d) economic performance of the fishery (Table 2). In addition, the different measures can be used singly or in combination; however, there may exist positive or negative interactions between these and other existing management actions (McConnaughey et al., 2020). Consequently, any potential interactions need to be considered systematically when any new measure is to be introduced. McConnaughey et al. (2020) assess such interactions and their consequences, which we present in Table 3.

To exemplify how the identification and assessment of available measures may promote discussions between stakeholders and subsequent actions in a region, we summarise the recent work that took place in the northeast Atlantic. The International Council for the Exploration of the Sea (ICES) is the organisation tasked with providing advice to the North-East Atlantic Fisheries Commission. On a workshop to discuss the recent work by McConnaughey et al. (2020), ICES gathered fishing representatives, non-governmental

organization (NGO) representatives and managers as part of a new request for "advice on a set of management options to reduce the impact of mobile bottom contacting fishing gears (MBCG) on seafloor habitats, and for each option provide a trade-off analysis between fisheries and the seafloor" (ICES, 2021). The purpose of the advice request was to provide a neutral analysis of potential costs or benefits to fisheries of achieving different levels of seafloor protection, based on the different management options identified. Outcomes of the workshop fed into a subsequent technical workshop where trade-offs were quantified, when tools and data were available, for each management option in different EU (sub-)regions and subdivisions (ICES, 2021). Participants from each group were asked to select five preferred management scenarios from a series of actions and discuss strengths and weaknesses. Additionally, participants were asked to comment on four options involving the removal of MBCG fishing effort (with nested options) by either starting from the least or most fished grid cells, and on freezing the MBCG fishing effort activity. Here, we summarise the outputs of this workshop in Table 4. In a similar manner, this exercise could be conducted by SIOFA to identify what is or not possible for their region.

AN ASSESSMENT OF SIGNIFICANT ADVERSE IMPACTS FROM FISHING ACTIVITIES IN SIOFA

BERTA RAMIRO SÁNCHEZ, BORIS LEROY

Project PAE2021-01

APRIL 2023

MUSEUM NATIONAL D'HISTORIE NATURELLE
43 Rue Cuvier, 75005 Paris, France

Funded by the European Union

Suggested citation:

Ramiro-Sanchez, B. and Leroy, B. (2023). *An assessment of significant adverse impacts from fishing activities in the SIOFA area (Project PAE2021-01)*. Muséum d'Histoire Naturelle, France.

Table of Contents

Executive summary

The present study assessed and quantified existing potential fishing impacts on predicted bioregions based on VME indicator taxa in the SIOFA area. Specifically, we investigated the intensity and spatial extent of fishing activities by gear type across bioregions, expressed as aggregated fishing effort (number of fishing events). We base the present study on the predictive bioregionalization method "group first, then predict" developed in project PAE2020-02.

In order to investigate how the existing fishing footprint could be potentially impacting predicted bioregions based on VME indicator taxa, we focused on understanding potential impacts in three ways: (1) how the fishing effort (split by gear type) was distributed over bioregions; (2) how the number of fishing events was distributed across the probability classes of the bioregions; and (3) how the spatial and intensity data could be integrated in an index that quantified the impact.

To assess the spatial extent and intensity of fishing per bioregion, we recovered the predictions of the habitat suitability scores of each bioregion and classified them in intervals of increasing suitability. For each gear type, fishing effort expressed as aggregated number of fishing events per pixel, was classified based on 10% quantiles. Next, we calculated the frequency of the number of fishing events per quantile and per suitability class.
We developed a fishing intensity impact index that measured how bioregions were affected by fishing by accounting for the fishing intensity (the higher the fishing intensity, the higher the index), and the location of fishing (the more fishing occurring in high suitability areas, the higher the index). In addition, the index was relative to the total surface area of each bioregion.

The overall picture indicated that fishing activities in SIOFA occur across all three main bioregions, although fishing types are spatially distributed and take place differently across nested subregions. While trawling, gillnet and trap fishing are only present in bioregion 1 and bioregion 3, only line fishing is present in bioregion 2. The intensity of fishing is also variable per gear type, but we found that, in many subregions, areas with maximum fishing intensity coincided with areas of high suitability for subregions. Our analyses suggested that bioregion 1 was affected with most of the fishing intensity by trawling, gillnets, and traps. For subregions, subregion 1.2 was the most impacted from bottom trawling, gillnet, and line fishing, while subregion 1.1 was the most at risk for traps.

The impact index offers a first approximation to assess potential impacts of fishing on bioregions. As new data becomes available in the form of better bioregion predictions and yearly fishing effort, the index can be updated accordingly. In combination with SIOFA's expert knowledge of their fisheries and area of competence, the index can be used to highlight particular areas of conservation interest around ambiguous areas and explore impacts of management measures.

1. Purpose of the report

This report aims to address Terms of Reference 5 "*Identify and update existing and potential Significant Adverse Impacts (SAIs) on vulnerable marine ecosystems (VMEs) within SIOFA area*" of project PAE2021-01.

2. Introduction

The main objectives of the management of deep-sea fisheries are to promote responsible fisheries that provide economic opportunities while ensuring the conservation of marine living resources and the protection of marine biodiversity, by ensuring the long-term conservation and sustainable use of marine living resources in the deep seas; and preventing significant adverse impacts on VMEs (Paragraph 11 of the FAO Guidelines; FAO, 2009).

Significant adverse impacts are those that compromise ecosystem integrity (i.e., ecosystem structure or function) in a manner that: (i) impairs the ability of affected populations to replace themselves; (ii) degrades the long-term natural productivity of habitats; or (iii) causes, on more than a temporary basis, significant loss of species richness, habitat or community types. Impacts should be evaluated individually, in combination, and cumulatively (Paragraph 17 of the FAO Guidelines; FAO, 2009).

FAO Guidelines indicate that when determining the scale and significance of an impact, there are six factors to be considered (FAO, 2009):

(i) the intensity or severity of the impact at the specific site being affected;
(ii) the spatial extent of the impact relative to the availability of the habitat type affected;
(iii) the sensitivity/vulnerability of the ecosystem to the impact;
(iv) the ability of an ecosystem to recover from harm, and the rate of such recovery;
(v) the extent to which ecosystem functions may be altered by the impact; and
(vi) the timing and duration of the impact relative to the period in which a species needs the habitat during one or more of its life-history stages.

In this study we focus on how the fishing footprint of SIOFA's fisheries potentially overlaps with the distribution of predicted bioregions of VME indicator taxa in the Convention Area to assess potential significant impacts. The present work builds on a bioregionalization scheme developed during project PAE2020-02, where three modelling techniques were used to develop three classifications (Ramiro-Sánchez & Leroy, 2022). We base the present study on the bioregionalization method "group first, then predict", which yielded the more robust results (see Ramiro-Sánchez & Leroy, 2022 for a detailed discussion of the methods). Furthermore, we investigate the intensity of the aggregated fishing effort (as number of fishing events) across bioregions and develop a fishing intensity impact index to quantify how each fishing gear type (trawling, gillnets, lines, and traps) may impact the different bioregions existing within SIOFA. The index takes into account the intensity of fishing in relation to the predicted probabilities of occurrence of each bioregion.

3. Methods

3.1 Data

Fishing data

Fishing effort was provided as aggregated number of fishing events and as tonnes of fish catch per grid cell (pixel) for bottom-trawling, gillnets, lines, and traps fishing. The spatial resolution of the layers was 10 arc min (~ 17.5 km at the equator).

Bioregions

Project PAE2020-02 developed three bioregionalization schemes based on VME indicator taxa for the Southern Indian Ocean (Ramiro-Sánchez & Leroy, 2022). Out of the three schemes, we exclusively use the predicted distributions of the bioregions produced with the "group first, then predict" modelling approach, deemed as the most robust in terms of results. A detailed description of the three modelling methods, advantages, and disadvantages were presented in the corresponding report[1] for the consultancy (Ramiro-Sánchez & Leroy, 2022). For context, we include here in Appendix A a brief overview of the methods of the "group first, then predict" approach, the interpretation of the bioregions, and the combined spatial predictions.

We recovered the individual predictions of the bioregions at the first hierarchical level (hereafter "fist-level bioregions") and at the second hierarchical level (hereafter "nested subregions") resulting from the bioregionalization analysis. Figure 1 shows the predictions of the first-level bioregions and Figure 2 shows the predictions of the nested subregions.

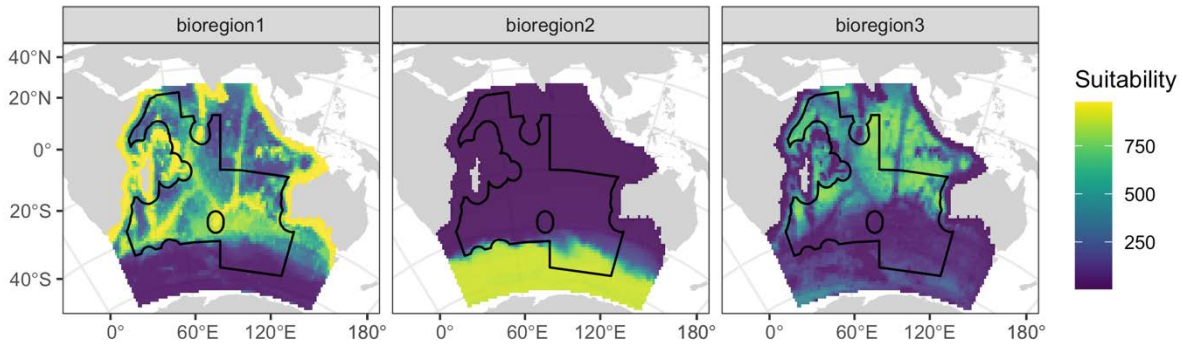

Figure 1. Predictive suitability models for the three bioregions detected at the first hierarchical level of the bioregionalization analysis "group first, then predict". Suitability values range from 0 (unsuitable environments) to 1000 (highly suitable environments). A description of each bioregion is included in Appendix A.

[1] https://siofa.org/sites/default/files/files/VMEMapping_FullReport.pdf

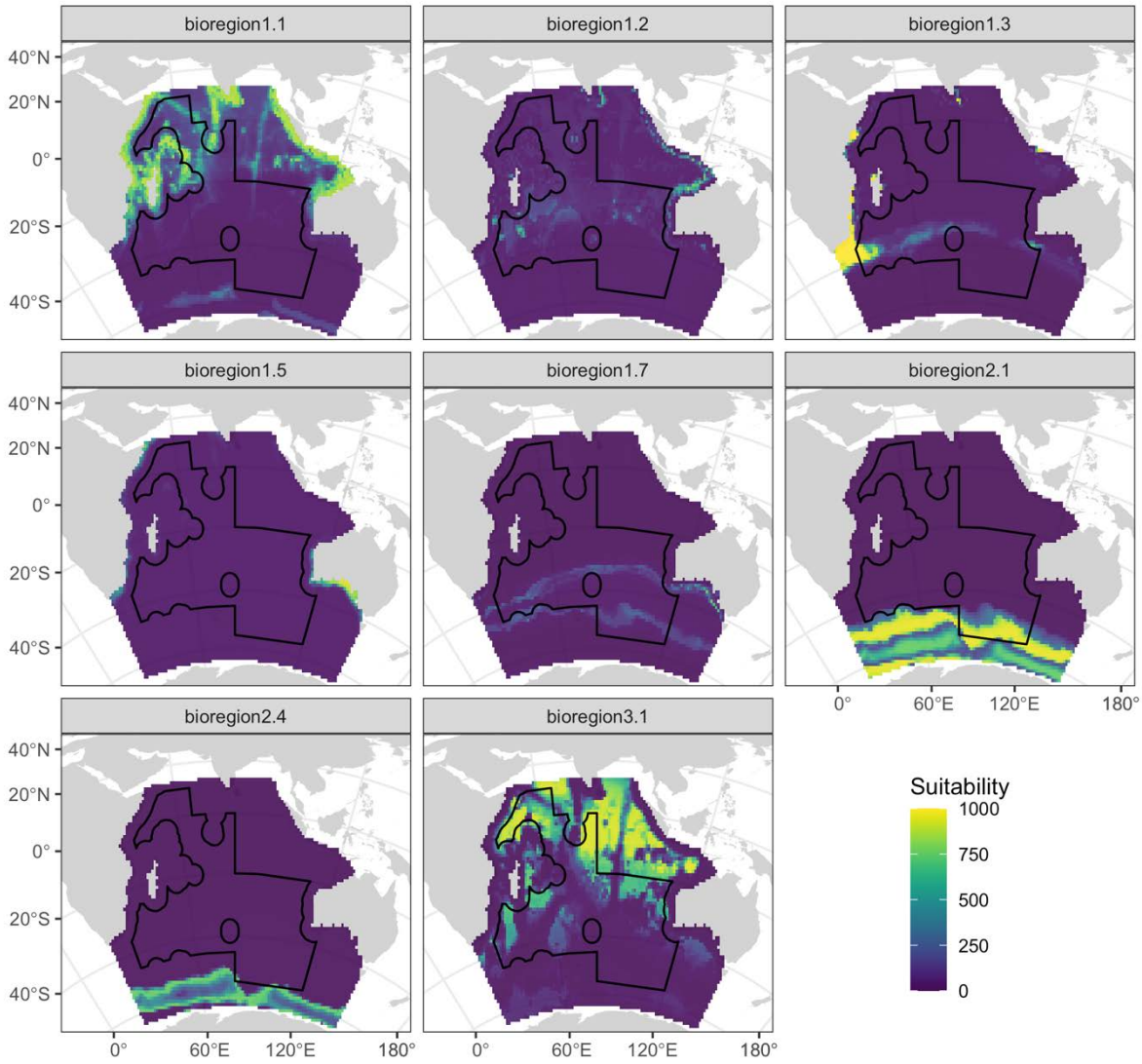

Figure 2. Predictive suitability models for the eight nested subregions detected at the second hierarchical level of the bioregionalization analysis "group first, then predict". Suitability values range from 0 (unsuitable environments) to 1000 (highly suitable environments). A description of each subregion is included in Appendix A.

3.2 Assessment of potential SAIs on bioregions of VME indicator taxa

In order to investigate how the existing fishing footprint could be potentially impacting bioregions based on VME indicator taxa, we focused on understanding potential impacts in three ways:

1) First, we focused on how the fishing footprint (split by gear type) is distributed over bioregions, i.e., over which bioregions each fishing type occurs.

2) Second, we focused on the fishing intensity across bioregions, i.e., how the number of fishing events is distributed across the probability classes of the bioregions.

3) We integrated fishing intensity and where it occurs (over which bioregion probability classes) in a "fishing intensity impact index" to assess how potentially impacted is each bioregion.

In the next sections we elaborate on the methods for each of the three objectives.

3.2.1 Spatial distribution and intensity of fishing in SIOFA's bioregions

We recovered the distribution of suitability scores for each of the main three bioregions (Figure 1) and the eight nested subregions (Figure 2) relative to the total suitability scores. We preferred to consider the total suitability scores rather than values beyond a given threshold of suitability because the latter suggests binary transformations, which are an oversimplification of the biological reality and generally misleading (Muscatello et al., 2021). We resampled the predictive maps from 1° to 10 arc min - the same resolution as the aggregated fishing effort.

Predicted bioregion suitability values, which ranged from 0 to 1000, were next classified in ten intervals by 100 suitability increases. Fishing effort for each gear type, expressed as aggregated number of fishing events per pixel, was classified based on 10% quantiles. Next, we extracted the suitability of each bioregion at the centroid coordinates of the fishing effort pixel for each gear type. We then calculated the frequency of the number of fishing events per 10% quantile. Alluvial plots were created with this summary information for each bioregion and gear type.

3.2.3 Fishing intensity impact index

To quantify the difference in fishing intensity across bioregions and gear type, we developed a *fishing intensity impact index* that weighs the intensity of fishing (as aggregated number of fishing events) according to the suitability class where it occurs and the total number of aggregated fishing events occurring per suitability class.

The overarching objective of this index is to summarize how each bioregion is impacted by fishing. We designed an index of impact such that regions that are fished with high intensity (i.e., large number of fishing events), especially in their areas predicted with highest suitability, will be considered as heavily impacted; whereas a low impact will be attributed to regions with lower intensity of fishing, or which are only fished in areas of low suitability. Therefore, the index has two main components and is thus calculated in two steps. First, we estimate the intensity of fishing on a per-pixel basis. This step is done by assigning weights to each pixel based on the fishing intensity. Second, the impact index is calculated for the entire region by combining fishing impact weights and bioregion suitability values for all pixels of the regions, such that pixels with a high suitability weigh more in the final index than pixels with a low suitability.

3.2.3.1 Step 1: Calculating fishing impact weights for each pixel

Because the distribution of fishing impact is highly skewed (i.e., a low number of pixels have extremely high values), it cannot be linearly included in an index, otherwise the index values would be highly skewed as well. To address this issue, we calculated fishing impact weights that are based on classes of fishing efforts. To define classes of fishing effort, we used 10% quantiles. In other words, the first class is composed of the 10% cells with the lowest fishing effort; the second class is composed of the next 10% cells with lowest fishing effort, and so on until the last class, which is composed of the 10% cells with the highest fishing effort. We assigned a rank to each class, starting from rank 1 which is always the class with no fishing effort, and followed by the classes of fishing effort until the maximum rank, which corresponds to a class with maximum fishing. The number of ranks varies depending on the distribution of fishing effort, because multiple classes can be merged together if they share the same amount of fishing events. For example, if the 20% of cells with the lowest fishing effort all have only one fishing event, then the 0-10% and 10-20% will be merged in a single class. Regardless of the number of ranks, the rank will always start at 1 (no fishing). To account for the variable number of ranks, our weighting scheme is relative to the maximum rank, such that the weight ranges from 0 (the pixel is not fished at all) to 1 (the pixel is fished with maximum intensity):

$$weight = \frac{1-\frac{1}{rank^{y}}}{1-\frac{1}{\max(rank)^{y}}} \qquad \textit{Equation 1}$$

The exponent *y* offers the user flexibility to assign different importance to the ranks: the exponent affects how rapidly and steep is in the change in weight relative to the rank. It can take any value above 0. Values of *y* above 1 will decrease the differences among classes, as most classes of fishing impact will have a high weight (Figure 3). Values of *y* below 1 will increase the differences among classes, until it becomes almost linearly proportional to the ranks of the classes (Figure 3). In other words, lower ranks will have more or less importance (i.e., weight) as higher ranks depending on the value of *y*. In this investigation, we used a value of *y* = 1 as a starting case to compute the fishing impact index for each bioregion.

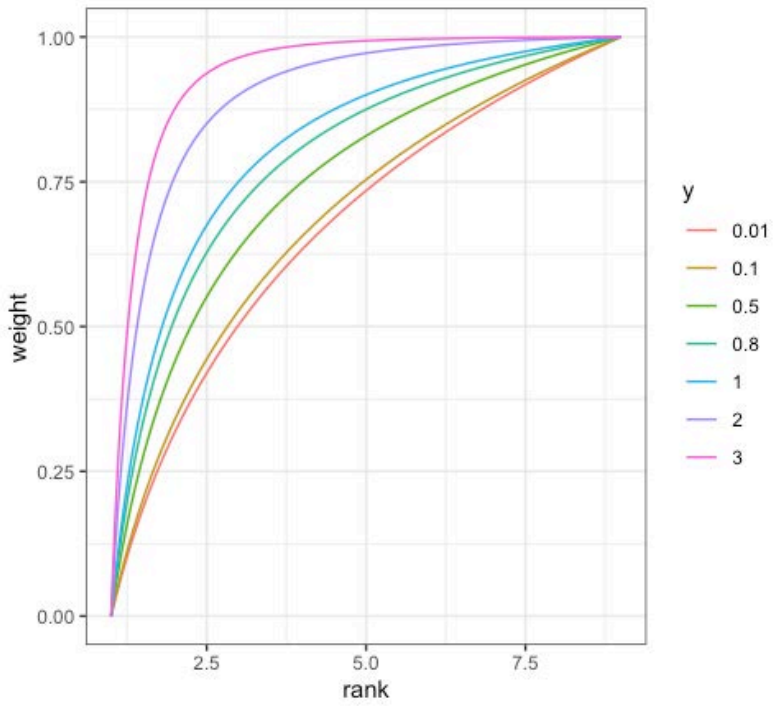

Figure 3. Changes in weight as a function of fishing effort ranks. The parameter y (Equation 1) influences the rate at which higher weight values are obtained for a given rank (derived from classifying fishing effort in 10% quantiles). For y < 1, weights are almost linearly proportional to the ranks; for y > 1 differences decrease among ranks.

3.2.3.2 Step 2. Calculating the impact of fishing per bioregion

Once all weights have been calculated for all pixels, we compute the impact index for each bioregion. To estimate how much a region is impacted by fishing, the basic calculation consists of summing the fishing impact weights of all pixels of the region. However, because we want to distinguish between high suitability areas and low suitability areas, we ponder the fishing impact of each pixel by its suitability, such that pixels with highest suitability weigh more in the final index than pixels with lowest suitability.

Suitability values range from 0 to 1000, which we split into 10 intervals of 100 increments (0-100, 100-200, …, 900-1000). For the lowest class (0-100) we defined a suitability weight of 0, because the number of pixels with a low suitability is expected to be disproportionately high for many regions. In other words, pixels which are unsuitable for a given region do not weigh in the index. For all classes between 100 and 1000, we used the upper boundary as the suitability weight divided by 1000, e.g., for the suitability class (100,200] the weight is 0.2, whereas for the maximum suitability class (900, 1000], the weight is 1. Thus, the weight is proportional to the suitability class where fishing is present. The *fishing intensity impact index* results as follows:

$$Impact\ index = \frac{\sum_j \frac{\sum_i w_{ij}}{N_j} \times H_j}{\sum_j H_j} \qquad \textit{Equation 2}$$

where *i* refers to pixel and *j* to habitat suitability class (H_j), w_{ij} is the weight of a pixel for a given suitability class, and N_j is the to total number of pixels of a given suitability class. The division of the sum of fishing impact weights by the total number of pixels for a given habitat suitability class ensures that the index is not disproportionately affected by classes with very large number of pixels– it works based on the proportion of fished pixels in each suitability class.

The impact index theoretically ranges from 0 (none of the pixels of a bioregion are fished) to 1 (all pixels of a bioregion are fished with the maximum intensity of impact). However, in practice, the actual values of the impact index will be very small, because the proportion of non-fished pixels is enormous relative to the number of fished pixels (i.e., fishing occupies a small area within SIOFA relative to the overall SIOFA's area of competence). For reference, we provide the proportion of fished pixels relative to the total number of pixels in a suitability class per bioregion and subregion in Table B1 and Table B2 of Appendix B, respectively. Thus, the purpose of the index is not to be interpreted as an absolute value – but rather to serve as an index of the evolution of fishing practices in the SIOFA area and explore which bioregions are the most affected by fishing impacts, according to gear type. For example, the index could be used to analyse the effect of fishing management measures, such as how limiting or fostering fishing in a particular area would affect the index of fishing impact for each bioregion.

All analyses were conducted in the R statistical software (R Core Team, 2021).

4. Results

4.1 Spatial distribution of fishing within bioregions

Bottom-trawling, gillnets, and line fishing activities occurred across all three first-level bioregions. Trawling dominated in bioregion 1 and bioregion 3 (Figure 4); fishing with gillnets dominated in bioregion 1 (Figure 5); and fishing with lines was predominant in bioregion 1 and bioregion 2 (Figure 6). In contrast, fishing with traps only took place in bioregion 1 and bioregion 3 (Figure 7).

Inspecting the aggregated fishing effort in the nested subregions indicated that trawling occurred across all suitability areas in all subregions, except for subregion 2.1 and subregion 2.4 (Figure 8). For subregions 1.1, 1.3, and 3.1, trawling occurred across all suitability classes: in low (< 300), moderate (between 300 and 700) and high suitability areas (> 700). Trawling occurred in areas of low suitability (< 300) of subregion 1.7, and in areas of moderate suitability (between 300 and 700) of subregion 1.2.

Fishing with gillnets occurred only in subregion 1.2 in low (< 300) and in moderate suitability areas (between 300 and 700) (Figure 9).

Fishing with lines occurred in moderate (between 300 and 700) and high suitability areas (> 700) of subregion 1.1; in low suitability areas (< 300) of subregion 1.3 and subregion 1.7; in low (< 300), moderate (between 300 and 700) and high suitability areas (> 700) of subregion 2.1; and in low (< 300), moderate (between 300 and 700) suitability areas of subregion 2.4 (Figure 10).

Fishing with traps occurred only in subregion 1.1 in low (<300) and high (> 700) suitability areas, and in subregion 3.1 in moderate (between 300 to 700) suitability areas (Figure 11).

We have included figures for subregion 1.5 in the report for reference but note that this subregion is not predicted within SIOFA (see Figure 2), therefore analyses do not apply to this subregion.

4.2 Fishing intensity within bioregions

First-level bioregions

Trawling

Fishing intensity ranged from 2 to 2550 aggregated number of events per pixel. The number of 1 and 2 fishing events are overrepresented, such that their classes cover 30% to 40 % of the pixels (Figure 4).

Bottom-trawl fishing occurred across all suitability classes of bioregion 1, but 75% of the fishing events occurred in the very high suitability classes (those with suitability > 700) (Figure 4). Fishing events in the top 90% quantile of the distribution (from 55 to 2550 events per pixel) also occurred more frequently in the higher suitability class. In Bioregion 2, virtually all fishing events occurred in areas predicted as low suitability for the bioregion (suitability class 0-100). Bioregion 3 showed an inverse trend to that of bioregion 1: most fishing occurred in areas of low suitability (0-100) (75% of the data); however, it also occurred in areas of high suitability, including fishing events in the top 10% most suitable cells.

Gillnets

Fishing intensity ranged from 3 to 131 aggregated number of events per pixel (Figure 5).

All of the aggregated gillnet fishing events occurred in areas of very high suitability (> 800) for bioregion 1, whereas for bioregion 2 and bioregion 3 these occurred in areas of very low suitability.

Lines

Fishing intensity ranged from 2 to 1089 aggregated fishing events per pixel (Figure 6).

For bioregion 1, about 50% of the line fishing events occurred in areas of very high suitability (> 900) and the other 50% across the remaining suitability classes, with each fishing intensity quantile class equally distributed among the suitability classes. Bioregion 2 displayed an

inverse pattern: 50% of the fishing events across all fishing intensities occurred in areas of low suitability for bioregion 2 (< 100) and the other 50% was distributed across the remainder of the suitability classes, with in total more than a quarter of fishing events located in high suitability areas (> 700). Bioregion 3 had all fishing events classes occurring within low suitability classes (< 300).

Traps

Fishing intensity ranged from 1 to 57 aggregated events, which occurred very infrequently (Figure 7).

All trap fishing activity in bioregion 1 occurred in areas of medium (500-600) and high suitability (> 900). Bioregion 2, in contrast, had all activity occurring in areas of very low suitability (< 100). Bioregion 3 had trap activity in areas of low (< 200) and medium (400-600) suitability.

Second-level bioregions

Trawling

In subregion 1.1 there was fishing activity across all suitability classes (Figure 8). Approximately 90% of the fishing events occurred in areas of very low suitability (< 100). We note that the highest fishing intensity classes (from 16 to 2550 events) also occurred in high (> 800) suitability classes.

Subregion 1.2 presented fishing activity in low to moderate suitability classes (100 – 600), although most the events occurred in the low suitability areas (Figure 8). The highest fishing intensity events occurred also mostly in the low suitability areas.

Bottom-trawling fishing occurred all suitability classes of subregion 1.3, but 75% of them occurred in the low class (< 100) (Figure 8). Very few events across all types of intensity occurred in the higher suitability classes.

In subregion 1.7, the lower suitability classes (< 300) presented all fishing activity, with the fishing intensity event classes similarly occurring across the three suitability classes (Figure 8).

Bottom-trawling in subregion 2.1 and subregion 2.4 occurred only in very low suitability classes (< 100) (Figure 8).

About 90% of the bottom-trawling activity in subregion 3.1 occurred in the very low suitability class (< 100), however, fishing occurred across all suitability types, with the 10% of the remaining events happening in the rest of suitability classes equally (Figure 8).

Gillnets

Fishing with gillnets occurred only in areas of very low suitability (< 100) in all subregions except for subregion 1.2 (Figure 9). In subregion 1.2, fishing occurred in low (0 to 300) to

moderate suitability areas (300 - 600), with more prevalence in the latter. Fishing intensity event classes were equally distributed across the moderate suitability classes.

Lines

Fishing with lines occurred across all subregions, except for subregion 1.3, where it occurred in areas of low suitability (< 200), and for subregion 3.1 where it only occurred in areas of very low (< 100) suitability class (Figure 10).

In subregion 1.1, about 75% of the fishing events occurred in very low suitability areas and the remaining 25% was distributed across areas of moderate to high suitability (500 -900) (Figure 10). All fishing intensity classes were present in the high suitability class.

In subregion 1.2, fishing was prevalent across low to moderate suitability areas, with 60% of the events occurring in the low suitability areas (Figure 10). Notably, the top 10% of highest fishing intensity events occurred almost entirely in the lowest suitability class, while the top 20% occur in the highest suitability class (moderate, in this case).

Line fishing in subregion 1.7 occurred across low suitability areas (100 - 300) and all fishing intensity classes distributed across them (Figure 10).

Subregion 2.1 had fishing across all suitability classes, with approximately 90% of the events occurring in the very low (< 100) and low (200-300) suitability class (Figure 10). The remaining 10% of fishing events were distributed across the higher suitability areas. All fishing intensity events were present in all the suitability areas.

Subregion 2.4 presented a similar trend to subregion 2.1. About 90% of the events occurred in areas of very low suitability (< 100) and the remaining 10% in areas of low to moderate suitability (300-600) (Figure 10). The different fishing intensity classes were present in all predicted suitability areas.

Traps

Fishing with traps occurred only in subregion 1.1 and subregion 3.1 (Figure 11) – for all other regions it happened in very low suitability areas (0-100). In subregion 1.1, the lowest fishing intensity events occurred in areas of very low suitability (< 300) whereas pixels fished at the highest fishing intensity occurred in the areas of high suitability (> 800) for subregion 1.1. In contrast, in subregion 3.1, fishing at lowest intensity occurred in moderate suitability (< 500) areas.

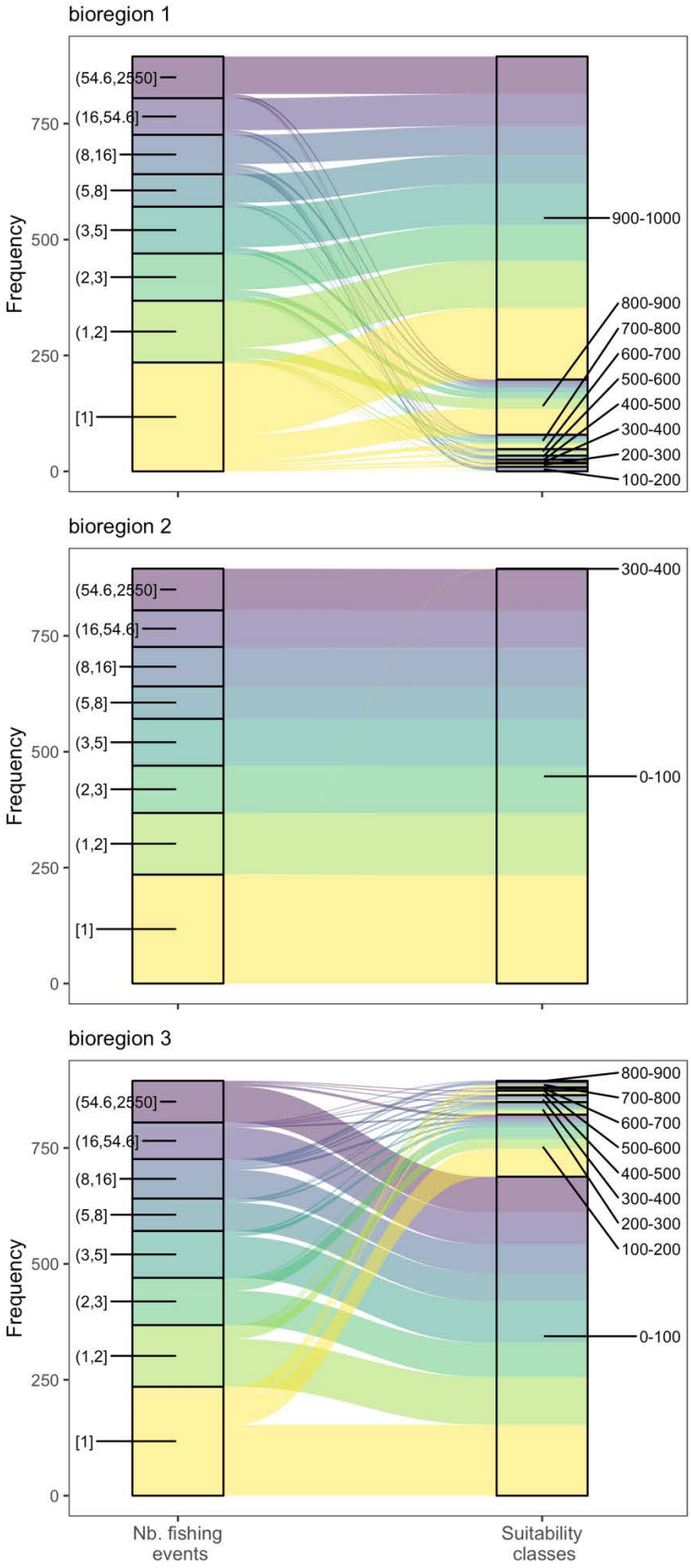

Figure 4. Alluvial diagram depicting the distribution of bottom-trawling fishing effort (as aggregated number of fishing events) across first-level bioregions. For each bioregion, the diagram also shows how the frequency of fishing events, classed in 10% quantiles, is distributed across the predicted habitat suitability classes.

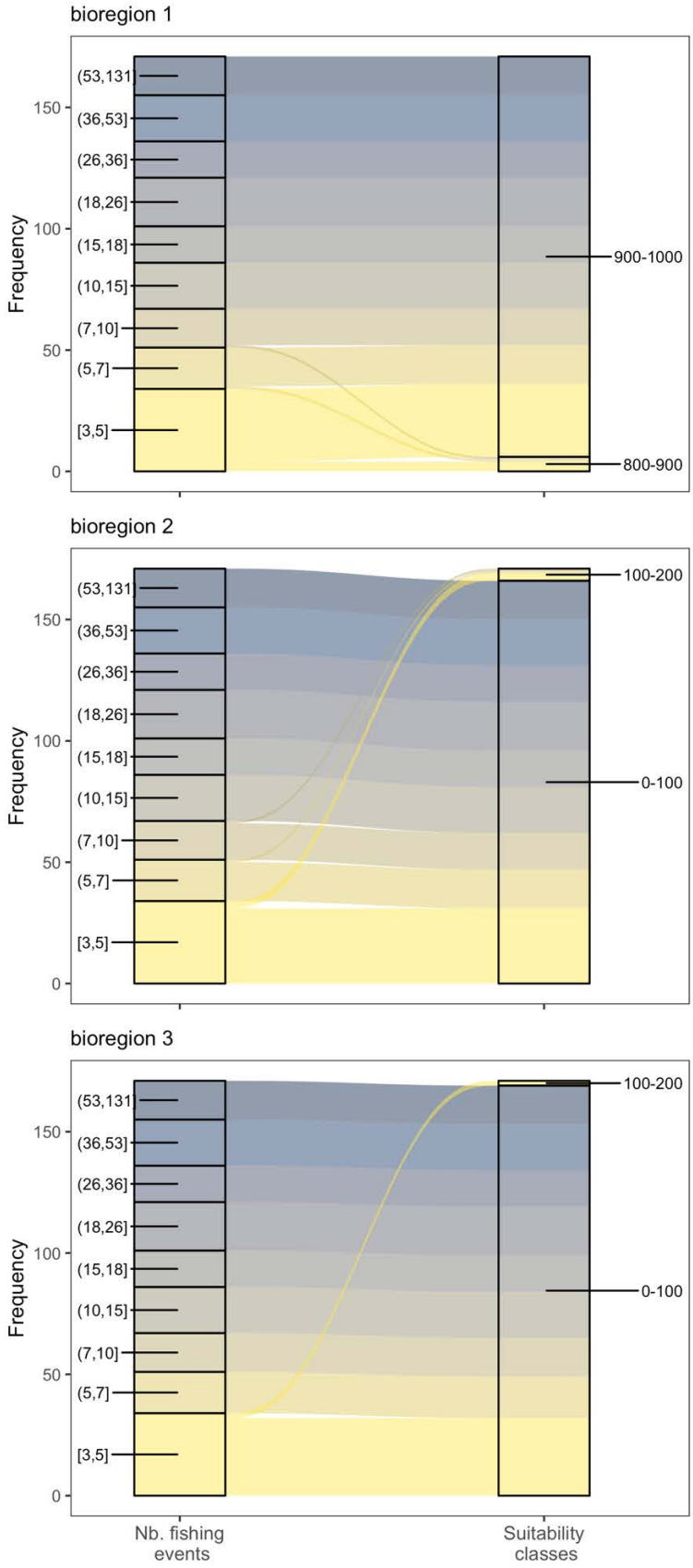

Figure 5. Alluvial diagram depicting the distribution of gillnets fishing effort (as aggregated number of fishing events) across first-level bioregions. For each bioregion, the diagram also shows how the frequency of fishing events, classed in 10% quantiles, is distributed across the predicted habitat suitability classes.

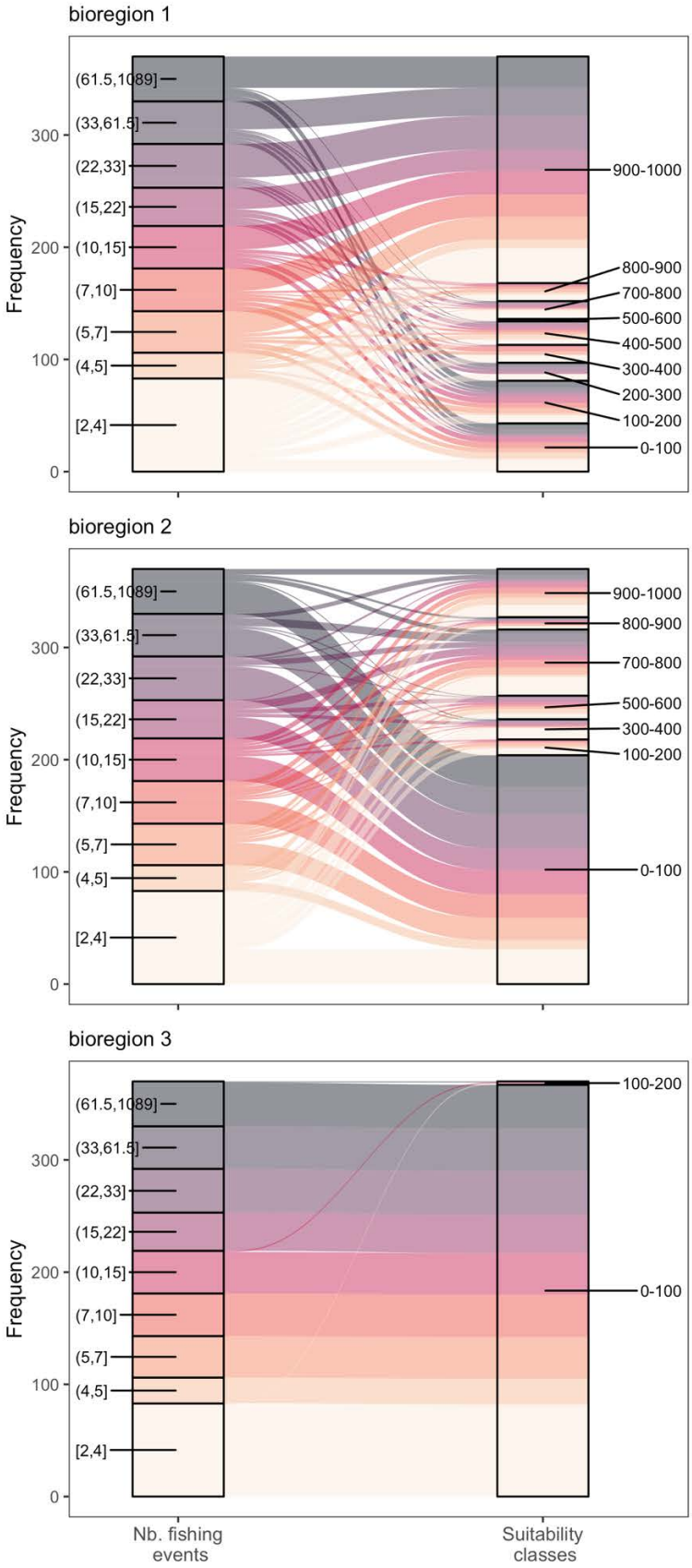

Figure 6. Alluvial diagram depicting the distribution of line fishing effort (as aggregated number of fishing events) across first-level bioregions. For each bioregion, the diagram also shows how the frequency of fishing events, classed in 10% quantiles, is distributed across the predicted habitat suitability classes.

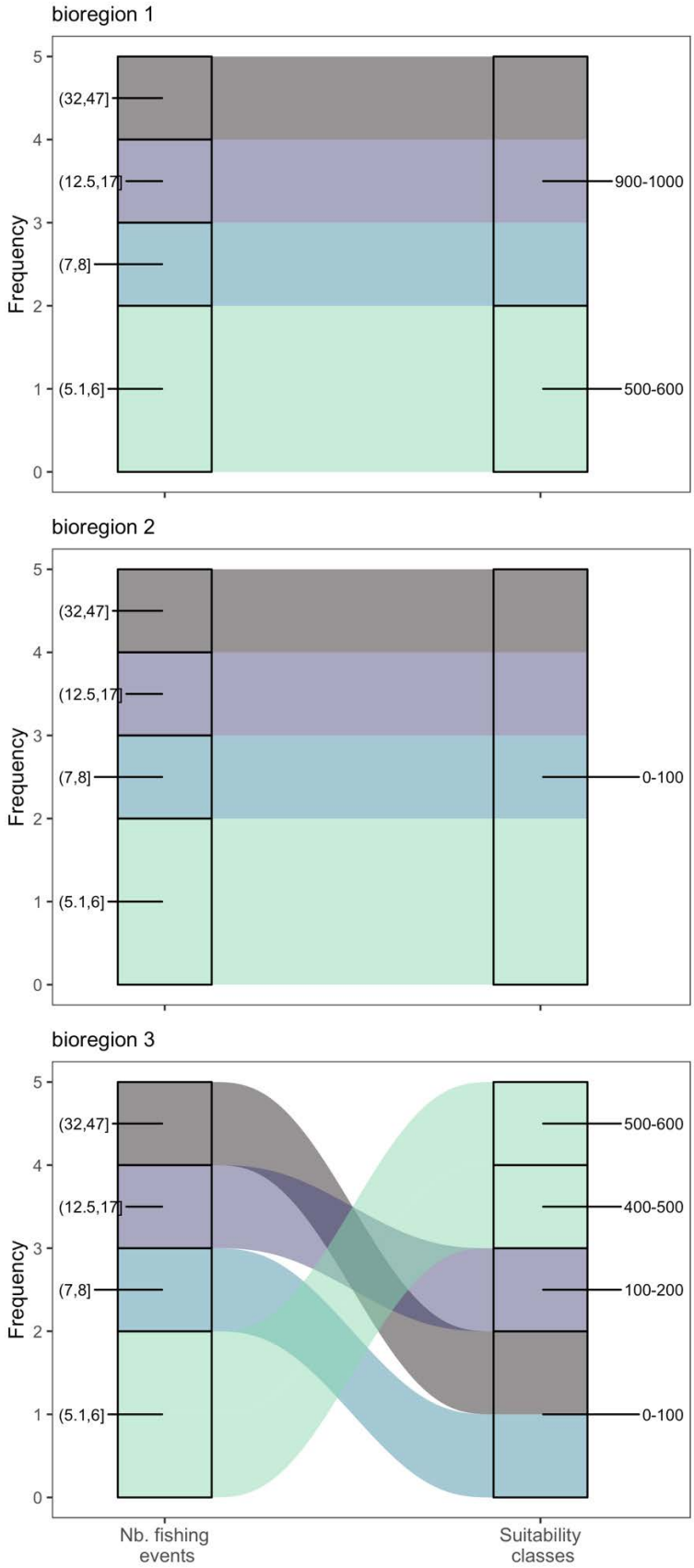

Figure 7. Alluvial diagram depicting the distribution of traps fishing effort (as aggregated number of fishing events) across first-level bioregions. For each bioregion, the diagram also shows how the frequency of fishing events, classed in 10% quantiles, is distributed across the predicted habitat suitability classes.

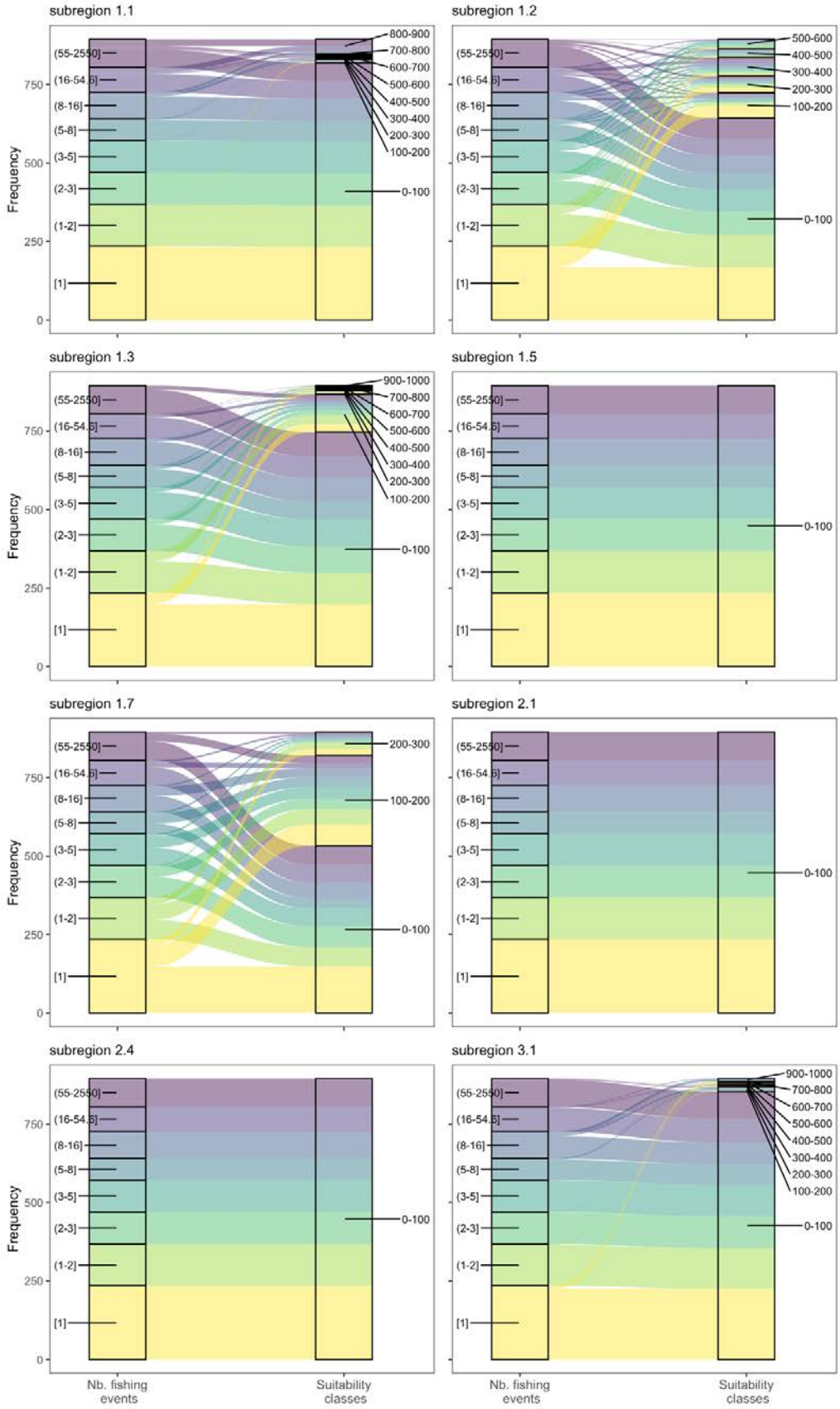

Figure 8. Alluvial diagram depicting the distribution of bottom-trawling fishing effort (as aggregated number of fishing events) across second-level nested subregions. For each bioregion, the diagram also shows how the frequency of fishing events, classed in 10% quantiles, is distributed across the predicted habitat suitability classes.

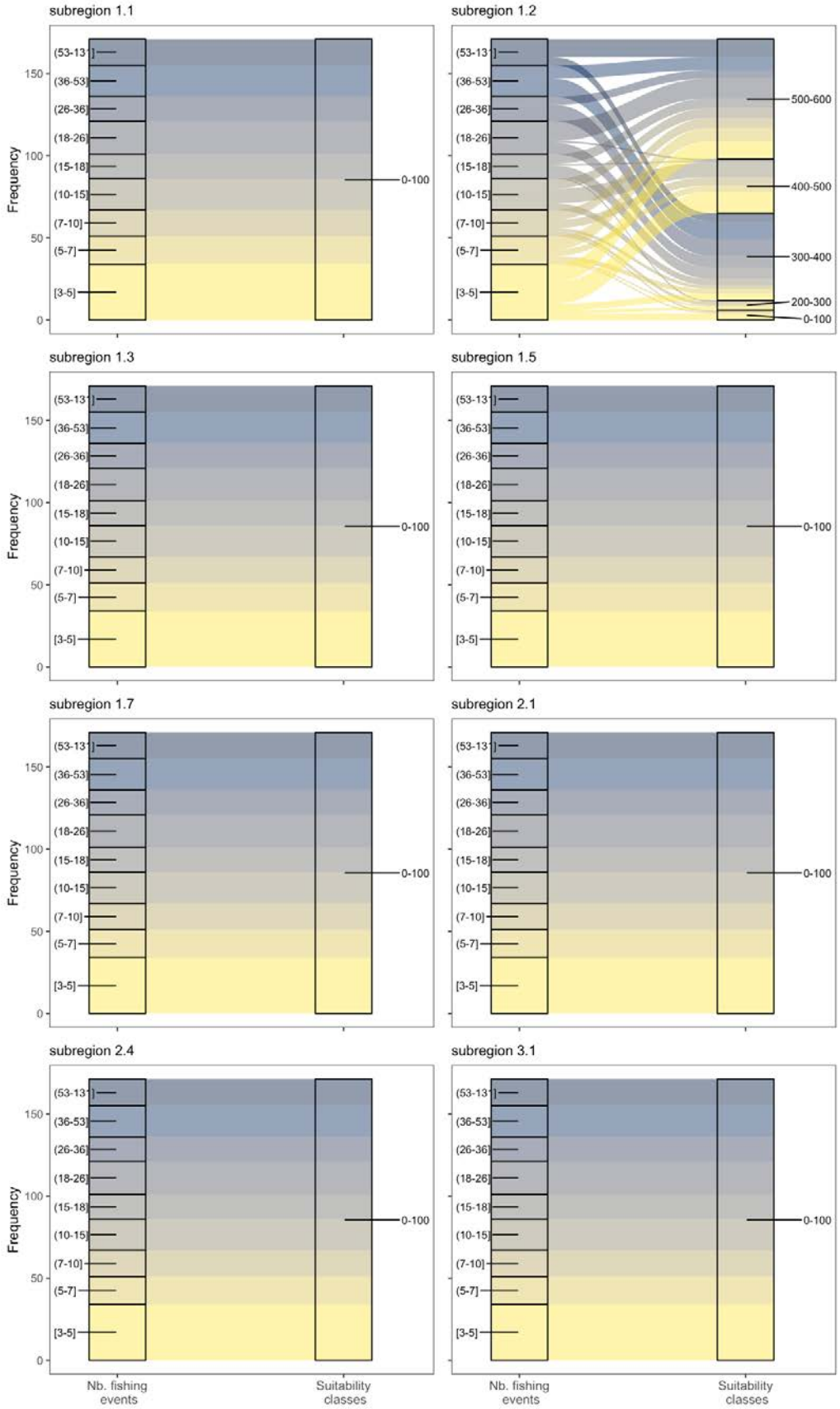

Figure 9. Alluvial diagram depicting the distribution of gillnets fishing effort (as aggregated number of fishing events) across second-level nested subregions. For each bioregion, the diagram also shows how the frequency of fishing events, classed in 10% quantiles, is distributed across the predicted habitat suitability classes.

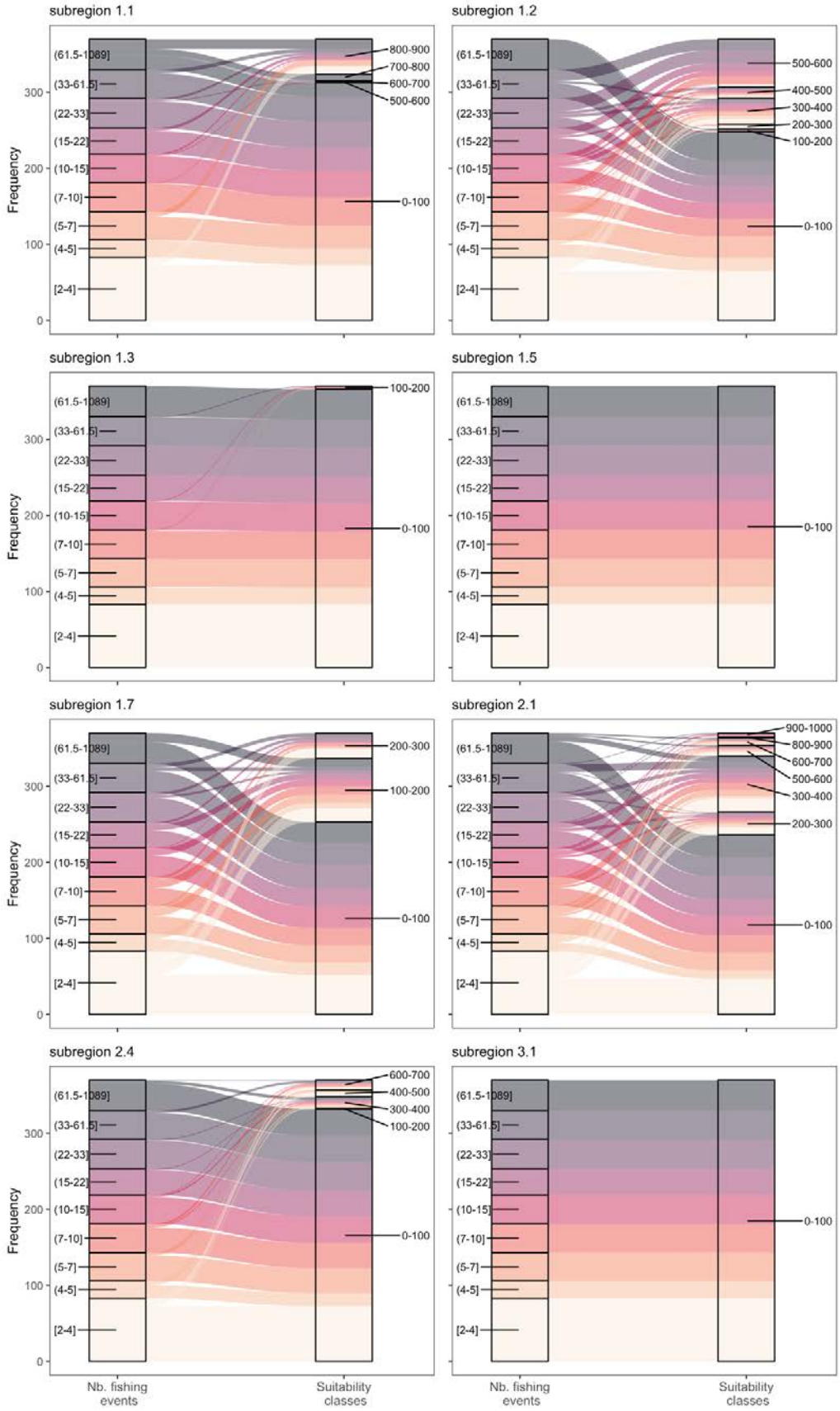

Figure 10. Alluvial diagram depicting the distribution of line fishing effort (as aggregated number of fishing events) across second-level nested subregions. For each bioregion, the diagram also shows how the frequency of fishing events, classed in 10% quantiles, is distributed across the predicted habitat suitability classes.

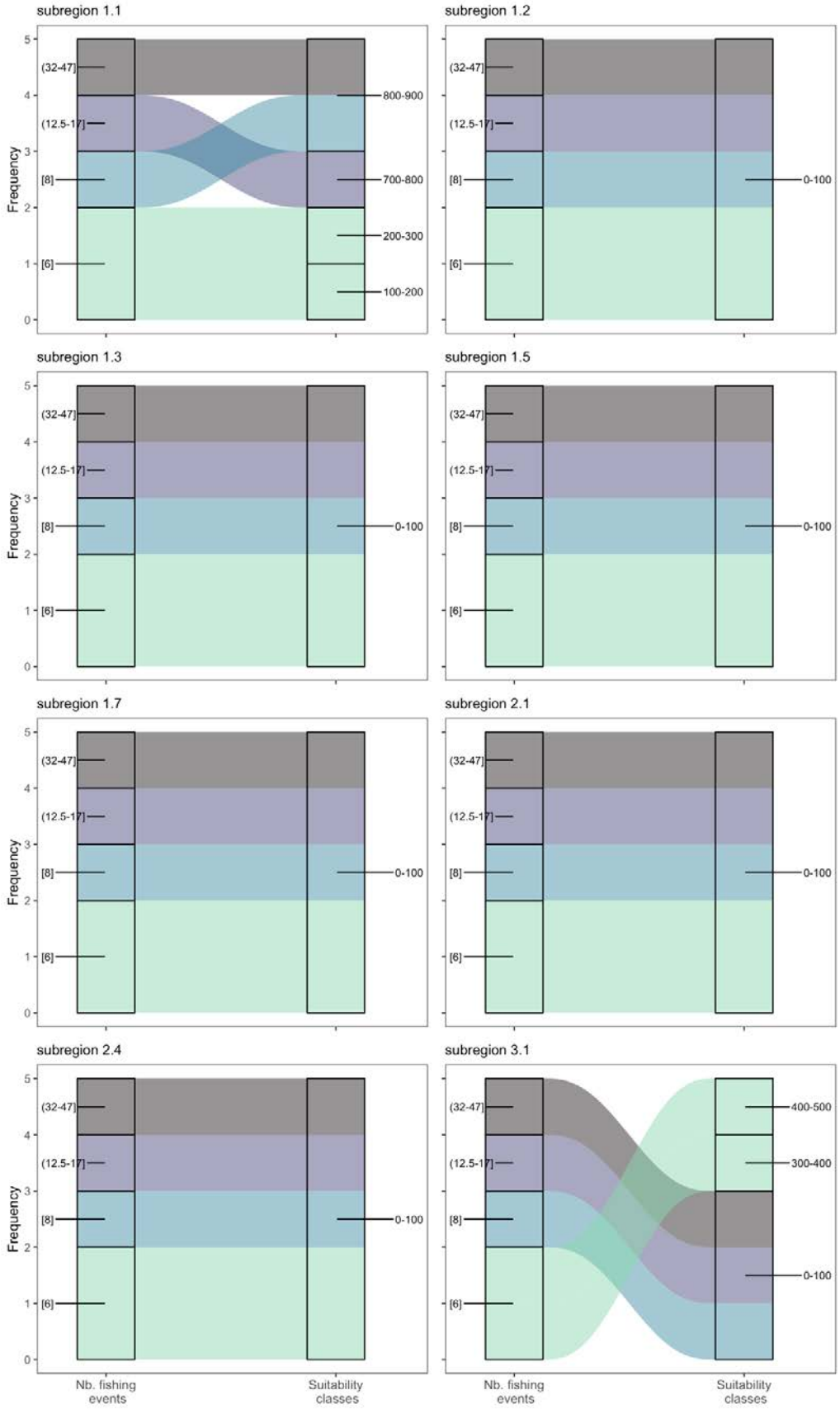

Figure 11. Alluvial diagram depicting the distribution of traps fishing effort (as aggregated number of fishing events) across second-level nested subregions. For each bioregion, the diagram also shows how the frequency of fishing events, classed in 10% quantiles, is distributed across the predicted habitat suitability classes.

4.3 Fishing intensity impact index

The *fishing intensity impact index* measures how bioregions are affected by fishing by accounting for two parameters: first, the fishing intensity (the higher the fishing intensity, the higher the index); second, the location of fishing (the more fishing occurs in high suitability areas, the higher the index). Therefore, values of the index will increase for a bioregion if fishing intensity is high in areas of high suitability.

The index is relative to the total surface area of each bioregion, hence the fishing intensity impact index is expected to have very low values because fishing is expected to only occur in a small proportion of the area, but also because the intensity of fishing is not expected to be maximal everywhere. To grasp the meaning of the index, we invite readers to compare its values with the figures above. For example, values of the index in Table 1, column Trawling, can be compared with Figure 4: it is obvious from Figure 4 that bioregion 1 is the most impacted of the three regions, which is reflected in the index. Note however that, as explained in the methods, the lowest suitability class (0-100) has a suitability weight of 0 in the index, and so it does not affect the index: index values are only affected by suitability classes > 100. In addition, unfished cells do not appear in Figures 4-11, but do weigh in the calculation of the index: the higher the number of unfished cells, the lower the index value (for the proportion of fished pixels relative to total number of pixels in a suitability class per bioregion and subregion, see Table B1 and Table B2 in Appendix B, respectively).

The *fishing intensity impact index* revealed differences between the three first-level bioregions per gear type and among gear types (Table 1). For bottom trawling, bioregion 1 was the most intensely fished of the three first-level bioregions, 14 times more intensely fished than bioregion 3 for instance, and with bioregion 2 the least impacted. For gillnets, bioregion 1 was again the most impacted, followed by bioregion 2 and, lastly, bioregion 3. In contrast, for line fishing, it was bioregion 2 the most impacted, twice impacted than bioregion 1, while bioregion 3 had a negligible impact in comparison. Finally, the risk of impact for trap fishing was generally low, null for bioregion 2 and higher for bioregion 1 than bioregion 3.

Table 1. Fishing intensity impact index for first-level bioregions at the of the bioregionalization analysis.

	Trawling	**Gillnets**	**Lines**	**Traps**
Bioregion 1	0.02066048	0.00488618	0.00701453	0.00012608
Bioregion 2	0.0000280	0.0000266	0.01242043	0
Bioregion 3	0.00144462	0.00000665	0.0000110	0.0000308

The impact assessment of the subregions revealed that for trawling, subregion 1.2 was the most impacted from fishing, followed by subregion 1.1 and subregion 1.7 (Table 2). There was minimal impact for subregion 1.3 and 3.1, and no risk for subregions 2.1 and 2.4. For gillnets, it was subregion 1.2 the most impacted by this activity, whereas for the remaining bioregions the impact was minimal in comparison. The impact of line fishing was evident across all

subregions except for subregion 3.1 and minimal for subregion 1.3. The most impacted was subregion 1.2, which was twice more impacted than subregion 2.4. Subregion 1.1, subregion 1.7 and subregion 2.4 were similarly impacted, but the impact was higher in subregion 1.1. Finally, the impact of fishing with traps was generally very low and negligible for all subregions.

Table 2. Fishing intensity impact index for the second-level subregions of the bioregionalization analysis.

	Trawling	Gillnets	Lines	Traps
Subregion 1.1	0.04230372	0	0.04303634	0.00248335
Subregion 1.2	0.17426158	0.27705476	0.2186114	0
Subregion 1.3	0.00563023	0	0.0000291	0
Subregion 1.5*	---	---	---	---
Subregion 1.7	0.03609041	0	0.01664393	0.0000237
Subregion 2.1	0	0	0.01066662	0
Subregion 2.4	0	0	0.10553208	0
Subregion 3.1	0.00096655	0	0	0.0000530

**Subregion 1.5 is not predicted in SIOFA's Convention Area*

5. Discussion

In this study we investigated potential SAIs on predicted bioregions based on VME indicator taxa for the SIOFA area. We analysed the spatial distribution of fishing in relation to the bioregion predictions, the intensity of fishing, and summarised both sources of information into an index of fishing intensity impact to quantify the vulnerability of the bioregions against each of the four fishing gear types. The overall picture indicates that fishing activities in SIOFA occur across all three bioregions, although fishing types are spatially distributed and take place differently across subregions. While trawling, gillnet and trap fishing are only present in bioregion 1 and bioregion 3, only line fishing is present in bioregion 2. The intensity of fishing is also variable per gear type, but we found that, in many subregions, areas with maximum fishing intensity coincided with areas of high suitability for subregions. Our analyses suggest that bioregion 1 is the most impacted, particularly by trawling, gillnets, and traps fishing. Subregion 1.2 is the most impacted by bottom trawling, gillnet, and line fishing, while subregion 1.1 is the most at risk by traps.

Spatial distribution and intensity of fishing in SIOFA

SIOFA's report[2] on the overview of its fisheries state the occurrence of fishing activities up to ~ 2000 m water depth. Our first-level bioregions represent the main depth and climatic zones

[2] https://siofa.org/sites/default/files/files/Overview%20of%20SIOFA%20Fisheries%202022_0.pdf

found in SIOFA (Ramiro-Sánchez & Leroy, 2022), where bioregion 1 dominates the shelf and bathyal depths (Figure A2, Annex A). Accordingly, fishing activities were found to be more prevalent in this bioregion than in bioregion 3, for instance (see Figure 4 – Figure 7). Bioregion 2 also had a distribution at bathyal depths (Figure A2, Annex A); however, this bioregion characterized the subantarctic zone (Figure A4, Annex A), where different environment characteristics prevail that result in different fish distributions and, as such, in different fisheries. Thus, first-level bioregions offer a complete overview of the distribution of fishing activities in SIOFA, in turn reflecting the different fisheries taking place in each management subarea.

The extent of the spatial distribution of the predicted subregions will determine the impacts of a concentrated fishing footprint and a varying fishing intensity. Subregion 1.1 displayed fishing activity by all gear types except for gillnets, with the highest fishing intensity classes taking place in areas highly suitable for the subregion (first panel in Figure 8 – Figure 11). Subregion 1.1 has a relatively small, predicted area in the northwest of SIOFA's Convention Area (Figure 2) and given its proximity to EEZs, it would be appropriate to consider additional impacts from bordering fishing nations in the area. In addition, a spatial exploration of SIOFA's observers' data (2017 – 2020) revealed large bycatch of sponges in that area (30 - 200 kg), highlighting the potential importance of the subregion for VME indicator taxa.

Subregion 1.2 had all its maximum suitability classes fished by all gears except for traps (second panel in Figure 8 – Figure 11). Notably, it was the top 20% fishing intensity class that always occurred in areas of highest moderate suitability (the maximum suitability for this subregion). This subregion is predicted over plateaus and shallower areas of SIOFA (mostly shelf and upper bathyal depths, 200-800 m; Figure A3, Appendix A), mainly from 20°S to 40°S and with a maximum suitability of ~ 750 (Figure 2), which coincides with SIOFA's fishing footprint at those latitudes. Thus, placed in the context of the maximum suitability predicted for this subregion, our analysis indicates that subregion 1.2 could indeed be heavily fished.

Subregion 1.3 was mainly affected by bottom-trawl fishing (third panel in Figure 8 – Figure 11). Although with very few events, including the highest intensity fishing classes, it took place in areas of high suitability for the subregion. Subregion 1.3 prediction is localised on the southwestern tip of SIOFA and over the Agulhas Plateau, where it shows maximum suitability, covering all bathyal depths (200 – 3500 m). There are some areas with moderate suitability eastwards along the 40°S latitude. Although few high intensity events occur on the areas of maximum suitability, it will require careful examination and interpretation since, similar to subregion 1.1, subregion 1.3's proximity to EEZs may exacerbate potential impacts to the subregion from bordering fishing nations.

Subregion 1.7 was mainly occupied by bottom-trawl and line fishing, but most of the fishing events, including the high intensity classes, occurred in areas predicted as of low suitability (fifth panel in Figure 8 – Figure 11). This subregion had its maximum suitability of prediction in the bathyal depths of the south of Australia. However, due to the strong eddies that spin off in the western south tip of Australia, water masses are carried out westward reaching the western Indian Ocean and potentially favouring the presence of subregion 1.7 along the 35°S - 40°S latitudes. Within SIOFA, it is on the westernmost part of the Southwest Indian Ocean

Ridge where we can find the maximum suitability for this subregion, outside the northeast of Saint Paul and Amsterdam's EEZ.

Subregion 2.1 and subregion 2.4 were solely affected by line fishing but, in both subregions, this activity occupied mostly areas of moderate to high suitability, including one third of the high fishing intensity classes (sixth and seventh panels in Figure 8 – Figure 11). Both subregions have a subantarctic distribution, where subregion 2.1 displays maximum suitability from 45°S to 50°S latitude, and subregion 2.4 from 50°S to 60°S west of 80°E and below 60°S east of 80°E. Within SIOFA, however, predictions are relatively small and localised (Figure 2), especially subregion 2.4, which only has a very small area predicted within SIOFA east of the Kerguelen Plateau. Fishing is also concentrated in that area; therefore, these subregions require careful consideration of protection, particularly so, because of fishing activities occurring within the bordering EEZ.

Finally, subregion 3.1 was mostly impacted by all fishing intensity classes of bottom-trawling, which occurred in areas of moderate to high suitability for the subregion (eighth panel in Figure 8 – Figure 11). This was the most unexpected finding because subregion 3.1 was described as strictly abyssal (Figure A3, Appendix A), and SIOFA's fisheries reports state fishing occurring at much shallower depths (~ 1500 m). We attribute this apparent inconsistency of potential fishing at depths > 2000 m in subregion 3.1 to the 1° latitude-longitude spatial resolution at which we carried out the bioregionalization analysis. The use of grid cells (pixels) at coarse resolution involves aggregating species data from multiple depths. As such, it is very likely that species from bathyal depths and species from the bordering abyssal depths have been clumped together in the same grid cell despite the bathyal and abyssal depth zones constituting fundamentally different habitats. This is an intrinsic disadvantage of using the "group first, then predict" modelling approach, which we extensively discussed in the report[3] for project PAE2020-02 (Ramiro-Sánchez & Leroy, 2022). Nevertheless, even though the network approach could only detect one subregion, bioregion 3 also encompasses bathyal depths (Figure A2, Appendix A). Thus, habitats of bioregion 3 are, in fact, trawled (Figure 4).

Fishing intensity impact index

The *fishing intensity impact index* reflected the trends observed in the analysis of the spatial distribution (in what suitability classes) and intensity (high or low number of fishing events) of fishing across regions. Bioregion 1, with a wide distribution across SIOFA and representing the bathyal depths of the Indian Ocean (Figure 1), was the most affected by trawling, gillnets, and line fishing. Bioregion 2, with a subantarctic distribution, was the next most affected by traps fishing. Because of the large predictions in SIOFA of each bioregion, the impact index was affected by the large proportion of unfished areas relative to the fished areas (i.e., index values were very low). The analysis of the subregions allowed to refine our understanding of the potential fishing impacts within the bioregions.

The occurrence of fishing in a consistent manner over areas of moderate suitability has a higher impact than few high intensity events occurring in areas of high suitability. This is the case for subregion 1.1 and subregion 1.2. Subregion 1.2 was the most impacted from trawling,

[3] https://siofa.org/sites/default/files/files/VMEMapping_FullReport.pdf

gillnets, and line fishing (Table 2), with all its maximum suitability classes (moderate classes) consistently fished at all fishing intensity levels (second panel in Figure 8 – Figure 11). The risk is especially significant for areas with gillnet fishing, as fishing appears to be extremely spatially concentrated in this subregion compared to the other subregions. In contrast, subregion 1.1, which is also fished at all intensities (first panel in Figure 8 – Figure 11), has higher predicted suitability classes (> 600) in SIOFA than subregion 1.2. However, it is four times at lower risk than subregion 1.2 (Table 2). This may seem surprising because subregion 1.1 has a narrower distribution than subregion 1.2 in SIOFA. However, because all of the maximum suitability classes of subregion 1.2 and pixels are being intensely fished, this results in a higher impact than for subregion 1.1.

The *fishing intensity impact index* will be exacerbated if a bioregion is disproportionately fished over its area (i.e., if the proportion of fished pixels over unfished pixels is high). For instance, subregion 2.4 was the second subregion most at risk by line fishing. Its reduced, predicted area within SIOFA and the occurrence of fishing in a large proportion of its areas of moderate suitability seem to be responsible for the vulnerability of this subregion. In contrast, subregion 2.1 has a wider predicted distribution (Figure 2). Yet, it displayed a similar distribution of the fishing intensity classes over the suitability classes, which also included a similar range of classes to those of subregion 2.4 (Figure 10). As a result, the difference in the index is ten times higher for subregion 2.4 than for subregion 2.1.

For some bioregions it will be important to take into consideration the potential effect of cumulative impacts from bordering nations; even for static gears like traps, which have less physical impact on the seabed than mobile gears. Subregion 1.1 and subregion 2.4 are good examples of potential cumulative impacts; SIOFA's fishing footprint occurs spatially concentrated in these subregions along the limits of SIOFA's area of competence. Thus, ensuring spatial connectivity among predicted areas may play a key role in future habitat recovery.

The interpretation of the *fishing intensity impact index* should go hand in hand with the understanding of the bioregionalization predictions (see Ramiro-Sánchez & Leroy, 2022). First, it is important to remember that only the predictive models of first-level bioregions were evaluated, whereas the models of the second-level subregions did not have enough occurrences to be evaluated via cross-validation: subregions represent only a small proportion of first-level bioregions whose detection depends on having sufficient species data. As a result, the validity of the nested levels remains uncertain to some extent, although their biogeographic patterns are supported by the distribution of water masses and ocean currents and by other biogeographic studies. Second, the possibility of overprediction exists for some subregions such as for subregion 2.1 and subregion 2.4, whose predictions were driven by the strong imprint of certain features, such as the Polar Front, the Subantarctic Front, and the Subtropical Front. The salinity and density properties drove the subregion predictions; however, it is unclear whether habitats of subregion 2.1 will occur at the depths of the Southeastern Indian Ridge, for instance. Finally, subregion 1.7 might not exist in SIOFA. This subregion was detected at bathyal depths of south of Australia and predicted westwards following the imprint of marked mesoscale eddies that shed from the Leewin Current along the band of the Eastern Gyral Current (Talley et al., 2011). Since no bioregion occurrences

were observed on the western Indian Ocean, the prediction of subregion 1.7 seemed to be driven by environmental conditions there too (Ramiro-Sánchez & Leroy, 2022).

Ideally, an index would integrate the six factors suggested by the FAO guidelines to determine the scale and significant of an impact, including knowledge of both gear selectivity and efficiency to draw conclusions from catch data. The integration of the last three FAO factors requires an in-depth knowledge of a species' biology, which includes the sensitivity of the species to the pressure in question. Our fishing impact index integrates not only the spatial distribution of fishing (i.e., whether it occurs in one bioregion or another), but also the intensity of fishing weighted by the suitability class of the bioregion where it is happening (vulnerability). In areas with robust sampling, other indices have been developed that, based on biological trait analysis, quantify the vulnerability of a species to trawling (González-Irusta et al., 2018) or longlining (de la Torriente Diez et al., 2022). At a global scale, Clark and Tittensor (2010) developed a risk index that quantified the vulnerability of seamounts using predictions of stony corals on seamounts and the likely distribution of seamount fish, informed by bottom-trawl fishery catches on seamounts and fisheries impacts.

At last, the fishing effort used for this study constitutes the cumulative fishing effort involving historical data. It may be the case that some of the patterns observed are relicts. In our index, the use of ranks to class fishing intensities prevents assigning a disproportionate weight to potential outliers, as it would be the case if the index used absolute numbers (i.e., original numbers). Yet, SIOFA will need to investigate and consider such possible discrepancies. The use of objective 10% quantiles as a basis for the ranks will permit the continuous update of the index results as more fishing data and better bioregion predictions become available. Furthermore, this index can be used to explore the impacts of management measures, such as how limiting or fostering fishing in a particular area would affect the significant adverse impacts for a given region.

6. Conclusions

In this report, we summarised how fishing activities may be impacting predicted bioregions of vulnerable marine ecosystem indicator taxa in the SIOFA area. Ideally, we would expect to observe higher fishing effort in areas of predicted low suitability for bioregions, potentially indicating a low impact. However, we found that for all four gear types and all bioregions, there was fishing occurring in predicted areas of high suitability. Bioregion 1 was the most impacted by trawling, gillnets, and line fishing. The marked spatial distribution of the four gear types in SIOFA results in subregions being affected differently by each gear type, where subregion 1.2 seems to be the most impacted from trawling, gillnets, and line fishing and subregion 2.4 by traps fishing. For the subregions, the interpretation of the index results will have to take into consideration the uncertainties of the predictions, particularly for subregion 2.1, subregion 2.4, and subregion 1.7. In addition, potential cumulative impacts, such as fishing along SIOFA's edges from bordering nations, will also need to be considered in order to ensure connectivity between and within bioregions to maintain populations. Therefore, the index offers a first approximation to assess potential impacts of fishing on the bioregions for each gear type. As new data becomes available in the form of better bioregion predictions

and fishing effort, the index can be updated. In combination with SIOFA's expert knowledge of their fisheries and area of competence, the index can be used to highlight particular areas of conservation interest around ambiguous areas and explore impacts of management measures.

7. Recommendations

We recommend the Scientific Committee to:

- Note the development of the *fishing intensity impact index* to aid in the consideration of bioregions (and specific areas within) potentially more impacted from fishing activities and the potential to update the index as new data becomes available.
- Note the potential risk of Bioregion 1 from fishing impacts, and particularly the risk of subregion 1.2.

References

Clark, M. R., & Tittensor, D. P. (2010). An index to assess the risk to stony corals from bottom trawling on seamounts. *Marine Ecology*, *31*(SUPPL. 1), 200–211. https://doi.org/10.1111/J.1439-0485.2010.00392.X

de la Torriente Diez, A., González-Irusta, J. M., Serrano, A., Aguilar, R., Sánchez, F., Blanco, M., & Punzón, A. (2022). Spatial assessment of benthic habitats vulnerability to bottom fishing in a Mediterranean seamount. *Marine Policy*, *135*, 104850. https://doi.org/10.1016/J.MARPOL.2021.104850

FAO. (2009). International Guidelines for the Management of Deep-Sea Fisheries in the High Seas Directives. In *FAO Fisheries and Aquaculture Circular. No. 1036* (Vol. 1036, Issue 1036).

González-Irusta, J. M., de La Torriente, A., Punzón, A., Blanco, M., & Serrano, A. (2018). Determining and mapping species sensitivity to trawling impacts: the BEnthos Sensitivity Index to Trawling Operations (BESITO). *ICES Journal of Marine Science*, *75*(5), 1710–1721. https://doi.org/10.1093/ICESJMS/FSY030

Muscatello, A., Elith, J., & Kujala, H. (2021). How decisions about fitting species distribution models affect conservation outcomes. *Conservation Biology*, *35*(4), 1309–1320. https://doi.org/10.1111/COBI.13669

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (v4.1.2). R Foundation for Statistical Computing.

Ramiro-Sánchez, B., & Leroy, B. (2022). *SIOFA bioregionalization and VME project*.

Talley, L. D., Pickard, G. L., Emery, W. J., & Swift, J. H. (2011). Indian Ocean. In *Descriptive Physical Oceanography: An Introduction* (6th ed., pp. 1–555). Elsevier Ltd. https://doi.org/10.1016/C2009-0-24322-4

Appendix A – Bioregionalization method "Group first, then predict"

Overview of the methods and description of bioregions

We summarise the procedure as follows:
(1) we delineated biogeographical regions at the species level using a network approach, on the basis of occurrence datasets, at a 1° x 1° spatial resolution;
(2) once we identified the bioregions, we individually modelled the relationship between each VME bioregion and environmental predictors using an ensemble modelling approach based on seven algorithms, obtaining predictive suitability maps for each bioregion;
(3) based on the suitability maps for each region, we identified in each pixel the bioregion most likely to occur.

To calibrate the bioregion distribution models, we used as presence data the known pixels of the regions, and as absence data the known pixels for other biogeographical regions. To evaluate the predictive accuracy of bioregion distribution models, we applied a block cross-validation procedure to limit over-optimistic performance evaluations due to autocorrelation between calibration and evaluation data. We did not apply block cross-validation for the sub-bioregions because the lack of data prevented such a procedure.

We detected two nested levels of biogeographical regions with the network approach. At the first level, we detected three large biogeographical regions in the Southern Indian Ocean (Figure A1-A). These three biogeographical regions were geographically cohesive, they had distinct species composition, and occurred at different depth zones and environmental conditions. In summary, Cluster 1 had a median depth range spanning from shallow waters to 2,800 m, Cluster 2 had a deeper median depth of ~ 2,000 m, and Cluster 3 was present in the lower bathyal (> 800 m) and abyssal areas (Figure A2).

Nested at a second level, we detected eight subregions: Cluster 1 subdivided into five subregions (1.1, 1.2, 1.3, 1.5 and 1.7), Cluster 2 into two subregions (2.1 and 2.4), and Cluster 3 into one subregion (3.1) (Figure A1-B). Cluster 1.1 had a median depth of 1,000 m, with spanning depths from 200 to 2,700 m; Cluster 1.2 had 75% of its observations in waters above 1,500m, with a median depth of 800 m and a maximum depth of 3,000 m; Cluster 1.3 (Figure A3) had a median depth of 400-500 m, ranging from 200 to 2,700 m; Cluster 1.5 had a shallow median depth of 200 m; Cluster 1.7 had a median depth of 400 m, with a maximum of 1,800 m; Cluster 2.1 had a median depth of 600 m, with depths spanning from 200 to 1,500 m; Cluster 2.4 was a deep cluster with a median depth of 2,900 m, where 50% of the observations fell between 2,100 and 4,000 m; Cluster 3.1 was the deepest of all subregions with a median depth of 5,000 m (Figure A3).

The predictive map of bioregions at the first level showed that the three bioregions encompassed the SIOFA area (Figure A4). Bioregions 1 and 3 prevailed from 20°N to 10°S latitude and below 40°S latitude, the Southern Ocean bioregion 2 dominated. We added on the map the degree of uncertainty (index of confidence from 0 to 1, see methods on report on project PAE2020-02) that we had in our predictions, such that we are less confident of predictions in darker areas. These areas are uncertain because none of the samples we

gathered were located in similar environmental conditions – thus, model predictions in these areas are extrapolations. Areas of uncertainty in predictions coincide with areas of heterogeneity in biogeographical composition and with areas where there was no initial data. We modelled the distribution of the eight detected subregions, although no formal evaluation of the model performance was carried out for predictions because of the low number of occurrence points. As for bioregions at the first level, we incorporated confidence levels in the predictions. Individual predictions are presented in Figure A5.

Bioregions at the first level depict the distribution of fauna across two distinct environments, the bathyal and abyssal depth zones, and a specific Southern Ocean bioregion. As discussed in the report for project PAE2020-02, existing global biogeographical classifications for the Southern Indian Ocean reflect these depth zones, specifically the Global Open Oceans and Deep Seabed (GOODS) classification. The subregions we detected further improve that scheme, reflecting the importance of water masses and topographic features in the Southern Indian Ocean. In the report, we extensively elaborated on the characteristics and drivers of each of the eight subregions and how these are supported by the literature. For this reason and because they provide a finer understanding of biogeographic patterns in the Southern Indian Ocean, we used this level to compare the overlap with existing fisheries within SIOFA's Convention Area.

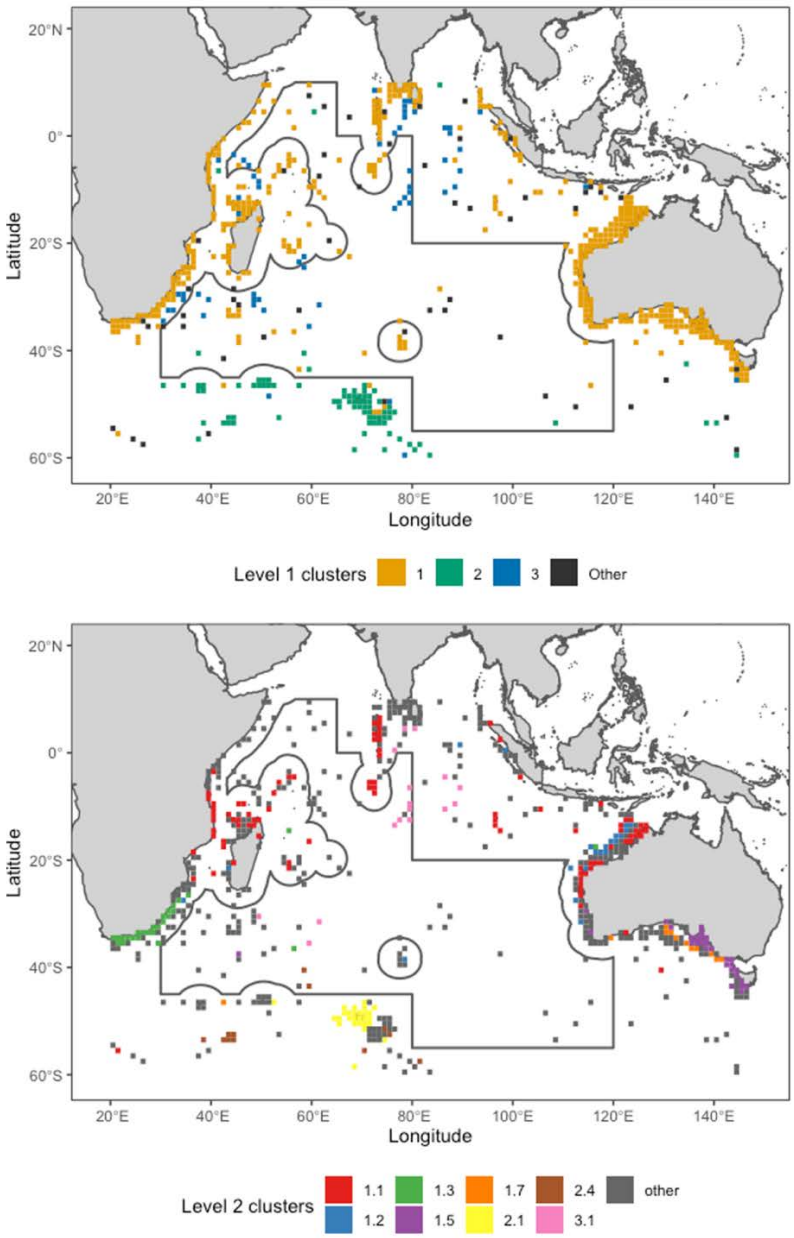

Figure A1. Biogeographical regions (clusters) of VME indicator taxa in the Southern Indian Ocean: (A) Three biogeographical regions of similar species composition detected at the first hierarchical level. (B) Finer sub-biogeographical regions detected at the second hierarchical level; black indicates no assigned subregion.

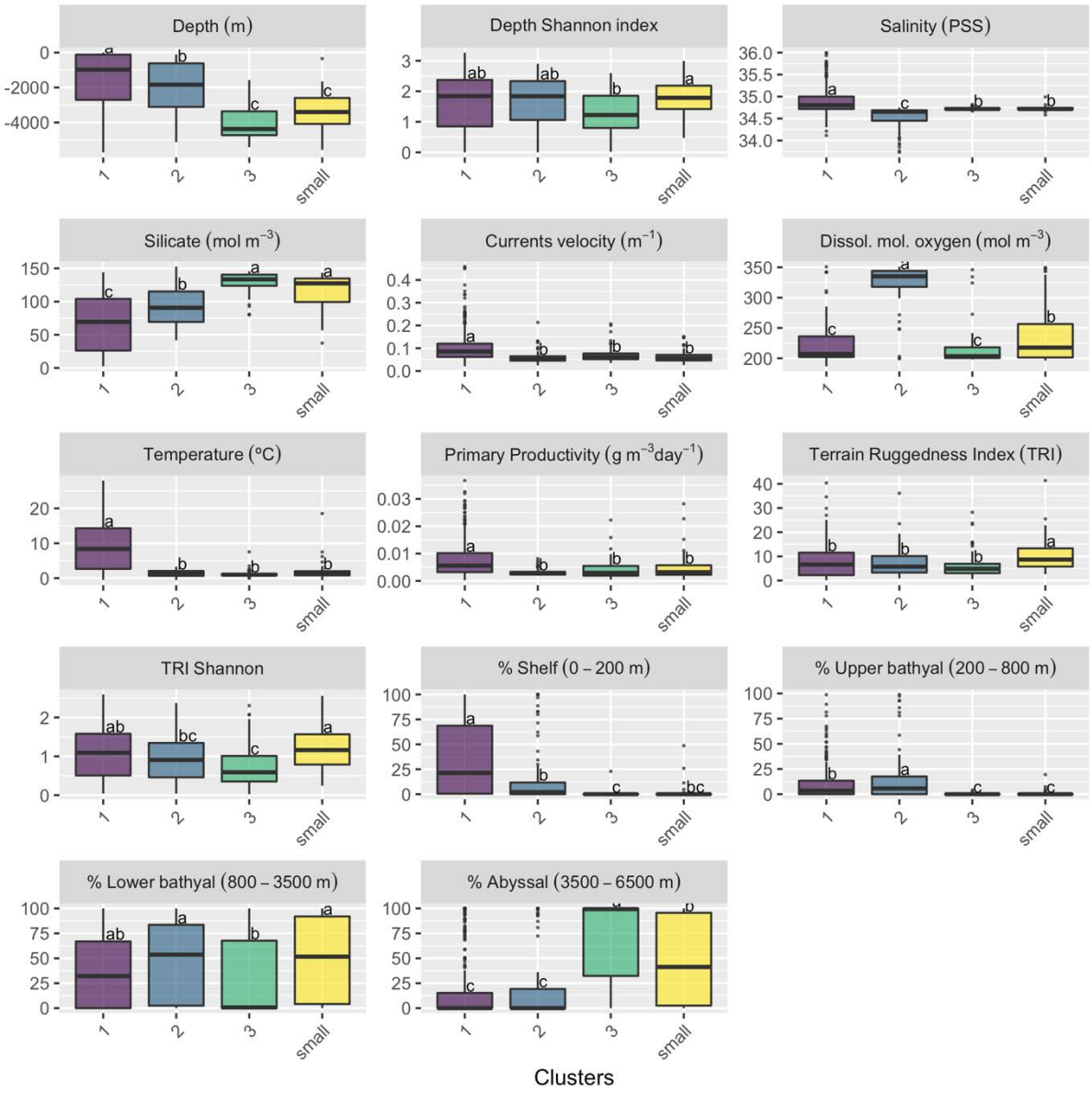

Figure A2. Environmental variables distribution for each of the three main biogeographical observed regions detected at the first hierarchical level of the bioregionalization analysis. Small clusters (labelled 'small' in the graph) are also shown for comparison. Variables from top left to bottom right: depth (m); depth Shannon index; salinity (PPS); silicate concentration (mol m-3); speed of currents (m-1); dissolved molecular oxygen (mol m-3); temperature (°C); surface primary productivity (g m-3 day-1); terrain ruggedness index (TRI); TRI Shannon index; percentage of shelf depth zone (0 – 200 m) covered; percentage of upper bathyal depth zone (200 – 800 m) covered; percentage of lower bathyal depth zone (800 – 3,500 m) covered; percentage of abyssal depth zone (3,500 – 6,500 m) covered. Differences between subclusters per environmental variable are shown through pairwise comparisons (post-hoc Tukey test) with letters on the boxplots: mean values with at least one common letter are not significantly different. From the report on project PAE2020-02, Figure 7.

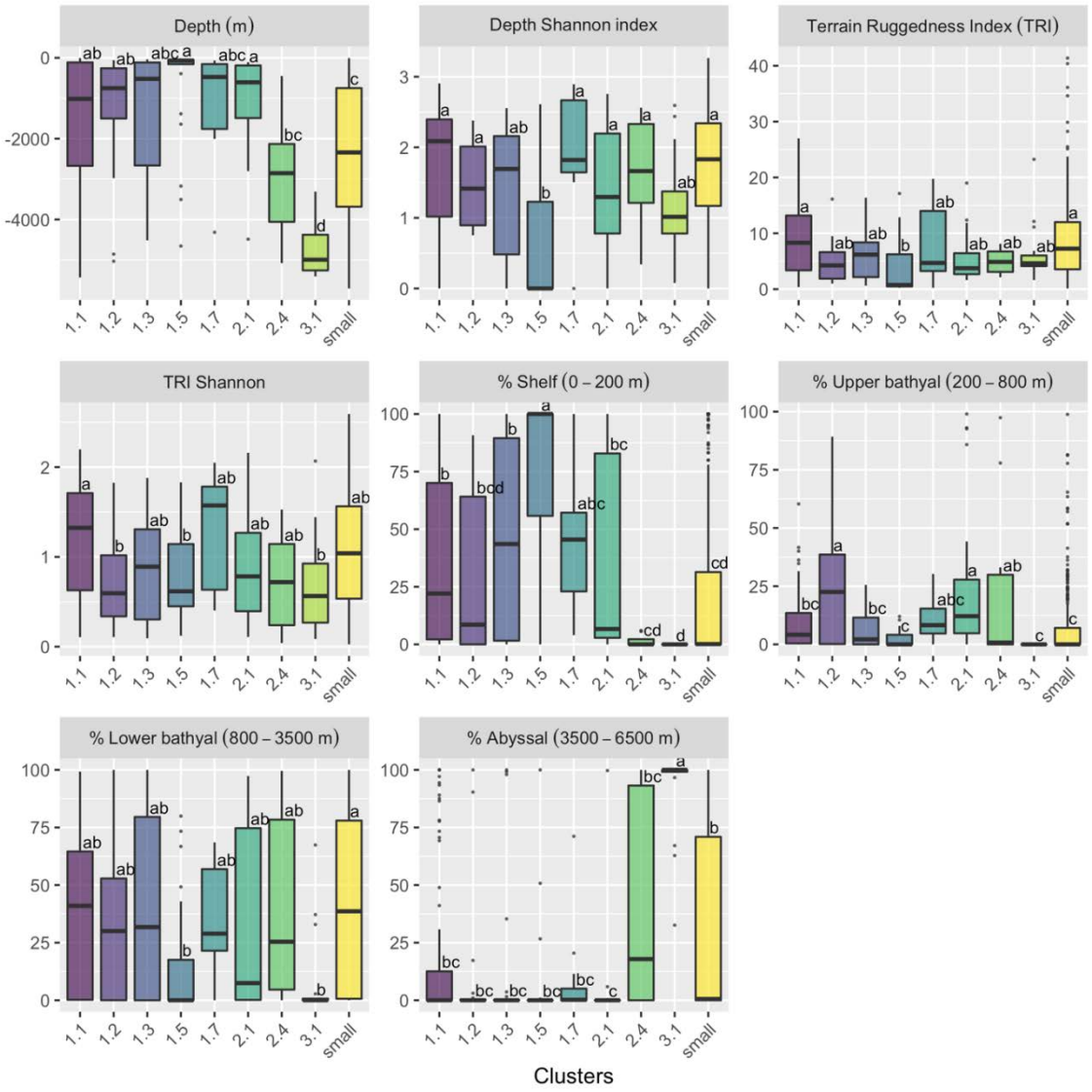

Figure A3. Environmental variables distribution for each of the eight nested biogeographical regions detected at the second hierarchical level of the bioregionalization analysis. Small clusters (labelled 'small' in the graph) are also shown for comparison. Variables from top left to bottom right: depth (m); depth Shannon index; terrain ruggedness index (TRI); TRI Shannon index; percentage of shelf depth zone (0 – 200 m) covered; percentage of upper bathyal depth zone (200 – 800 m) covered; percentage of lower bathyal depth zone (800 – 3,500 m) covered; percentage of abyssal depth zone (3,500 – 6,500 m) covered. Differences between subclusters per environmental variable are shown through pairwise comparisons (post-hoc Tukey test) with letters on the boxplots: mean values with at least one common letter are not significantly different. From the report on project PAE2020-02, Figure 9.

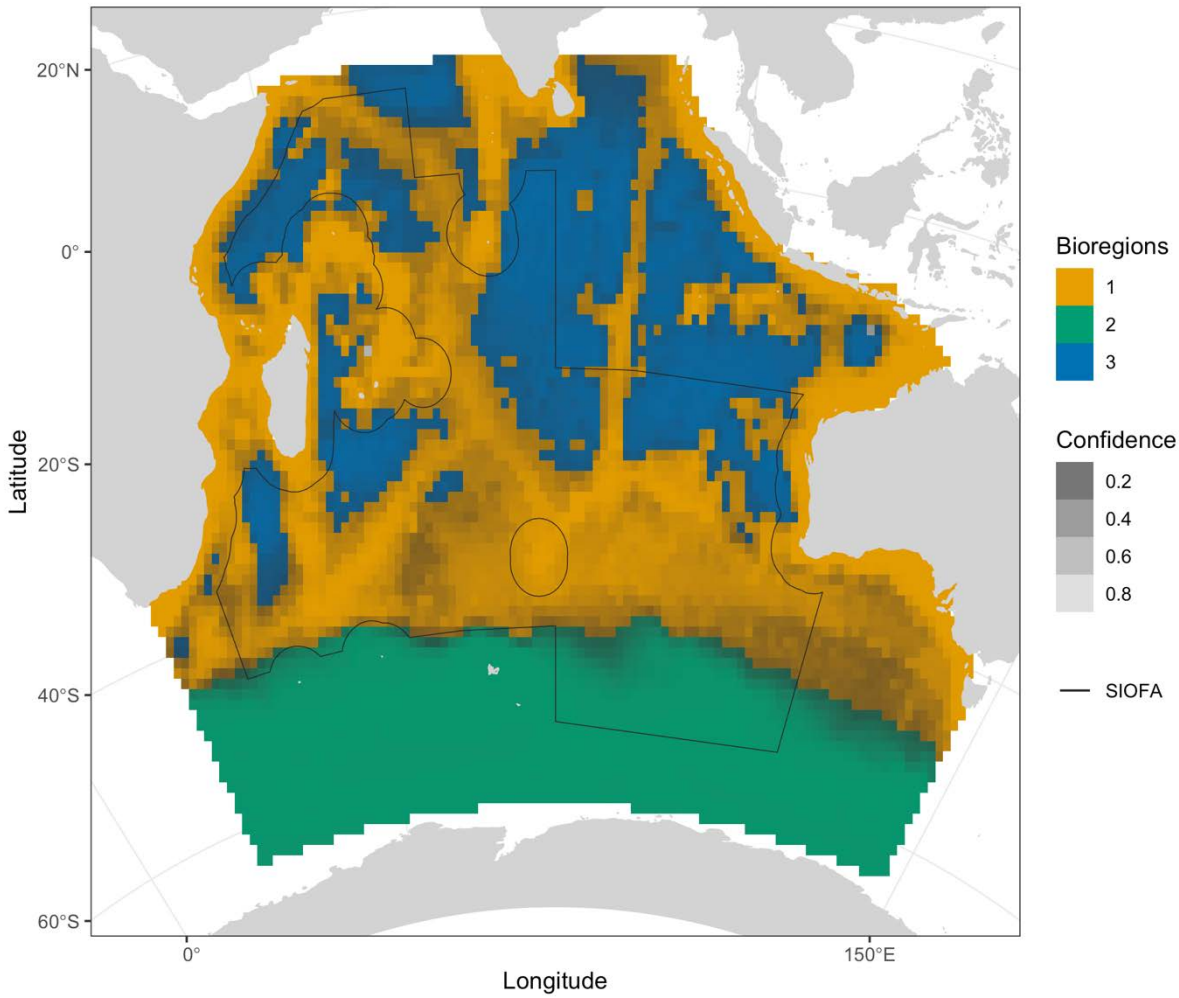

Figure A4. Predicted biogeographical regions of VME indicator taxa in the Southern Indian Ocean at the first level of the hierarchy of the bioregionalization analysis. Areas with low confidence in the prediction are shown in darker shades of grey.

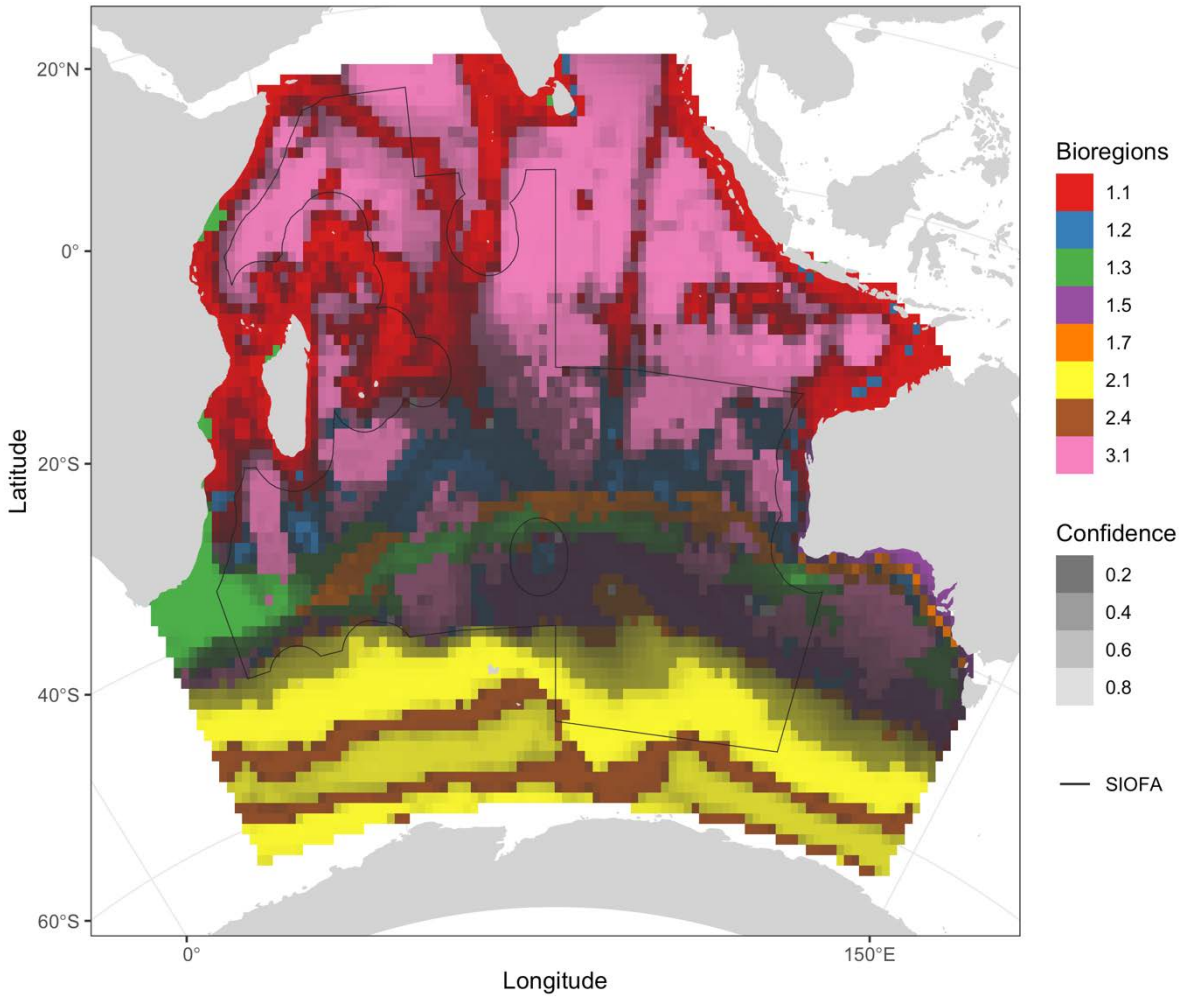

Figure A5. Predicted biogeographical regions of VME indicator taxa in the Southern Indian Ocean at the second level of the hierarchy of the bioregionalization analysis. Areas with low confidence in the prediction are shown in darker shades of grey. Note that, because of the low number of points, we could not reliably evaluate these predictions.

Appendix B – Proportion of fished cells per habitat suitability class and gear type

Table B1. Proportion of fished and non-fished cells per habitat suitability (HS) class per bioregion and gear type.

		Bioregion1		Bioregion2		Bioregion3	
	HS class	**Fished**	**Non-fished**	**Fished**	**Non-fished**	**Fished**	**Non-fished**
Trawling	[0,100]	0	9552	894	72554	688	37770
	(100,200]	10	7600	0	5173	134	12524
	(200,300]	7	6913	0	1848	27	6997
	(300,400]	2	8865	1	1486	15	6666
	(400,500]	6	10316	0	946	10	6945
	(500,600]	9	10982	0	970	5	6800
	(600,700]	14	10433	0	1212	2	8156
	(700,800]	31	14292	0	1625	12	7052
	(800,900]	119	10015	0	7054	2	1780
	(900,1000]	697	5722	0	1822	0	0
		Bioregion1		**Bioregion2**		**Bioregion3**	
	HS class	**Fished**	**Non-fished**	**Fished**	**Non-fished**	**Fished**	**Non-fished**
Gillnets	[0,100]	0	9552	171	72554	173	37770
	(100,200]	0	7600	5	5173	3	12524
	(200,300]	0	6913	0	1848	0	6997
	(300,400]	0	8865	0	1486	0	6666
	(400,500]	0	10316	0	946	0	6945
	(500,600]	0	10982	0	970	0	6800
	(600,700]	0	10433	0	1212	0	8156
	(700,800]	0	14292	0	1625	0	7052
	(800,900]	6	10015	0	7054	0	1780
	(900,1000]	170	5722	0	1822	0	0
		Bioregion1		**Bioregion2**		**Bioregion3**	
	HS class	**Fished**	**Non-fished**	**Fished**	**Non-fished**	**Fished**	**Non-fished**
Lines	[0,100]	46	9552	206	72554	391	37770
	(100,200]	41	7600	17	5173	4	12524
	(200,300]	16	6913	0	1848	0	6997
	(300,400]	19	8865	24	1486	0	6666
	(400,500]	26	10316	0	946	0	6945
	(500,600]	5	10982	26	970	0	6800
	(600,700]	2	10433	0	1212	0	8156
	(700,800]	17	14292	63	1625	0	7052
	(800,900]	20	10015	13	7054	0	1780
	(900,1000]	203	5722	46	1822	0	0
		Bioregion1		**Bioregion2**		**Bioregion3**	
	HS class	**Fished**	**Non-fished**	**Fished**	**Non-fished**	**Fished**	**Non-fished**
Traps	[0,100]	0	9552	6	72554	3	37770
	(100,200]	0	7600	0	5173	1	12524
	(200,300]	0	6913	0	1848	0	6997
	(300,400]	0	8865	0	1486	0	6666
	(400,500]	0	10316	0	946	1	6945
	(500,600]	2	10982	0	970	1	6800
	(600,700]	0	10433	0	1212	0	8156
	(700,800]	0	14292	0	1625	0	7052
	(800,900]	0	10015	0	7054	0	1780
	(900,1000]	4	5722	0	1822	0	0