

CONTRIBUTED PAPER

Structured machine learning modeling to support conservation of deep-sea benthic biodiversity

Gustavo Fonseca¹  | Danilo C. Vieira¹  | Juliane C. Carneiro⁵ | Renato S. Carreira² | Milena Ceccopieri² | Thais N. Corbisier³ | Adriana Galindo Dalto⁴ | Alberto G. Figueiredo Jr.⁵ | Fabiane Gallucci¹ | Paula Gheller³ | Simone Brito de Jesus¹ | Helena Passeri Lavrado⁴ | Letícia Lazzari² | Eduardo Hilzendeger Marcon⁷ | Daniel Leite Moreira⁷ | Rafael Bendayan de Moura⁴ | Ellen Pape⁶ | Ana Cláudia Aoki Santarosa³ | João Regis dos Santos Filho⁸ | Silvia Helena de Mello e Sousa³ | Thaisa Marques Vicente³ | Luciana Erika Yaginuma¹  | Cinthia Yamashita³ | Wandrey Watanabe¹

¹Instituto do Mar, UNIFESP, Santos, Brazil

²Departamento de Química, PUC-Rio, Rio de Janeiro, Brazil

³Instituto Oceanográfico, USP, São Paulo, Brazil

⁴Departamento de Biologia Marinha, IB, UFRJ, Rio de Janeiro, Brazil

⁵Geology and Geophysics Department, Fluminense Federal University, Niteroi, Brazil

⁶Marine Biology Research Group, Ghent University, Ghent, Belgium

⁷PETROBRAS Research Center, Centro de Pesquisas Leopoldo Américo Miguez de Mello (CENPES), Rio de Janeiro, Brazil

⁸Faculdade de Oceanografia, UERJ, Rio de Janeiro, Brazil

Correspondence

Gustavo Fonseca, Instituto do Mar, UNIFESP, Carvalho de Mendonça-144, Santos - SP, 11070-100, Brazil.
Email: gfonseca@unifesp.br

Article impact statement: Structured models predict deep-sea biodiversity from simulated data, enhancing conservation efforts and monitoring efficiency.

Funding information

Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPQ, Grant/Award Number: 306780/2022-4; Agência Nacional do Petróleo, Gás Natural e Biocombustíveis

Abstract

Biodiversity monitoring programs need to deliver accurate, timely, and actionable predictions. To establish a predictive monitoring program for deep-sea benthos of the Santos Basin, Brazil, we developed a two-stage structured model that allowed comparison of biodiversity predictions obtained from environmental simulations (2M-Sim). We also modeled the environmental variables as a function of spatial and temporal variables and compared this model's predictions with predictions obtained from real environmental data (2M). We built unstructured models (1M) as references to evaluate whether the proposed structured approach was reliable. We expected no significant differences between 1M and 2M or between 2M and 2M-Sim. Data were obtained from 100 stations at depths of 25–2400 m during two surveys (2019 and 2021). In our model framework, we used 12 benthic macro- and meiofaunal variables, 44 sediment and water column environmental variables, and four spatial and temporal variables. We applied a random forest algorithm to the structured and unstructured models. All comparisons were performed with 20% of the dataset set aside for validation. The average accuracy was 72%, 69%, and 68% for the 1M, 2M, and 2M-Sim models, respectively. Accuracies of 2M ranged from 38% to 84% and were generally higher for macrofauna. The observed accuracy loss from 1M to 2M (3%) and from 2M to 2M-Sim

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Conservation Biology* published by Wiley Periodicals LLC on behalf of Society for Conservation Biology.

(1%) was not significant for any biodiversity variable. The 2M model identified 30 significant environmental variables; bottom water parameters and sedimentary phytopigment and carbonate concentrations were the best predictors. Our approach supports biodiversity conservation by optimizing data needs and future sampling and by guiding data-driven management decisions for benthic biodiversity.

KEYWORDS

structure modeling, benthos, continental margin, random forest, environmental essential variables, monitoring

INTRODUCTION

Conservation aims to protect and manage natural environments by preserving their biodiversity, functionality, and resilience against anthropogenic disturbances. This requires characterization of the ecosystem baseline conditions through observation and analysis of multiple environmental parameters such that natural variability is distinguishable from variability caused by climate and other anthropogenic changes (Linder et al., 2017; Sparrow et al., 2020). Given the complexity of ecosystems and their vulnerability to human pressures, environmental monitoring programs must quickly obtain and process data and make accurate predictions to support effective management decisions. Most environmental monitoring programs collect extensive data but still fail to model them effectively (Lovett et al., 2007; Rowland et al., 2021; Schreiber et al., 2022). Traditionally, ecological studies have modeled biodiversity indicators with all the environmental, spatial, and temporal variables simultaneously (hereafter *unstructured approach*). This approach can be used to determine the proportion of variance explained by each of the three variables (Dray et al., 2012). Its disadvantages include a limited capacity to project biodiversity measures without collecting data on the entire set of predictors and the inability to extrapolate the biodiversity measures to less studied areas or the future (Simmonds et al., 2024). An unstructured approach is ineffective for ecosystem monitoring programs that generally involve multiple biodiversity indicators, need to anticipate unwanted ecosystem changes, and require timely management decisions. An alternative is to use a structured modeling framework, in which the modeling steps are explicitly linked based on cause–effect relationships (Schuwirth et al., 2019). The choice of modeling approach is thus fundamental to making robust inferences related to spatial and temporal patterns, to minimizing the amount of data needed to make inferences, and to guiding the best management decisions (Gerhardt et al., 2024; Walling & Vaneeckhaute, 2020; Zurell et al., 2022).

A critical aspect to be considered when implementing a model-oriented ecosystem monitoring program is the amount of information needed to make predictions. Collecting and processing biodiversity data requires significant resources, especially for offshore marine biodiversity (Nigard et al., 2016). Efficient monitoring programs for marine ecosystems should be based on essential environmental variables (EEVs) that are easily obtained and processed and capable of providing accurate predictions of the target variables (Ditria et al., 2022). It

includes biological and physicochemical variables that can be obtained remotely or in situ, preferably with sensors to speed data processing (Muller-Karger et al., 2018). Particularly for the benthos along continental margins, EEVs operate across spatial and temporal scales in pelagic and benthic systems. Priority should be given to those related to food availability, oxygen concentration, and mean grain sizes, which vary with depth and influence the community structure and lead to distinct zonation patterns from shelf to abyssal plains (Levin et al., 2001). Additional EVVs to be considered are related to water masses. Water masses have unique physicochemical signatures that regulate ocean productivity and determine the carbon flux to the seafloor, which has a positive effect on benthic biodiversity (Danovaro et al., 2020; Rafaelli et al., 2003; Smith et al., 2008). Therefore, understanding how these drivers collectively structure biodiversity and which can be used as proxies of target indicators is critical for designing monitoring frameworks that balance the complexity of the oceanographic processes with feasibility, particularly in areas with the potential for extractive industries (Reyers et al., 2017).

We sought to develop a structured modeling framework grounded in a hierarchical framework in which spatial and temporal drivers (e.g., bathymetry, seasons) shape environmental conditions, which in turn regulate the benthic biodiversity. With spatial drivers, structured modeling increases analytical complexity, which can decrease performance relative to an unstructured approach, but it makes data acquisition easier and improves the generalization potential of its predictions (Eisenhauer et al., 2015). Explicitly incorporating this hierarchy allows simulation of environmental conditions based on the spatial and temporal drivers. The results of the simulations can then be used to estimate biodiversity patterns. We focused on the practical trade-off between model accuracy and monitoring efficiency by specifically testing whether a structured model generates predictions as good as an unstructured model. We also examined whether the performance of a structured model based on the simulation of the environmental conditions is comparable to the performance of a model generated from real data.

We used machine learning (ML) algorithms as the analytical foundation for all models. For oceanography, ML models offer advantages over classical statistical methods, including flexibility in handling various data types, ability to address complex problems and confirm predictions on unseen data, and potential to rank predictors in a rigorous hypothesis testing framework (Fonseca & Vieira, 2023). Unlike classical linear models, which

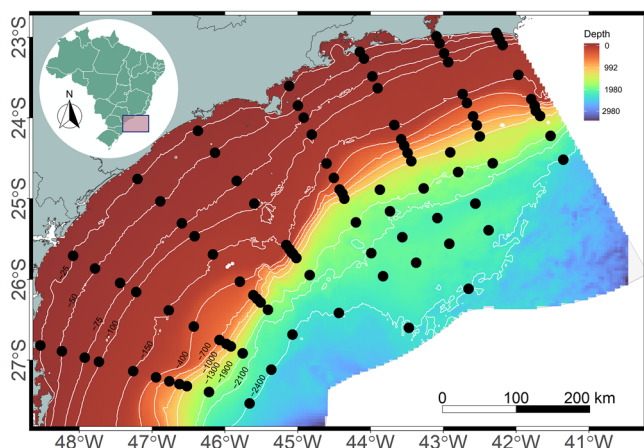


FIGURE 1 Sampling stations (black dots), isobaths, and the bathymetric color gradient of the Santos Basin, Brazil (grid resolution 2×2 km).

rely on strict statistical assumptions (e.g., linearity, normality, homoscedasticity), the core premise of ML is that the training and test data are exchangeable—that is, the training set is representative of the population from which the test set is drawn. Structured ML approaches advance monitoring methods by demonstrating how simulations rooted in EEVs can streamline conservation efforts without compromising insights.

METHODS

Study area and sampling design

The Santos Basin is in southeastern Brazil between the Campos Basin and the Pelotas Basin. It is delimited to the north by Cabo Frio High (22°S) and to the south by Florianópolis High (28.5°S) (Moreira et al., 2023). The basin occupies approximately $350,000 \text{ km}^2$ and borders four Brazilian states along 271 km of the southeastern coast. It reaches 3000 m water depth in the São Paulo Plateau. The continental shelf is narrower (70 km) in the Cabo Frio region (Rio de Janeiro state, RJ) and wider off the city of Santos (230 km), in São Paulo state (SP). Mean declivity is $\sim 0.08^{\circ}$, and shelf-break depth varies from 120 to 180 m (Figueiredo Jr. et al., 2023).

Abiotic and biotic benthic data were obtained from the Santos Project (Santos Basin Environmental Characterization) (Moreira et al., 2023). These data were used by the Regional Environmental Characterization Project in support of Petrobras' Santos Basin drilling licensing process (operation license 1006/2011) led by the Brazilian Environmental Agency (IBAMA). A total of 88 sampling stations were distributed along eight transects perpendicular to the coast and at 11 isobaths (25, 50, 75, 100, 150, 400, 1000, 1300, 1900, and 2400 m) (Figure 1). Twelve additional stations were sampled from 1900 to 2400 m in the São Paulo Plateau, where most of the oil and gas production takes place. Sampling cruises were conducted during 2019 and 2021 (Moreira et al., 2023). Hereafter, we refer to stations at <200 m as being on the continental shelf: 400–700 m,

upper slope; 100–1300 m, lower slope; and >1300 m, São Paulo Plateau. In addition to the sampling stations, a 2×2 -km bathymetrical grid was generated that had 100,555 data points for each survey.

Sampling and sample processing

Sediment samples were taken in three replicates with a spade-type box corer (0.25 m^2 surface area), or a modified Van Veen grab (231 L, 0.75 m^2 surface area) was used when the substrate was coarser (sixth station down from the southern- and northern-most transects). To describe the benthos, macrofauna (animals $>300 \mu\text{m}$) and meiofauna (animals $>0.40 \mu\text{m}$) were sampled. At each station, macrofauna samples were obtained from nine juxtaposed cylindrical corers of 10 cm^2 (total 900 cm^2) inserted in the top 10 cm of the sediment. The meiofauna samples were taken with a cylindrical corer of 5 cm diameter and 10 cm height (19.63 cm^2 area). All samples were fixed onboard in a 10% formaldehyde solution buffered with borax. Twelve benthic community variables were considered: total macrofauna abundance (individuals/ m^2), meiofauna abundance (individuals/ 10 cm^2), macrofauna family richness, meiofauna taxa richness, total macro- and meiofaunal polychaeta abundance, total macrofaunal annelid abundance, total abundance of macrofaunal mollusks and crustaceans, and total abundance of meiofaunal nematodes, copepods, and kinorhynchans. A total of 44 environmental variables were analyzed: grain size, parameters of the organic matter, and parameters of the water column 10 m above the bottom (Appendix S1). All the variables listed here are required by IBAMA in the environmental characterization program (IBAMA, 2019). Further details on sampling and analyses of the environmental variables are available in Figueiredo Jr. et al. (2023), Gallucci et al. (2023), Moreira et al. (2023), and Moura et al. (2023).

Modeling frameworks

We used two distinct modeling frameworks to predict benthic biodiversity patterns from environmental drivers (Appendix S2). This allowed the comparison of a novel structured approach against a traditional unstructured approach commonly used in benthic ecology. The unstructured model (1M) directly linked measured macrofauna and meiofauna data to in situ measurements of environmental variables (organic matter, sediment properties, and bottom water parameters) and geographical coordinates in a single random forest model. In contrast, the structured model had two stages. In the first stage, we developed individual environmental models to predict the spatial distribution of the environmental variables, such as organic matter, sediment, and bottom water parameters. In the second stage, outputs from the first stage—both the original in situ measurements and their model-predicted (simulated) values—were used as predictors in separate random forest models for biodiversity. This two-stage design allowed us to generate and compare two distinct types of biodiversity predictions: one based on

real environmental data (2M) and the other based on simulated environmental data (2M-Sim). The latter prediction type allows predictions in areas that lack actual environmental data. We evaluated all models with cross-validation (CV) and analyzed feature importance (FI) to identify the primary environmental drivers. For FI, all potential predictor variables were input directly to the random forest algorithm. The statistical significance (p) of FI was assessed using a permutation-based approach in which each predictor variable was permuted multiple times to sever its relationship with the response variable. We compared the resulting decay in model performance with the full model (Altman et al., 2010).

Analytical workflow

Prior to the analyses, replicates were averaged per station to better characterize the local conditions for the purpose of modeling the regional patterns (Gallucci et al., 2023), and biodiversity measures were transformed to $\log_{10}$. To enable a comprehensive comparison across the modeling structures and directly quantify how well a model trained on one set of data performs relative to another, unseen set, the dataset was split into two parts: 80% of the data for training models and 20% for model validation (i.e., 38 sampling stations). Model performance was compared using the unseen 20% of the data. Folds were created by randomly splitting the entire dataset (all 100 stations from each survey), not by grouping data by survey or transect. This means that in each iteration of the model training, the training set and the test set contained distinct random subsets of points from both surveys. All models were performed with the random forest regression algorithm. Although we are aware of the importance of testing multiple ML algorithms (Fonseca & Vieira, 2023), we have decided to perform all models with random forest to allow for comparisons across the models. In addition, random forest is generally among the top-rated models for comparing multiple algorithms (de Jesus et al., 2023; Yaginuma et al., 2025). The parameterization of all the models was the same: a training phase with 500 trees, tune length of 5, a fivefold CV, and five repetitions. The remaining parameters were kept in the default format (Vieira et al., 2025). For each biodiversity model, the FI analysis was performed to retrieve the significant environmental variables selected by the models ($p < 0.05$) and the number at which each variable appeared as a root when all trees were considered. We visualized these results in a Sankey diagram. To verify that the use of random CV did not introduce bias due to spatial autocorrelation and returned over-fitted models, we tested the spatial independence of residuals from the validation sets with Moran's I with k nearest-neighbor spatial weights. Significance was evaluated after applying the Bonferroni correction to account for multiple comparisons across biodiversity descriptors.

Model comparison

All model comparisons were performed using 20% of the data that were not used in the training phase of the random forest models. First, the differences between the absolute errors of the 1M and 2M were assessed for normality. If the differences were normally distributed, a one-tailed paired t -test was used under the assumption that significant differences would be detected if the errors increased from 1M to 2M. Otherwise, a paired Wilcoxon test was performed. Such an increase in error can be due to processes related to the effects of time or space on biodiversity not being captured by the environmental variables (Eisenhauer et al., 2015). The second comparison evaluated the differences between the errors of the 2M and 2M-Sim models, following the same hypothesis testing procedure. Significant changes at this stage would indicate that biodiversity estimates derived from the environmental simulations had higher prediction errors than those based directly on environmental data. All paired tests considered a significant threshold of $p < 0.05$. In addition, a bootstrap resampling procedure (10,000 iterations) was applied to estimate empirical confidence intervals and one-tailed p values. This procedure complemented the paired tests with a nonparametric assessment of model differences.

Uncertainty propagation analyses

To explicitly quantify the propagation of uncertainty from the environmental simulations to the biodiversity predictions (2M-Sim), we used a Monte Carlo resampling procedure. For each validation site, we first calculated residuals between observed and simulated environmental predictors obtained from the 2M-stage models. These residuals represented the discrepancies between measured conditions (used in 2M) and their simulated counterparts (used in 2M-Sim). In the Monte Carlo procedure, 1000 perturbed versions of the simulated predictors were generated for each site by randomly resampling residuals with replacement. These variable environmental conditions were then propagated through the biodiversity models trained with observed environmental predictors. To account for the intrinsic predictive variability of the biodiversity models, residuals from the biodiversity predictions themselves were also resampled and added to each simulation draw. As a result, each site and descriptor produced an ensemble of predictions that reflected both sources of uncertainty: environmental simulation errors and biodiversity model errors. The resulting ensembles were summarized by their mean, standard deviation, and empirical prediction intervals (PIs) at 90% confidence. To ensure nominal coverage, a conformal calibration step was applied. Absolute errors between observed and mean predictions were used to adjust the interval limits until empirical coverage reached the desired level (Zhou et al., 2025). Interval widths and coverage were then aggregated across descriptors to evaluate and compare the uncertainties across the models.

Spatial prediction and mapping

Our models generated raster maps in which each pixel's biodiversity value was predicted by a random forest model based on its unique environmental data. These point-by-point predictions created spatial patterns based solely on the learned environment–biodiversity relationship. Consequently, areas with similar environmental conditions—regardless of geographic proximity—were predicted to support similar biodiversity.

Software

We used iMESc (Vieira et al., 2025), an interactive ML application built in R, to perform all data analyses. This platform was selected because it was designed specifically for use with environmental data. It integrates data preprocessing, ML, spatial analysis, and visualization in a reproducible framework. The iMESc application was designed for managing multidisciplinary datasets (e.g., numeric, spatial, categorical) in a unified structure (Datalist), and its dedicated tools for ecological modeling (e.g., Biodiversity tools) and spatial prediction make it uniquely suited for our complex analytical workflow, from model training to the generation of biodiversity prediction maps. The Monte Carlo simulations and conformal calibration procedures for uncertainty propagation were implemented directly in R (R Core Team, 2025). All one-tailed paired *t*-tests and Wilcoxon tests were also conducted in R.

RESULTS

Model comparisons

The random forest analyses for the 12 descriptors of the fauna in the 1M unstructured model showed that the average model performance was 72% (Table 1). The performance varied from 54% for meiofaunal kinorhynch abundance to 85% for the total abundance of macrofauna. Overall, the macrofauna variables performed better than the meiofauna variables. The accuracy of the structured models (2M) decreased on average by 3% relative to the unstructured 1M (Table 1), but this difference was mostly not significant (Appendices S3–S5). Results from the 2M-Sim returned, on average, an accuracy of 68%. The best model was for total macrofauna abundance and mollusks (81% accuracy). The paired tests between 2M and 2M-Sim were all nonsignificant (Appendices S3–S5). Bootstrap resampling confirmed these nonsignificant results, with the exception of Mollusca and Crustacea between 1M and 2M (Appendices S6 & S7). Residuals from the validation sets showed that biodiversity descriptors were not spatially autocorrelated after Bonferroni correction (Moran's *I*) (Appendix S8).

The environmental models used to generate predictions of the 2M-Sim reached, on average, 64% accuracy with the test data (Appendix S1). The models achieved over 80% accuracy for

total gravel, very fine gravel, fine sand, and carbonates, and for water temperature, salinity, density, and dissolved oxygen. The best models of organic matter ranged from 78% to 79% accuracy for the chlorophyll-a-to-phaeopigments ratio, total organic matter, nitrogen, and inorganic phosphorus.

Feature importance

From 44 environmental variables, the FI analysis from the 2M selected 30 variables as significant for at least one benthic community descriptor (Figure 2). Water density, water temperature, and water dissolved oxygen were the three variables that occurred most often in the regression models. The sedimentary chlorophyll-a-to-phaeopigments ratio, concentration of phaeopigments, and percentage of carbonates were also important variables.

Biodiversity and environmental patterns

The abundance of macrofauna varied on average from 178 to 16,167 individuals per 1 m² (Appendix S9). The number of macrofauna taxa varied on average from 6.3 to 17.3 per station. Polychaetes were the most abundant group of the macrofauna (63%), followed by crustaceans (21%) and mollusks (9%). The abundance of meiofauna varied on average from 29.88 to 2623.89 individuals per 10 cm² (Appendix S9). Nematoda was the dominant group (83%), followed by copepods (12%), polychaetes (2%), and kinorhynchs (1%).

Meiofauna and nematode abundance were highest down to the shelf break (25–150 m) and at the southern portion of the continental shelf during the first survey (Figure 3). These temporal changes were more evident for meiofauna richness and the abundance of the meiofauna taxa. All parameters of the macrofauna showed a clear bathymetrical gradient; values were highest along the continental shelf and upper slope and lowest at the deeper stations (Figure 3). The richness of the macrofauna was particularly high along the slope compared with the continental shelf. Different from the meiofauna, the macrofauna descriptors did not show clear temporal trends.

Water density at the deep seafloor was above 1028 kg/m³ during both surveys. The southern portion of the continental shelf had much lower water density during the first survey (<1023 kg/m³) (Figure 4). Water temperature, dissolved oxygen, and salinity also showed clear spatial and temporal patterns. They all differed among the continental shelf stations, the slope, and São Paulo Plateau stations. Although temperature and salinity varied temporally along the continental shelf, dissolved oxygen varied temporally in the deeper waters.

The concentration of phaeopigments, the chlorophyll-a-to-phaeopigments ratio, and carbonate content in the sediment did not show clear temporal trends but had marked spatial variability (Figure 5). Phaeopigments were higher from 100- to 150-m depths. The chlorophyll-a-to-phaeopigments ratio was highest at the deep-sea stations, and carbonates were higher at the shelf break and in the deep sea.

TABLE 1 The R^2 and mean absolute error (MAE) results of regression models of deep-sea biodiversity variables as a function of the environmental variable for each model structure.

Biodiversity group	Variables	1M ^a		2M ^b		2M-Sim ^c	
		R^2	MAE	R^2	MAE	R^2	MAE
Meiofauna	n	0.75	0.16	0.71	0.16	0.71	0.17
	nema	0.74	0.16	0.72	0.17	0.70	0.17
	cop	0.66	0.20	0.57	0.22	0.56	0.20
	kino	0.54	0.42	0.46	0.45	0.54	0.43
	poly	0.69	0.23	0.70	0.23	0.67	0.22
	s	0.53	0.04	0.38	0.05	0.44	0.05
Macrofauna	n	0.85	0.12	0.84	0.13	0.81	0.13
	annel	0.81	0.14	0.79	0.14	0.78	0.14
	poly	0.81	0.14	0.80	0.14	0.78	0.14
	crust	0.61	0.22	0.61	0.23	0.56	0.23
	moll	0.82	0.24	0.81	0.24	0.81	0.25
	s	0.83	0.07	0.84	0.07	0.75	0.08
Average		0.72	0.18	0.69	0.18	0.68	0.18

Abbreviations: annel, Annelida; cop, Copepoda; crust, Crustacea; kino, Kinorhyncha; moll, Mollusca; n, abundance; nema, Nematoda; poly, Polychaete; S, taxonomic richness.

^aUnstructured model.

^bStructured model.

^cStructured model with simulations.

Uncertainty

Results from the Monte Carlo simulations propagated the inherent uncertainty from the simulated environmental predictors into the final biodiversity forecasts. For all biodiversity descriptors, the empirical coverage of the 90% PIs (PI90) met or exceeded the nominal target of 90% (Appendix S12a), confirming that the combined uncertainty from both the environmental simulations and the biodiversity models was accurately quantified. The average width of these PI90 intervals, however, varied among descriptors (Appendix S12b). Wider intervals were associated with kinorhynchs and mollusks, and narrower intervals were observed for richness metrics. At the individual site level, the PIs successfully captured the observed biodiversity values for most descriptors (Appendix S13).

DISCUSSION

Oceanographic processes and their influence on spatiotemporal patterns in benthic communities

The oceanographic aspects of the Santos Basin have been investigated thoroughly in recent years in response to industrial interests (Fonseca et al., 2023). Our results confirmed some previous findings. At the basin scale, benthos spatial gradients, especially those related to bathymetry, were much stronger than those related to temporal changes. Temporal fluctuations in biodiversity did occur but were location- and process-specific. For instance, one important oceanographic process at the southeast Brazilian continental shelf was the

intrusion of a cold nutrient-rich water mass with low salinity from the Sub-Antarctic Argentinian shelf, which brings the runoff from the La Plata River and the Patos Lagoon (Brandini et al., 2018; Sasaki et al., 2024). These waters are carried by the Brazilian Coastal Current, which flows northward and is more intense during austral winters (de Souza & Robinson, 2004). The observed low-density bottom waters were a strong indication that those waters were present during the sampling in 2019; they reached up to 25°S. The influence of the southern waters promotes primary production in the pelagic system (Brandini, 1990; Brandini et al., 2018), and, as we observed, it sustains a high concentration of phaeopigments in the sediments in both periods. This oceanographic process is indicated mainly by a particular set of benthic biodiversity descriptors. Meiofauna total abundance and taxonomic richness and copepod, kinorhynch, and polychaete abundances were high in the presence of this water mass. It is thus plausible to assume that the observed temporal change in benthic community parameters could be related to their rapid response to organic enrichment, as shown in other studies (Moodley et al., 2005; Witte et al., 2003).

The northern portion of the continental shelf in the Santos Basin is under the influence of coastal upwelling events. During these events, colder and nutrient-rich South Atlantic Central Water (SACW) rises to the surface, promoting phytoplankton productivity and the deposition of high quantities of organic matter to the bottom (Brandini et al., 2018), which affects benthic systems (De Léo & Pires-Vanin, 2006; Sumida et al., 2005). The upwelling events are more frequent during austral summers, when northeast winds predominate (Valentin, 2001), and their influence can reach 300 km southward. Despite being perennial on the surface, its effects seem long-lasting on the seafloor,

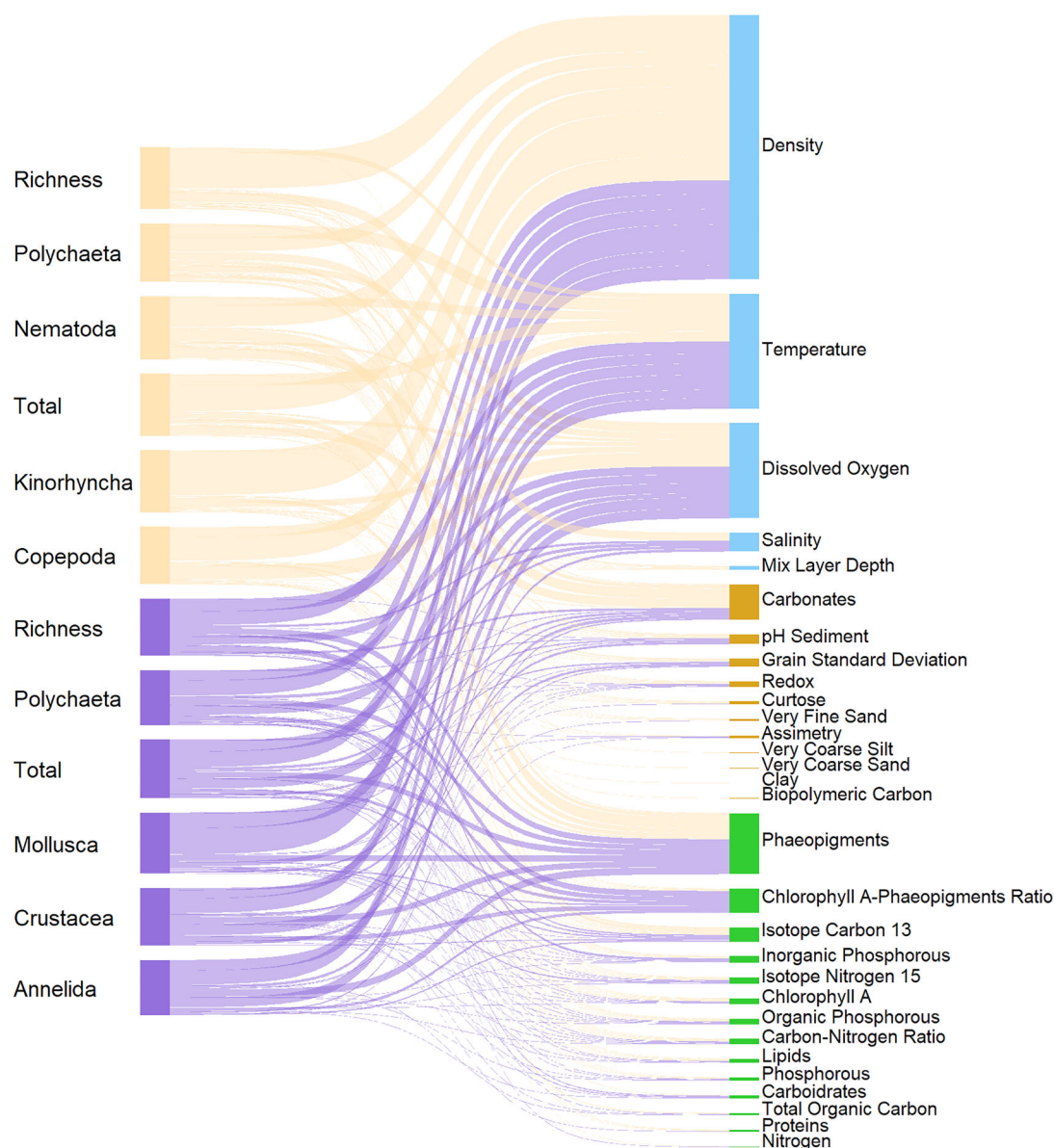


FIGURE 2 Results of the feature importance analysis for the significant biodiversity variables ($p < 0.05$) of the 2M model (biodiversity predictions obtained from environmental simulations) (link width, proportional to the number of times the variable appeared as the root of the random forest model; node size, relative cumulative importance of each environmental variable across the models; purple nodes, macrofauna; beige nodes, meiofauna; blue nodes, water; green nodes, organic material; yellow nodes, granulometry).

promoting temporal biodiversity stability (Sumida et al., 2005). During both surveys, higher concentrations of phaeopigments and higher densities of all benthic community parameters were observed at the northern portion of the continental shelf in the Santos Basin relative to the other stations. Additionally, during the surveys, the SACW occupied more than 50% of the water volume in the inner and mid-shelf of the northern region (Dottori et al., 2023).

At the lower slope (deeper than 400 m) and above the São Paulo Plateau, the parameters of the sediments and water were spatially and temporally homogeneous. Across the Santos Basin, the lower slope was characterized by high concentrations of clay in the sediments and a water mass with low salinity, which is a

typical signature of Antarctic Intermediate Water (Stramma & England, 1999). As expected for an oligotrophic and deep-sea region, very little organic matter and phytodetritus reached the deep-sea seafloor before or during our surveys. However, during the first survey, the chlorophyll–phaeopigments ratio was higher, suggesting a faster depositional rate of the phytodetritus. This increase in the quality of phytodetritus reaching the seafloor was reflected in higher abundances of all meiofauna taxa and of the macrofaunal crustaceans. As observed for other deep-sea regions, even small increments of organic material can increase benthic abundances (Billet et al., 1983; Brown et al., 2001).

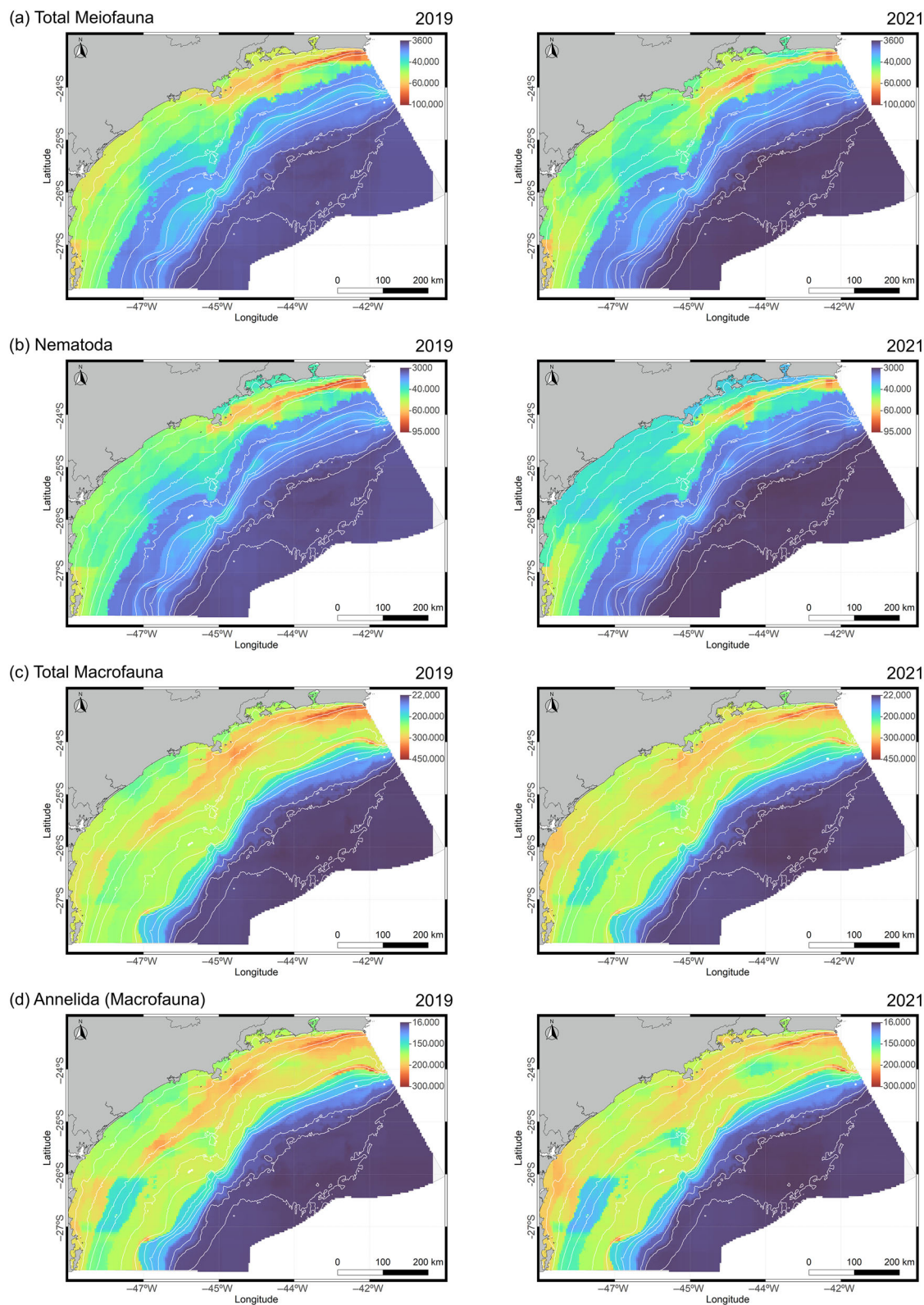


FIGURE 3 Machine learning model predictions of the abundance of total meiofauna, total macrofauna, Nematoda, and macrofaunal Annelida in the Santos Basin, Brazil, in two survey periods. The remaining maps for biodiversity parameters of meiofauna and macrofauna are in Appendices S10 and S11, respectively.

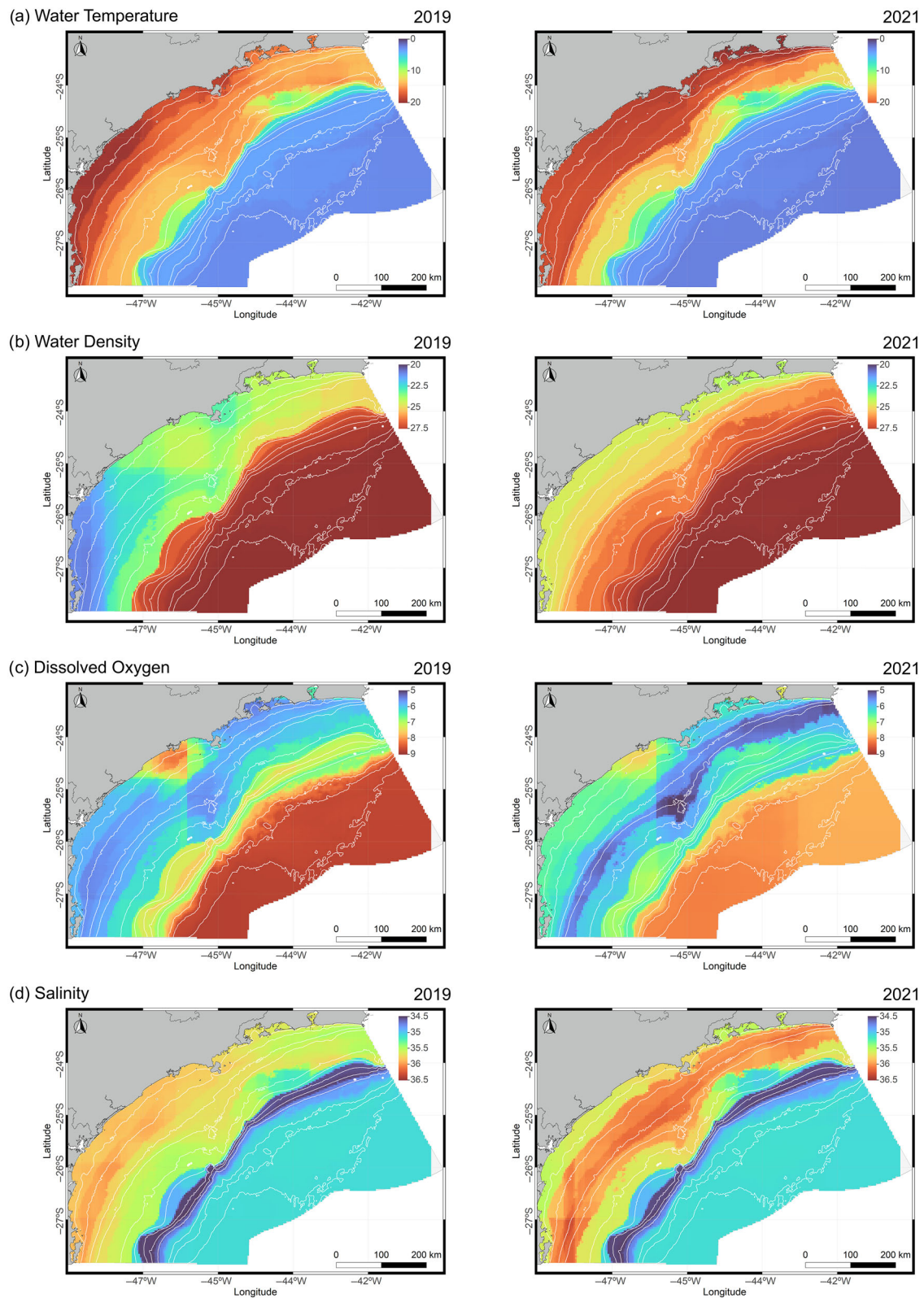


FIGURE 4 Machine learning model predictions of the most important bottom water parameters (see Figure 3) for predicting the benthic biodiversity variables at the Santos Basin, Brazil, during two surveys. Water density is expressed as density plus 1000 for scaling purposes. Units of the environmental variables are in Appendix S1.

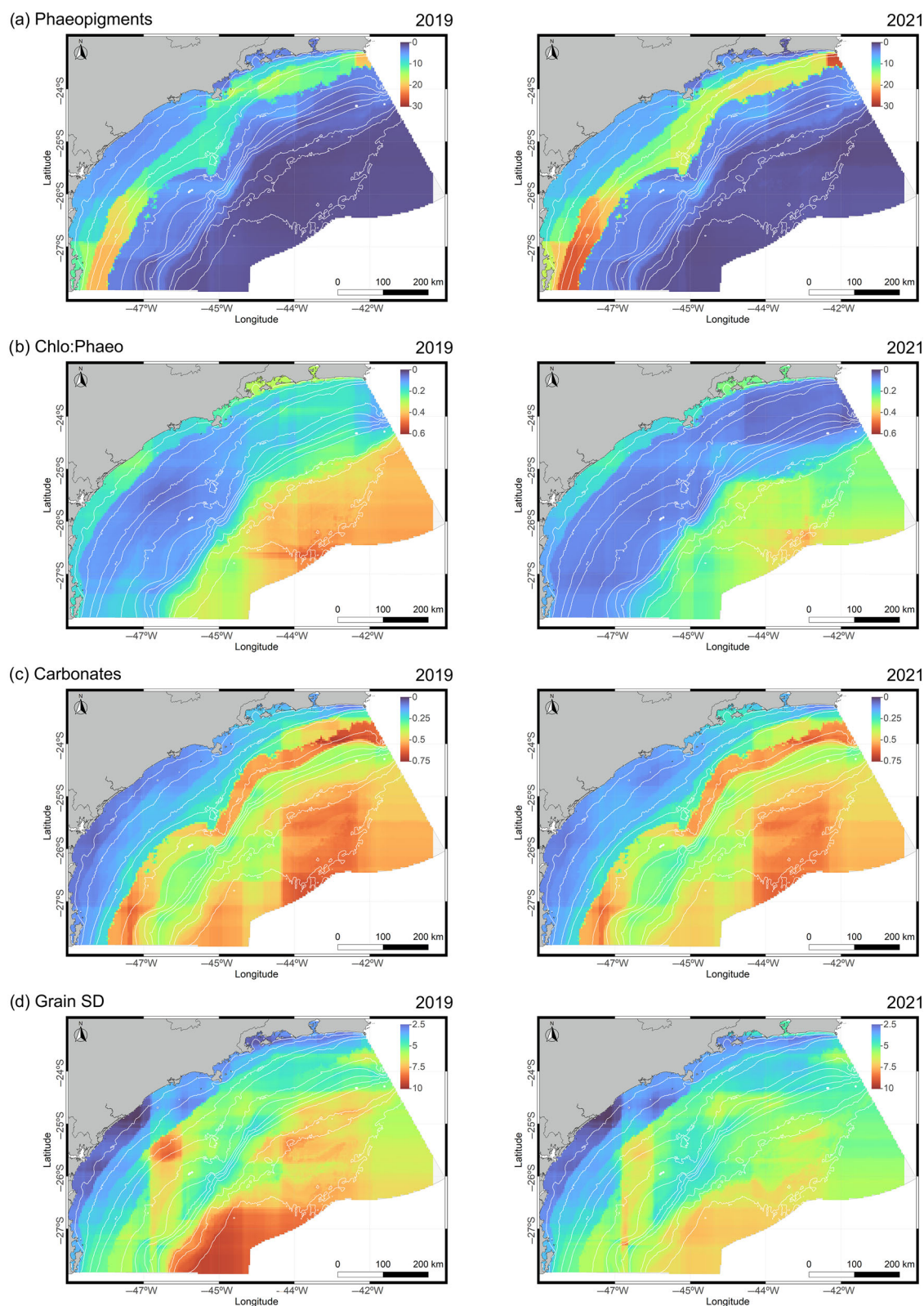


FIGURE 5 Machine learning model predictions of the most important sediment environmental parameters (see Figure 3) for predicting benthic biodiversity variables at the Santos Basin, Brazil, during two surveys: phaeopigments, (b) Chlo:Phaeo (ratio between chlorophyll-a and phaeopigments); (c) carbonates, and (d) grain SD (sediment grain sizes standard deviation; units are in Appendix S1).

Meiofauna and nematodes, in particular, occurred in low abundances in sediments with high carbonate content. At the mesophotic shelf break, the main source of bioclastic facies is linked to the presence of bioconstructions, which represent an ecologically important area as a refuge, breeding ground, feeding site, and habitat for benthic and pelagic species (Santos Filho et al., 2022). On the São Paulo Plateau, carbonates are associated with deposits of planktonic shell deposits (mainly pteropods) mixed with hemipelagic muddy terrigenous sediments (Figueiredo Jr. et al., 2023). The larger interstitial space of carbonated sediments does not favor the occurrence of meiofauna groups (Gallucci et al., 2023; Semprucci et al., 2013).

Applicability of the modeling framework for environmental monitoring programs

We found that applying a structured modeling framework allowed more information to be gathered from the data and made biodiversity measures generalizable across the spatial and temporal scales of the Santos Basin. The lack of significant differences between the structured and unstructured models suggests that the main environmental gradients influencing biodiversity in the region were adequately captured by the modeling framework. Together with the uncertainty analysis, these results confirmed that the structured two-stage models (2M-Sim) had a minimum loss of predictive accuracy that was comparable to models based on observed environmental data and provided well-calibrated estimates of predictive uncertainty. The structured approach allowed identification of cause-and-effect relationships among predictors, simulation of the predictions across space and time, and, thus, optimization of data acquisition in a future monitoring program.

By incorporating ML algorithms into the framework, it was possible to select the best model based on the minimum set of EEVs and, most importantly, to evaluate the model's performance on unseen data. Geographical coordinates, bathymetry, time, and water parameters are easily obtained and, at the Santos Basin, can be used to forecast environmental conditions and spatiotemporal benthic community patterns. The next step for implementing a predictive monitoring program that will guide management decisions is to investigate the tipping points of biodiversity in relation to anthropogenic and environmental change (Lewis et al., 2023), keeping in mind that gradual changes and lags in response cannot be discharged (Hillerbrand et al., 2023). These tipping points are important aspects to anticipate and are the basis for optimization of sampling (Henry et al., 2024).

For the benthos of the Santos Basin, we found that if the 30 environmental variables we investigated are continuously monitored (i.e., simulated and collected to validate the models), biodiversity estimates will be as precise as those obtained from a synoptic sampling. Such findings reinforce the importance of constructing robust analytical frameworks based on sound baseline data for the conservation of marine ecosystems (Gerhardt et al., 2024; Walling & Vaneeckhaute, 2020; Zurell et al., 2022).

For complex and heterogeneous datasets, it is expected that the biodiversity predictions made from the generalizations of the environmental models (2M-Sim) will be less precise than those based on the actual data (2M). In theory, this loss can be associated with an error propagation across the modeling structure (Simmonds et al., 2024). Error propagation means that a poor modeling performance in the first stage (e.g., environmental models) will result in low-quality estimates being used as predictors. However, in our dataset, this propagation was neither constant nor cumulative. The average performance of the environmental models in the 2M modeling structure was relatively low (<56%). Yet, the difference between 2M and 2M-Sim was only 1% and not significant. This means that even if we improved each environmental model up to 100%, it would still result in a maximum gain of 1% in the biodiversity models. Such findings imply that the current environmental models already capture the major environmental gradients in the basin and that the responses of the biodiversity descriptors to these gradients are probably nonlinear, with critical thresholds (Ehrnsten et al., 2020; Hiddink et al., 2023).

In this context, improvement of the accuracy of the biodiversity models would be achieved by having a larger dataset and by adding other environmental predictors. Additional temporal sampling events to gain more information on spatial and temporal variability will certainly contribute to reducing uncertainty. Incorporating EEVs from the pelagic zone, especially those obtainable through remote sensing, such as surface primary productivity and sedimentation rates (Kavanaugh et al., 2021), might also enhance model performance and speed data acquisition. Another fundamental aspect needed to implement an effective monitoring program is the inclusion of the variables related to the potential disturbances of human activities on the benthos, such as concentrations of metals and hydrocarbons (Nygard et al., 2020). A complete understanding of ecosystem functioning must also consider additional benthic indicators, such as community descriptors of the mega- and micro-benthos (Danovaro et al., 2020; Ingels et al., 2021). These improvements will strengthen understanding of the dynamics at continental margins and the ability to conserve them. Data-driven monitoring programs, grounded in robust statistical models, are fundamental to keeping pace with the rapid environmental changes induced by human activities.

We used ML to model deep-sea benthic diversity and thus moved beyond the limitations of classical spatial interpolation (e.g., Kriging), which primarily relies on spatial autocorrelation (Gazis et al., 2024). Our ML structured approach captured the complex, nonlinear relationships between biodiversity and multifaceted environmental drivers (e.g., depth, productivity, sediment properties). This facilitates a shift from purely spatial prediction to process-based prediction, essential for extrapolating into unsampled areas of similar environmental space (Elith & Leathwick, 2009). Unlike spatial overlay techniques requiring a priori layer weighting (e.g., Xiang, 2001), our model automatically determines the relative importance of each driver and their interactions, providing key ecological insights. The framework is suited for synthesizing multidisciplinary oceanographic datasets (Fonseca & Vieira, 2023). By learning about

the environment–response relationship, the ML model makes informed predictions where environmental conditions are within the training range, offering a powerful tool for predicting biodiversity patterns and testing ecological hypotheses.

From a methodological standpoint, the structured modeling framework proved valuable for environmental monitoring programs. The models effectively captured key environmental gradients and enabled the generalization of findings without losing accuracy. Expanding the dataset size and integrating predictors from the pelagic zone could further improve predictive capabilities. The proposed hierarchical approach highlights the interplay between oceanographic processes and benthic communities in the region. Spatial and temporal changes in environmental conditions, such as water masses, sediment properties, and organic matter fluxes, influenced meiofauna and macrofauna distributions. Our results showed that even small increases in phytodetritus in the deep sea led to higher benthic abundance. We concluded that data-driven monitoring programs, based on structured models and robust baseline data, are crucial for anticipating biodiversity shifts caused by environmental and anthropogenic changes. This approach is critical for the sustainable management and conservation of the Santos Basin ecosystem.

ACKNOWLEDGMENTS

We are deeply grateful to the members of the Benthos Team for their dedication and invaluable support in sample processing. Finally, we thank C. Mara Hercos for providing the bathymetrical data. Financial support was provided by the Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPQ to G.F. (306780/2022-4). Data used in this study were provided by the Santos Project—PCR-BS, which was funded, promoted, and executed by Petrobras.

The Article Processing Charge for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614).

DATA AVAILABILITY STATEMENT

All the data to perform the analyses and the results of the analyses can be assessed as savepoints in the GitHub repository “DaniloCVieira/imesc_savepoints” under the path: https://github.com/DaniloCVieira/imesc_savepoints/tree/1a985ef732eff5e27494ba2b64856a1c9e80870c/Fonseca_2026. The instructions to open the savepoints in iMESc are provided in Appendix S14.

ORCID

Gustavo Fonseca  <https://orcid.org/0000-0001-8625-4279>

Danilo C. Vieira  <https://orcid.org/0000-0002-7989-6553>

Luciana Erika Yaginuma  <https://orcid.org/0000-0002-8311-7364>

REFERENCES

Altmann, A., Tološi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, 26(10), 1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>

- Billett, D. S. M., Lampitt, R. S., Rice, A. L., & Mantoura, R. F. C. (1983). Seasonal sedimentation of phytoplankton to the deep-sea benthos. *Nature*, 302, 520–522. <https://doi.org/10.1038/302520a0>
- Brandini, F. P. (1990). Hydrography and characteristics of the phytoplankton in shelf and oceanic waters off southeastern Brazil during winter (July/August 1982) and summer (February/March 1984). *Hydrobiologia*, 196, 111–148. <https://doi.org/10.1007/BF00006105>
- Brandini, F. P., Tura, P. M., & Santos, P. P. (2018). Ecosystem responses to biogeochemical fronts in the South Brazil Bight. *Progress in Oceanography*, 164, 52–62. <https://doi.org/10.1016/j.pocean.2018.04.012>
- de Jesus, S. B., Vieira, D., Gheller, P., Cunha, B. P., Gallucci, F., & Fonseca, G. (2023). Machine learning algorithms accurately identify free-living marine nematode species. *PeerJ*, 11, e16216. <https://doi.org/10.7717/peerj.16216>
- Brown, C. J., Lamshead, P. J., Smith, C. R., Hawkins, L. E., & Farley, R. (2001). Phytodetritus and the abundance and biomass of abyssal nematodes in the central, equatorial Pacific. *Deep Sea Research Part I: Oceanographic Research Papers*, 48, 555–565. [https://doi.org/10.1016/S0967-0637\(00\)00049-2](https://doi.org/10.1016/S0967-0637(00)00049-2)
- Danovaro, R., Fanelli, E., Aguzzi, J., Billett, D., Carugati, L., Corinaldesi, C., Dell'Anno, A., Gjerde, K., Jamieson, A. J., Kark, S., McClain, C., Levin, L., Levin, N., Ramirez-Llodra, E., Ruhl, H., Smith, C. R., Snelgrove, P. V. R., Thomsen, L., Van Dover, C. L., & Yasuhara, M. (2020). Ecological variables for developing a global deep-ocean monitoring and conservation strategy. *Nature Ecology & Evolution*, 4, 181–192. <https://doi.org/10.1038/s41559-019-1091-z>
- De Léo, F. C., & Pires-Vanin, A. M. S. (2006). Benthic megafauna communities under the influence of the South Atlantic Central Water intrusion onto the Brazilian SE shelf: A comparison between an upwelling and a non-upwelling ecosystem. *Journal of Marine Systems*, 60(3–4), 268–284. <https://doi.org/10.1016/j.jmarsys.2006.02.002>
- de Souza, R. B., & Robinson, I. S. (2004). Lagrangian and satellite observations of the Brazilian Coastal Current. *Continental Shelf Research*, 24, 241–262. <https://doi.org/10.1016/j.csr.2003.10.001>
- Ditria, E. M., Buelow, C. A., Gonzalez-Rivero, M., & Connolly, R. M. (2022). Artificial intelligence and automated monitoring for assisting conservation of marine ecosystems: A perspective. *Frontiers in Marine Science*, 9, Article 918104. <https://doi.org/10.3389/fmars.2022.918104>
- Dottori, M., Sasaki, D. K., Silva, D. A., Del-Giovannino, S. R., Pinto, A. P., Gnamah, M., Santos, A. D., Silveira, I. C. A. D., Belo, W. C., Martins, R. P., & Moreira, D. L. (2023). Hydrographic structure of the continental shelf in Santos Basin and its causes: The SANAGU and SANSED campaigns (2019). *Ocean and Coastal Research*, 71(Suppl 3), Article e23013. <https://doi.org/10.1590/2675-2824071.22062md>
- Dray, S., Pélessier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P. R., Bellier, E., Bivand, R., Blanchet, F. G., De Cáceres, M., Dufour, A.-B., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J., & Wagner, H. H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, 82, 257–275. <https://doi.org/10.1890/11-1183.1>
- Ehrnsten, E., Norkko, A., Müller-Karulis, B., Gustafsson, E., & Gustafsson, B. G. (2020). The meager future of benthic fauna in a coastal sea—Benthic responses to recovery from eutrophication in a changing climate. *Global Change Biology*, 26, 2235–2250. <https://doi.org/10.1111/gcb.15014>
- Eisenhauer, N., Bowker, M. A., Grace, J. B., & Powell, J. R. (2015). From patterns to causal understanding: Structural equation modeling (SEM) in soil ecology. *Pedobiologia*, 58, 65–72. <https://doi.org/10.1016/j.pedobi.2015.03.002>
- Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Figueiredo, A. G., Jr., Carneiro, J. C., & Santos Filho, J. R. D. (2023). Santos Basin continental shelf morphology, sedimentology, and slope sediment distribution. *Ocean and Coastal Research*, 71(Suppl 3), Article e23007. <https://doi.org/10.1590/2675-2824071.22064agfj>
- Fonseca, G., Carreira, R. S., & Sousa, S. H. M. (2023). Subsidies for ecosystem-based management in an offshore oil and gas exploration area, Santos Basin, Brazil. *Ocean and Coastal Research*, 71(Suppl 3), Article e23061. <https://doi.org/10.1590/2675-2824071.23194gf>

- Fonseca, G., & Vieira, D. C. (2023). Overcoming the challenges of data integration in ecosystem studies with machine learning workflows: An example from the Santos project. *Ocean and Coastal Research*, 71(Suppl 3), Article e23021. <https://doi.org/10.1590/2675-2824071.22044gf>
- Gallucci, F., Fonseca, G., Vieira, D. C., Yaginuma, L. E., Gheller, P. F., Brito, S., & Corbisier, T. N. (2023). Predicting large-scale spatial patterns of marine meiofauna: Implications for environmental monitoring. *Ocean and Coastal Research*, 71(Suppl 3), Article e23037. <https://doi.org/10.1590/2675-2824071.22070fg>
- Gaziz, I. Z., Charlet, F., & Greinert, J. (2024). An interpretable multi-model machine learning approach for spatial mapping of deep-sea polymetallic nodule occurrences. *Natural Resources Research*, 33, 2473–2501. <https://doi.org/10.1007/s11053-024-10393-7>
- Gebhardt, S., Assis, J. C., Lacayo-Emery, M., Scherpenisse, A., Van Den Broek, K., Speelman, E., Wassen, M. J., Bakker, M., & Van Dijk, J. (2024). Supporting stakeholder dialogue on ecosystem service tradeoffs with a simulation tool for land use configuration effects. *Environmental Modelling & Software*, 179, Article 106097. <https://doi.org/10.1016/j.envsoft.2024.106097>
- Henry, P. A., Mondain-Monval, T. O., & Jarvis, S. G. (2024). Adaptive sampling in ecology: Key challenges and future opportunities. *Methods in Ecology and Evolution*, 15, 1483–1496. <https://doi.org/10.1111/2041-210X.14393>
- Hiddink, J. G., Valanko, S., Delargy, A. J., & Van Denderen, P. D. (2023). Setting thresholds for good ecosystem state in marine seabed systems and beyond. *ICES Journal of Marine Science*, 80, 698–709. <https://doi.org/10.1093/icesjms/fsad035>
- Hillebrand, H., Kuczyński, L., Kunze, C., Rillo, M. C., & Dajka, J. C. (2023). Thresholds and tipping points are tempting but not necessarily suitable concepts to address anthropogenic biodiversity change—An intervention. *Marine Biodiversity*, 53, Article 43. <https://doi.org/10.1007/s12526-023-01342-3>
- IBAMA. (2019). *Nota Técnica nº 4/2019/COEXP/CGMAC/DILIC*. Acessado via SEI/IBAMA - 6580026 em 2024.
- Ingels, J., Vanreusel, A., Pape, E., Pasotti, F., Macheriotou, L., Arbizu, P. M., Sørensen, M. V., Edgcomb, V. P., Sharma, J., Sánchez, N., Homoky, W. B., Wouds, C., Leduc, D., Gooday, A. J., Pawłowski, J., Dolan, J. R., Schratzberger, M., Gollner, S., Schoenle, A., ... Zeppilli, D. (2021). Ecological variables for deep-ocean monitoring must include microbiota and meiofauna for effective conservation. *Nature Ecology & Evolution*, 5, 27–29. <https://doi.org/10.1038/s41559-020-01335-6>
- Kavanaugh, M., Bell, T., Catlett, D., Cimino, M., Doney, S., Klajbor, W., Messić, M., Montes, E., Muller Karger, F., Otis, D., Santora, J., Schroeder, I., Triñanes, J., & Siegel, D. (2021). Satellite remote sensing and the marine biodiversity observation network. *Oceanography*, 34, 62–79. <https://doi.org/10.5670/oceanog.2021.215>
- Levin, L. A., Etter, R. J., Rex, M. A., Gooday, A. J., Smith, C. R., Pineda, J., Stuart, C. T., Hessler, R. R., & Pawson, D. (2001). Environmental Influences on Regional Deep-Sea Species Diversity. *Annual Review of Ecology and Systematics*, 32(1), 51–93. <https://doi.org/10.1146/annurev.ecolsys.32.081501.114002>
- Lewis, A. S. L., Rollinson, C. R., Allyn, A. J., Ashander, J., Brodie, S., Brookson, C. B., Collins, E., Dietze, M. C., Galliani, A. S., Juvigny-Khenafou, N., Koren, G., Meglinn, D. J., Moustahfid, H., Peters, J. A., Record, N. R., Robbins, C. J., Tonkin, J., & Wardle, G. M. (2023). The power of forecasts to advance ecological theory. *Methods in Ecology and Evolution*, 14, 746–756. <https://doi.org/10.1111/2041-210X.13955>
- Linder, H. L., Horne, J. K., & Ward, E. J. (2017). Modeling baseline conditions of ecological indicators: Marine renewable energy environmental monitoring. *Ecological Indicators*, 83, 178–191. <https://doi.org/10.1016/j.ecolind.2017.07.015>
- Lovett, G. M., Burns, D. A., Driscoll, C. T., Jenkins, J. C., Mitchell, M. J., Rustad, L., Shanley, J. B., Likens, G. E., & Haeuber, R. (2007). Who needs environmental monitoring? *Frontiers in Ecology and the Environment*, 5, 253–260. [https://doi.org/10.1890/1540-9295\(2007\)5\[253:WNEM\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2007)5[253:WNEM]2.0.CO;2)
- Moodley, L., Middelburg, J. J., Soetaert, K., Boschker, H. T. S., Herman, P. M. J., & Heip, C. H. R. (2005). Similar rapid response to phytodetritus deposition in shallow and deep-sea sediments. *Journal of Marine Research*, 63, 457–469. https://elischolar.library.yale.edu/journal_of_marine_research/87
- Moreira, D. L., Dalto, A. G., Figueiredo Jr., A. G., Valerio, A. M., Detoni, A. M. S., Bonecker, A. C. T., Signori, C. N., Namiki, C., Sasaki, D. K., Pupo, D. V., Silva, D. A., Kutner, D. S., Duque-Castaño, D. C., Marcon, E. H., Gallotta, F. D. C., Paula, F. S., Gallucci, F., Roque, G. C. F., Campos, G. S., ... Sousa, S. H. M. (2023). Multidisciplinary scientific cruises for environmental characterization in the Santos Basin—Methods and sampling design. *Ocean and Coastal Research*, 71(suppl 3), Article e23022. <https://doi.org/10.1590/2675-2824071.22072dlm>
- Moura, R. B. D., Dalto, A. G., Sallorenzo, I. D. A., Moreira, D. L., & Lavrado, H. P. (2023). Community structure of the benthic macrofauna along the continental slope of Santos Basin and São Paulo Plateau, SW Atlantic. *Ocean and Coastal Research*, 71(suppl 3), Article e23032. <https://doi.org/10.1590/2675-2824071.22091rbdm>
- Muller-Karger, F. E., Miloslavich, P., Bax, N. J., Simmons, S., Costello, M. J., Sousa Pinto, I., Canonico, G., Turner, W., Gill, M., Montes, E., Best, B. D., Pearlman, J., Halpin, P., Dunn, D., Benson, A., Martin, C. S., Weatherdon, L. V., Appeltans, W., Provoost, P., ... Geller, G. (2018). Advancing marine biological observations and data requirements of the complementary essential ocean variables (EOVs) and essential biodiversity variables (EBVs) frameworks. *Frontiers in Marine Science*, 5, Article 211. <https://doi.org/10.3389/fmars.2018.00211>
- Nygård, H., Keinänen, M., Mattila, J., Aro, E., Lappalainen, A., & Setälä, O. (2016). Price vs. value of marine monitoring. *Frontiers in Marine Science*, 3, Article 205. <https://doi.org/10.3389/fmars.2016.00205>
- Nygård, H., Lindegarth, M., Darr, A., Dinesen, G. E., Eigaard, O. R., & Lips, I. (2020). Developing benthic monitoring programmes to support precise and representative status assessments: A case study from the Baltic Sea. *Environmental Monitoring and Assessment*, 192, Article 795. <https://doi.org/10.1007/s10661-020-08764-7>
- R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Raffaelli, D., Bell, E., Weithoff, G., Matsumoto, A., Cruz-Motta, J. J., Kershaw, P., Parker, R., Parry, D., & Jones, M. (2003). The ups and downs of benthic ecology: Considerations of scale, heterogeneity and surveillance for benthic–pelagic coupling. *Journal of Experimental Marine Biology and Ecology*, 285–286, 191–203. [https://doi.org/10.1016/S0022-0981\(02\)00527-0](https://doi.org/10.1016/S0022-0981(02)00527-0)
- Reyers, B., Stafford-Smith, M., Erb, K.-H., Blyth, S., & Fischer, G. (2017). Essential variables help to focus sustainable development goals monitoring. *Current Opinion in Environmental Sustainability*, 26–27, 97–105. <https://doi.org/10.1016/j.cosust.2017.05.003>
- Rowland, J. A., Bland, L. M., James, S., & Nicholson, E. (2021). A guide to representing variability and uncertainty in biodiversity indicators. *Conservation Biology*, 35, 1669–1682. <https://doi.org/10.1111/cobi.13699>
- Santos Filho, J. R. D., Figueiredo, A. G., Carneiro, J. C., Dias, G. T. D. M., & Ramalho, A. S. (2022). Mesophotic bioclastics and bioconstructions at the continental shelf of Santos Basin, Brazil. *Sedimentary Geology*, 442, Article 106294. <https://doi.org/10.1016/j.sedgeo.2022.106294>
- Sasaki, D. K., Silva, D., Del Giovannino Júnior, S. R., Almeida Da Silveira, I. C., Belo, W. C., Martins, R. P., & Dottori, M. (2024). Hydrographic climatology of the South Brazil Bight continental shelf and slope. *Theoretical and Applied Climatology*, 155, 9407–9425. <https://doi.org/10.1007/s00704-024-05144-w>
- Schreiber, S. G., Schreiber, S., Tanna, R. N., Roberts, D. R., & Arciszewski, T. J. (2022). Statistical tools for water quality assessment and monitoring in river ecosystems—A scoping review and recommendations for data analysis. *Water Quality Research Journal*, 57, 40–57. <https://doi.org/10.2166/wqrj.2022.028>
- Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D., Kuemmerlen, M., Langhans, S. D., Martínez-López, J., & Vermeiren, P. (2019). How to make ecological models useful for environmental management. *Ecological Modelling*, 411, Article 108784. <https://doi.org/10.1016/j.ecolmodel.2019.108784>
- Semprucci, F., Colantoni, P., Baldelli, G., Sbrocca, C., Rocchi, M., & Balsamo, M. (2013). Meiofauna associated with coral sediments in the Maldivian subtidal habitats (Indian Ocean). *Marine Biodiversity*, 43, 189–198. <https://doi.org/10.1007/s12526-013-0146-7>
- Simmonds, E. G., Adjei, K. P., Cretois, B., Dickel, L., González-Gil, R., Laverick, J. H., Mandeville, C. P., Mandeville, E. G., Ovaskainen, O., Sicacha-Parada, J., Skarstein, E. S., & O'Hara, B. (2024). Recommendations for quantita-

- tive uncertainty consideration in ecology and evolution. *Trends in Ecology & Evolution*, 39, 328–337. <https://doi.org/10.1016/j.tree.2023.10.012>
- Sparrow, B. D., Edwards, W., Munroe, S. E. M., Wardle, G. M., Guerin, G. R., Bastin, J.-F., Morris, B., Christensen, R., Phinn, S., & Lowe, A. J. (2020). Effective ecosystem monitoring requires a multi-scaled approach. *Biological Reviews*, 95, 1706–1719. <https://doi.org/10.1111/brev.12636>
- Stramma, L., & England, M. (1999). On the water masses and mean circulation of the South Atlantic Ocean. *Journal of Geophysical Research*, 104, 20863–20883. <https://doi.org/10.1029/1999JC900139>
- Sumida, P. Y. G., Yoshinaga, M. Y., Ciotti, A. M., & Gaeta, S. A. (2005). Benthic response to upwelling events off the SE Brazilian coast. *Marine Ecology Progress Series*, 291, 35–42. <https://doi.org/10.3354/meps291035>
- Valentin, J. L. (2001). The Cabo Frio upwelling system, Brazil. In U. Seeliger & B. Kjerfve (Eds.), *Coastal marine ecosystems of Latin America. Ecological studies, Vol. 144* (pp. 97–105). Springer. https://doi.org/10.1007/978-3-662-04482-7_8
- Vieira, D. C., Paula, F. S., Yaginuma, L. E., & Fonseca, G. (2025). iMESc—An interactive machine learning app for environmental sciences. *Frontiers in Environmental Science*, 13, Article 1533292. <https://doi.org/10.3389/fenvs.2025.1533292>
- Walling, E., & Vaneeckhaute, C. (2020). Developing successful environmental decision support systems: Challenges and best practices. *Journal of Environmental Management*, 264, Article 110513. <https://doi.org/10.1016/j.jenvman.2020.110513>
- Witte, U., Aberle, N., Sand, M., & Wenzhöfer, F. (2003). Rapid response of a deep-sea benthic community to POM enrichment: An in situ experimental study. *Marine Ecology Progress Series*, 251, 27–36. <https://doi.org/10.3354/meps251027>
- Xiang, W.-N. (2001). Weighting-by-choosing: A weight elicitation method for map overlays. *Landscape and Urban Planning*, 56(1–2), 61–73. [https://doi.org/10.1016/S0169-2046\(01\)00169-4](https://doi.org/10.1016/S0169-2046(01)00169-4)
- Yaginuma, L. E., Gallucci, F., Vieira, D. C., Gheller, P. F., Brito de Jesus, S., Corbisier, T. N., & Fonseca, G. (2025). Hybrid machine learning algorithms accurately predict marine ecological communities. *Frontiers in Marine Science*, 12, Article 1458014. <https://doi.org/10.3389/fmars.2025.1458014>
- Zhou, X., Chen, B., Gui, Y., & Cheng, L. (2025). Conformal prediction: A data perspective. *ACM Computing Surveys*, 58(2), 1–37. <https://doi.org/10.1145/3736575>
- Zurell, D., König, C., Malchow, A.-K., Bozinovic, F., Colchero, F., & Schurr, F. M. (2022). Spatially explicit models for decision-making in animal conservation and restoration. *Ecography*, 2022(4), Article e05787. <https://doi.org/10.1111/ecog.05787>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fonseca, G., Vieira, D. C., Carneiro, J. C., Carreira, R. S., Ceccopieri, M., Corbisier, T. N., Dalto, A. G., Figueiredo Jr., A. G., Gallucci, F., Gheller, P., de Jesus, S. B., Lavrado, H. P., Lazzari, L., Marcon, E. H., Moreira, D. L., de Moura, R. B., Pape, E., Santarosa, A. C. A., dos Santos Filho, J. R., de Mello e Sousa, S. H., ..., Watanabe, W. (2026). Structured machine learning modeling to support conservation of deep-sea benthic biodiversity. *Conservation Biology*, 40, e70255. <https://doi.org/10.1111/cobi.70255>

