Distribution models of the branched hexactinellid Sarostegia oculata in Rio Grande Rise (SW Atlantic)

Paulo Corrêa (correapvf@gmail.com)
Instituto Oceanográfico da Universidade de São Paulo https://orcid.org/0000-0001-6474-7355
Paulo Sumida

Article

Keywords:

Posted Date: June 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1663413/v1

Abstract

The mining of ferromanganese (Fe-Mn) crusts in the deep sea have gained more attention in the last decade due to increased demand for rare earth elements that are critical for low-carbon technologies, which makes exploitation of this resource feasible and profitable. The Rio Grande Rise (RGR) is a distinct feature located in the South Atlantic and has become a region of great commercial and scientific interest because of its potential for mining Fe-Mn crusts. Extraction of Fe-Mn deposits may cause irreversible changes by removal of substrate, creation of sediment plumes, among others. Here, we use species distribution models (SDMs) to predict the occurrence of *Sarostegia oculata*, a branched hexactinellid that mimics the 3D skeletal framework of actual corals. It is the dominant organism in areas rich in Fe-Mn crusts and has relevant ecological importance in RGR. The models had excellent or good performance statistics and a high discrimination power between presence and absence sites. Our results support the relationship between *S. oculata*, the Fe-Mn crusts, and the rift at RGR, and may help to create a management plan and preserve unique marine habitats and biodiversity from the deep sea.

Introduction

Deep-sea sponges have importance as organisms that can provide multiple ecosystem goods and services, and thus have increasingly been mentioned in the literature in recent years[1]. Habitats dominated by sponges are globally distributed, forming structurally complex and often highly diverse communities, along with corals[2,3]. They have key roles in ecosystem functions, including habitat provision for associated fauna[4,5], increasing biodiversity[6–9], promoting silicon and organic carbon cycling[10–12], and potentially acting as nursery habitat for many species[13].

Vulnerable Marine Ecosystems (VMEs) are characterized as areas with low resilience and slow recovery from anthropogenic disturbances such as bottom trawling and mining (extraction of fossil fuels, gas, or minerals)[14]. Some taxonomic groups, including sponges, are considered indicators of VMEs as a consequence of having slow growth rates, longevity, late maturity, fragility, as they form three-dimensional structures associated with diverse communities[15]. They have been used to assist agencies responsible for the protection of particular ocean regions, such as slopes, seamounts, and canyons[16]. Increasing knowledge of the distribution and biology of species is necessary to create a consistent network of marine protected areas (MPAs), which can diminish human impacts on the ocean and prevent loss of biodiversity.

The Rio Grande Rise (RGR) is a major positive feature in the South Atlantic, between the parallels 28°–35° South and meridians 28°–39° West, located about 1200 km east off the Brazilian coast, and forming a volcanic plateau that rises from depths of about 5000 m to above 1000 m in some areas[17]. The western region of RGR (WRGR) forms an elliptical bulge of 500 km × 300 km that includes seamounts and guyots towering over or juxtaposing with this massive platform[18]; it is intersected by a NW-SE trending submarine rift between 20–40 km wide (also referred to as ‘graben’ or Cruzeiro do Sul Lineament). RGR has gained special attention from researchers and the Brazilian government in recent years[19,20] due to prospecting surveys that revealed its potential for the mining of Fe-Mn crusts[18,21,22]. Impacts created by mining activities may range from direct destruction of habitats and formation of toxic and particle-rich sediment plumes[23–25] to noise, vibration, and light, which may lead to significant (potentially irreversible) damage to deep-sea communities and ecosystems and loss of biodiversity on mining sites and surrounding areas[19,26–29].

Sarostegia oculata is a branched hexactinellid that forms an erect, complex surface, similar to the shape of a stony coral. It has a close association with a zoanthid, likely *Thoractis topsenti*[30], growing on and within the skeleton in a similar way to coral polyps, thus mimicking the 3D skeletal framework of actual corals[31]. It was first described at Cape Verde by Topsent[32], and further sampled throughout the Atlantic[33,34] and Indian[35] oceans. More recently, Hajdu et al.[31] recorded extensive sponge gardens in RGR, being the dominant organism in areas rich in Fe-Mn crusts that play a fundamental role in supporting communities in the area.

Species distribution models (SDMs; also known as habitat suitability models) are a method to predict the likelihood of a species to occur in a given location, based on their observed relationship with environmental conditions. SDM is widely used in ecology and conservation, especially in the deep sea where data is scarce, and such models are useful for the development of conservation measures and marine ecosystem management by predicting species distributions beyond surveyed areas[36–38].

In the present study, we built such models to predict the distribution of *S. oculata* in two areas with available high-resolution multibeam data at RGR, surrounding the rift within the bulge of WRGR. The growing use of mineral resources in our global economy will surely lead to future commercial extraction of mineral resources in the deep sea, including, but not limited to, Fe-Mn crusts[39–41]. Such models can be used to inform spatial management planning for protecting VMEs and environmental impact assessments that minimize the effects of mining on deep-sea ecosystems[42,43]. Our study is the first to create a high-resolution species distribution model of a VME indicator species in Rio Grande Rise. *Sarostegia oculata* is a dominant organism on substrates rich in Fe-Mn crusts, which play a fundamental role to support communities and provide key ecosystem functions in RGR. The five models built here predicted a high likelihood of *S. oculata* along with the rift borders with a low degree of uncertainty, in an area aimed at exploration of Fe-Mn crust deposits. This reinforces the relationship between *S. oculata*, the Fe-Mn crusts, and the rift at RGR, implicating that direct destruction of habitats caused by the mining of Fe-Mn crusts could lead to a severe impact in the communities sustained by *S. oculata*, which should be addressed in the management and MPA network design for RGR. There are still large gaps in the knowledge of the benthic fauna distribution and ecosystem functions on RGR, but our study provides expert advice on the occurrence of deep-sea sponges in the area and represents initial steps to guide future surveys and conservation plans in the region.

Methods

Data collection

Two surveyed areas near the center of WRGR were used in this study. The first area (here named Area 1) was delimited between latitudes 30º35’S–31°03’S and longitudes 35º36’W–36°16’W (Fig. 1a). This area is ~ 2,000 km², ranging from 600 m to 1,800 m depth, and encompasses a small segment of the rift and a plateau on both sides, here named NE and SW plateaus (Fig. 1b). It also contains a terrace in the NE plateau, which runs almost parallel to the rift border, a slide south of the SW plateau, and a canyon that splits a section of the SW plateau in two, one next to the rift border and another named the inner SW plateau. The south and west of the SW plateau are neighbored by a lower plane, which extends beyond the MBES data. The second area (Area 2) is smaller, located about 30 km northwest of Area 1, delimited between latitudes 30°19’S–30°25’S and longitudes 35°57’W–36°06’W, is ~ 77 km², ranging from 600 to 1400 m depth. It encompasses a small area of the northeastern plateau and the rift (Fig. 1c).

Area 1 (Figs. 1b and 2) was surveyed with the *N/Oc. Alpha Crucis*, from January 30 to February 21, 2018, and *RSS Discovery*, from October 26 to November 8, 2018, as part of the FAPESP/RCUK funded project Marine E-Tech. *N/Oc. Alpha Crucis* used the Reason 7160 multibeam echosounder (MBES) operating at 41 kHz. *RSS Discovery* was equipped with a ship-mounted Kongsberg EM122, and multibeam was processed using Caris HIPS and SIPS v9.1.8. The data in both products were gridded at 25 m using NaviModel Producer 4.3.2, then fused with ArcGIS Pro 2.9, using the Mosaic tool with the “Blend” mosaic operator. The survey in the *RSS Discovery* included 13 dives (codes HY31–43) of the Robotic Underwater Vehicle, *RUV HyBIS* (Hydraulic Benthic Interactive Sampler). *HyBIS* was equipped with a Sony Full HD camera that recorded over 36 hours and 26 km of video footage of the seafloor and it was connected to the ship's USBL system to record its position (latitude, longitude, and depth).

Area 2 (Figs. 1c and 3) was surveyed with the submersible *Shinkai 6500* and *Yokosuka* support vessel, from April 13 to May 5, 2013, as part of the ‘Iata-Piuna’ expedition. Yokosuka was equipped with multibeam Sea Beam 2112.004, operating at 12 kHz. The data was also gridded at 25 m using NaviModel Producer. *Shinkai* 6500 was equipped with two HD video cameras. One of them was fixed at the bow, which recorded 4.5 hours and 2,282 m of footage from the seabed during the dive 6K1338. The other video camera was mobile and was ignored in this study so the analysis would be more consistent with the *RUV HyBIS*.

The *HyBIS* and *Shinkai* videos were annotated and counted for benthic fauna while classifying them to the lowest possible taxonomic group. The branched hexactinellid *Sarostegia oculata* was the most commonly observed habitat-forming VME indicator species in both Area 1 and a 2. The timestamp of the video was used to obtain the coordinates of the observations from the *HyBIS* and *Shinkai* track data. Grid cells where at least one individual of *S. oculata* were classified as “presence”, and grid cells with overlapping track and no observations of *S. oculata* were classified as “absence”. Each of these grid cells are hereafter called records or sites, used to train and evaluate the models.

A set of additional seafloor variables (Figs. 2 and 3) was derived from the bathymetric data using the Benthic Terrain Model 3.0 tool[44] in ArcGIS. These variables were slope, aspect (measured in terms of northness and eastness), roughness (11 neighborhood size), curvature, fine-scale (3/30 radius) Bathymetric Position Index (BPI), and broad-scale (20/200 radius) BPI. The rift is a major feature in both areas that may play an important role in the distribution of the species in RGR. As such, we included the distance from the rift border as another environment variable in our analysis. This was achieved by creating a contour at value 0 from the broad-scale BPI raster. This will result in two contours outlining the NE and SW rift borders, which were used as source for the “Euclidean Distance” tool in ArcGIS. Before data analysis, collinearity between variables was checked using Pearson’s correlation coefficient and the Variance Inflation Factors (VIFs)[45] using the *usdm* package[46].

Modeling methods

The distribution of *S. oculata* was modeled using five algorithms widely used in species distribution models: Random Forest (RF), Boosted Regression Trees (BRT), MaxEnt, Generalized Additive Models (GAMs), and Artificial Neural Networks (ANN). In general, these models differ in the way they determine the fitted function, and how they handle interactions, model complexity, and overfitting[47]. Models were developed using R[48] with the framework caret[49]. The caret package is a set of functions that streamline the process of creating predictive models and provides a uniform interface for data splitting, variable selection, and model tuning. It uses other packages internally to create the models. We used default parameters for each package, except when stated in the text below or Table 1. The function *rfe* was used to perform recursive feature elimination in the models, but no advantage was gained in model performance from removing variables from the models, so we kept all variables in the final models.

RF builds multiple classification trees by taking many bootstrap samples from the training data and making an average prediction over all fitted trees[50]. RF uses only a randomly smaller set of the predictor variables (parameter **mtry**) on each split while growing each tree, which reduces the variance of the final model, with subsequent gains for predictive performance. We set the number of trees to grow (parameter **ntree**) as 1000.

BRT builds several regression trees (parameter **n.trees**) in a forward stagewise fashion. At each step of model fitting, the algorithm fits each new tree to the residuals of the previously fitted trees and uses only a random subset of data [51]. This gradually increases emphasis on observations modeled poorly by the existing collection of trees. The contribution of each tree is controlled by the **shrinkage** parameter, as the model-building process usually performs best if it moves slowly down the gradient. **Interaction.depth** controls the maximum depth of the individual trees and **n.minobsinnode** controls the minimum number of observations in terminal nodes.

Maxent estimates species distribution by minimizing the relative entropy between two probability densities[52,53]. Although it was initially developed as a machine-learning algorithm, MaxEnt has known links to Poisson point process models[54] and it can be categorized as a “regression-based” model. MaxEnt has the ability to fit complex models, controlled by the use of transformed features of the predictor covariates (parameter **feature types**). **Regularization multiplier** can also be used to smooth and avoid fitting overly complex models. In this work, we employed real absences derived from the dives instead of more commonly used pseudo-absences and used the Java implementation to train the MaxEnt models, version 3.4.4.

GAMs are an extension of General Linear Models (GLMs) but use nonparametric smoothing functions which allow modeling of non-linear relationships between the response and explanatory functions. Multiple smoothing parameter estimation methods are available in the package (parameter **method**)[55]. It can also add an extra penalty to each term (parameter **select**), meaning that the smoothing parameter estimation can completely remove terms from the model. We used a binomial distribution with the logit function and thin plate regression splines for the smoothing basis of all terms.

ANNs are a deep learning approach that consist of a large number of nodes and connections, typically organized in layers[56]. The input layer contains the environmental data, with each input node representing one environmental variable. The information from each input node is fed into the hidden layers, which may contain multiple layers with a variable number of nodes. Data from the last hidden layer is then fed to the output layer which represents the result of the model. The connections between the nodes from one layer to the next, also known as weights, are optimized using a back-propagation process. Our model has a single hidden layer, with 1 to 21 nodes (parameter **size**). **Decay** is used as a penalty for these connections that both help the optimization process and avoid over-fitting. The feed-forward and back-propagation processes are repeated until the model reaches a pre-defined accuracy, or a maximum set number of runs controlled by the parameter maxit (in our model we set **maxit** as 500).

Table 1
Modeling methods, their implementations during tuning, and the tuned parameters used in the final models.

Method	R package	Parameter	Values	Tuned
RF	randomForest	mtry	1–9	1
BRT	gbm	n.trees	from 100 to 3,000 by 50	2600
		interaction.depth	from 1 to 16 by 3	10
		shrinkage	10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}	10^{-3}
		n.minobsinnode	5, 15	5
Maxent	custom[1]	feature types	l, lq, lqp, lqph, lqh, lqpt, lqt, lqpth[2]	lqpth
		regularization multiplier	from 0.2 to 2 by 0.2, 2.5, 3, 3.5, 4, 4.5, 5	1
GAM	mgcv	method	REML, ML	ML
		select	TRUE, FALSE	TRUE
ANN	nnet	size	1, 3, 5, 7, 9, 11, 13, 21	21
		decay	0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}	0.1

[1] Custom script was used to run the MaxEnt java application and read its output within caret.

[2] l = linear, q = quadratic, h = hinge, p = product and t = threshold.

The relationship between the environmental layers and the predicted probability of presence was analyzed using response curves produced using the methodology described by Elith et al.[57]. All variables were held constant at their mean except the target variable, which varied at 100 points across its range. Then, the prediction of each algorithm was computed for each of the 100 values of the target variable and used to produce the response curves. The advantage of this method is that it can be applied to any model and be compared against different model techniques, but it does not account for interactions between variables. The importance of each variable in the final output was computed using the methodology described by Thuiller et al.[58]. The target variable was shuffled, while the others were left untouched. The Pearson's correlation coefficient between the reference and shuffled predictions was calculated and the score *1 - correlation* is returned. Scores near one mean the variable has a high influence on the model, and a score of zero assumes no influence of the variable on the model. The process was repeated 100 times for each variable.

An ensemble model is a popular technique for reducing the uncertainty of model predictions[59]. They may reduce prediction errors and decrease mean bias due to the choice of method[60]. This is usually achieved by computing the weighted or unweighted averages of the different model outputs. Ensemble models also permit uncertainty caused by the different predictions to be calculated. In this study, all model predictions were rescaled between 0 and 1, and their unweighted average was used to build the ensemble model. Spatial measures of uncertainty were calculated using a bootstrap technique[15,61]. A random sample of the presence-absence data, of equal size, was drawn with replacement, and the five models were constructed with the same settings as the originals. Predictions were then made to the study area, and this process was repeated 200 times. Model uncertainty was represented as the coefficient of variation (CV) of the bootstrap output, i.e, the standard deviation divided by the mean.

Model evaluation

The data in this study was divided into train, validation (also referred to as development in the literature), and test sets. The train set is used to train the model, while the validation set is used to evaluate the model during tuning. After the best parameters are selected, both train and validation sets are used to train a final model, which is evaluated using the test set. Data obtained from Area 1 with *HyBIS* was used as train and validation sets. Such data inherits a high spatial dependency, as records are available next to each other throughout each dive, and can lead to overestimation of model performance[62]. To provide a larger level of spatial independence between the training and the validation data, the first or last 20% records in each dive were selected to be used as a validation set. Consequently, the remaining 80% were used to train the model. The region of the dive which was cut from training, whether from the beginning or towards the end, was selected at random for each dive, and this process was repeated 25 times. During model tuning, the AUC_{ROC} metric was used to select the best models. Data obtained from Area 2 with *Shinkai* is completely independent from Area 1 and was used exclusively as a test set for the final model. The evaluation was performed 25 times using a random selection of 70% of the test data (subsampling) in each interaction.

Models were evaluated using five metrics to cover different aspects of the modeling performance: (1) area under the receiver operating characteristic curve (AUC_{ROC}), (2) area under the precision-recall gain curve (AUC_{PRG}), (3) Sensitivity, (4) Specificity, and (5) True Skill Statistics (TSS). AUC_{ROC} and AUC_{PRG} are threshold-independent measures, while Sensitivity, Specificity, and TSS need to convert the predicted likelihood of the model (a value between 0 and 1) into a presence/absence classification using a threshold. Here we used the threshold which balances sensitivity and specificity, which minimizes both commission and omission errors, and is recommended for SDMs model[63]. Note that thresholds were selected using the validation data set.

AUC_{ROC} assesses the ability of models to discriminate presence from absence sites and is widely used in species distribution models. The ROC curve is created by calculating the true positive rate (the proportion of presences correctly predicted, also known as sensitivity or recall) and the false positive rate (the proportion of absences falsely predicted, also calculated as 1 - specificity) at various threshold settings. AUC_{ROC} ranges from 0 to 1, where values above 0.9 indicate excellent performance, values between 0.7 and 0.9 indicate good performance, and a value of 0.5 or below indicates no better discrimination than a random classification.

The area under the precision-recall curve (AUC_{PR}) measures the ability to capture true presences and does not include false positives or absences in its calculation. AUC_{PR} is recommended when the user is more interested in an accurate prediction of the presences or when the number of absences is much larger than presences[64]. Like the ROC curve, the precision-recall curve calculates precision (the proportion between correctly predicted presences and all predicted presences) and recall across multiple thresholds. The baseline in the AUC_{PR} depends on the prevalence in the testing data [65], making it difficult to compare between species or studies. As such, the precision-recall curve is plotted in a new coordinate system, called the precision-recall gain curve, as described by Flach and Kull[66]. AUC_{PRG} values from 0 to 1 indicate a performance better than random, where 1 is a perfect discrimination model. Negative values indicate predictions worse than random.

Specificity is the proportion of absences correctly predicted. Both sensitivity and specificity are considered high with values above 0.8[63]. TSS normalizes the overall accuracy and accounts for both omission and commission errors[67]. TSS is calculated as sensitivity plus specificity minus one and ranges from − 1 to 1, where + 1 indicates perfect agreement and values of zero or less indicate a performance no better than random. TSS is also considered less sensitive to prevalence compared to Cohen's Kappa statistic.

Because we used presence-absence data to train the models, we also produced calibration plots to evaluate the agreement between predicted probabilities of occurrence and observation of presence and absence. If the plotted tend is lined up close to the diagonal, we conclude that the model is well-calibrated. Calibration plots were drawn using the scripts in Phillips and Elith[68]. Statistical significance of differences between the models for the five statistics were tested using Friedman's Aligned Rank test ($p < 0.05$)[69], using the R package *scmamp* v0.2.55[70]. We also applied the post hoc test to conduct pairwise comparisons for each metric, using the Shaffer correction.

Results

From the 1193 records (grid cells) that were observed by *HyBIS* in Area 1, 231 had the presence of at least one *Sarostegia oculata*. In Area 2, 29 out of 255 records were observed with the presence of *S. oculata* by *Shinkai*. The hexactinellid showed a higher abundance near the rift walls in both areas and was found only on hard substrates, of which the majority had a ferromanganese crust.

The performance of the models was good for each metric calculated, for both validation and test data. AUC_{ROC} (Fig. 4a) values were > 0.9, except for ANN which performed slightly worse. All models had good AUC_{PRG} scores > 0.8 (Fig. 4b) and the RF, BRT, MaxEnt, GAM, and Ensemble models had nearly perfect values close to one for the test data, indicating a good discrimination ability to detect presence records. Similarly, all models showed high values for Sensitivity (Fig. 4c) and Specificity (Fig. 4d) for the validation data, suggesting a high proportion of observed presences and absences correctly predicted, respectively. Models RF, BRT, MaxEnt, and Ensemble had a higher sensitivity for the test data compared to the validation data. In contrast, GAM and ANN models had worse performance. Models had high specificity for test data as well, although the RF model was slightly worse. For the TSS metric, the ANN model had the overall worst performance, while models BRT, MaxEnt, and Ensemble had the highest values for test data (Fig. 4e). Friedman's Aligned Rank test showed a significant difference between models for the five statistics. The post-hoc test showed that the most significant differences were found when comparing GAM and ANN with other models (Appendix D). The threshold values used to discriminate between presence and absence were 0.248, 0.126, 0.304, 0.121, 0.187, and 0.225 for the RF, BRT, MaxEnt, ANN, GAM, and Ensemble models, respectively.

The calibration plots show that the true probability of presence compared to the predicted presence of the five models are badly calibrated (Fig. 5). The ideal curve (dotted line) is below the lower confidence interval of the fitted calibration curve, indicating that the true probability of presence is much larger than the estimate given by the models. Only models RF (Fig. 5a) and MaxEnt (Fig. 5c) had the ideal curves above the higher confidence interval for low probability values. These results indicate models have a high discrimination power, i.e. the ability of a model to correctly distinguish between occupied and unoccupied sites, but the model output should not be interpreted as estimates of *conditional* probability of presence.

Generally, all models predicted a suitable habitat along with the full extent of both NE and SW rift borders in Area 1, and the NE rift border in Area 2. For Area 1 (Fig. 6), RF model predicted the distribution to be more extensive, with a relatively higher likelihood on the bottom of the rift compared to the middle of the plateaus. Other models had a relatively low (< 0.1) likelihood on the bottom of the rift, on top of the plateaus, on the south canyon, and the lower plane regions at southwest. All models predicted a larger area with a high likelihood around the north end of the NE rift border. However, only models RF and ANN extended this area throughout the small terrace at the NE plateau. Models predicted high suitability nearby the area between the east side of the canyon and the SW plateau. The predicted likelihood extended to the other side of the canyon for the MaxEnt model, and even further around the inner SW plateau for models RF and BRT. The south slide of the SW plateau showed high predicted suitability as well, except for the GAM model. For Area 2 (Fig. 7), models predicted a high likelihood near the slope of the NE rift border, between 700 and 1000 m. All models, except for RF, showed low (< 0.1) suitability at the rift bottom and the top to the plateau.

The predicted suitable habitat by the ensemble model reflected the average of all five models accordingly. This predicted distribution reflected environmental variables included in the model, namely depth and slope. The region with bottom depths between 700 and 1000 m that had nearby slopes > 20 degrees contained a continuous band of high prediction of suitable habitat. The spatial patterns of low-modeled uncertainty corresponded to the main areas predicted as highly suitable on the rift borders in both Area 1 and Area 2 (Fig. 8), together with the majority of the SW plateau and around the inner SW plateau. Regions of high uncertainty were obtained for the rift bottom, the canyon, the lower plane regions, and in some areas on top of the plateaus.

The importance of the environmental variables varied across the modeling algorithms, but depth, fine BPI, and northness usually had a high influence across all models (Table 2). For the RF model, the variables showed a low importance index (< 0.1), indicating that this model uses all variables to predict the presence of *S. oculata*, and changing a single variable has little effect on its output. Only depth and fine BPI had higher importance relative to other variables. For BRT, MaxEnt, and GAM models, the variables depth, fine BPI, northness, and curvature had a high influence in the likelihood of *S. oculata*. However, depth had a higher influence for BRT and GAM compared to MaxEnt, and northness had a higher influence for MaxEnt and GAM compared to BRT models. For the ANN model, most variables showed a high influence in its output, except for broad BPI and eastness.

Table 2
Mean index of the importance of each predictor variable across 100 permutations in the training dataset, for the Random Forest (RF), Boosted Regression Trees (BRT), MaxEnt, Generalized Additive Models (GAM), and Artificial Neural Networks (ANN) models.

Variables	**RF**	**BRT**	**MaxEnt**	**GAM**	**ANN**
depth	0.09	0.388	0.113	0.572	0.553
slope	0.035	0.009	0.034	0.051	0.183
broad BPI	0.016	0.002	0.000	0.000	0.066
fine BPI	0.072	0.296	0.441	0.588	0.44
rift distance	0.049	0.008	0.004	0.000	0.349
northness	0.044	0.112	0.435	0.762	0.338
eastness	0.031	0.003	0.007	0.000	0.081
rugosity	0.036	0.011	0.008	0.000	0.19
curvature	0.037	0.046	0.106	0.094	0.304

In general, models showed similar response patterns across the gradient of each environmental variable. Depth (Fig. 9a): MaxEnt, GAM, and ANN models showed a low response at higher depths below ~ 1000 m and shallower waters above ~ 700 m, but showed a peak in the predicted likelihood of *S. oculata* within ~ 700–1000 m. For RF and BRT models, the response remained high from deeper sites until ~ 800 m where it reached a peak, and then the response was low at shallow depths. Slope (Fig. 9b): models had a low response at flat sites, which increased at steeper slopes. The biggest variation in the response for slope was found in the ANN model. Broad BPI (Fig. 9c): it had a high response around 70 for RF, BRT, and MaxEnt models, but GAM and ANN predicted likelihood was higher at lower broad BPI values (< 0), and lower at broad BPI > 100. Fine BPI (Fig. 9d): RF, BRT, and ANN showed a lower response at fine BPI < 0, and a higher response at values > 0, with a larger variation produced by the ANN model. MaxEnt response was higher with negative fine BPI, while GAM was unresponsive for this variable, characterized by a flat horizontal line in the plot. Rift distance (Fig. 9e): All models except ANN showed a similar pattern for the rift distance. A peak in response near the rift (< 2000 m) and from 4500 to 7000 m, along with low response between 2000 and 4500 m and regions more distant than 7000 m. The ANN model was different, with a high response near the rift, lowering constantly as it moves away until 10,000 m. The MaxEnt, GAM, and ANN had very low predicted outputs when far away from the rift (> 20,000 m). Northness (Fig. 9f): peaks in response were produced at sides facing north and south in models RF, BRT, and MaxEnt. GAM generated a slightly higher response in sites facing north than south. ANN, instead, generated a higher response in sites facing south than north. Eastness (Fig. 9g): models predicted a slightly higher response on sites facing either east or west. However, the GAM was unresponsive for this variable as well. Rugosity (Fig. 9h): models had a low response at sites with low rugosity, that increased at areas with higher values. Curvature (Fig. 9i): only ANN model had a large variation in response for curvature, with a peak in sites with curvature close to zero. RF and BRT, instead, showed a smaller response in these sites. MaxEnt and GAM were unresponsive for this variable.

Discussion

The Rio Grande Rise contains diverse benthic communities which include several VME indicator species such as sponges, scleractinians, octocorals, and black corals (Corrêa et al., unpublished data). Although RGR is located in oligotrophic waters, this diversity is likely related to strong currents and complex geomorphology with multiple habitats[19]. Our study provides high-resolution distribution maps (DM) for an important VME indicator species recorded in RGR, the branched hexactinellid *Sarostegia oculata*. This sponge was the most abundant species found in the footage near the rift[31], closely associated with Fe-Mn crust substrates. A few fragments of *S. oculata* sampled by dredges[71] revealed 18 species associated with this sponge, the most notorious being a zoanthid and an annelid, *Thoracactis topsenti* and *Hermadion fauveli cf.* Gravier[30] respectively, which also emphasize the importance *S. oculata* has in this environment. Our work is the first to create a species distribution model of *S. oculata*, and the first to produce high-resolution DMs at RGR, with potential use for deep-sea conservation and management in this area.

Deep-sea sponges have a high conservation and management significance because of their low resilience[43]. These species and their associated fauna are particularly vulnerable to anthropogenic impacts such as fishing and deep-sea mining due to their slow growth rates and low or unpredictable recruitment[72], proving to have a very slow or nonexistent recovery[5]. For being suspension filters, they will likely suffer from sediment loads caused by deep-sea mining[29]. The water canal system of such sponges are at risk of becoming clogged[73] by high loads of suspended particulate matter. These mining plumes can be of low or no nutritional value, further threatening the maintenance of these animals. Hence, *S. oculata* not only will be primarily affected by crust removal, but also in the area surrounding mining operations

High-resolution distribution models of benthic species, using seafloor camera imagery and bathymetric data obtained from multibeam surveys, have proven useful in the last two decades. Similar models for VME indicator taxa have been developed for conservation and management of areas of interest in the deep sea[15,43,74–80]. Such models are able to provide expert advice on the occurrence of VME indicator taxa in efforts to limit anthropogenic threats from future marine resource exploitation

Species distribution model

The five models trained in this study predicted a high likelihood of *S. oculata* along with all extent of both rift borders in the multibeam survey. These findings reinforce the idea that *S. oculata* distribution is highly related to the NE-SW rift border that runs throughout RGR, as suggested by Hajdu et al.[31] found in our study. This area is known to be an optimal place for Fe-Mn crusts formation[19], highly supported by the first claims by CPRM to ISA, which consists of 150 blocks of 20 km^2 each, the maximum allowed by Regulations on Prospecting and Exploration for Fe-Mn crusts[20]. The majority of the blocks are situated along the plateaus of WRGR, on both sides of the rift[19], where CPRM likely prospected areas of high Fe-Mn deposits before submitting the claim. In addition, *S. oculata* is somehow more closely associated with Fe-Mn crusts pavements in the plateaus near the rift and along the rift walls compared to other substrates[31]. Our study reinforces this relationship between *S. oculata*, Fe-Mn crusts, and the rift at RGR. A second notable suitable area was predicted between the southwestern of the SW plateau and on the top northeastern canyon side. However, we have no footage of this area that can confirm or not the presence of *S. oculata* there.

Sampled Fe-Mn crusts near the rift suggest that they were eroded by the strong currents that impacted the RGR plateau[21], and bottom currents of more than 0.2 m/s may be capable of eroding Fe-Mn crust surfaces[81]. These currents may provide a high food supply for the development of *S. oculata*, and can be an important key factor in the development of this organism. Unfortunately, we do not have a hydrodynamic model in the study area that could provide currents velocity and direction, which could be used as predictors in the distribution models. Future studies should focus on obtaining and applying such variables in the models, as they likely play an important role in the distribution of species in RGR. The complex outlines of the rift walls may generate vortices[31] and areas of slope may facilitate the propagation of internal tidal waves[1,82], which can cause resuspensions favoring greater development of sponges. This may explain why the response for rift distance and fine BPI had such importance to explain the distribution of *Sarostegia*. Steeper slopes also predicted higher response in the models, although this variable did not output great importance in the bootstrapped correlation test. The aspect of the seabed, in terms of northness, predicted the highest likelihood on the northern and southern slopes, which may correspond with the direction of currents in RGR as well.

Another variable considered important in our results was depth. The prediction of *S. oculata* was restricted between ~ 700–1000 m, which corresponds to the plateaus and upper rift walls. Depth is usually an important variable that predicts the occurrence of VME indicator species and can act as a surrogate for other important variables such as temperature, salinity, oxygen, nutrients, water masses, exported surface production, and aragonite saturation[15,83,84]. *S. oculata* was collected/observed at 598–1311 m depth in Cape Verde[32], at 745 m in east of Miami Terrace, south-west of Bimini[33], at 900–790 m in the Vitória Trindade seamounts chain[34], at 738—1040 m in RGR by Hajdu el al.[31], and in our study, from 681 to 1203 m. The similarity in depth ranges in the Atlantic between these records suggests that this sponge occurs in bathymetrically constrained bands. However, it is important to note that this bathymetric range alone is not sufficient to predict the distribution, as the inner SW plateau is within this range, but the models had a low likelihood in this region.

There are a few global distribution models of VME indicator taxa, namely scleractinians, octocorals, and black corals, that predicted their occurrence on RGR[19,85–87]. Similar models were built using data exclusively from the Brazilian continental margin by Barbosa et al.[88], who addressed the distribution of scleractinians and octocorals in RGR. Overall, they predicted a high suitability of VME indicator taxa on the plateau of RGR, especially in regions near the rift, somewhat similar to the predicted output of our models. However, these models use coarse resolution data (> = 0.0083°, ~ 1 km) as environmental predictors, and none had access to biological records from RGR to train or evaluate the model. High-resolution bathymetric data can better represent seabed physiographic features and improve regional and local suitability models[89].

Implications for Spatial Management and Conservation

There is a growing number of examples in the literature that demonstrate the potential of using species distribution models to predict the occurrence of deep-sea sponges for their conservation and management from impacts caused by anthropogenic activities such as bottom trawling and mining[1,15]. They provide fundamental ecosystem functions, and even in low densities, hexactinellid sponges may create a suitable substrate for colonization and development of several invertebrate taxa, serving as island habitats on the deep-sea floor[90]. Thus, increasing knowledge of their distribution and environmental conditions responsible for their formation and persistence are key factors to ensure the protection of the marine environment, especially from harmful effects resulting from human activities.

The models performance were excellent or good in most cases, correctly identifying 85.4% of sites for validation and 88.8% for testing. Our study uses a completely different dataset to test the models (Area 2), independent from the data used to train and validate the model (Area 1). This method was intended to simulate how models would perform in case they were used to predict the distribution of *S. oculata* in a neighboring, unexplored region compared to the original area where models were built. The high performance models had in the test dataset suggests they could be used to predict the distribution of *S. oculata* in unsurveyed areas, at least to some extent. In addition, our study suggests an overlap in the potential distribution of Fe-Mn crusts and *S. oculata*, which should be addressed in the management of mining to minimize impacts in this community and diminish the loss of biodiversity in RGR.

There are a few limitations that should be considered during the modeling approach. Species occurrence in different areas can be difficult to model, as SDMs can fail to account for biotic processes, such as competition and predator-prey interactions, and due to shortfalls in the available data, such as sampling bias and lack of key drivers of habitat suitability[91]. The absolute uncertainty of the model prediction is unknown, but the bootstrap procedures provided a measure of internal consistency across the models[61]. Uncertainty maps are a key resource when applying predictions of distribution models to management measures. The uncertainty of the ensemble model was the lowest near the rift border on both sides, and higher below 1,000 m depth or in some areas on top of the plateaus. Predictions in the rift border had more confidence, which could be explained by a higher sampling effort in this region. Thus, it is advised to use our models carefully to predict areas away from the rift. Increase in model performance and reduction in uncertainty may be achieved with (a) a regularly spaced sampling regime that covers the entirety of the environmental conditions observed in the region of interest[92], (b) inclusion of key environmental drivers for sponges, such as sediment type, current regimes, and nutrients[76,77], (c) and broader high-resolution surveys, that are still scarce for the area[71].

Rio Grange Rise had been treated as an area beyond national jurisdictions (ABNJ) and under regulations of the International Seabed Authority (ISA). In December 2018, Brazil presented a partial revised submission to the Commission on Limits of The Continental Shelf (CLCS), which includes RGR as part of its continental margin[20]. If approved, RGR will become a region under the jurisdiction and sovereign rights of Brazil, along with its mineral resources. This creates unforeseen implications for the management of RGR, as the regulations that will govern this area are still uncertain. Nevertheless, this issue should not undermine research that produces data and results which could be used to inform management and conservation planning in the area.

Declarations

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

Code is available at the GitHub repositories https://github.com/ correapvf/SDM-RGR-2022 and https://github.com/correapvf/caretSDM.

Acknowledgements

This study was funded by the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP, BR) grant number 2014/50820-7 and the Natural Environment Research Council (NERC, UK), as part of the Brazil-UK joint project “Marine ferromanganese deposits: a major resource of E-tech elements”. PVFC was funded by FAPESP grant number 2017/11884-8 and PYGS was funded by CNPq grant number

301554/2019-6. We are grateful to the Captain and Crew Members of the N/Oc. AlphaCrucis and *RSS Discovery* for their professionalism and dedication to data acquisition.

Author contributions

The authors contributed equally for this work.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Beazley, L. *et al.* Predicted distribution of the glass sponge Vazella pourtalesi on the Scotian Shelf and its persistence in the face of climatic variability. PLOS ONE **13**, e0205505 (2018).
2. Hogg, M. M. *et al. Deep-sea Sponge Grounds: Reservoirs of Biodiversity*. (UNEP-WCMC, 2010).
3. Maldonado, M. *et al.* Sponge Grounds as Key Marine Habitats: A Synthetic Review of Types, Structure, Functional Roles, and Conservation Concerns. in *Marine Animal Forests: The Ecology of Benthic Biodiversity Hotspots* (eds. Rossi, S., Bramanti, L., Gori, A. & Orejas Saco del Valle, C.) 1–39 (Springer International Publishing, 2016). doi:10.1007/978-3-319-17001-5_24-1.
4. Cook, S. E., Conway, K. W. & Burd, B. Status of the glass sponge reefs in the Georgia Basin. Marine Environmental Research **66**, S80–S86 (2008).
5. Klitgaard, A. B. The fauna associated with outer shelf and upper slope sponges (Porifera, Demospongiae) at the Faroe Islands, northeastern Atlantic. Sarsia **80**, 1–22 (1995).
6. Beazley, L., Kenchington, E., Yashayaev, I. & Murillo, F. J. Drivers of epibenthic megafaunal composition in the sponge grounds of the Sackville Spur, northwest Atlantic. Deep Sea Research Part I: Oceanographic Research Papers **98**, 102–114 (2015).
7. Bett, B. J. & Rice, A. L. The influenceof hexactinellid sponge (Pheronema carpenteri) spicules on the patchy distribution of macrobenthos in the porcupine seabight (bathyal ne atlantic). Ophelia **36**, 217–226 (1992).
8. Bo, M. *et al.* Role of deep sponge grounds in the Mediterranean Sea: a case study in southern Italy. in *Ancient Animals, New Challenges: Developments in Sponge Research* (eds. Maldonado, M., Turon, X., Becerro, M. & Jesús Uriz, M.) 163–177 (Springer Netherlands, 2012). doi:10.1007/978-94-007-4688-6_14.
9. Marliave, J. B., Conway, K. W., Gibbs, D. M., Lamb, A. & Gibbs, C. Biodiversity and rockfish recruitment in sponge gardens and bioherms of southern British Columbia, Canada. Mar Biol **156**, 2247–2254 (2009).
10. Cathalot, C. *et al.* Cold-water coral reefs and adjacent sponge grounds: hotspots of benthic respiration and organic carbon cycling in the deep sea. Frontiers in Marine Science **2**, (2015).

11. Maldonado, M., Navarro, L., Grasa, A., Gonzalez, A. & Vaquerizo, I. Silicon uptake by sponges: a twist to understanding nutrient cycling on continental margins. Sci Rep **1**, 30 (2011).
12. Pile, A. J. & Young, C. M. The natural diet of a hexactinellid sponge: Benthic–pelagic coupling in a deep-sea microbial food web. Deep Sea Research Part I: Oceanographic Research Papers **53**, 1148–1156 (2006).
13. Rossi, S. The destruction of the 'animal forests' in the oceans: Towards an over-simplification of the benthic ecosystems. Ocean & Coastal Management **84**, 77–85 (2013).
14. Ramirez-Llodra, E. *et al.* Man and the Last Great Wilderness: Human Impact on the Deep Sea. PLOS ONE **6**, e22588 (2011).
15. Rowden, A. A. *et al.* High-Resolution Habitat Suitability Models for the Conservation and Management of Vulnerable Marine Ecosystems on the Louisville Seamount Chain, South Pacific Ocean. Front. Mar. Sci. **4**, 335 (2017).
16. *International Guidelines for the Management of Deep-Sea Fisheries in the High-Seas*. (Food and Agriculture Organization of the United Nations, 2009).
17. Mohriak, W. Genesis and evolution of the South Atlantic volcanic islands offshore Brazil. Geo-Mar Lett **40**, 1–33 (2020).
18. Cavalcanti, J. A. D., Santos, R. V., Lacasse, C. M., Rojas, J. N. L. & Nobrega, M. Potential Mineral Resources of Phosphates and Trace Elements on the Rio Grande Rise, South Atlantic Ocean. *44th Underwater Mining Conference* (2015).
19. Montserrat, F. *et al.* Deep-sea mining on the Rio Grande Rise (Southwestern Atlantic): A review on environmental baseline, ecosystem services and potential impacts. Deep Sea Research Part I: Oceanographic Research Papers **145**, 31–58 (2019).
20. da Silva, A. P. Brazil advances over the Area: The inclusion of the Rio Grande Rise within the Brazilian outer continental shelf and its consequences for other states and for the common heritage of mankind. Marine Policy **125**, 104399 (2021).
21. Benites, M., Hein, J. R., Mizell, K., Blackburn, T. & Jovane, L. Genesis and Evolution of Ferromanganese Crusts from the Summit of Rio Grande Rise, Southwest Atlantic Ocean. Minerals **10**, 349 (2020).
22. Sousa, I. M. C. *et al.* Mineralogy and chemical composition of ferromanganese crusts from the Cruzeiro do Sul Lineament - Rio Grande Rise, South Atlantic. Journal of South American Earth Sciences **108**, 103207 (2021).
23. Christiansen, B., Denda, A. & Christiansen, S. Potential effects of deep seabed mining on pelagic and benthopelagic biota. Marine Policy **114**, 103442 (2020).
24. Levin, L. A. *et al.* Defining "serious harm" to the marine environment in the context of deep-seabed mining. Marine Policy **74**, 245–259 (2016).
25. Schmidt, C. W. Going Deep: Cautious Steps toward Seabed Mining. Environmental Health Perspectives **123**, A234–A241 (2015).

26. Dunn, D. C. *et al.* A strategy for the conservation of biodiversity on mid-ocean ridges from deep-sea mining. Science Advances **4**, eaar4313 (2018).
27. Jones, D. O. B., Ardron, J. A., Colaço, A. & Durden, J. M. Environmental considerations for impact and preservation reference zones for deep-sea polymetallic nodule mining. Marine Policy **118**, (2020).
28. Lodge, M. W. & Verlaan, P. A. Deep-Sea Mining: International Regulatory Challenges and Responses. Elements **14**, 331–336 (2018).
29. Miller, K. A., Thompson, K. F., Johnston, P. & Santillo, D. An Overview of Seabed Mining Including the Current State of Development, Environmental Impacts, and Knowledge Gaps. Front. Mar. Sci. **4**, 418 (2018).
30. Gravier, Ch. Note sur une Actinie (*Thoracactis* n. g., *Topsenti* n. sp.) et un Annélide Polychète (*Hermadion Faiweli* n. sp.), commensaux d'une Éponge siliceuse (*Sarostegia oculata* Topsent). *Bulletin du Musée Océanographique de Monaco* **344**, 1–20 (1918).
31. Hajdu, E., Castello-Branco, C., Lopes, D. A., Sumida, P. Y. G. & Perez, J. A. A. Deep-sea dives reveal an unexpected hexactinellid sponge garden on the Rio Grande Rise (SW Atlantic). A mimicking habitat? Deep Sea Research Part II: Topical Studies in Oceanography **146**, 93–100 (2017).
32. Topsent, E. *Sarostegia oculata*, Hexactinellide nouvelle des îles du Cap-Vert. Bulletin du Musée Océanographique de Monaco **10**, 1–8 (1904).
33. Dohrmann, M. *et al.* Systematics and spicule evolution in dictyonal sponges (Hexactinellida: Sceptrulophora) with description of two new species. Zoological Journal of the Linnean Society **163**, 1003–1025 (2011).
34. Tabachnick, K. R., Menshenina, L. L., Lopes, D. A. & Hajdu, E. Two new *Hyalonema* species (Hyalonematidae: Amphidiscosida) from eastern and south-eastern Brazil, and further Hexactinellida (Porifera) collected from seamounts off south-eastern Brazil by the RV 'Marion Dufresne' MD55 expedition. J. Mar. Biol. Ass. **89**, 1243–1250 (2009).
35. Reiswig, H. M. Family Farreidae Gray, 1872. in *Systema Porifera: a guide to the classification of sponges* (eds. Hooper, J. N. A., Van Soest, R. W. M. & Willenz, P.) 1332–1340 (Springer, 2002). doi:10.1007/978-1-4615-0747-5_136.
36. Howell, K.-L., Piechaud, N., Downie, A.-L. & Kenny, A. The distribution of deep-sea sponge aggregations in the North Atlantic and implications for their effective spatial management. Deep Sea Research Part I: Oceanographic Research Papers **115**, 309–320 (2016).
37. Marshall, C. E., Glegg, G. A. & Howell, K. L. Species distribution modelling to support marine conservation planning: The next steps. Marine Policy **45**, 330–332 (2014).
38. Reiss, H. *et al.* Benthos distribution modelling and its relevance for marine ecosystem management. ICES Journal of Marine Science **72**, 297–315 (2015).
39. Guilhon, M., Montserrat, F. & Turra, A. Recognition of ecosystem-based management principles in key documents of the seabed mining regime: implications and further recommendations. ICES Journal of Marine Science **78**, 884–899 (2021).
40. Wedding, L. M. *et al.* Managing mining of the deep seabed. Science **349**, 144–145 (2015).

41. Wedding, L. M. *et al.* From principles to practice: a spatial approach to systematic conservation planning in the deep sea. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20131684–20131684 (2013).
42. Kenchington, E. L. *et al.* Use of Species Distribution Modeling in the Deep Sea. *Canadian Technical Report of Fisheries and Aquatic Sciences* **3296**, 1–76 (2019).
43. Ramiro-Sánchez, B. *et al.* Characterization and Mapping of a Deep-Sea Sponge Ground on the Tropic Seamount (Northeast Tropical Atlantic): Implications for Spatial Management in the High Seas. Frontiers in Marine Science **6**, (2019).
44. Walbridge, S., Slocum, N., Marjean Pobuda & Wright, D. Unified Geomorphological Analysis Workflows with Benthic Terrain Modeler. Geosciences **8**, 94 (2018).
45. Guisan, A., Thuiller, W. & Zimmermann, N. E. *Habitat Suitability and Distribution Models: With Applications in R*. (Cambridge University Press, 2017). doi:10.1017/9781139028271.
46. Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. Where is positional uncertainty a problem for species distribution modelling? Ecography **37**, 191–203 (2014).
47. Merow, C. *et al.* What do we gain from simplicity versus complexity in species distribution models? Ecography **37**, 1267–1281 (2014).
48. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2020).
49. Kuhn, M. Building Predictive Models in *R* Using the **caret** Package. J. Stat. Soft. **28**, (2008).
50. Strobl, C., Malley, J. & Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods **14**, 323–348 (2009).
51. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. Journal of Animal Ecology **77**, 802–813 (2008).
52. Elith, J. *et al.* A statistical explanation of MaxEnt for ecologists: Statistical explanation of MaxEnt. Diversity and Distributions **17**, 43–57 (2011).
53. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**, 231–259 (2006).
54. Renner, I. W. & Warton, D. I. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. Biometrics **69**, 274–281 (2013).
55. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **73**, 3–36 (2011).
56. Basheer, I. A. & Hajmeer, M. Artificial neural networks: fundamentals, computing, design, and application. Journal of Microbiological Methods **43**, 3–31 (2000).
57. Elith, J., Ferrier, S., Huettmann, F. & Leathwick, J. The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. Ecological Modelling **186**, 280–289

(2005).

58. Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M. B. BIOMOD – a platform for ensemble forecasting of species distributions. Ecography **32**, 369–373 (2009).
59. Dormann, C. F. *et al.* Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. Ecological Monographs **88**, 485–504 (2018).
60. Araújo, M. B. & New, M. Ensemble forecasting of species distributions. Trends in Ecology & Evolution **22**, 42–47 (2007).
61. Anderson, O. F. *et al.* Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand. Deep Sea Research Part I: Oceanographic Research Papers **115**, 265–292 (2016).
62. Fourcade, Y., Besnard, A. G. & Secondi, J. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography **27**, 245–256 (2018).
63. Liu, C., Berry, P. M., Dawson, T. P. & Pearson, R. G. Selecting thresholds of occurrence in the prediction of species distributions. Ecography **28**, 385–393 (2005).
64. Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J. & Elith, J. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs **92**, e01486 (2022).
65. Saito, T. & Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLOS ONE **10**, e0118432 (2015).
66. Flach, P. & Kull, M. Precision-Recall-Gain Curves: PR Analysis Done Right. in *Advances in Neural Information Processing Systems* (eds. Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. & Garnett, R.) vol. 28 838–846 (Massachusetts Institute of Technology (MIT) Press, 2015).
67. Allouche, O., Tsoar, A. & Kadmon, R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS): Assessing the accuracy of distribution models. Journal of Applied Ecology **43**, 1223–1232 (2006).
68. Phillips, S. J. & Elith, J. POC plots: calibrating species distribution models with presence-only data. Ecology **91**, 2476–2484 (2010).
69. García, S., Fernández, A., Luengo, J. & Herrera, F. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. Information Sciences **180**, 2044–2064 (2010).
70. Calvo, B. & Santafé, G. scmamp: Statistical Comparison of Multiple Algorithms in Multiple Problems. The R Journal **8**, 248–256 (2016).
71. Jovane, L. *et al.* Multidisciplinary Scientific Cruise to the Rio Grande Rise. Front. Mar. Sci. **6**, 252 (2019).
72. Gollner, S. *et al.* Resilience of benthic deep-sea fauna to mining activities. Marine Environmental Research **129**, 76–101 (2017).

73. Abdul Wahab, M. A., Fromont, J., Gomez, O., Fisher, R. & Jones, R. Comparisons of benthic filter feeder communities before and after a large-scale capital dredging program. Marine Pollution Bulletin **122**, 176–193 (2017).
74. Dolan, M. F. J., Grehan, A. J., Guinan, J. C. & Brown, C. Modelling the local distribution of cold-water corals in relation to bathymetric variables: Adding spatial context to deep-sea video data. Deep Sea Research Part I: Oceanographic Research Papers **55**, 1564–1579 (2008).
75. Howell, K. L., Holt, R., Endrino, I. P. & Stewart, H. When the species is also a habitat: Comparing the predictively modelled distributions of Lophelia pertusa and the reef habitat it forms. Biological Conservation **144**, 2656–2665 (2011).
76. Rengstorf, A. M., Mohn, C., Brown, C., Wisz, M. S. & Grehan, A. J. Predicting the distribution of deep-sea vulnerable marine ecosystems using high-resolution data: Considerations and novel approaches. Deep Sea Research Part I: Oceanographic Research Papers **93**, 72–82 (2014).
77. Rengstorf, A. M., Yesson, C., Brown, C. & Grehan, A. J. High-resolution habitat suitability modelling can improve conservation of vulnerable marine ecosystems in the deep sea. Journal of Biogeography **40**, 1702–1714 (2013).
78. Robert, K., Jones, D. O. B., Roberts, J. M. & Huvenne, V. A. I. Improving predictive mapping of deep-water habitats: Considering multiple model outputs and ensemble techniques. Deep Sea Research Part I: Oceanographic Research Papers **113**, 80–89 (2016).
79. Rodríguez-Basalo, A. *et al.* High Resolution Spatial Distribution for the Hexactinellid Sponges Asconema setubalense and Pheronema carpenteri in the Central Cantabrian Sea. Front. Mar. Sci. **8**, 612761 (2021).
80. Tong, R., Purser, A., Guinan, J. & Unnithan, V. Modeling the habitat suitability for deep-water gorgonian corals based on terrain variables. Ecological Informatics **13**, 123–132 (2013).
81. Yeo, I. A. *et al.* Distribution of and hydrographic controls on ferromanganese crusts: Tropic Seamount, Atlantic. Ore Geology Reviews **114**, 103131 (2019).
82. Klitgaard, A. B. & Tendal, O. S. Distribution and species composition of mass occurrences of large-sized sponges in the northeast Atlantic. Progress in Oceanography **61**, 57–98 (2004).
83. Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T. & Taylor, P. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Marine Ecology Progress Series **321**, 267–281 (2006).
84. Thresher, R. *et al.* Strong Depth-Related Zonation of Megabenthos on a Rocky Continental Margin (∼700–4000 m) off Southern Tasmania, Australia. PLOS ONE **9**, e85872 (2014).
85. Davies, A. J. & Guinotte, J. M. Global Habitat Suitability for Framework-Forming Cold-Water Corals. PLoS ONE **6**, e18483 (2011).
86. Yesson, C., Bedford, F., Rogers, A. D. & Taylor, M. L. The global distribution of deep-water Antipatharia habitat. Deep Sea Research Part II: Topical Studies in Oceanography (2015) doi:10.1016/j.dsr2.2015.12.004.

87. Yesson, C. *et al.* Global habitat suitability of cold-water octocorals: Global distribution of deep-sea octocorals. Journal of Biogeography **39**, 1278–1292 (2012).
88. Barbosa, R. V., Davies, A. J. & Sumida, P. Y. G. Habitat suitability and environmental niche comparison of cold-water coral species along the Brazilian continental margin. Deep Sea Research Part I: Oceanographic Research Papers **155**, 103147 (2020).
89. Ross, L. K., Ross, R. E., Stewart, H. A. & Howell, K. L. The Influence of Data Resolution on Predicted Distribution and Estimates of Extent of Current Protection of Three 'Listed' Deep-Sea Habitats. PLoS ONE **10**, e0140061 (2015).
90. Beaulieu, S. E. Life on glass houses: sponge stalk communities in the deep sea. Marine Biology **138**, 803–817 (2001).
91. Vierod, A. D. T., Guinotte, J. M. & Davies, A. J. Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. Deep Sea Research Part II: Topical Studies in Oceanography **99**, 6–18 (2014).
92. Hirzel, A. & Guisan, A. Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling **157**, 331–341 (2002).

Appendix D

Appendix D is not available with this version.

Figures

Figure 1

(a) Rio Grande Rise, showing the locations of panels (b) and (c). **(b)** MBES of Area 1, used for model training and validation. **(c)** MBES of Area 2, used for model testing. Red lines represent dives from (a) *HyBIS* and (c) *Shinkai*. Maps (b) and (c) are in the same scale.

Figure 2

The bathymetry-derived variables in Area 1, which were used as explanatory variables in the models for training and validation. **(a)** Depth. **(b)** Slope. **(c)** Broad-scale Bathymetric Position Index (BPI). **(d)** Fine-scale BPI. **(e)** Euclidian distance from the border of the rift. **(f)** Northness. **(g)** Eastness. **(h)** Rugosity. **(i)** Curvature**.**

Figure 3

The bathymetry-derived variables in Area 2, which were used as explanatory variables in the models for testing. **(a)** Depth. **(b)** Slope. **(c)** Broad-scale Bathymetric Position Index (BPI). **(d)** Fine-scale BPI. **(e)** Euclidian distance from the border of the rift. **(f)** Northness. **(g)** Eastness. **(h)** Rugosity. **(i)** Curvature**.**

Figure 4

Mean and standard error of model performance statistics for both validation (orange) and test (blue) data across replicates. **(a)** AUC_{ROC}, **(b)** AUC_{PRG}, **(c)** Sensitivity, **(d)** Specificity and **(e)** TSS.

Figure 5

Presence-absence calibration plots for each model fitted with natural splines. **(a)** Random Forest, **(b)** Boosted Regression Trees, **(c)** MaxEnt, **(d)** Generalized Additive Models, and **(e)** Artificial Neural Networks. The calibration curve is in blue and a confidence interval of ± 2 SD is in orange. The rug plots show model predictions at presences (orange) and absences (blue). The dotted line indicates a 1:1 relationship representing ‘perfect’ calibration.

Figure 6

Prediction maps of *Sarostegia oculata* in Area 1, using **(a)** Random Forest, **(b)** Boosted Regression Trees, **(c)** MaxEnt, **(d)** Generalized Additive Models, **(e)** Artificial Neural Networks, and **(f)** Ensemble models.

Figure 7

Prediction maps of *Sarostegia oculata* in Area 2, using **(a)** Random Forest, **(b)** Boosted Regression Trees, **(c)** MaxEnt, **(d)** Generalized Additive Models, **(e)** Artificial Neural Networks, and **(f)** Ensemble models.

Figure 8

Uncertainty (CV) for the ensemble distribution model of **(a)** Area 1 and **(b)** Area 2.

Figure 9

Response curves of the likelihood of presence of *Sarostegia oculata* for each predictor variable for the Random Forest (RF), Boosted Regression Trees (BRT), MaxEnt, Generalized Additive Models (GAM), and Artificial Neural Networks (ANN) models. Rug plot inside bottom of panels show distribution of sites across that variable, in percentiles.