PAEWG-03-02

3rd Meeting of the Protected Areas and Ecosystems Working Group (PAEWG3)
01-04 March 2021
(online)

SIOFA Vulnerable Marine Ecosystem Mapping

Relates to agenda item: 3.3. VME mapping (ongoing consultancy from BOREA Laboratory, Biology of Aquatic Organisms and Ecosystems)

Working paper ☒ Info paper ☐

Consultancy from BOREA Laboratory, Biology of Aquatic Organisms and Ecosystems

SIOFA Conservation and Management Measures (CMM) 2019/01 directs the Scientific Committee (SC) to provide maps of where vulnerable marine ecosystems (VMEs) are known to occur, or likely to occur, in the Agreement Area. The Meeting of the Parties shall act on the advice of the SC in regards with the VME habitat mapping and update its interim bottom fisheries measures. Here we present the first steps towards completing the task of predicting the distribution of VME indicator taxa in the SIOFA.

Recommendations *(proposals and working papers only)*

This document is to be discussed within the PAEWG-03 meeting.

Funded by the European Union

Introduction

Typically, deep-sea species are slow-growth, long-lived, and late maturing; traits that limit their potential for resilience and recovery from human disturbances such as bottom-contact fishing, climate change and, potentially in the future, deep-sea mining (Ramirez-Llodra et al., 2011). Some of the species and habitats found in the deep seafloor are recognised as vulnerable marine ecosystems (VMEs) in need of protective management and conservation measures (UNGA, 2006). Several taxa have been acknowledged to be VME indicators because they are the most likely to be found in habitats meeting the criteria for VMEs (NAFO, 2019; NEAFC, 2014), amongst them cold-water corals (CWCs) encompassing reef framework forming and solitary scleractinians, octocoral aggregations (corals gardens) and deep-sea sponge aggregations (NAFO, 2019; NEAFC, 2014).

The VME concept started to gain recognition after the United Nations General Assembly (UNGA) 61/105 resolution was passed (UNGA, 2006), and called upon Member States to minimise damage to these habitats. Resolutions passed by the UNGA are not binding, but the Regional Fisheries Management Organisations (RFMOs), which nations are contracting, have the right to adopt legally binding measures.

The FAO guidelines for fishing activities in the High Seas fisheries indicate actions to prevent adverse impacts on VMEs. According to the guidelines, management actions are to be designated to ensure that High Seas ecosystems structure and functioning is not impaired. An inherent condition for the application of these measures is the spatial delineation of where these habitats reside (Rice et al., 2011). Accordingly, the objective of this project is to map bioregions based on VME indicator species distribution data. In order to characterise bioregions, we will first model the distributions of VME indicator taxa. In this working paper, we therefore present the progress up to date in collating data, curating it, and modelling distribution of indicator taxa for which a working example of outputs is shown.

Methods

Biological and environmental data

The occurrence database of VME indicator taxa compiled records from different repositories: the Global Biodiversity Information Facility (GBIF, gbif.org, extracted on 9/11/2020), the Ocean Biogeographic Information System (OBIS, obis.org, extracted on 9/11/2020), NOAA Deep-Sea Coral Data Portal (extracted on 16/11/2020), Smithsonian National Museum of Natural History (extracted on 16/11/2020), and SIOFA Secretariat. The initial search of records was limited to a wider polygon encompassing the SIOFA area and further restricted to records in the Indian Ocean (www.marineregions.org).

Environmental variables hypothesised to drive biogeography of deep-sea habitat forming species were downloaded from several publicly available global scale sources (Table 2): Bio-Oracle (https://www.bio-oracle.org/), National Oceanic and Atmospheric Administration (NOAA; https://www.ncei.noaa.gov/access/oads/), Global Marine Environmental Datasets (GMED; http://gmed.auckland.ac.nz/) and General Bathymetric Charts of the Oceans (GEBCO; https://www.gebco.net/). Seafloor variables included bathymetry and derivatives (slope, bathymetry position index and roughness), temperature, dissolved oxygen, salinity, currents, nitrogen, phosphate and silicate. Primary productivity at surface was also collated.

Table 1. SIOFA VME indicator taxa

VME indicator taxa	
Cnidaria	Gorgonacea (Order)
	Anthoathecatae (Order)
	Stylasteridae (Family)
	Scleractinia (Order)
	Antipatharia (Order)
	Zoantharia (Order)
	Actiniaria (Order)
	Alcyonacea (Order)
	Pennatulacea (Order)
Porifera	Hexactinellida (Class)
	Demospongiae (Class)
Ascidiacea (Class)	
Bryozoa (Phylum)	
Brachiopoda (Phylum)	
Pterobranchia	
Serpulidae (Family)	
Xenophyophora (Phylum)	
Bathylasmatidae (Family)	
Stalked crinoids (Class)	
Euryalida (Order)	
Cidaroida (Order)	

Table 2. List of environmental variables collated for the prediction of VME indicator taxa.

Environmental variables	Unit	Resolution	Source
Maximum, Minimum temperature at seafloor	Degree Celsius	5 arcmin (0.083 degree; ca. 9.2 km)	Bio-Oracle
Maximum, Minimum salinity at seafloor	pss		
Maximum, Minimum dissolved oxygen at seafloor	mol.m-3		
Maximum, Minimum nitrate at seafloor	mol.m-3		
Maximum, Minimum phosphate at seafloor	mol.m-3		
Maximum, Minimum silicate at seafloor	mol.m-3		
Maximum, Minimum primary productivity at surface	g.m-3.day-1		
Maximum, Minimum currents speed at seafloor	m-1		
Particle Organic Carbon (POC)	mg.m-3	2.5 arcmin (0.0417 degree; ca. 4.6 km)	GMED
Distance to land	km	5 arcmin (0.083 degree; ca. 9.2 km)	
Calcite	mol.m-3		
Depth	m	0.004 degree (ca. 0.4km)	GEBCO 2020
Total alkalinity	µmol.m-3	1 degree (ca. 111km)	NOAA's Ocean Carbon and Acidification Data Portal
Omega Aragonite	omegaAR		
Omega Calcite	omegaCA		

Data analysis

Indicators of data quality

Verification procedures were applied 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.

For duplicates of GBIF and OBIS, we used the catalogue number and geographical coordinates to filter out potential duplicates. In addition, records with no original depth recorded or with no available depth from GEBCO bathymetry were removed from further analysis.

The working dataset comprised 77,727 occurrence points after the curation of records. From these: 36,506 records were found deeper than 50 m water depth; 28,004 records deeper than

100 m; 23,661 records below 150 m, and 21,503 deeper than 200 m. We used the 100 m water depth threshold to perform exploration analyses of the occurrence data (Table 3).

Table 3. Proportion of records per database below 100 m water depth kept for analysis.

Database	Number of records
GBIF	15,969 (57%)
OBIS	11,314 (40.4%)
SIOFA	401 (1.4%)
NOAA	206 (0.7%)
Smithsonian Institution	114 (0.4%)

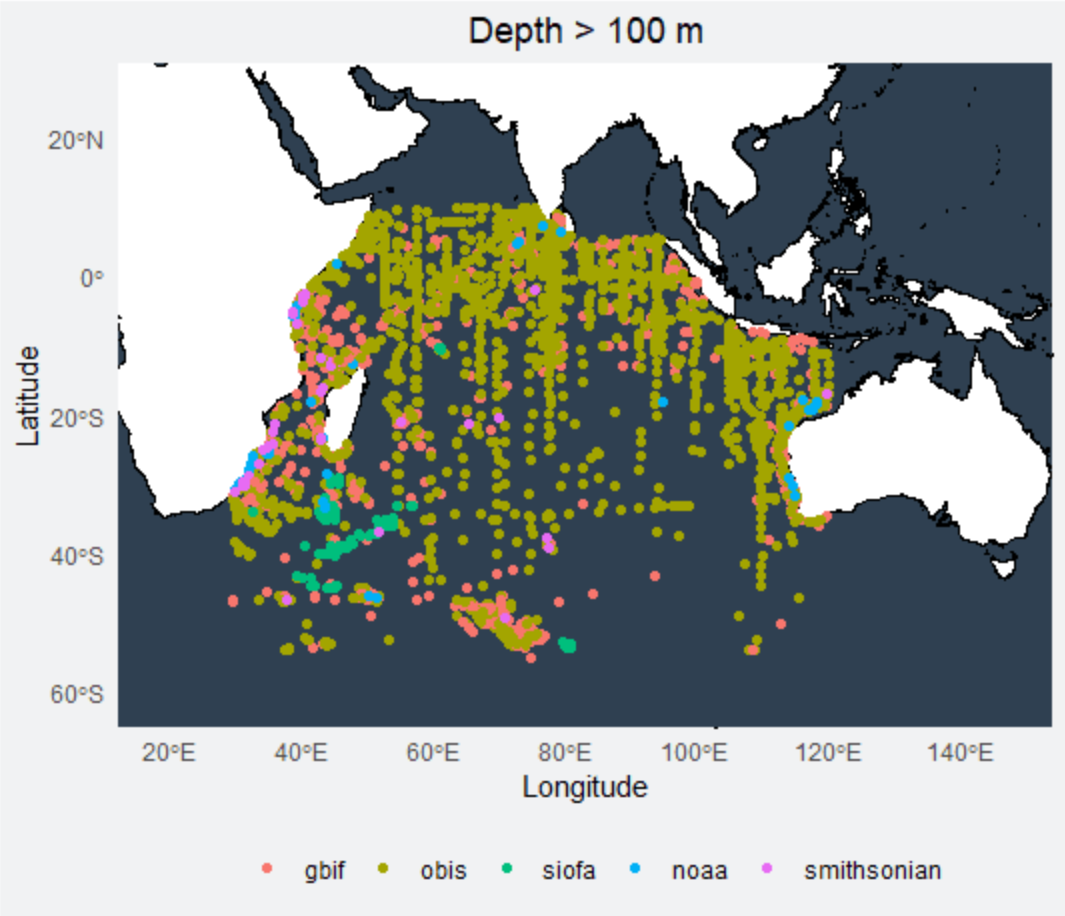

Figure 1. Distribution of records at depths below 100 m water depth coloured by their source of data.

To investigate the accuracy of the consolidated occurrence dataset, their completeness, and uncertainty levels, we produced several maps of distribution of VME indicator taxa and indicators of data quality (Soberón et al., 2007; Leroy et al., 2017). Specifically, we first mapped the point-locality distributions of VME indicator taxa and the gridded observed richness. Second, in each grid cell, we estimated the theoretical total species richness with three non parametric incidence-based richness estimators: ICE (Gotelli & Colwell, 2011), Chao2 and Jackknife (Chiu et al., 2014). These methods project the estimated total number of species on the basis of the observed number of species in the sample plus a correction based on the number of rare species (typically, singletons or doubletons), to account for the unobserved fraction of rare species. We chose non-parametric richness estimators because they have been proven to perform better than other richness estimators (Walther & Moore, 2005). Third, we produced maps of estimated sampling completeness by dividing the observed species richness in each cell by the estimated total richness. Because species richness estimators can be biased when the sampling intensity and observed richness are low, resulting in overestimations of data completeness (Leroy, 2012), we defined a richness threshold of 30 below which the completeness was set to zero. We produced these different maps at the 1 x 1 degree spatial resolution, and other resolutions are currently being investigated in order to determine an optimal spatial scale for analysis. . Finally, we explored the environmental gradients of the records for the selected environmental variables.

Species Distribution Modelling

We will model the potential distribution of indicator taxa for which we have enough occurrence records (at least 30 spatially independent records). We will undertake an ensemble modelling procedure which consists in producing multiple realizations of predictions (combining different state-of-the-art model classes with multiple sets of initial conditions, parameterizations), from which a consensus can be derived (e.g., average trend), and uncertainties quantified (Araújo & New, 2007). Ensemble modelling allows to account for spatially explicit uncertainties in maps by applying methods from information gap decision theory to fully inform management decisions. The nature of data is “presence-only”, i.e. with no reliable information on taxa absence; therefore, we will adapt our data preparation (selection of pseudoabsences), modelling protocols and model evaluation techniques accordingly (Guillera-Arroita et al., 2015; Leroy et al., 2018; Schickele et al., 2020).

Because we have no *a priori* assumption on the shapes of species responses to environmental variables, we will apply several classes of modelling techniques available in the biomod2 R package (Thuiller et al., 2009): (1) generalized linear models (McCullagh, 1984); (2) generalized additive models (Hastie & Tibshirani, 2004); (3) multivariate adaptive regression splines (Friedman, 1991); (4) artificial neural networks (Ripley, 2007); (5) random forests (Breiman, 2001); (6) boosted regression trees (generalized boosted models, Elith et al. 2008); (7) flexible discrimnant analysis (Hastie, Tibshirani, & Buja, 1994). We will investigate different methods to account for sampling bias, such as geographical filtering and environmental filtering methods (as described in Louppe et al., 2019, Schickele et al., 2020). To summarize, these methods aim at reducing sampling bias by reducing the weight of over-sampled areas and adapting the pseudoabsence sampling procedure. To adequately assess the performance of our models, we will implement a block cross-validation procedure (Roberts et al., 2017) which will minimize the degree of autocorrelation between the calibration and evaluation

datasets. The number of folds for the evaluation procedure will be adapted to the quantity and spatial distribution of data for each taxon, with a minimum of two folds. We will evaluate our models with two methods: (1) the Boyce index designed for presence only data (Boyce et al., 2002) and (2) with a critical, expert-based assessment of response curves (e.g., Schickele et al., 2020). On the basis of the selected models, we will produce two ensemble modelling maps: the average environmental suitability map, and an uncertainty map based on the standard deviation of environmental suitability.

For the purposes of this paper, we provided a case study on the Order Alcyonacea (Cnidaria: Anthozoa: Octocorallia) to demonstrate the above.

Results and discussion

Data distribution and quality

Most of the available records were obtained from the publicly available repositories GBIF and OBIS, where we could find information for most taxonomic resolutions. Records collated from NOAA and Smithsonian also stored information up to species level, whereas records coming from the SIOFA Secretariat presented the coarser taxonomic rank, sometimes up to Order level.

Records were mainly distributed around the coastal inshore areas, while large areas of scant information dominated the centre of the SIOFA range (Figure 1). The distribution of individual VME indicator taxa varied within the study area with groups such as Alcyonacea, Scleractinia, Antipatharia and Anthoathecata among the most widely distributed and frequent (Figure 2). Indeed, the observed species richness for the compiled records reflected a poor understanding of the area (Figure 3) as it did also suggest the sampling intensity (Figure 4). This was further confirmed by the estimation of the completeness index (Table 4). The completeness index, the ratio between observed richness and estimated richness, indicated that only 63% of the species richness of the area is known. This index varied for each of the phyla, with Foraminifera, Bryozoa, Brachiopoda and Echinodermanta amongst the best sampled and Porifera and Annelida (family Serpulidae) amongst the worst sampled. The phylum Cnidaria, which contains the subclasses Hexacorallia, Octocorallia and Hydrozoa, presented an index of 64%. However, for all taxa, the spatial distribution of the completeness index suggested that the SIOFA area is critically under-sampled (Figure 5). Other grid resolutions than 1 degree latitude and longitude are currently explored to define an optimal resolution of analysis, which will be a trade-off between spatial resolution and adequate completeness.

The majority of the records occurred above 1,000 m water depth, reflecting the aggregated distribution of records in coastal areas (Figure 6). The environmental variables presented certain variability at their original resolutions (Figure 6-7).

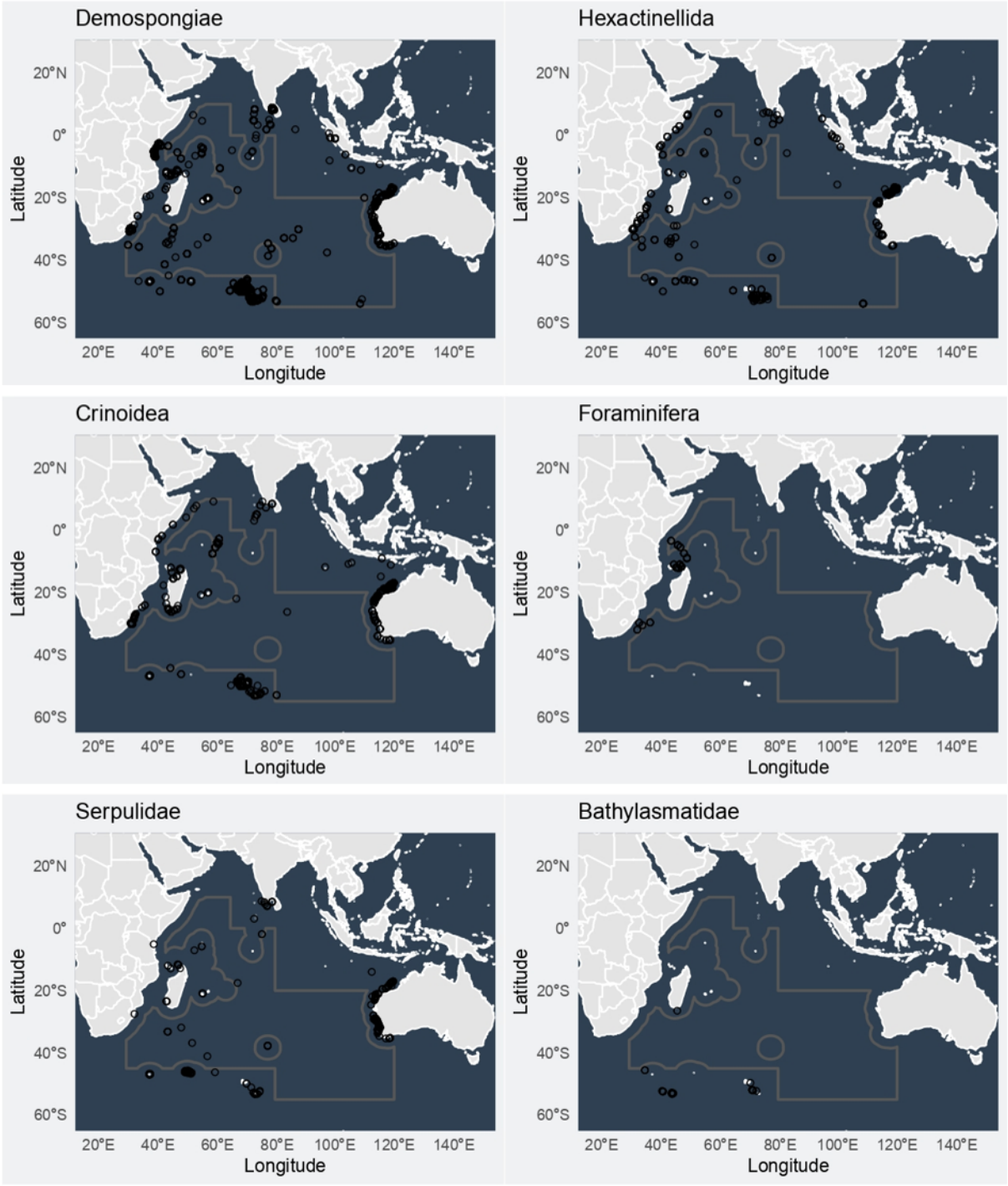

Figure 2. Occurrence records of indicator taxa in the Indian Ocean.

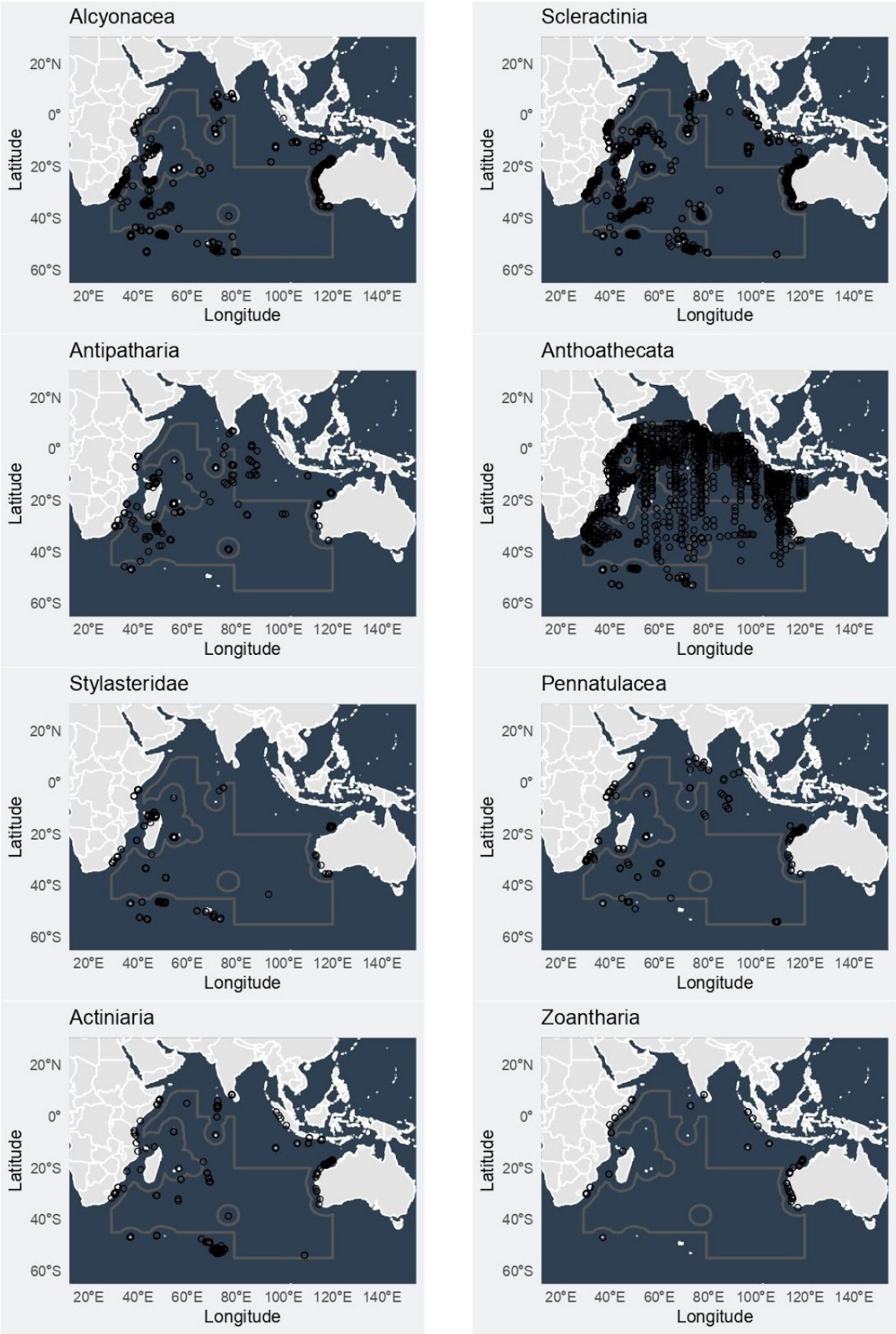

Figure 2. (Continued). Occurrence records of indicator taxa in the Indian Ocean.

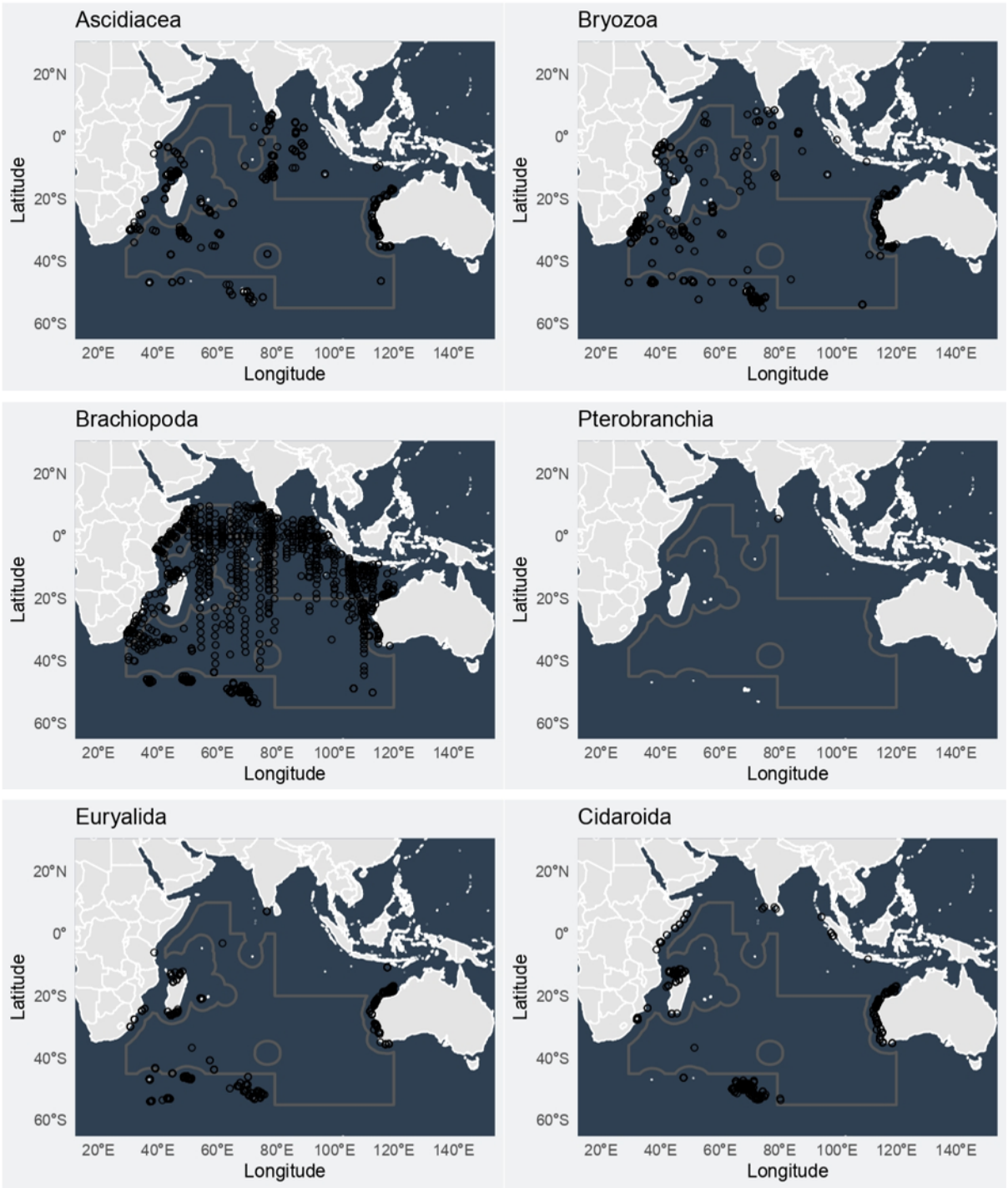

Figure 2. (Continued). Occurrence records of indicator taxa in the Indian Ocean.

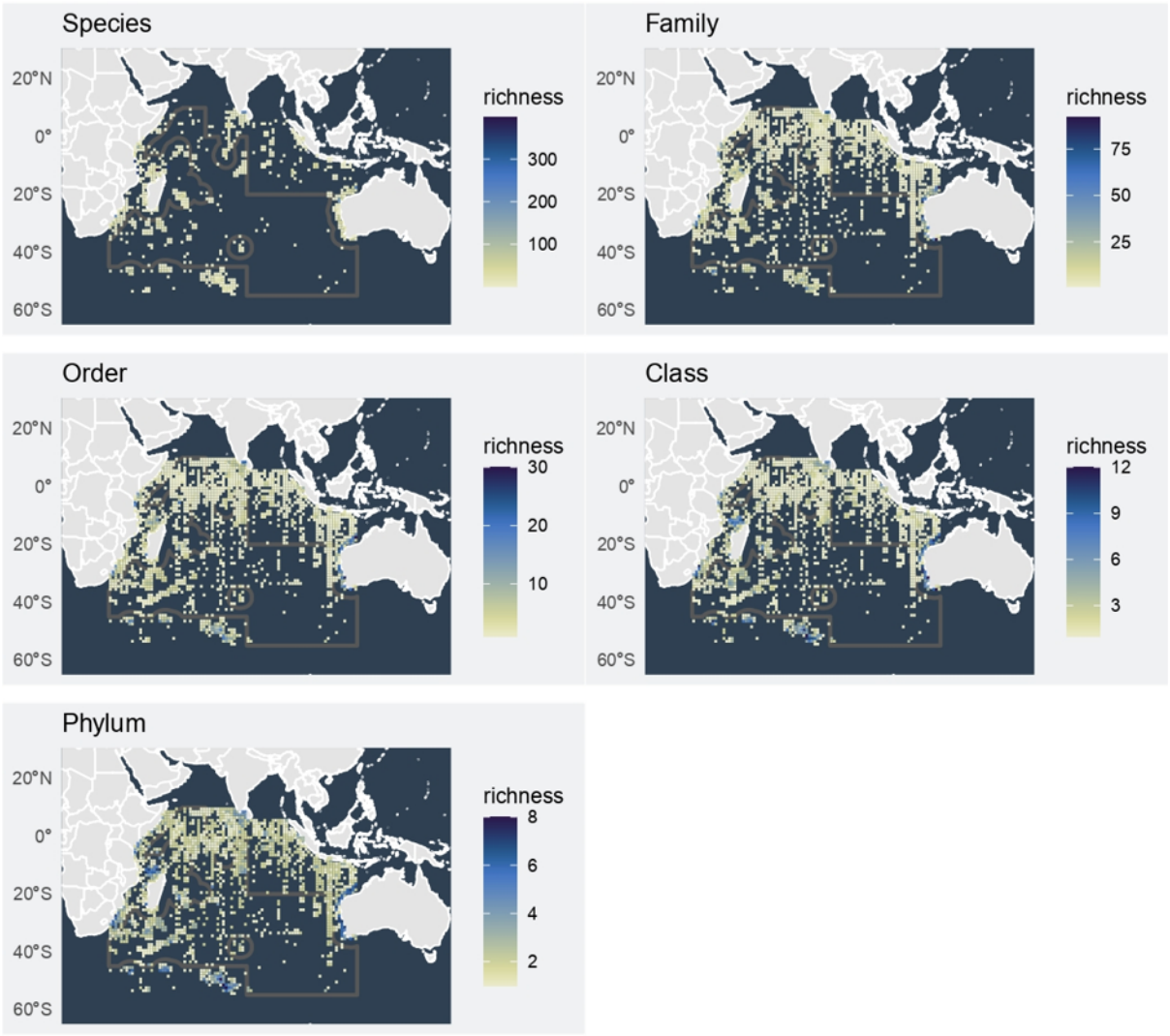

Figure 3. Observed richness per taxonomic level of all occurrence records.

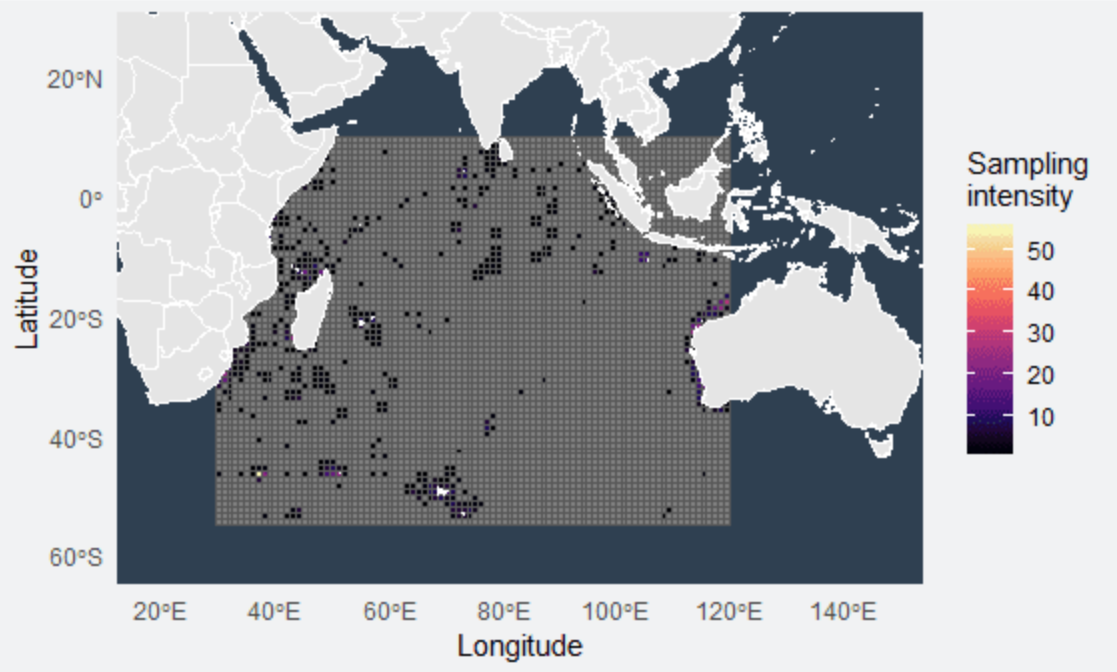

Figure 4. Sampling intensity in the study area as the number of sampling events per grid cell at 1 x 1 degree spatial resolution.

Table 4. Completeness and species richness of each phylum of the database. Completeness is the ratio between observed and estimated species richness (Soberón et al., 2007). For each phylum, the completeness is based on three estimators (Chao2, ICE and Jack1) and is averaged across all sites.

	Species richness	Average estimated richness	Completeness index
Database	1409	2251	0.63
Cnidaria	634	988	0.64
Echinodermata	86	115	0.75
Chordata	144	250	0.58
Brachiopoda	41	58	0.71
Bryozoa	238	330	0.72
Porifera	221	791	0.28
Annelida	30	88	0.34
Foraminifera	13	17	0.76

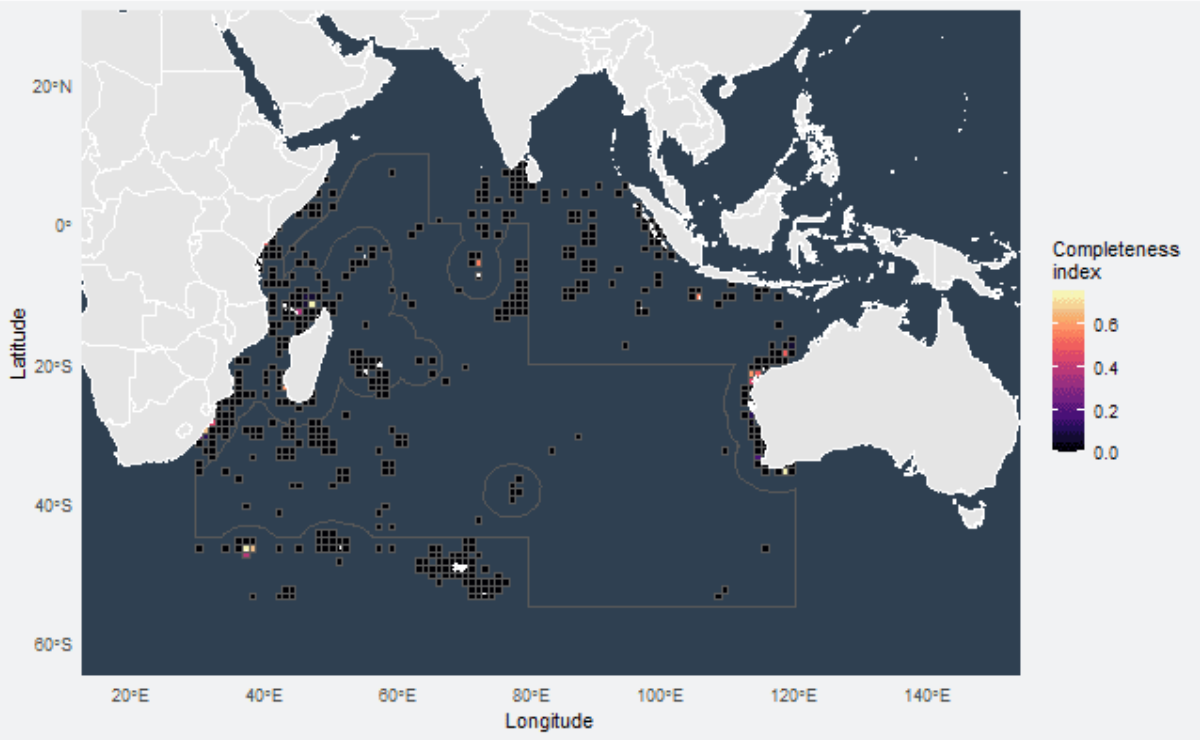

Figure 5. Completeness index. The ratio between the estimated richness by three estimators (ICE, Jackknife and Chao) and the observed richness. A higher index indicates less sampling bias.

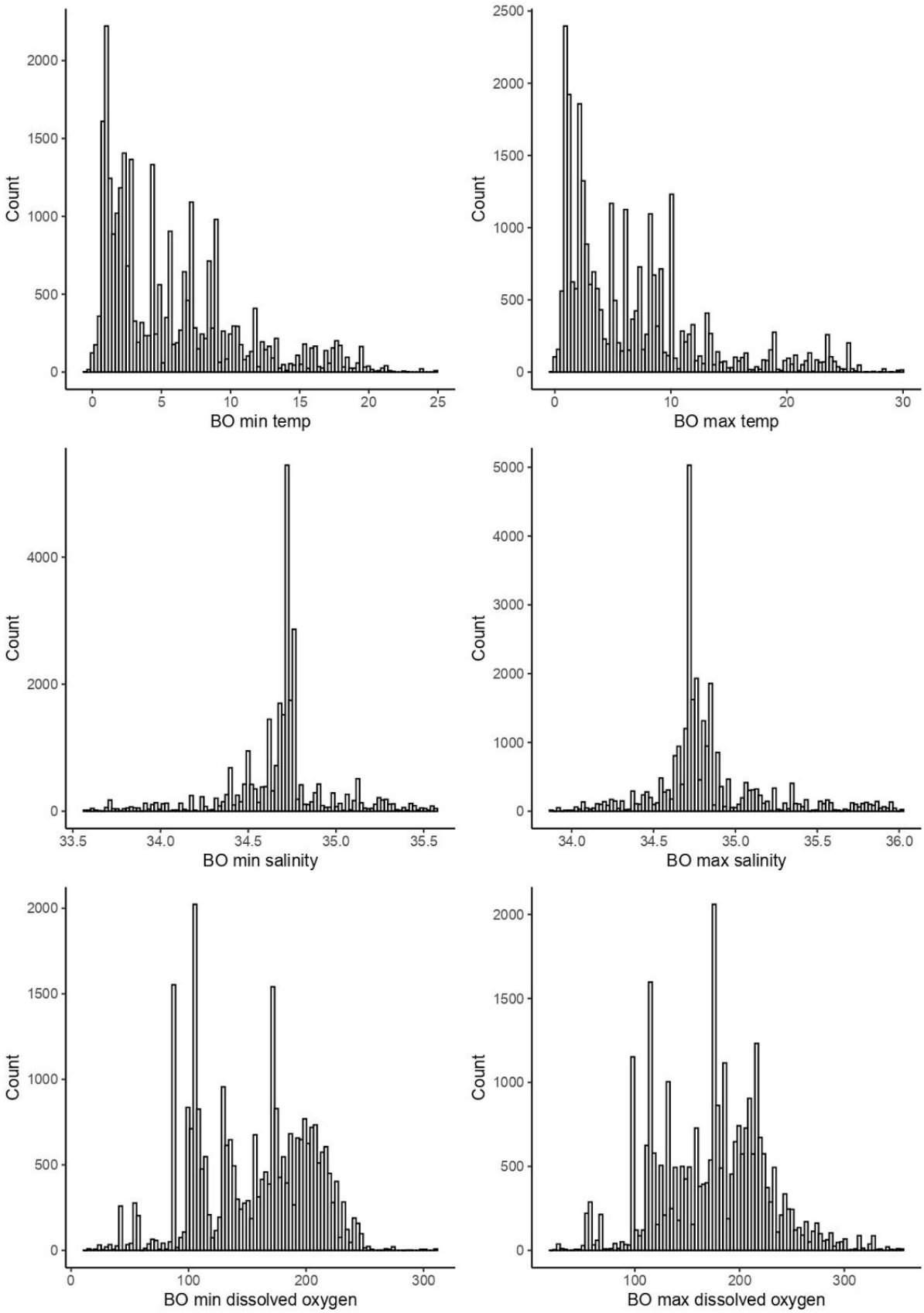

Figure 6. Environmental gradients of occurrence records at the native resolution of the environmental variables. The variables explored included the maximum (max) and minimum (min) of temperature, salinity, dissolved oxygen. 'BO' indicates 'Bio-Oracle dataset'.

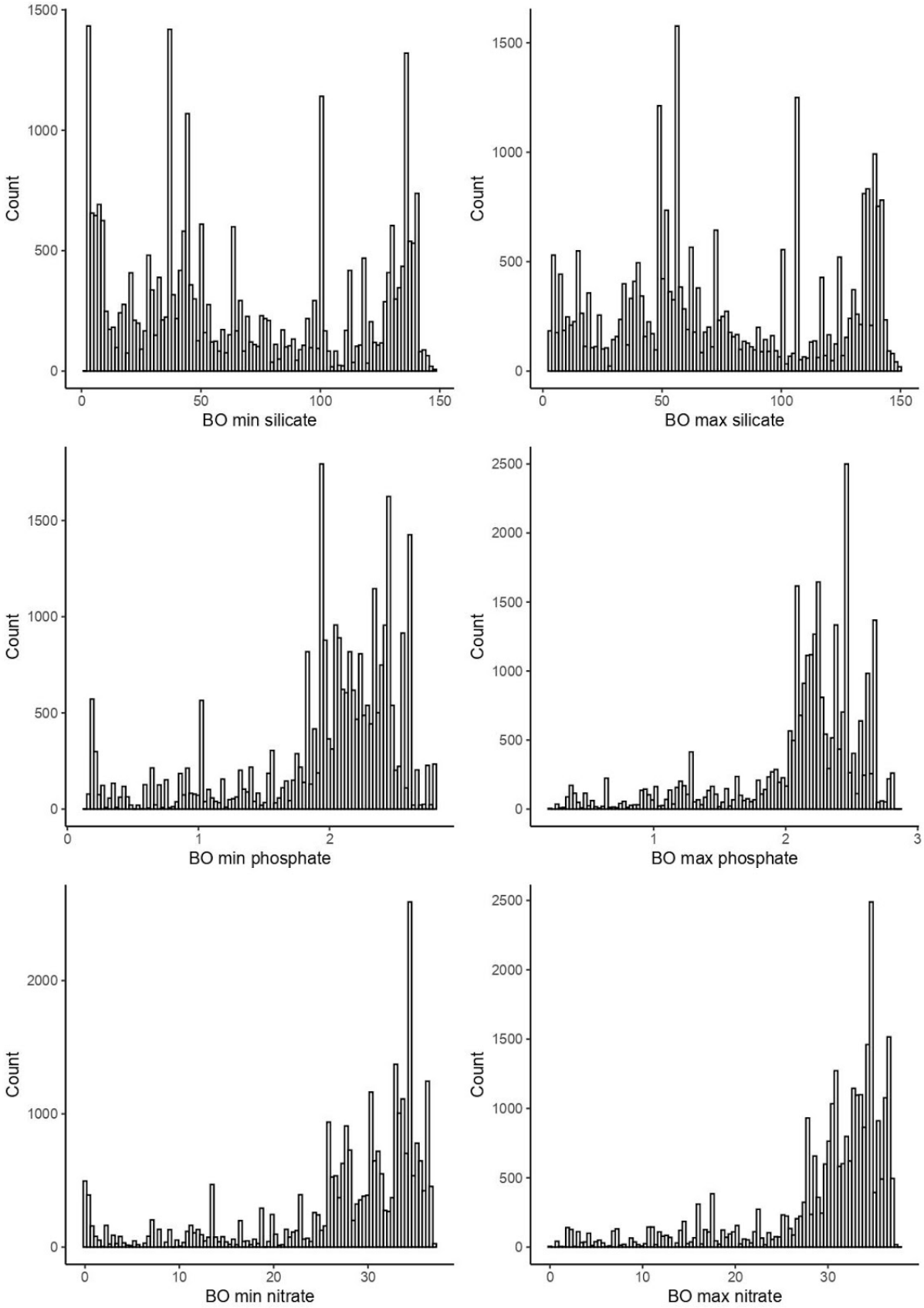

Figure 6. Environmental gradients of occurrence records at the native resolution of the environmental variables. The variables explored included the maximum (max) and minimum (min) of silicate, phosphate, nitrate concentrations. 'BO' indicates 'Bio-Oracle dataset'.

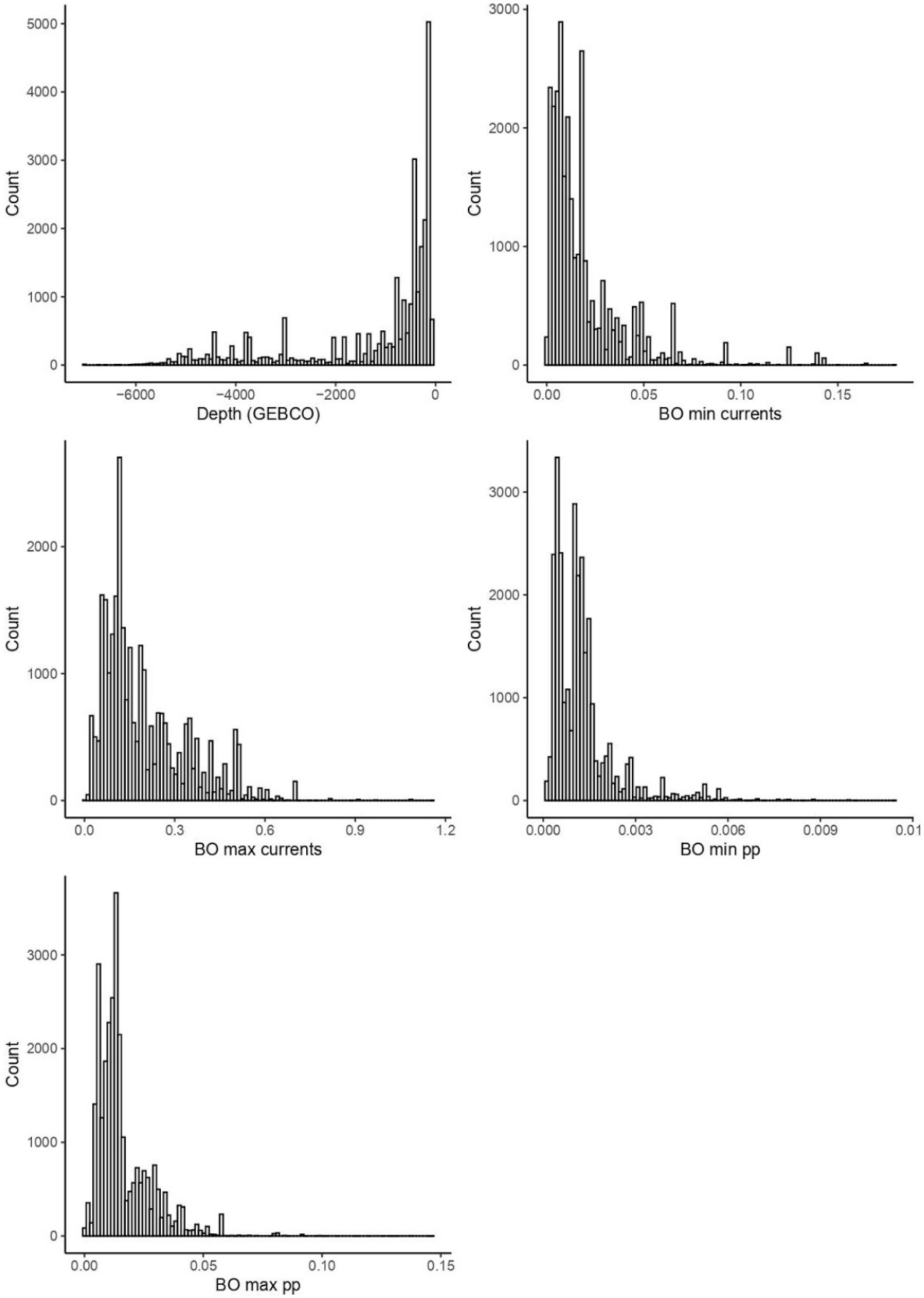

Figure 6. (Continued). Environmental gradients of occurrence records at the native resolution of the environmental variables. The variables explored included bathymetry, the maximum max) and minimum (min) of currents speed, and primary productivity at surface. 'BO' indicates 'Bio-Oracle dataset'.

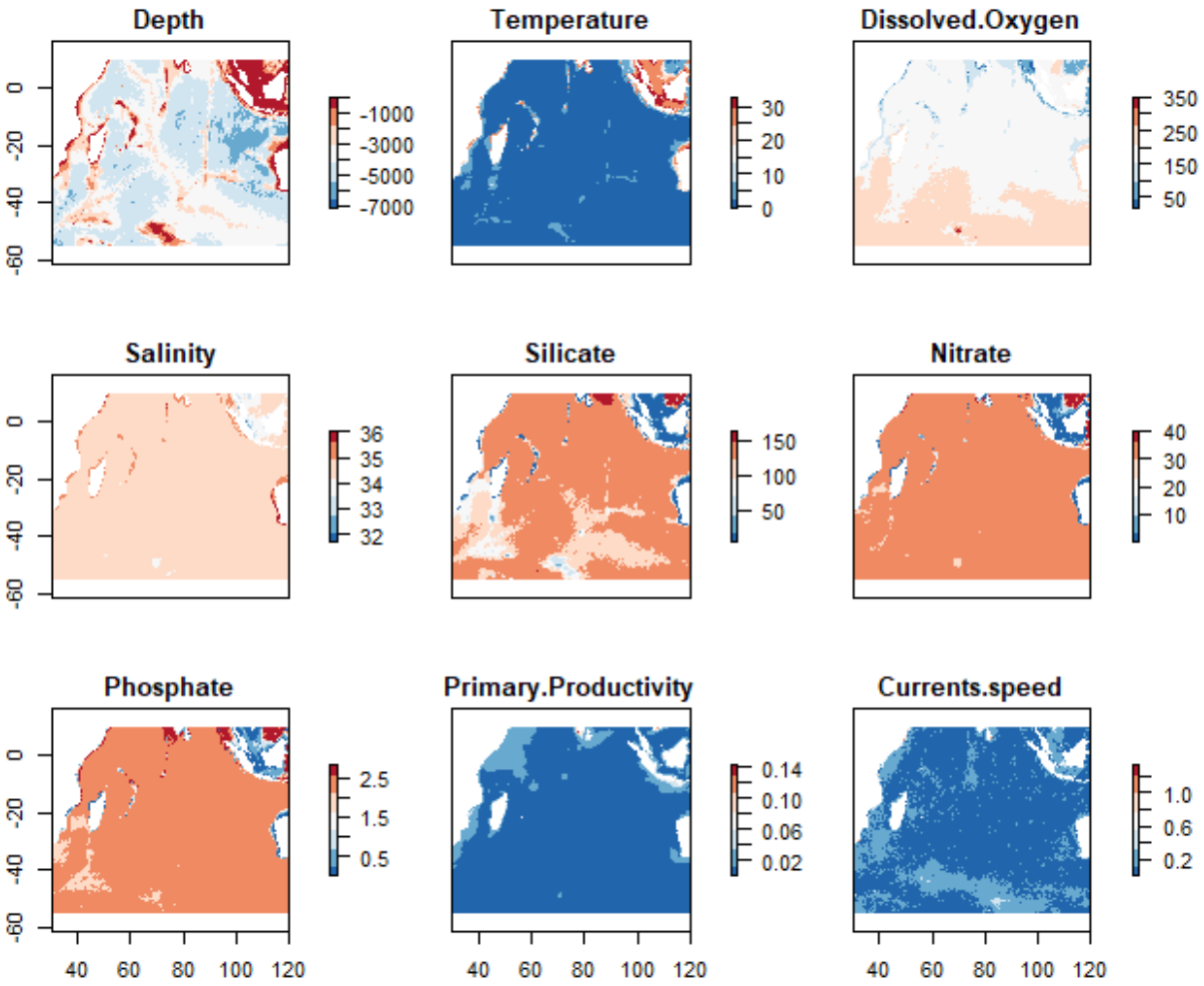

Figure 7. Environmental variables gathered for this study at their native resolutions.

Explorative predictive modelling

We tested the methodology to predict habitat suitability of indicator taxa on the Order Alcyonacea for which we had at least 30 rasterized observations, i.e. one observation per grid cell, for a total of 30 grid cells. We considered only a set of environmental variables as predictor variables for the purposes of the example, although a selection procedure based on Spearman correlation followed by a hierarchical clustering will be performed. Here we included seafloor dissolved oxygen concentration, temperature, salinity, currents speed, primary productivity at surface and depth.

Because the data was presence-only, the same amount of pseudo-absences as presence points were randomly generated using the 'biomod' R package. The models used in calibration were: general linear model (GLM), general additive model (GAM), artificial neural network (ANN), generalised multiple boosting models (GBM), flexible discriminant analysis (FDA) and random forest (RF). The performance of the models was assessed by cross-validation with the Boyce and Jaccard indices that are threshold-independent.

Once the models were calibrated and evaluated, we inspected the response curves to the environmental variables. The response curves show how the environmental suitability of the taxon responds to every individual calibrated model as well as the average trend for all models

(Figure 8). In this case, depth and seafloor temperature seemed to have an effect on the prediction of the taxon.

The ensemble model was then generated using the average prediction of all evaluated models (Figure 9A). The ensemble model can then be penalised with the standard deviation of the prediction of all models, a measure of uncertainty (Figure 9B). In the example, there are areas where all models predict habitat suitability with good certainty, such as the ocean ridges and continental shelves. However, on the other hand, there is uncertainty at the boundary of these suitable areas, and also uncertainty in some other areas of the SIOFA which might be suitable according to some models, but not for some other models, hence higher uncertainty.

Finally, the ensemble model can also be converted to a binary output of presence and absence of a taxon by using a threshold to determine the cut-off for a presence or an absence. Binary maps in combination of uncertainty measures are particularly useful as decision making tools in management scenarios.

Assessment of the implications of data quality and completeness for the bioregionalisation procedures

To briefly summarize the planned bioregionalisation scheme, we will provide three main types of bioregionalisation maps. First, we will provide maps of *observed* bioregions based on the *observed* distribution of VME indicator taxa. Second, we will provide *predictive* bioregions based on the individual *modelled* distributions of VME indicator taxa. Third, we will provide an alternative set of *predictive* bioregions based on the modelled relationship between *observed* bioregions and the environment. It is important to note the difference between the two types of *predictive* bioregions. The predictive bioregions based on individually modelled VME indicator taxa will account for the respective individual requirements (processes) of VME indicator taxa, but will omit the taxa for which there are not enough observations. The predictive bioregions based on the relationship between observed bioregions and the environment will include all VME indicator taxa, but will be based on a coarser process distinguishing between the major types of bioregions.

Overall, our first results evidence the paucity of distribution data available for VME indicator taxa in the SIOFA area. This finding has two main implications for the planned bioregionalisation scheme. First, the maps of *observed* bioregions based on observed data will likely be limited to a restricted subset of the overall SIOFA area, as illustrated in Figure 5. Second, mapping the *predictive* bioregions over the entire SIOFA area will be strongly based on the environmental data gathered here (Figure 7) and the chosen models. Consequently, adequate modelling procedures to estimate the environmental requirements of VME indicator taxa appear to be critical here and will require a thorough development and implementation, with a particular emphasis on evaluation of the uncertainties.

Given the number of records available for VME indicator taxa, we will explore modelling techniques dedicated to cases with small number of observations, such as Ensembles of Small Models, which have been reported to be superior to standard models in terms of performance and transferability (Breiner et al., 2018). However, regardless of the modelling procedure, the distribution of some VME indicator taxa will likely not be modelled because of

the lack of observations, which confirms our initial expectation that predictive bioregions based individually modelled VME taxa map will omit some of the VME indicator taxa.

This impossibility of accounting for all VME indicator taxa in the *predictive* bioregions based on individually modelled taxa emphasizes the importance of the other *predictive* approach which will model the relationship between bioregions and the environment. Altogether, these two approaches will provide complementary pictures of the distribution of bioregions for VME indicator taxa in the SIOFA area. Nevertheless, the shortfalls of available distribution data for VME indicator taxa suggest that both approaches will have important uncertainties, which we will estimate during the modelling procedure. In case the uncertainties are too important, we will consider a third type of predictive approach relying more on environmental data than the aforementioned ones. This third type of approach consists in simultaneously analysing biological data and environmental data in the process of generating bioregions (Hill et al., 2020), and constitutes a promising alternative in such data-poor contexts.

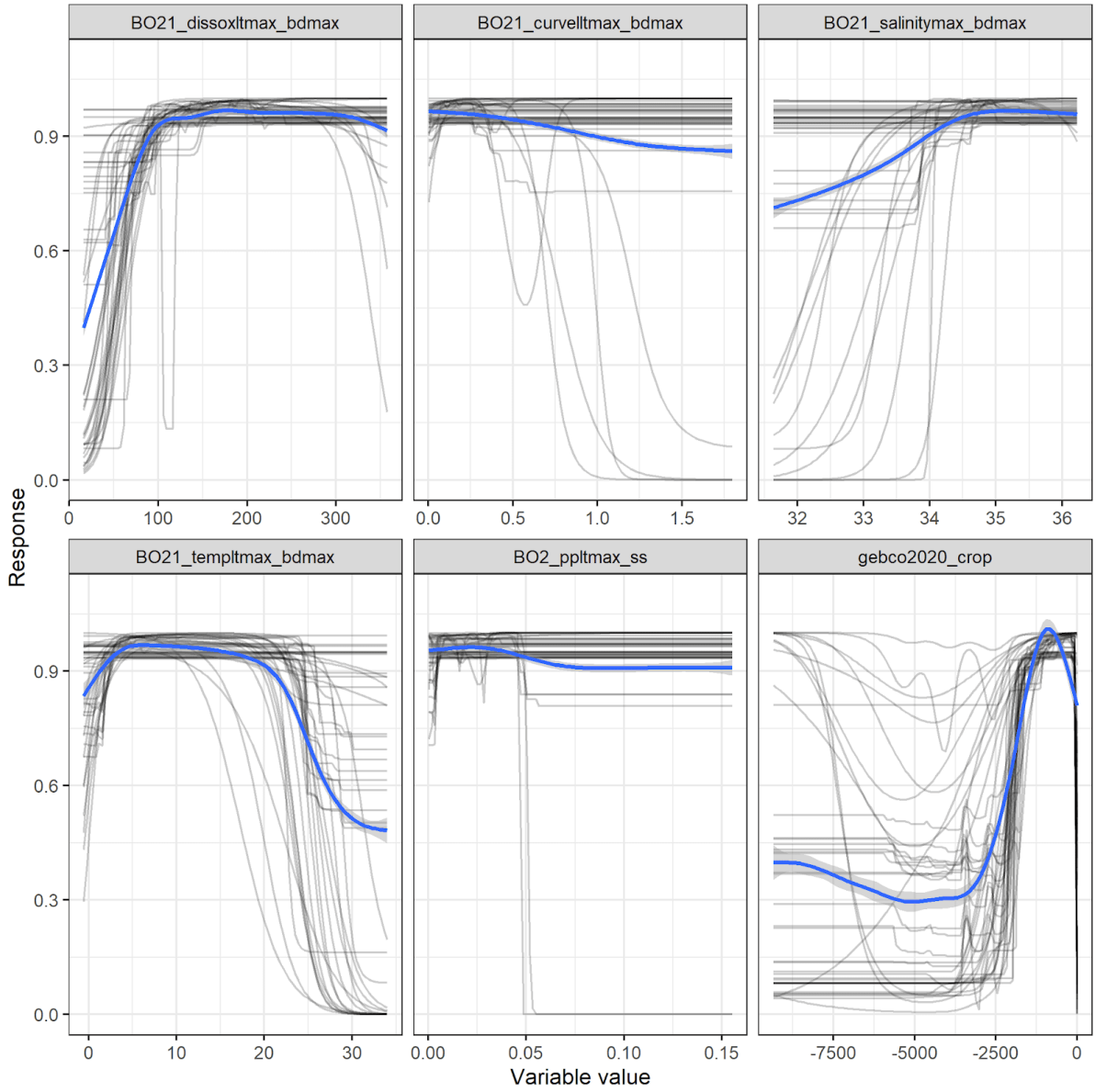

Figure 8. Response curves for all calibrated models. The graphs show the predicted probability of presence of the indicator taxa for each of the environmental variables selected. The thick blue line represents the average of all models. The response curve is the dependence of the probability of the presence on one predictor variable after averaging out the effects of the other predictor variables in the model.

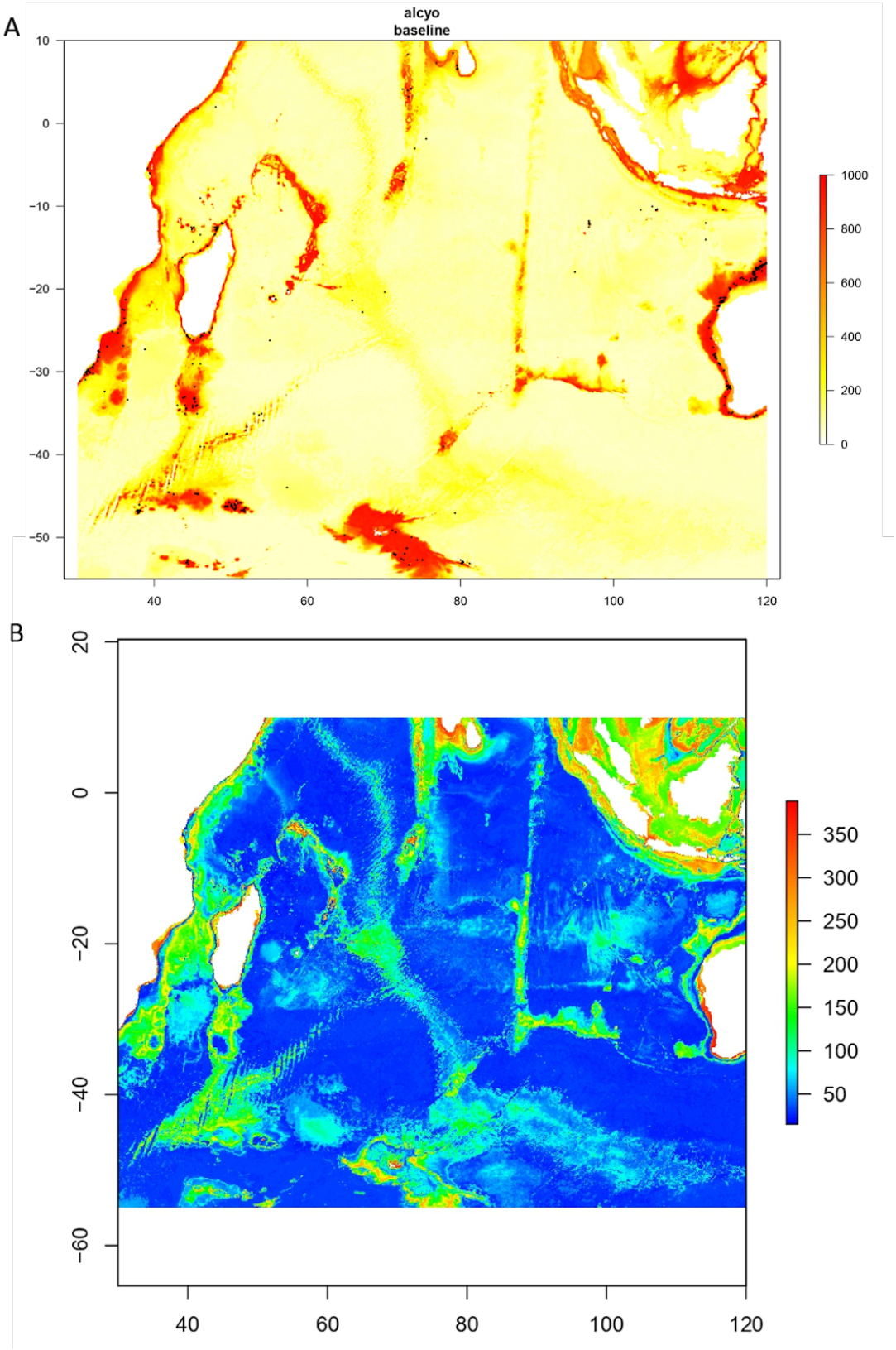

Figure 9. Ensemble models for the probability of presence of Alcyonacea. A. Continuous ensemble model. Higher suitability is represented by red colors. B. Standard deviation of the continuous model (A), with higher uncertainty in red colors.

References

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. Trends in ecology & evolution, 22(1), 42-47.

Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. Ecological Modelling, 157, 281–300.

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.

Breiner, F. T., Nobis, M. P., Bergamini, A., & Guisan, A. (2018). Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods in Ecology and Evolution*, *9*(4), 802-808.

Chiu, C. H., Wang, Y. T., Walther, B. A., & Chao, A. (2014). An improved nonparametric lower bound of species richness via a modified good–turing frequency formula. Biometrics, 70, 671–682.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802-813.

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. The Annals of Statistics, 19, 839–67.

Gotelli, N. J., & Colwell, R. K. (2011). Estimating species richness. In A. E. Magurran, & B. J. McGil (Eds.), Biological Diversity: Frontiers in measurement and assessment (pp. 39–54). Oxford: Oxford University Press.

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., ... & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography, 24(3), 276-292.

Hastie, T., & Tibshirani, R. (2004). Generalized additive models. Encyclopedia of statistical sciences. Chichester, UK: John Wiley & Sons Inc.

Hastie, T., Tibshirani, R., & Buja, A. (1994). Flexible discriminant analysis by optimal Scoring. Journal of the American Statistical Association, 89, 1255–1270.

Hill, N., Woolley, S. N., Foster, S., 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.

Leroy, B. (2012). *Utilisation des bases de données biodiversité pour la conservation des taxons d'invertébrés: indices de rareté des assemblages d'espèces et modèles de prédiction de répartition d'espèces* (Doctoral dissertation, Paris, Muséum national d'histoire naturelle).

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.

Leroy, B., Gallon, R., Feunteun, E., Robuchon, M., & Ysnel, F. (2017). Cross-taxon congruence in the rarity of subtidal rocky marine assemblages: No taxonomic shortcut for conservation monitoring. Ecological Indicators, 77, 239-249.

Louppe, V., Leroy, B., Herrel, A., & Veron, G. (2019). Current and future climatic regions favourable for a globally introduced wild carnivore, the raccoon Procyon lotor. *Scientific reports*, *9*(1), 1-13.

McCullagh, P. (1984). Generalized linear models. European Journal of Operational Research, 16, 285–292.

Northwest Atlantic Fisheries Organisation. (2019). Conservation and Enforcement Measures. NAFO/COM Doc. 19-01.

North East Atlantic Fisheries Organisation. (2014). Recommendation 19:2014. Protection of VMEs in NEAFC Regulatory Area.

Ramirez-Llodra, E., Tyler, P. A., Baker, M. C., Bergstad, O. A., Clark, M. R., Escobar, E., ... & Van Dover, C. L. (2011). Man and the last great wilderness: human impact on the deep sea. PLoS one, 6(8), e22588.

Rice, J., Gjerde, K. M., Ardron, J., Arico, S., Cresswell, I., Escobar, E., ... & 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.

Ripley, B. D. (2007). Pattern recognition and neural networks. Cambridge, UK: Cambridge University Press.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., ... & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913-929.

Schickele, A., Leroy, B., Beaugrand, G., Goberville, E., Hattab, T., Francour, P., & Raybaud, V. (2020). Modelling European small pelagic fish distribution: Methodological insights. Ecological Modelling, 416, 108902.

Soberón, J., Jiménez, R., Golubov, J., & Koleff, P. (2007). Assessing completeness of biodiversity databases at different spatial scales. Ecography, 30(1), 152-160.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369-373.

UNGA (2006). *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.* New York, NY: United Nations General Assembly.

Walther, B. A., & Moore, J. L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, *28*(6), 815-829.

WoRMS Editorial Board (2021). World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. Accessed 2021-01-29. doi:10.14284/170

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., ... & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods in Ecology and Evolution, 10(5), 744-751.