Diversity and Distributions WILEY

Ensemble modelling to predict the distribution of vulnerable marine ecosystems indicator taxa on data-limited seamounts of Cabo Verde (NW Africa)

Beatriz Vinha[1] | Francisco Javier Murillo[2] | Mia Schumacher[3] | Thor H. Hansteen[3] | Franziska U. Schwarzkopf[3] | Arne Biastoch[3,4] | Ellen Kenchington[2] | Stefano Piraino[1,5,6] | Covadonga Orejas[7] | Veerle A. I. Huvenne[8]

[1]Dipartimento di Scienze e Tecnologie Biologiche e Ambientali (DiSTeBA), Università del Salento, Lecce, Italy

[2]Ocean and Ecosystem Sciences Division, Department of Fisheries and Oceans, Bedford Institute of Oceanography, Halifax, Canada

[3]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

[4]Kiel University, Kiel, Germany

[5]Consorzio Nazionale Interuniversitario per le Scienze del Mare (CoNISMa), Rome, Italy

[6]National Biodiversity Future Center (NBFC), Palermo, Italy

[7]Instituto Español de Oceanografía, Centro Oceanográfico de Gijón (IEO-CSIC), Gijón, Spain

[8]Ocean BioGeosciences, National Oceanography Centre, Southampton, UK

Correspondence
Veerle A. I. Huvenne, Ocean BioGeosciences, National Oceanography Centre, Southampton, UK.
Email: vaih@noc.ac.uk

Funding information
Horizon 2020 Framework Programme, Grant/Award Number: 818123; Ministerio de Ciencia e Innovación; Regione Puglia, Grant/Award Number: POR Puglia FESR FSE 2014-2020

Editor: Enrico Pirotta

Abstract

Aim: Seamounts are conspicuous geological features with an important ecological role and can be considered vulnerable marine ecosystems (VMEs). Since many deep-sea regions remain largely unexplored, investigating the occurrence of VME taxa on seamounts is challenging. Our study aimed to predict the distribution of four cold-water coral (CWC) taxa, indicators for VMEs, in a region where occurrence data are scarce.

Location: Seamounts around the Cabo Verde archipelago (NW Africa).

Methods: We used species presence–absence data obtained from remotely operated vehicle (ROV) footage collected during two research expeditions. Terrain variables calculated using a multiscale approach from a 100-m-resolution bathymetry grid, as well as physical oceanographical data from the VIKING20X model, at a native resolution of 1/20°, were used as environmental predictors. Two modelling techniques (generalized additive model and random forest) were employed and single-model predictions were combined into a final weighted-average ensemble model. Model performance was validated using different metrics through cross-validation.

Results: Terrain orientation, at broad scale, presented one of the highest relative variable contributions to the distribution models of all CWC taxa, suggesting that hydrodynamic–topographic interactions on the seamounts could benefit CWCs by maximizing food supply. However, changes at finer scales in terrain morphology and bottom salinity were important for driving differences in the distribution of specific CWCs. The ensemble model predicted the presence of VME taxa on all seamounts and consistently achieved the highest performance metrics, outperforming individual models. Nonetheless, model extrapolation and uncertainty, measured as the coefficient of variation, were high, particularly, in least surveyed areas across seamounts, highlighting the need to collect more data in future surveys.

Main Conclusions: Our study shows how data-poor areas may be assessed for the likelihood of VMEs and provides important information to guide future research in Cabo Verde, which is fundamental to advise ongoing conservation planning.

KEYWORDS

Cabo Verde, cold-water corals, deep-sea ecosystems, ensemble modelling, seamounts, species distribution models, vulnerable marine ecosystems

Resumo

Objetivo: Montes submarinos são importantes formações geológicas com um notável papel ecológico e podem ser considerados Ecossistemas Marinhos Vulneráveis (VMEs). Dado que muitas regiões do mar profundo permanecem inexploradas, investigar a ocorrência de espécies indicadoras de VMEs é um desafio. O nosso estudo teve como objetivo prever a distribuição de quatro taxa de corais de água fria (CWC), indicadores de VMEs, numa região onde dados de ocorrência são escassos.

Localização: Montes submarinos no Arquipélago de Cabo Verde (NO África).

Métodos: Utilizamos dados de presença-ausência de CWC obtidos a partir de imagens de um Veículo de Operação Remota (ROV) durante duas expedições científicas. Como dados ambientais foram utilizados variáveis de terreno calculadas com uma abordagem multi-escala a partir de uma grelha de batimetria com 100 m de resolução, e dados de oceanografia física obtidos com o modelo VIKING20X, a uma resolução nativa de 1/20°. Duas técnicas de modelação (Generalized Additive Models (GAM) e Random Forest) foram usadas e resultados de modelos individuais foram combinados, através da média ponderada, num modelo final Ensemble. O desempenho dos modelos foi validado usando diferentes métricas através de técnicas de validação cruzada.

Resultados: A orientação do terreno, a larga escala, apresentou uma das maiores contribuições relativas para os modelos de distribuição de todos os CWCs, sugerindo que interações de hidrodinâmica com topografia beneficiam os corais, possivelmente pelo aumento da disponibilidade de alimento. No entanto, mudanças na morfologia de terreno a escalas mais finas e salinidade foram importantes para diferenças entre a distribuição de espécies específicas. O modelo Ensemble projetou a presença de indicadores de VMEs em todos os montes submarinos e, consistentemente, apresentou métricas de desempenho mais altas, superando modelos individuais. No entanto, medidas de extrapolação e incerteza foram elevadas, especialmente em áreas menos estudadas, destacando claramente a necessidade de recolher mais dados.

Conclusão: O nosso estudo mostra como áreas com poucos dados podem ser avaliadas quanto à probabilidade de VMEs e fornece informações importantes para guiar futuras investigações em Cabo Verde, sendo fundamental para aconselhar planos de conservação em curso.

1 | INTRODUCTION

Seamounts typically have enhanced primary productivity and structural complexity, making them suitable habitats for the presence of cold-water corals (CWC) (Davies & Guinotte, 2011; Rowden et al., 2010, 2017; Tracey et al., 2011; Yesson et al., 2012). However, their ecological and geological prominence exposes seamounts to anthropogenic disturbances, related to the exploration for fisheries (Clark et al., 2012; Goode et al., 2020; Kerry et al., 2022; Morato et al., 2006; Santos et al., 2021) and mineral resources (Leitner et al., 2021; Washburn et al., 2023). Many seamounts host populations of fragile long-lived species that form biogenic habitats with high vulnerability to bottom-contact fishing gears (Baco et al., 2020; Bergstad et al., 2019; Davies et al., 2015; Williams et al., 2020), and

are considered vulnerable marine ecosystems (VMEs) under the sustainable fishing resolutions adopted by the United Nations General Assembly (e.g. A/RES/61/105). Identifying and mapping indicator taxa that meet the established criteria (FAO, 2009) is the first step for the designation of a VME (Ardron et al., 2014). However, the benthic communities of most deep-water seamounts in the Central and South Atlantic Ocean – both within economic exclusive zones (EZZ) and in areas beyond national jurisdiction (ABNJ) – remain largely uncharacterized (Bridges, Howell, et al., 2023). This hampers our ability to evaluate potential impacts of anthropogenic activities on CWCs on specific seamounts, which is necessary to design and implement effective management plans for the sustainable use of resources and the conservation of those regions.

Different seafloor habitat mapping techniques can be used to map unexplored deep-sea regions. For example, seafloor habitat maps for data-limited regions can be created relying solely on physical properties of the seafloor, readily obtained from open-source databases, where different clusters of abiotic data allow the identification of areas with ecological importance based on the environmental conditions they present (Hogg et al., 2016; McQuaid et al., 2020, 2023; Schumacher et al., 2022; Swanborn et al., 2023; Van Audenhaege et al., 2021). However, incorporating species occurrence data, through supervised mapping techniques such as species distribution models (SDMs), is useful for identifying ecological relationships between species and the environment where they occur (Franklin, 2010; Guisan & Zimmermann, 2000).

Species distribution models (SDMs) have been widely used to predict the distribution of deep-sea benthic communities (e.g. Beazley et al., 2018; Howell et al., 2022; Kenchington et al., 2016; Knudby et al., 2013; Rooper et al., 2016; Taranto et al., 2023; Tittensor et al., 2009). However, in deep-sea ecosystems, including CWC habitats on seamounts, species data are limited due to constraints in sampling accessibility and/or are biased due to the ad hoc nature of deep-sea surveys, for example, video surveys with remotely operated vehicles (ROVs). Predicting the distribution of species using limited data with high sampling bias presents a high risk of model overfitting and lower accuracy (Bean et al., 2012; Hernandez et al., 2006; Jeliazkov et al., 2022; van Proosdij et al., 2016; Wisz et al., 2008) since these datasets are less capable of capturing the full range of environmental conditions where species are present.

Different modelling approaches are available to model the distribution of species (Franklin, 2010; Guisan et al., 2017). However, it can be challenging for modellers to choose the best approach to use, given inherent bias associated with different methods, especially when limited datasets are available. Ensemble models combine multiple predictions obtained from other modelling methods (e.g. generalized additive models (GAM), random forest (RF) and boosted regression trees (BRT), among others) to create a single-model output (Araújo & New, 2007). The incorporation of multi-model predictions can improve predictive power and accuracy (Araújo & New, 2007) and mitigate single-model limitations (Meller et al., 2014; Ramirez-Reyes et al., 2021). Ensembles have been successfully used to predict the distribution of CWCs, with multiple studies reporting higher model performance and accuracy in comparison to individual model predictions (Georgian et al., 2019; Pearman et al., 2020; Ramiro-Sánchez et al., 2019; Robert et al., 2016; Rowden et al., 2017). Ensemble approaches can also perform well for model transfer, allowing to create new predictions across new environments in space and time (Crimmins et al., 2013; Hao et al., 2019; Jones & Cheung, 2015), making it suitable to model species distribution in data-poor regions (Breiner et al., 2015; Liu et al., 2019; Qazi et al., 2022). Model transfer from areas where more data records are available to areas with few data points could be a good starting point to tackle characteristic data limitations of deep-sea habitats (Bridges, Barnes, et al., 2023), as long as regions to be modelled are sufficiently similar and levels of uncertainty are reported (Buisson et al., 2010; Pearson et al., 2006).

The deep-sea region around the Cabo Verde archipelago, in the Central Atlantic Ocean, remains largely unexplored. With respect to deep-sea biodiversity, only a few studies are available (Chi et al., 2020; Hoving et al., 2020; Menezes et al., 2015; Stenvers et al., 2021). Nonetheless, the bathymetry of the islands, showing the presence of several seamounts, indicates that Cabo Verde is an area where VME indicator taxa may be present. Indeed, video records from two deep-sea research expeditions confirmed the presence of CWCs on the seamounts around the archipelago (Hansteen et al., 2014; Orejas et al., 2022).

In this study, we used an ensemble modelling approach, based on two modelling methods (GAM and RF), to predict the distribution of four CWC taxa, all indicators of VME, on five volcanic seamounts of Cabo Verde, where few records of species presence–absence are available. Our study aims to provide information on the environmental variables driving the distribution of the selected CWCs, with predictive habitat maps being useful to guide future research expeditions to the seamounts of the archipelago, as a basis for management and conservation planning.

2 | METHODS

2.1 | Study area

The Cabo Verde archipelago (14°–18° N; 21°–26° W) is a group of 10 islands and 5 islets of volcanic origin, located in the Eastern Central Atlantic Ocean, off the coast of West Africa (Figure 1). In 2010, the M80/3 Meteor Research Expedition (Hansteen et al., 2014) explored the geomorphology and volcanic origin of the seamounts of Cabo Verde (Kwasnitschka et al., 2024). Although the main scope of the expedition focused on geological studies, images collected during ROV dives gave some insights into the deep-sea benthic biodiversity present on the seamounts. One decade later, in 2021, the multidisciplinary research expedition iMirabilis2 (Orejas et al., 2022), part of the Horizon 2020 project iAtlantic, on board the R/V *Sarmiento de Gamboa* (UTM-CSIC), was the first expedition dedicated to exploring the deep-sea benthic communities of SW Cabo Verde. The recorded video transects, conducted using an ROV, revealed pristine CWC
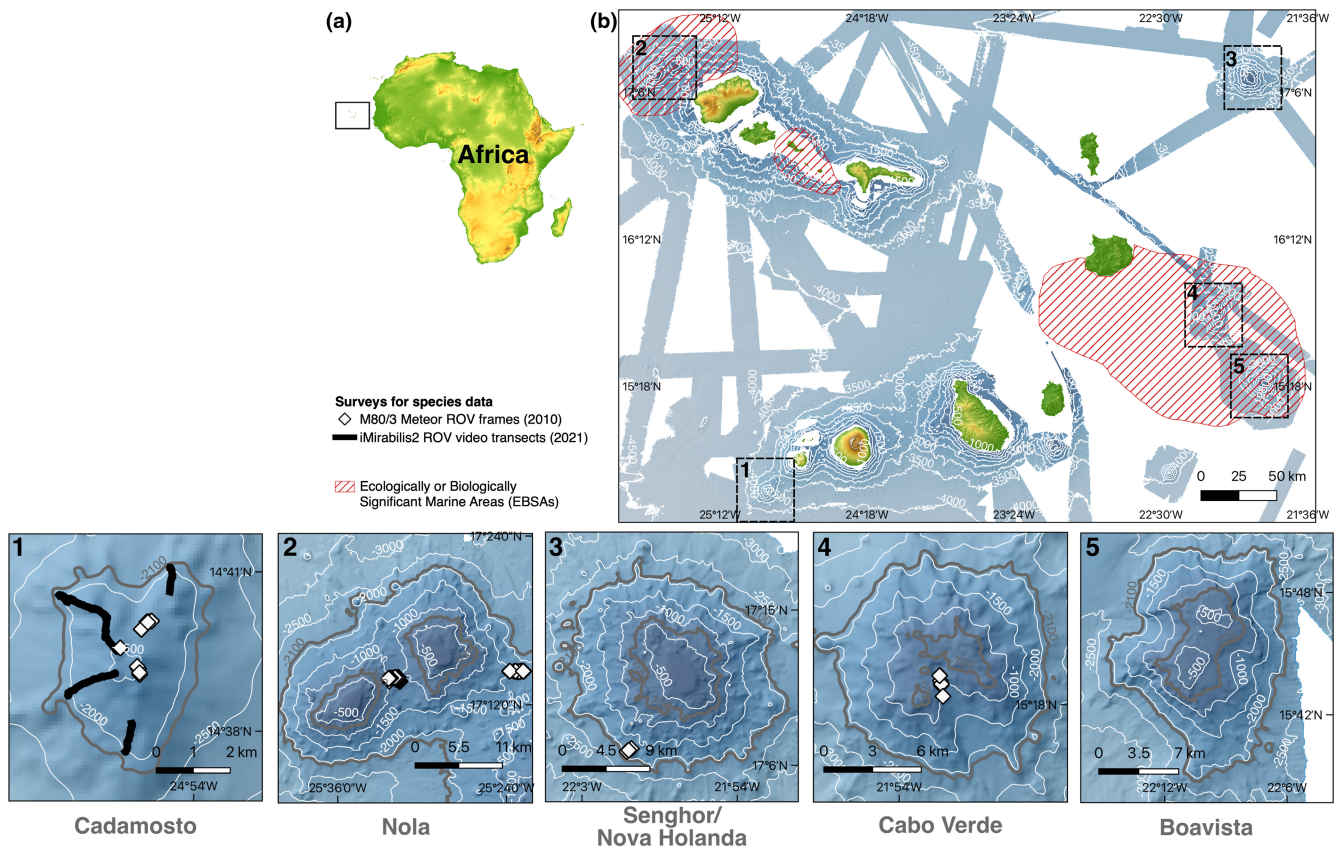
FIGURE 1 Study area map. (a) Location of the Cabo Verde archipelago in NW Africa (black square). (b) Map of Cabo Verde with the location of the five seamounts (Cadamosto, Nola, Senghor/Nova Holanda, Cabo Verde and Boavista) highlighted by the squares outlined with dashed lines, and of the designated ecologically or biologically significant marine areas (EBSAs) for the archipelago. The dark grey contour lines on the seamount's maps (1, 2, 3, 4 and 5) represent the surveyed depth range considered for modelling. Surveyed areas where species presence–absence data were collected are highlighted with black lines and white symbols. Black lines correspond to continuous video transects conducted in 2021 during the iMirabilis2 expedition and white symbols correspond to the locations of the video frames from GEOMAR's M80/3 meteor expedition, in 2010.

communities at Cadamosto Seamount (SW of Brava Island), including over 60 CWC morphospecies (Vinha et al., 2022).

In this study, the distribution of four VME indicator taxa was modelled for five seamounts of Cabo Verde (see Figure 1). SDMs were limited to a depth range between 750 and 2100 m on those seamounts, corresponding to the depth interval in which species data are currently available. When few species presence records are available, limiting predictions to the known surveyed range of species and environmental predictors results in more reliable and stable predictions (Grenouillet et al., 2011; Hernandez et al., 2006; Qazi et al., 2022). The five seamounts were chosen because they occur within the targeted depth of the models. A detailed description of the five seamounts is presented in Supplementary Text S1.

2.2 | Species data

The target taxa of this study (Figure 2) were the octocorals *Acanella arbuscula* (Johnson, 1862), *Metallogorgia* spp. and *Paramuricea* spp.

and the scleractinian coral *Enallopsammia rostrata* (Pourtalès, 1878), representing the most widely distributed and abundant VME indicator taxa observed in the available data for Cabo Verde.

Species presence–absence records were gathered from ROV footage collected during the M80/3 Meteor (2010) (Hansteen et al., 2014) and the iMirabilis2 (2021) (Orejas et al., 2022) research expeditions. Data from the two expeditions were available for four of the five targeted seamounts in this study (Vinha et al., 2023), with no available data on the distribution of these CWCs for Boavista Seamount since this seamount was not included in any of the ROV surveys of the two cruises.

Video data from the iMirabilis2 expedition were analysed, quantitively, using the open-source software BIIGLE (biigle.de) (Langenkämper et al., 2017). Observations from five continuous 1- to 2-km-long video transects between 2000 and 1400 m depth at Cadamosto Seamount were converted into presence–absence data points. Similar data were not available for the seamounts explored during M80/3 Meteor. However, all the available images and short video clips from that expedition were analysed to identify presence and absence points for each of the four target CWCs.
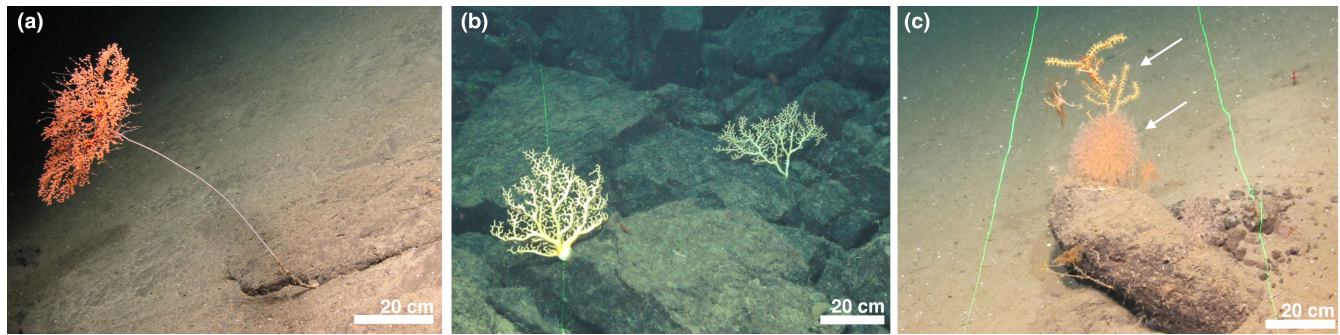
FIGURE 2 Selected VME indicator taxa for species distribution models of the seamounts of Cabo Verde. (a) *Metallogorgia* spp.; (b) *Enallopsammia rostrata*; (c) *Paramuricea* spp. (top arrow) and *Acanella arbuscula* (bottom arrow). Photo credits: iMirabilis2 (IEO, CSIC)/EMPEC/iAtlantic project.

TABLE 1 Total number of presence and absences records, for each VME indicator taxa, on each seamount.

Seamounts	Cadamosto	Nola	Senghor/Nova Holanda	Cabo Verde	Boavista
VME indicator taxa	**Depth range (m) of presence records/seamount**				
Acanella arbuscula	1890–1990	750–1330	1820	870–930	
Enallopsammia rostrata	1480–1900	1330–1770	1680–1720		
Metallogorgia spp.	1470–2100	1330–1770	1680–1720	980	
Paramuricea spp.	1450–1900		1730		
VME indicator taxa	**Total number of presences**		**Number of presence records/seamount**		
Acanella arbuscula	27		2	20	1	4	0
Enallopsammia rostrata	46		31	13	2	0	0
Metallogorgia spp.	46		35	8	2	1	0
Paramuricea spp.	29		28	0	1	0	0
VME indicator taxa	**Total number of absences**		**Number of absence records/seamount**		
Acanella arbuscula	141		96	20	9	16	0
Enallopsammia rostrata	122		67	27	8	20	0
Metallogorgia spp.	122		63	32	8	19	0
Paramuricea spp.	139		70	40	9	20	0

Note: For each taxon, a total of 82 records were collected in 2010 during the M80/3 ROV Meteor expedition across four of the five modelled seamounts, whereas a total of 86 records were collected in 2021 during the iMirabilis2 expedition, on Cadamosto seamount.

When there is a limited number of species occurrences and/or when species' absence locations are not available, it is common to use pseudo-absences (Lobo & Tognelli, 2011), that is, simulated absence points representing locations where a species is presumed not to occur. However, in our study, despite the low number of absences on each seamount, the decision to use true absence data, instead of pseudo-absence data, was made after running several model trials and observing low performance and accuracy in the models using pseudo-absence response data. Indeed, models built with true absences often result in improved predictive power and accuracy (Wisz & Guisan, 2009). Finally, all the available presence–absence data (Table 1) from the two expeditions were transformed into one-point per grid cell of the 100-m-resolution bathymetry grid, with the prevalence of the presence records over the absence records, in grid cells where both categories overlapped. Sampling effort, while variable across grid cells, was not accounted for in the models.

2.3 | Environmental data

Terrain variables were derived from a 100-m-resolution bathymetry grid, created from a compilation of all available bathymetry data collected by multibeam echosounder (MBES) in the Cabo Verde region. We used an analytical multiscale approach to calculate terrain variables by considering, when possible, different neighbourhood sizes (i.e. number of grid-cells (n)) for calculations. Using different neighbourhood (or window) sizes to calculate terrain variables allows to better capture different levels of terrain detail and heterogeneity (Lecours et al., 2015; Wilson et al., 2007). For example, the

distribution of CWCs can be influenced by variations in fine-scale terrain features, such as small mounds (Dolan et al., 2008), or broad-scale terrain features, such as seamount flanks (Guinan et al., 2009; Lecours et al., 2015; Wilson et al., 2007). Considering different analytical scales in terrain derivatives provides, therefore, a more comprehensive characterization of the environmental predictions driving species distribution, ultimately leading to improved model accuracy (Lecours et al., 2016; Wilson et al., 2007). Hence, in this study, slope, aspect (converted to eastness and northness) and three types of terrain curvature (plan, profile and mean) were calculated following a Fibonacci sequence of four increasing n values ($n = 3$, 9, 17, 33) (Dolan et al., 2008). For this, the functions 'SlpAsp' and 'Qfit' of the 'Multiscale DTM' library (Ilich et al., 2023) were used in R Studio Version 4.1.1 (R Studio Team, 2022). Topographic position index (TPI) and vector ruggedness measure (VRM) were calculated at two scales, both fine and broad scales ($n = 3, 33$), using the 'tpi' and 'vrm' functions, respectively, of the 'spatialEco' R Package (Evans & Ram, 2021). Roughness and terrain ruggedness index (TRI) were calculated using the 'terrain' function from the 'raster' R package (Hijmans et al., 2015), using the default $n = 3$. Final terrain variables and scales considered in the models were chosen after investigating collinearity between variables (see next section on initial variable selection).

The monthly averages of bottom temperature, bottom salinity and bottom zonal (U) and meridional (V) velocity components for the period of 2009 to 2019 were obtained from a hindcast simulation in the high-resolution VIKING20X ocean general circulation model (VIKING20X-JRA-OMIP described in Biastoch et al., 2021), with a native horizontal resolution of 1/20° (~5.3 km). Bottom U and V were converted into mean bottom current speed.

Each oceanographic parameter was converted into a single grid that represents the mean value for the time period considered. The final grid was then resampled to match the resolution of the final spatial scale of the other predictors (100 m), using a bilinear interpolation method, by applying the 'resample' function in the R package 'raster'.

All the final data layers (Table 2) were masked to consider the spatial extent of the five seamounts; and a depth mask, corresponding to the depth interval of the species data available (750–2100 m), was also used to represent the spatial extent used for modelling. All predictor layers were reprojected to a UTM26N coordinate system.

2.4 | Initial variable selection

An initial variable selection was done to fulfil model assumptions of independence between variables by assessing collinearity and variance inflation factor (VIF) of environmental variables. Variables with a Pearson's correlation coefficient >0.5 and VIF > 5 (Zuur, 2012) were removed from the initial set of variables considered for modelling. A preliminary RF model with all the initial set of environmental variables was used to determine which of the variables to remove between a pair of correlated variables, by assessing the mean decrease in Gini coefficient as a measure of variable importance. From the initial set of 34 environmental variables, 14 had collinearity issues (Figure S1) and were removed: mean curvature at all scales ($n = 3, 9, 17, 33$), broad- and fine-scale TPI, broad-scale VRM ($n = 33$), TRI, eastness at intermediate scales ($n = 9, 17$), northness at intermediate ($n = 9$) and broad scale ($n = 33$), slope at fine ($n = 3$) and intermediate scales ($n = 17$), plan curvature at intermediate scale ($n = 17$) and mean bottom current speed. Moreover, depth and mean bottom temperature had a VIF > 5 and were also removed.

TABLE 2 Environmental variables considered for species distribution modelling, with the resolution or neighbourhood size in grid cells (n), used, when applicable, and the respective unit.

Variable type	Variable	Spatial resolution	n	Unit
Water column depth	Depth	100 m		m
Terrain Slope	Slope		3, 9, 17, 33	degrees
Terrain orientation	Eastness		3, 9, 17, 33	
Terrain orientation	Northness		3, 9, 17, 33	
Terrain morphology	Profile curvature		3, 9, 17, 33	
Terrain morphology	Plan curvature		3, 9, 17, 33	
Terrain morphology	Mean Curvature		3, 9, 17, 33	
Terrain morphology	Topographic position index (TPI)		3, 15	
Terrain roughness	Roughness		3	
Terrain ruggedness	Terrain ruggedness index (TRI)		3	
Terrain rugosity	Vector ruggedness measure (VRM)		3, 33	
Oceanography	Mean bottom temperature	Native resolution of 1/20° and rescaled to 100 m		°C
Oceanography	Mean bottom salinity			psu
Oceanography	Mean bottom current speed			m/s

2.5 | Modelling methods

Two modelling methods (RF and GAM), commonly used in SDMs, were used to predict the probability of presence of four VME indicator taxa on the seamounts of Cabo Verde, including unexplored areas.

2.5.1 | Random forest (RF)

Random forest (RF) is a machine learning algorithm, where decision trees are used to make predictions on randomly selected subsets of the data, resulting in a final model that is a prediction of all trees combined (Breiman, 2001). In this study, an RF regression was implemented in R using the 'randomForest' package (Liaw & Wiener, 2002). To fine-tune the implementation of the model, the function 'TuneRF' of the same package was used to find out the optimal parameters, resulting in the implementation of an RF model with 500 trees and four variables at each split.

2.5.2 | Generalized additive model (GAM)

Generalized additive models (GAM) are an extension of generalized linear models (GLM), where a smoothing function is used to construct one best model of the relationship between the response and the predictor variables, in a flexible, non-linear way, being thus able to capture complex non-linear relationships between variables (Zuur, 2012). In this study, we used the 'mgcv' package in R (Wood, 2015) to fit a GAM using presence–absence data with a binomial distribution and using the restricted marginal-likelihood (REML) method for smoothing parameter estimation. The parameters used to tune each GAM for each taxon are presented in Table S1, and the smoothers plot, for the models of each taxon, is presented in Figure S2.

2.6 | Final variable selection

Variable selection is important to avoid models that are more susceptible to overfitting (Zuur, 2012) and harder to interpret, especially in regions where data records are limited (Qazi et al., 2022).

Therefore, final variable selection for RF was done using the Boruta algorithm (Kursa & Rudnicki, 2010), implemented in R using the 'Boruta' R Package (Kursa & Rudnicki, 2015). The Boruta algorithm evaluates variable importance by iteratively comparing the importance of a target variable against shuffled values of the same variable. Because it considers multi-variable relationships, its implementation is suitable both for classification and regression tasks (Borokini et al., 2023; Downie et al., 2021).

To fit GAMs, variable selection was done using forward stepwise variable selection. For this procedure, we started with an empty model, and variables were added based on their statistical significance (p-value < 0.05) until no more variables could be added without decreasing model performance. Model fit was then compared using the Akaike information criteria (AIC) score, and the final set of variables, resulting in the lowest AIC and most parsimonious model, were used to fit the model. The final set of variables selected for each model of each taxon is presented in Table S2.

2.7 | Model performance

Model performance was evaluated using a 5-fold cross-validation procedure. The method involves partitioning the data into five subsets (folds), where the model is trained on four of the folds and tested on the remaining fold. The process is repeated five times, and, at each time, a different fold is held for testing the model. Using fewer k folds (e.g. 5 folds instead of 10 folds) works better for small datasets by ensuring that each fold contains enough data for training and validation and by reducing variance in the estimated performance metrics, given that each fold will contain larger portions of data. In addition, given that a class-imbalanced dataset can result in biased model evaluation metrics, an equal proportion of presences and absences was kept at the split of each fold to account for class imbalance.

A threshold value above which the predicted probability is considered presence was calculated using the function 'optimal.thresholds' in the 'PresenceAbsence' R Package (Freeman & Moisen, 2008). We used a threshold that maximizes the sum of sensitivity (i.e. the proportion of accurately predicted presences) and specificity (i.e. the proportion of accurately predicted absences) to generate the confusion matrix and extract model evaluation metrics. A description of the evaluation metrics used to measure the prediction capacity of the model on the test data, at each fold repetition, is presented in Table S3. The final evaluation metric for each model was calculated as the mean value and standard deviation generated at each fold repetition.

Finally, the degree of spatial autocorrelation (SAC) was assessed in the final model residuals, at each fold, using the Moran's index, a statistical metric that measures the similarity between adjacent observations in a spatial distribution by comparing each observation to the average value of its neighbouring observations. SAC was not significant in the resulting models (Table S4) and, therefore, no additional method to account for SAC in the modelling framework was applied.

2.8 | Model predictions

The models fitted for the seamounts where species presence–absence data were available were transferred to make predictions on areas within seamounts where no data were available and on seamounts where no surveys were conducted (i.e. Boavista Seamount). Predictions of the probability of presence of each species for the

whole study area extent were generated at each fold repetition, for each modelling technique. The final prediction of each model represents the mean probability of presence considering all repetitions, with the standard deviation representing the prediction confidence (Figures S3–S5).

2.9 | Ensemble models

A weighted average ensemble model was calculated based on predictions generated at each 5-fold repetition of both model techniques (RF and GAM), following the method described in Nephin et al. (2020). This resulted in 10 different model predictions, weighed by the AUC score, considered in the final weighted average calculation. Ensemble model predictions were validated with the same withheld testing dataset generated across the 5-fold cross-validation procedure, as described in Section 2.7. Likewise, ensemble model uncertainty was measured with the coefficient of variation, calculated as the standard deviation divided by the mean, between model predictions at each fold repetition. For data visualization purposes, the values of the coefficient of variation were transformed into five categories of percentage (%) of uncertainty (0–20, 20–40, 40–60, 60–80 and 80–100).

Although wrapper R packages are available for ensemble modelling (e.g. 'biomod2' (Thuiller et al., 2016)), in our study, manually calculating the weighted averages and coefficient of variation was useful to avoid a 'black box' in the application of the ensemble approach.

2.10 | Model outputs and interpretation

2.10.1 | Binary maps

The mean probability of predicted presences obtained for each model was converted into binary maps, for all modelling techniques considered (i.e. GAM, RF and Ensemble). The optimal threshold obtained to evaluate model performance (see Section 2.7) was used as the criteria to classify a grid cell as presence or absence, that is, if probability > threshold, then presence (Table S5). Different thresholds were used to create the binary outputs, meaning that the predicted areas of presence of each CWC depend on the probability threshold used. For instance, a probability value classified as "presence" for one taxon might be classified as "absence" for another taxon with a higher threshold value. For the ensemble binary maps, presence probability thresholds were set at 30% for *A. arbuscula*, 44% for *E. rostrata*, 34% for *Metallogorgia* ssp. and 32% for *Paramuricea* spp. (see Table S5).

2.10.2 | Extrapolation analysis

Identifying areas where models are extrapolating is important to interpret model results and to communicate transparent model outputs. The recognition of regions of model extrapolation, obtained from making predictions in areas where environmental data are outside the environmental ranges of observed data, allows to gain insights on model limitations by identifying areas where targeted efforts should occur (i.e. focus of new surveys should be in extrapolated areas) or for conservation measures (i.e. if models are to be used for conservation, caution should be taken regarding model predictions in extrapolated areas). In this study, the extent and magnitude of extrapolation in model predictions were identified using the function 'compute.extrapolation' from the 'dsmextra' package (Bouchet et al., 2020). The function is based on the extrapolation detection (ExDet) tool (Mesgaran et al., 2014) and measures two types of extrapolations: univariate extrapolation (when ExDet values < 0, representing conditions outside the range of individual predictors in the reference sample) and combinatorial extrapolation (when ExDet > 1, representing new combinations of values within the univariate range of reference predictors).

2.10.3 | Response curves

The relationship between each environmental variable and the predicted values was analysed by plotting the functional response curves for each taxon (Figure S6). This was done by extracting the environmental data values at each obtained predicted probability value and by plotting the curves with a smoothing function with a 95% confidence interval, following the method described in Lopes et al. (2019). The functional response curves were plotted for the two modelling techniques and the ensemble model.

2.10.4 | Variable importance

The relative importance of each environmental variable was calculated to identify the most important variables for the distribution of each taxon and to compare results between models. For this, we applied the same method used by the 'biomod2' R package (Thuiller et al., 2016), where (1) a standard prediction with all data were made, then (2) the targeted variable was randomized and (3) a new prediction was made with the randomized dataset. Next, the (4) correlation score between the new prediction and the standard prediction was calculated and the score was considered to give an estimation of the variable importance to the model. A good correlation score between two predictions shows that the randomized variable has little importance (in other words, predictions are only slightly different), whereas low correlation score shows that the variable is important to the model (Thuiller et al., 2016).

3 | RESULTS

3.1 | Model performance

In general, the ensemble model consistently presented the best model performance for all taxa, while in terms of single-model performance, GAMs presented slightly higher performance than RF models, for all performance metrics (Figure 3).
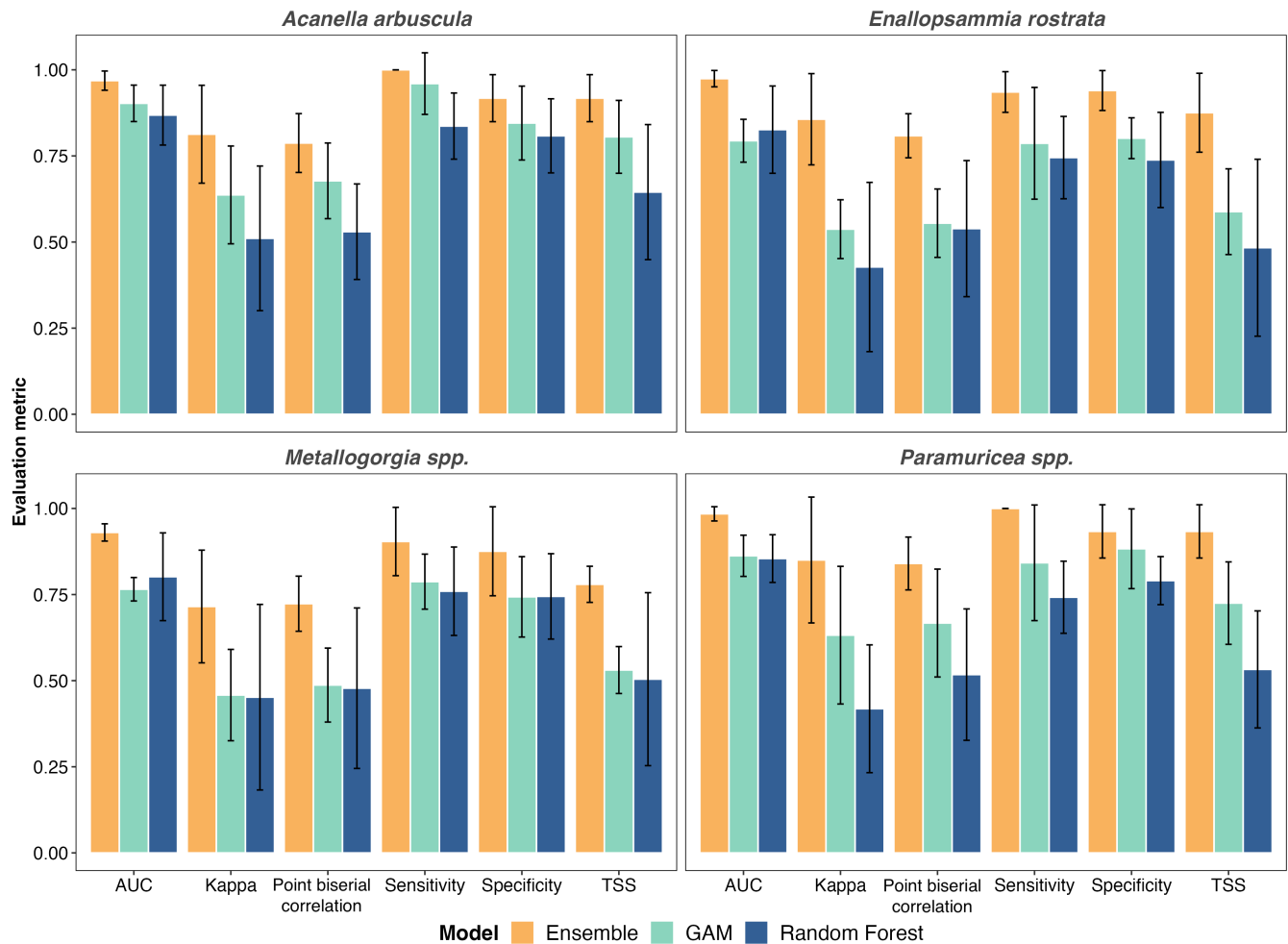
FIGURE 3 Model performance metrics (AUC, kappa, point-biserial correlation, sensitivity, specificity and true skill statistic (TSS)) for each of the modelled VME indicator taxa (*Acanella arbuscula*, *Enallopsammia rostrata*, *Metallogorgia* spp. and *Paramuricea* spp.) according to each modelling method (ensemble, GAM and random forest).

According to AUC scores, all models performed well (AUC > 0.76), but the ensemble models for all taxa showed the best overall model performance (AUC > 0.93). Random forest models yielded higher AUC scores for *E. rostrata* and *Metallogorgia* spp. Conversely, GAM models performed better than RF models for *A. arbuscula* and *Paramuricea* spp., with the GAM models of *A. arbuscula* exhibiting the highest AUC values of all RF and GAM models.

Similarly, Kappa values showed better performance with the ensemble models (Kappa > 0.85) in comparison to the individual GAM and RF models, indicating a good agreement between model predictions and the observed data. According to Kappa, GAM presented a higher performance than RF for all taxa, expect for *Metallogorgia* spp., where similar kappa values were obtained for both models but with a higher standard deviation for RF. A similar model performance pattern was also observed considering the point-biserial correlation values, with the ensemble models outperforming the two individual models (point-biserial correlation > 0.72) for all four CWCs.

In terms of sensitivity (i.e. model capacity to correctly identify presences) and specificity (i.e. model capacity to correctly identify

absences), all ensemble models demonstrated, consistently, the highest performance across all taxa (sensitivity and specificity > 0.88). For all taxa, GAMs presented higher sensitivity and specificity than RF models. Furthermore, this was also reflected in the high true skill statistic (TSS) values of the ensemble models (TSS > 0.77 for all taxa), suggesting a good discrimination power between presences and absences.

3.2 | Most contributing variables

The relative contribution of variables (Figure 4) for the distribution of each modelled CWC consistently showed a higher importance of broad-scale terrain features (across 33 grid cells or 3300 m) than finer-scale (across three grid cells or 300 m), although variable importance differed between the two modelling methods.

Both RF and GAM agreed on the importance of mean bottom salinity as one of the most important variables for the distribution of *A. arbuscula*. In addition, terrain curvature (profile and plan) and
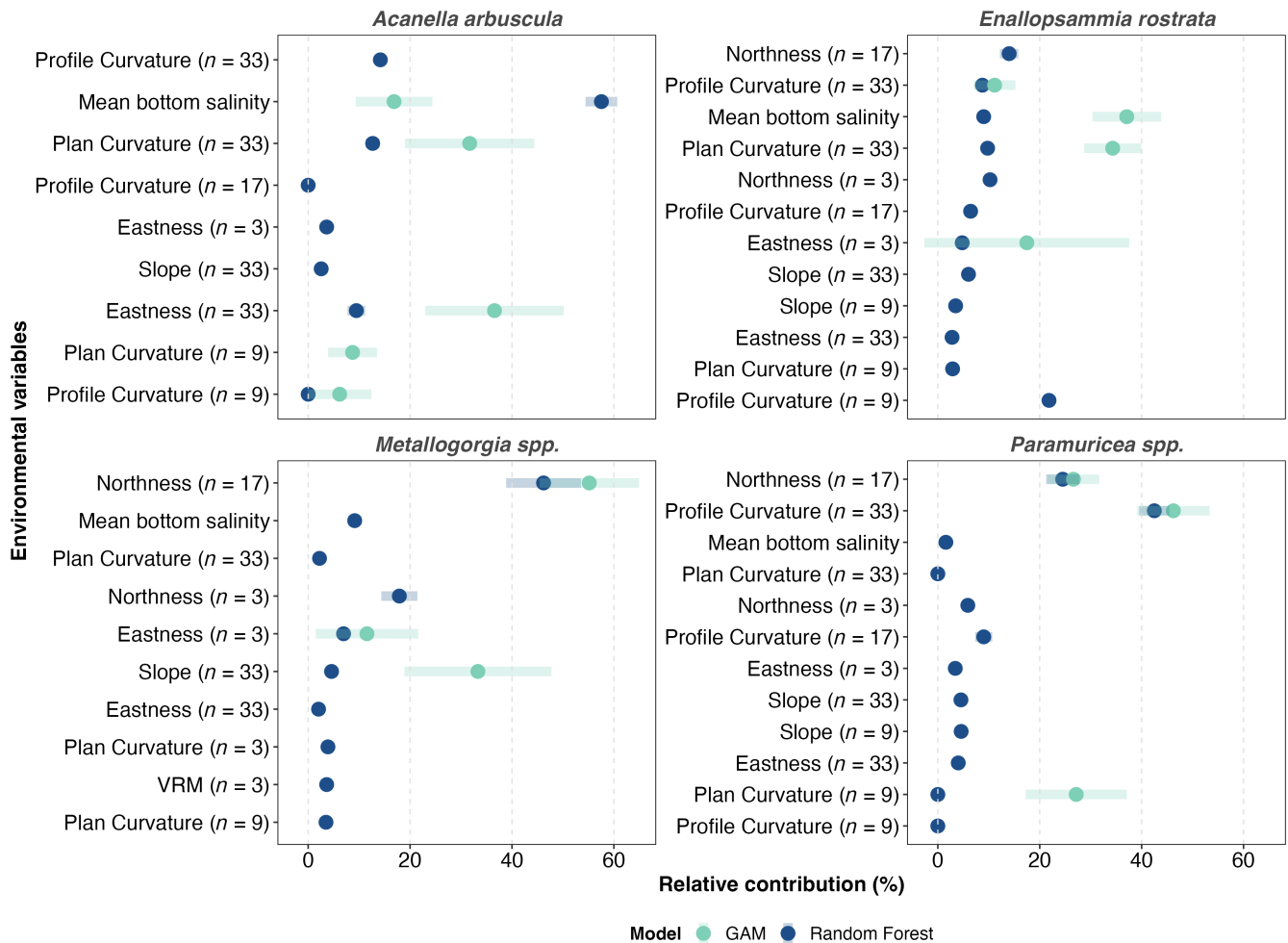
FIGURE 4 Relative contribution of each environmental variable, with *n* representing the neighbourhood size (in grid cells) used for multi-scale terrain variables calculation, for all the modelled VME indicator taxa. For each model (GAM and RF), dots represent the mean obtained value across the five folds in the cross-validation procedure and the horizontal bars represent the respective standard deviation.

orientation (eastness), both measured at broad scale (*n*=33), were also considered important variables according to both models.

For the scleractinian coral *E. rostrata*, variables related to terrain curvature (profile curvature) and orientation (northness), calculated at intermediate scales (*n*=9 and 17), were among the most contributing variables for RF models, whereas mean bottom salinity was the highest contributing variable for the GAM models, followed by broad-scale plan curvature (*n*=33).

Terrain orientation at intermediate scale (i.e. northness (*n*=17)) was consistently the most contributing variable for both RF and GAM models of *Metallogorgia* spp., while variables of fine-scale terrain orientation (*n*=3) were also among the most contributing ones for both individual distribution models of this CWC. In addition, broad-scale slope (*n*=33) presented the second highest relative variable for the GAM models of *Metallogorgia* spp.

For the octocorals from the genus *Paramuricea* spp., terrain morphology (i.e. profile curvature (*n*=33)) and orientation (i.e. northness (*n*=17)) at intermediate to broader scales were the most important variables for the distribution of this CWC, according to both models.

3.3 | Response curves

The most suitable conditions for *A. arbuscula* were observed within a mean bottom salinity range between 34.85 and 35.90 psu (Figure 5). The probability of presence decreased at 34.95 psu, coinciding with the value where the probability of presence of other taxa started to increase. Moreover, higher probability of presence of *A. arbuscula* was associated with negative values of terrain plan curvature, that is, areas of downslope concavity.

The highest probability of presence of *E. rostrata* was associated with values of mean bottom salinity over 35.00 psu, on moderate sloping terrain (maximum at 10°) linked with broad-scale terrain features with positive profile curvature (i.e. with upwardly concave shape).

The highest probability of presence of *Metallogorgia* spp. coincided with steeper slope values (broad-scale slope higher than 15°), facing a northward direction.

The most suitable habitat for *Paramuricea* spp. was associated with positive values of profile and plan curvature, at intermediate and broad scales.
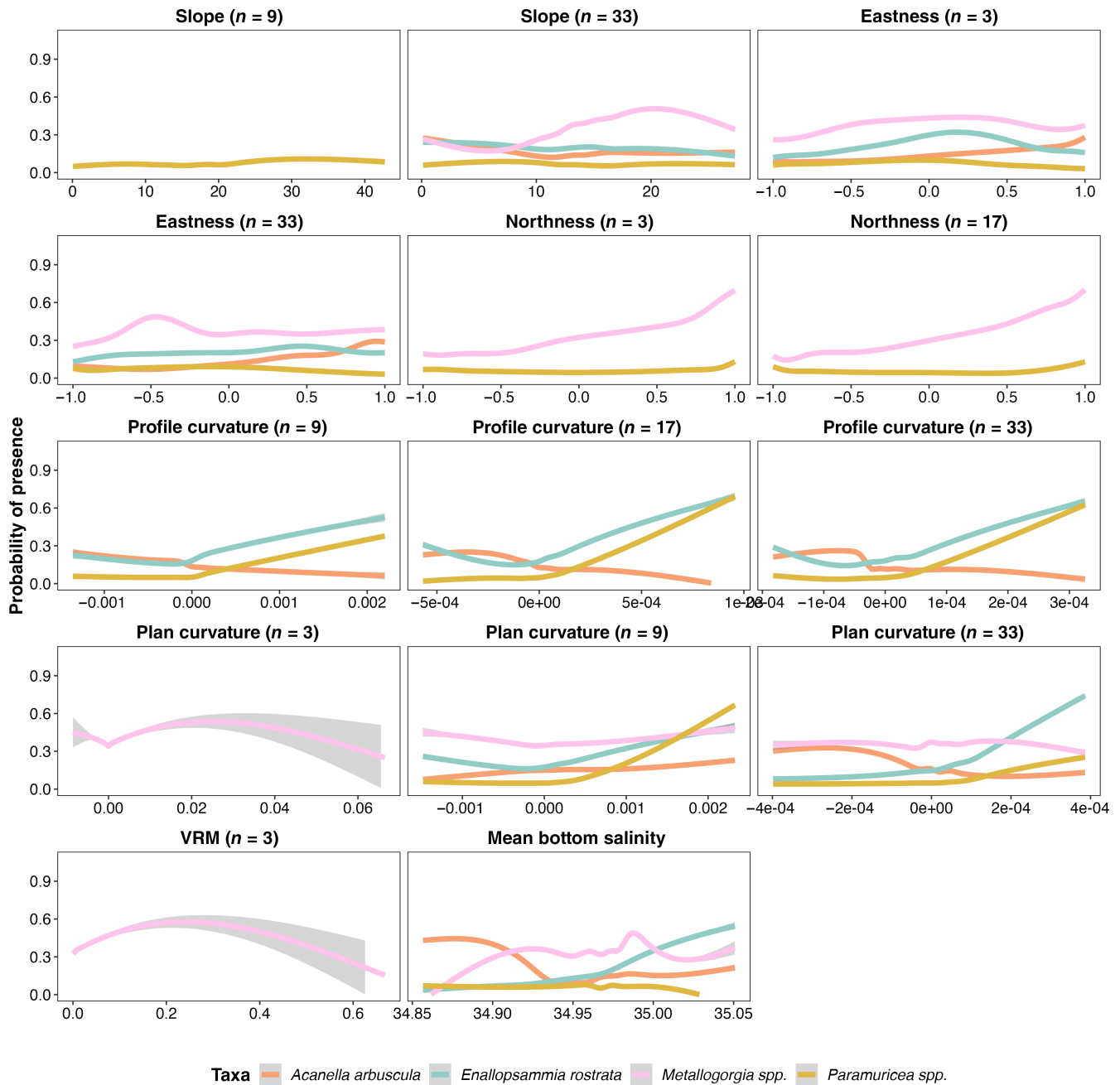
FIGURE 5 Functional response curves of each environmental variable used to fit the models, for each VME indicator taxa, based on the probability of predicted presence of the ensemble models.

3.4 | Observed and predicted distribution of CWCs

The available observational data on the distribution of the four CWCs showed that *Metallogorgia* spp. and *A. arbuscula* were the only taxa where presence records were observed on all four surveyed seamounts (Figure 6). However, the observed presence–absence data covered a limited spatial extent and depth range (from 750 to 2100 m depth) on each seamount and, therefore, it is likely that the range of environmental conditions occupied by each CWC was not represented in the response dataset.

All modelled taxa presented areas with relatively high predicted probability of presence (Figure 7), however, the octocoral *Metallogorgia* spp. was the taxon with the largest suitable area across the five seamounts, with a predicted area of presence covering a total of 487.05 km² (Figure 8, Table 3). Conversely, *Paramuricea* spp. had the smallest predicted areas of occurrence, with only 24.83 km² of suitable habitat identified across the five seamounts.

The bamboo coral *A. arbuscula* was observed on the four seamounts where species data were available, with the deepest observed record at 2000 m (on Cadamosto) and the shallowest at
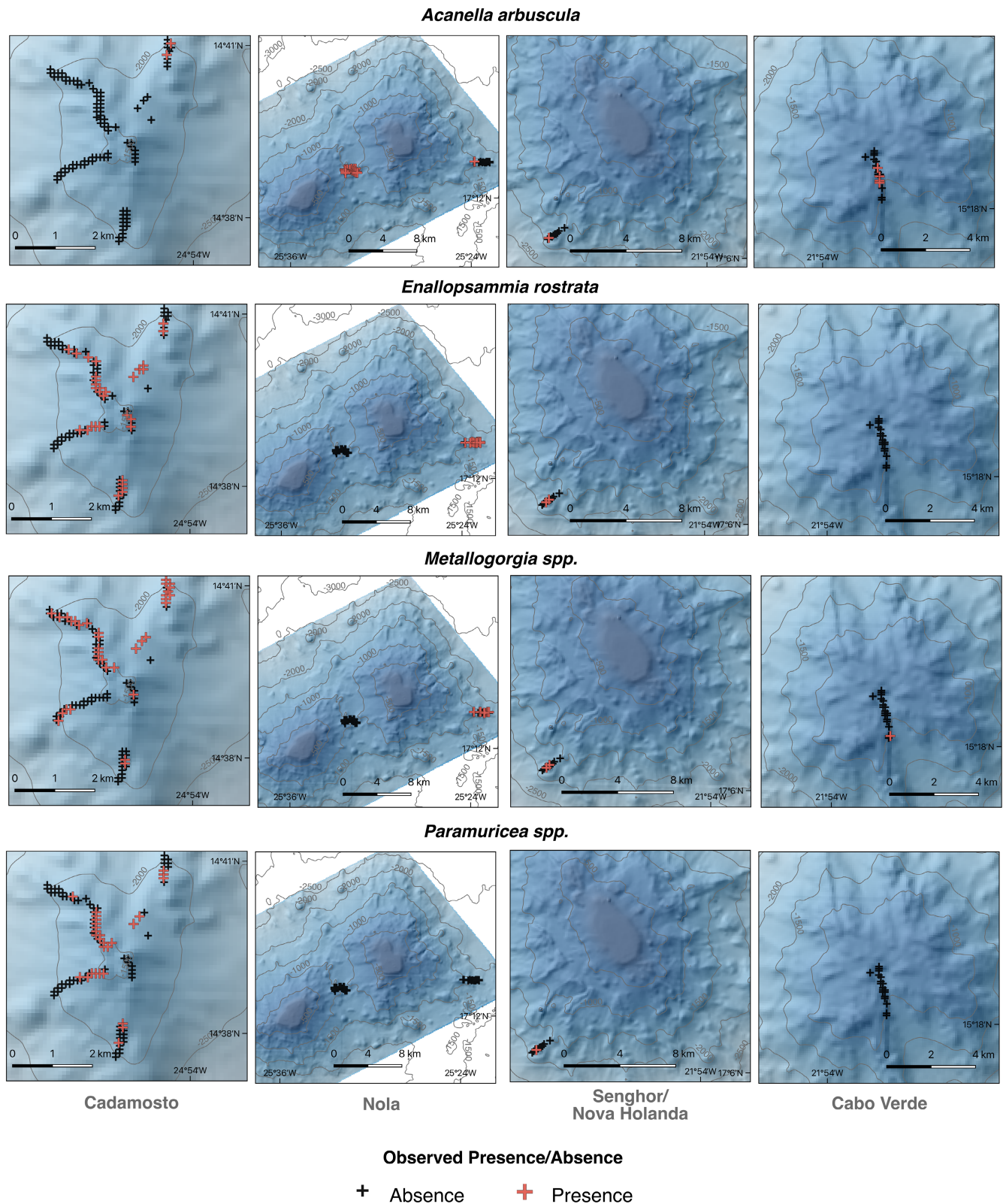
FIGURE 6 Observed presence–absence records, for the modelled VME indicator taxa, during ROV surveys on each seamount. No surveys were conducted on Boavista Seamount.

750 m depth (both on Nola and Cabo Verde seamounts). Most presence records of *A. arbuscula* were found on Nola Seamount in the area between the two seamount's summits (Figure 6). Similarly,

model predictions of *A. arbuscula* showed that Nola was the seamount with the largest predicted area of presence (61.95 km²) and the areas with the highest probability of occurrence were between
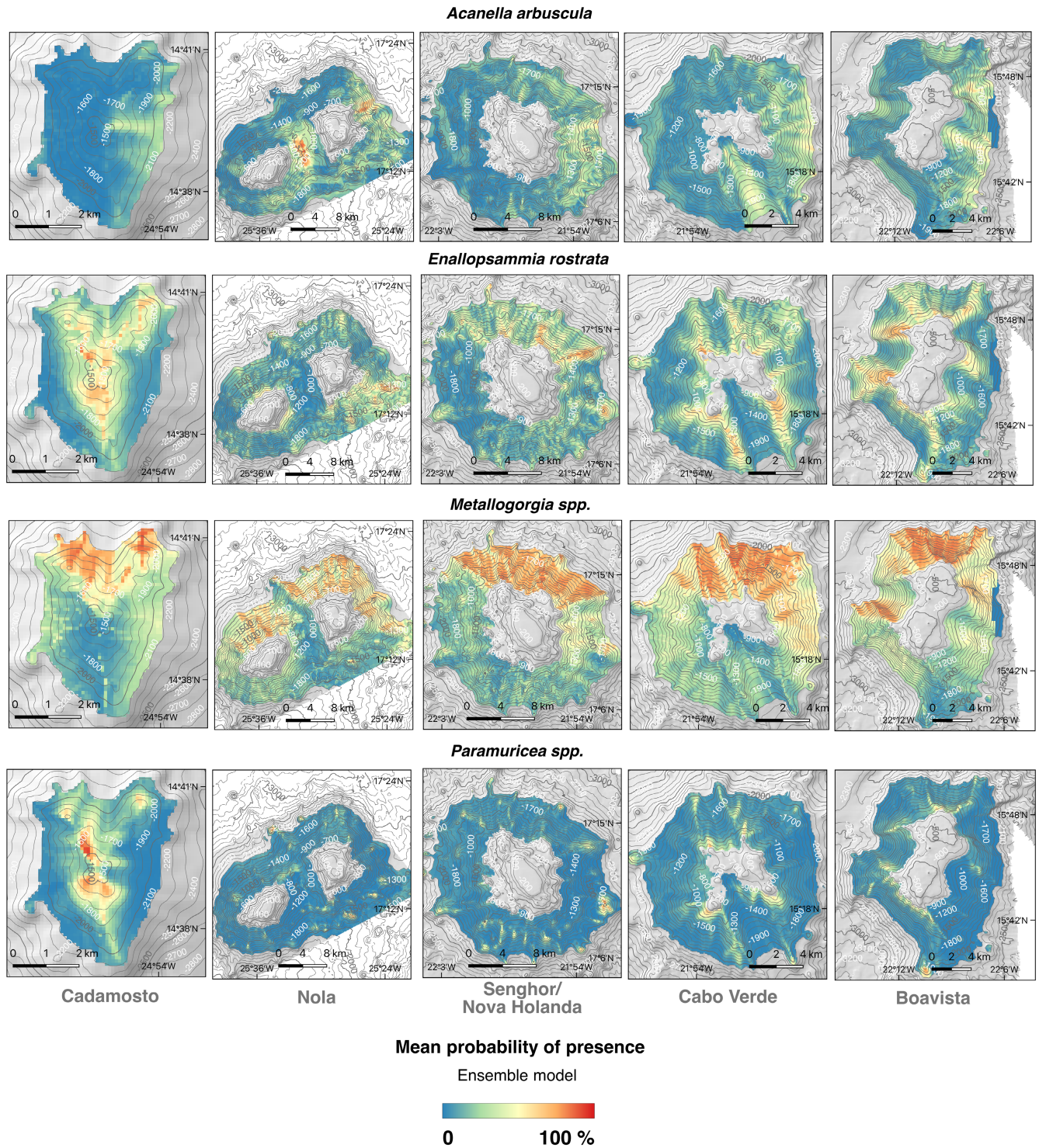
FIGURE 7 Mean probability maps of presence, according to the Ensemble model predictions, for each VME indicator taxa, at each seamount.

the two summits of the seamount (Figure 7), following the patterns of the observed data. On the other seamounts, the probability of presence of *A. arbuscula* was usually associated with the eastern sides of the seamounts. Cadamosto Seamount presented the lowest modelled probability of occurrence for *A. arbuscula*, with few predicted areas of presence. This is consistent with video observations where only two specimens of *A. arbuscula* were observed.

Observational data showed the presence of the scleractinian coral *E. rostrata* on Cadamosto, Nola and Senghor seamounts. Species distribution observations of *E. rostrata* ranged from 1700 to 1500 m on Cadamosto and Senghor and from 1600 to 1000 m depth on Nola. However, according to the modelled distribution, Boavista Seamount was the seamount with the largest predicted area of presence of *E. rostrata* (predicted area of presence of
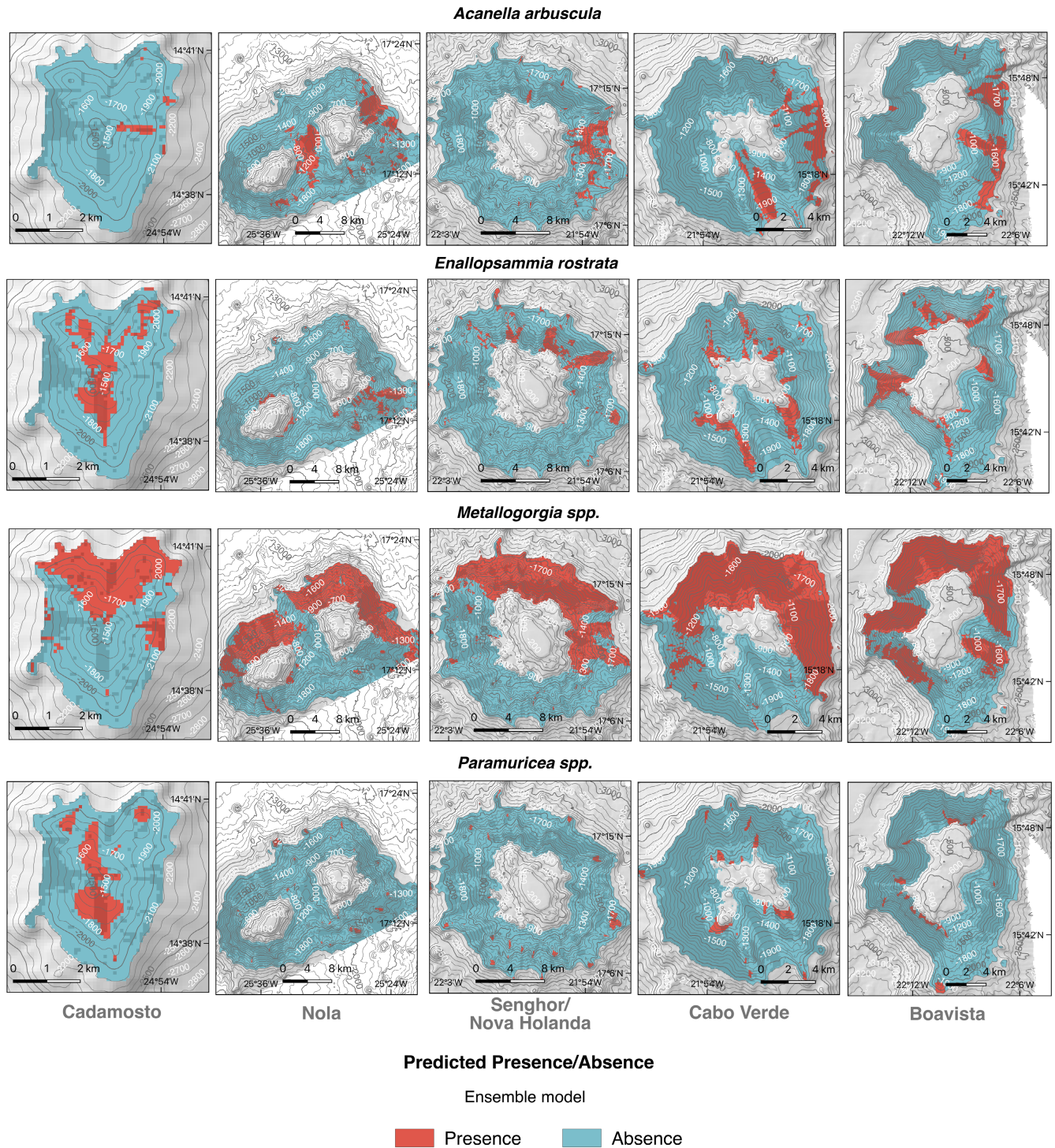
FIGURE 8 Categorical maps of the predicted presence and absence, according to the ensemble model predictions of each VME indicator taxa, at each seamount. Shaded grey areas indicate extrapolated areas.

28.05 km²), followed by Nola and Senghor seamounts. On Boavista, the ensemble model predicted the presence of *E. rostrata* along the most pronounced ridges on the seamount flanks, throughout the whole bathymetric range modelled in this study. A similar modelled distribution pattern was observed on Cabo Verde Seamount. On Nola, the areas with the highest probability of presence of *E. rostrata* were located on the SE side from 1700 to 1300 m depth.

Although with a lower probability of presence, there were some suitable areas for the species on the NE side, within the same depth range. On Senghor, areas with high probability of presence for *E. rostrata* were located on the northward-facing side, between 2000 and 1200 m depth. On this seamount, the presence of *E. rostrata* was also predicted on smaller mounds at 2000 m depth. At Cadamosto Seamount, the predicted distribution of *E. rostrata*

TABLE 3 Estimated predicted suitable area (in km²), according to the ensemble model, for each VME indicator taxa at each seamount.

VME indicator taxa	Predicted suitable area (km²)					Total suitable area (km²)
	Cadamosto	Nola	Senghor/ Nova Holanda	Cabo Verde	Boavista	
Acanella arbuscula	0.40	61.35	20.73	14.74	23.85	121.08
Enallopsammia rostrata	2.99	25.26	24.09	13.07	28.05	93.46
Metallogorgia spp.	6.25	199.82	111.37	69.10	100.50	487.05
Paramuricea spp.	2.75	5.24	7.90	3.99	4.94	24.83

covered a wider spatial distribution, although the areas with the highest probability of presence were above 1700m depth.

The available observational data for *Metallogorgia* spp. showed a widespread spatial distribution on Cadamosto Seamount, whereas this was not the case for the distribution of this CWC on the other seamounts. For example, on Nola, *Metallogorgia* spp. was present on the east side but absent on the west side on the seamount. The observed depth range of *Metallogorgia* spp. was from 2100m (on Cadamosto Seamount) to 1000m (on Cabo Verde Seamount). In terms of model predictions, *Metallogorgia* spp. presented significant predicted areas of presence on all five seamounts, with the highest probabilities on northward seamount slopes. Nola and Senghor were the seamounts with the highest probability of presence for *Metallogorgia* spp., followed by Boavista. On the first two seamounts, the most suitable areas of presence were located on the N and NE sides, while on Boavista, the modelled distribution of *Metallogorgia* spp. spanned across the whole seamount. The predicted areas of presence for this CWC were also large on Cabo Verde Seamount where the predicted distribution of *Metallogorgia* spp. also favoured the northward-facing slopes. Similarly, the most suitable areas for *Metallogorgia* spp. on Cadamosto were located on the north side of the seamount, with the highest probability of presence around 1700m depth.

Finally, the octocoral *Paramuricea* spp. was observed on Cadamosto and Senghor seamounts, along a narrow bathymetric distribution from 1800 to 1500m of water depth. On Cadamosto, most presence records of this CWC were observed near the summit, at 1400m of depth (see Figure 6). This was the CWC with the smallest predicted areas of presence on the five seamounts. According to the ensemble model, Cadamosto was the seamount with the highest probability of presence. The areas predicted to be most suitable for *Paramuricea* spp. corresponded to small seafloor topographical elevations near the summit of Cadamosto at 1500m depth. Although there was no observation presence data of *Paramuricea* on Nola and Cabo Verde, the models predicted the presence of this CWC on all seamounts.

3.5 | Uncertainty

The model predictions with the highest uncertainty (Figure 9) were the ones for *Paramuricea* spp., where over 50% of the total modelled area (i.e. total area of the five seamounts) showed an uncertainty above 60%. On the other hand, *Metallogorgia* spp. was the taxon

with the overall highest prediction confidence (in other words, with the lowest uncertainty).

In general, for the four taxa on the five seamounts, the predicted areas of absence had higher uncertainty, compared with the predicted areas of presence. Cabo Verde and Boavista were the seamounts with the highest prediction uncertainty for all taxa or, in other words, with the highest model disagreement between single-model predictions; whereas Cadamosto was the seamount with the highest confidence of predictions. The first two seamounts were the least surveyed seamounts, while Cadamosto was the seamount where the highest number of species observations were available. Similarly, the extrapolation analyses highlighted these differences in sampling effort, with Boavista and Cabo Verde presenting the highest proportions of extrapolated area (over 60% of the modelled area was extrapolated) (Table 4), while the proportion of extrapolated area decreased on the seamounts with more observations.

For all seamounts, extrapolated areas occurred where no species' presence–absence data were available, overlapping with areas of higher uncertainty (see Figures 8 and 9). Slope at broad scale and mean bottom salinity were the most influential variables to the extrapolation (Table S6), with the first variable contributing to over 20% of the extrapolated area. The environmental conditions of the extrapolation represent areas on the seamounts with flatter-to-intermediate slopes (average slope of 12±3°) and with profile curvature values close to 0, indicating that relatively flat terrain features were the most undersampled environmental conditions on the seamounts. This may reflect the tendency to target rugged features when planning ROV surveys.

For *A. arbuscula*, Cabo Verde and Boavista were the seamounts with the highest uncertainty, in which 35% of the modelled area presented an uncertainty category higher than 60%. The areas with the highest uncertainty predictions corresponded to the western side of these seamounts where the ensemble models predicted absence areas for the species (see Figure 7). A similar pattern was observed for Senghor Seamount. However, over 70% of the total modelled area of Cadamosto and Nola presented an uncertainty of less than 60%, with the areas of highest confidence (uncertainty <20%) in predicted presence areas for *A. arbuscula*.

Models for *E. rostrata* and *Metallogorgia* spp. showed predictions with the highest confidence on Cadamosto Seamount, where most of the modelled area corresponded to an uncertainty of less than 60% for both species. Nola was the seamount with the highest
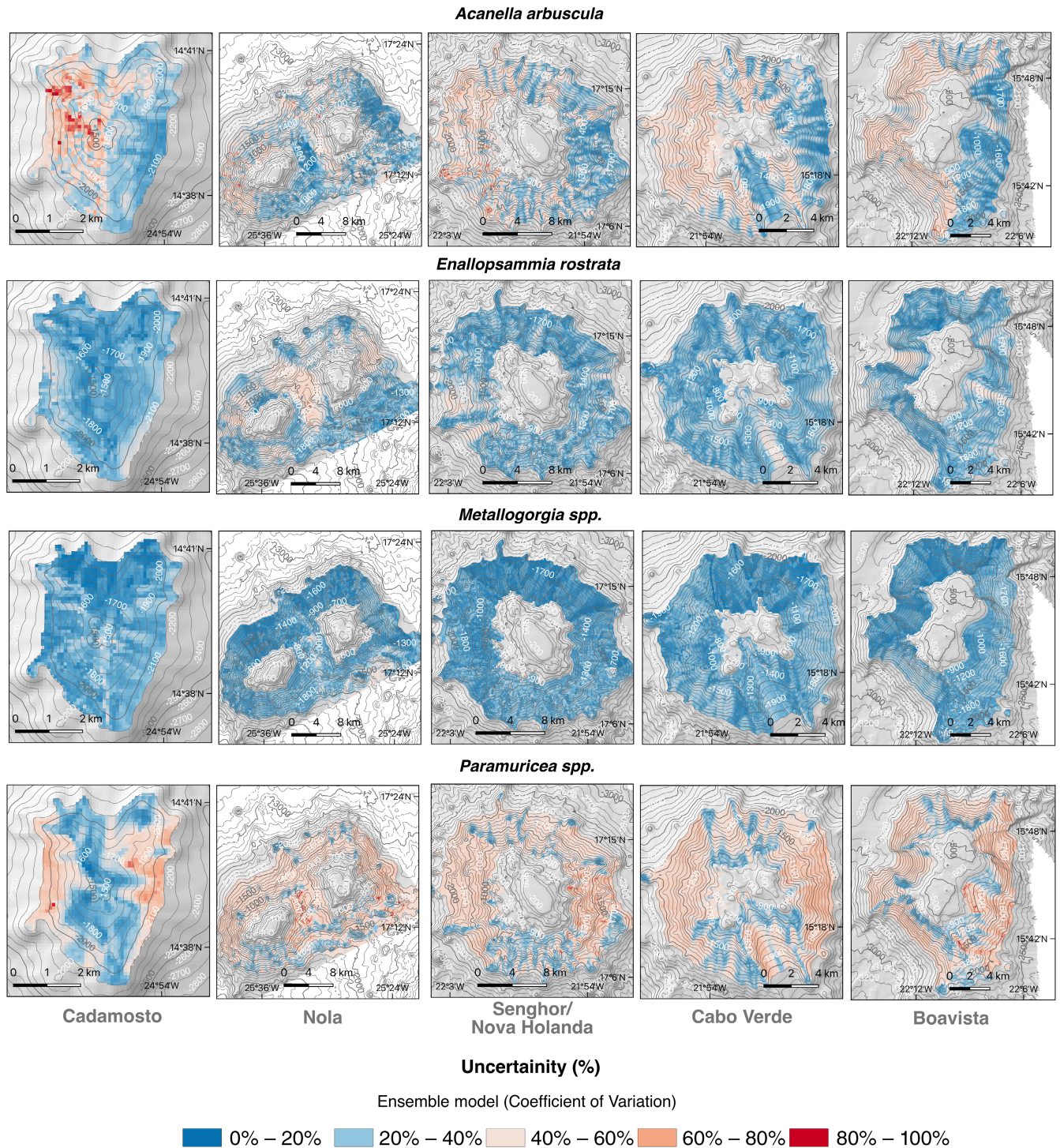
FIGURE 9 Categorical uncertainty maps, based on the coefficient of variation, for the predictions of the ensemble models of each VME indicator taxa, at each seamount.

uncertainty for the distribution models of _E. rostrata_, with the area with the highest model disagreement located between the two summits of Nola, where the ensemble model predicted an absence area. Approximately 30% of the modelled area of Senghor, Cabo Verde and Boavista presented an uncertainty category greater than 40% for _E. rostrata_.

The low uncertainty in the predicted distribution of _Metallogorgia_ spp. was consistent across the five seamounts, and there were no areas on any of the seamounts with an uncertainty category above 80%. Nola and Cabo Verde were the seamounts with the highest predicted uncertainty with 20% of the modelled area in these seamounts within an uncertainty category of 40%–60%. The area with

TABLE 4 Percentage (%) of the total modelled area that is extrapolated for each VME indicator taxa at each seamount.

VME indicator taxa	% of extrapolated area of the total modelled area				
	Cadamosto	Nola	Senghor/Nova Holanda	Cabo Verde	Boavista
Acanella arbuscula	18	32	31	68	62
Enallopsammia rostrata	18	34	31	68	62
Metallogorgia spp.	18	46	34	67	66
Paramuricea spp.	20	50	36	70	67

highest uncertainty for the predicted distribution of *Metallogorgia* spp. on Nola corresponded to a predicted absence area in the deepest part of the SE side of the seamount.

For *Paramuricea* spp., Cadamosto was the seamount with the lowest uncertainty predictions, however, on the other four seamounts, over 60% of the modelled area presented an uncertainty greater than 60%. Similar to what was observed for the other taxa on the other seamounts, the areas with the highest uncertainty represent areas of predicted absence for this CWC.

4 | DISCUSSION

This study used species distribution models (SDMs) to predict the distribution of VME indicator taxa in a poorly explored deep-sea region of the Central Atlantic Ocean. The use of an ensemble model proved to be a good approach to address the data limitations of our study, being capable of predicting new CWC presence areas on the seamounts of Cabo Verde and exhibiting higher model performance metrics in comparison to single-model GAM and RF predictions. However, the data limitations of our study have also contributed to high model uncertainty and large areas of extrapolation, especially in undersampled areas and on seamounts where fewer species observations were available. Nonetheless, these results are essential to guide future surveys in the study area by highlighting areas where sampling effort should focus. Below, we identify and discuss the main environmental patterns correlated with the distribution of the modelled taxa as well as associated model caveats and limitations.

4.1 | Environmental conditions for the presence of VME indicator taxa on the seamounts of Cabo Verde

Broad-scale terrain features on the seamounts of Cabo Verde were considered important for the distribution models of the four CWCs, with terrain orientation (northness and eastness) being consistently among the most contributing variables for the distribution of all taxa. Measures of terrain orientation can be used as surrogates to hydrodynamic flow (Guinan et al., 2009; Lecours et al., 2016; Wilson et al., 2007), suggesting the exposure of CWCs to local and regional current directions (Guinan et al., 2009). In fact, on seamounts, interactions between topography and hydrodynamics are a

key mechanism for food supply of benthic suspension feeders, since steeper seamount slopes promote current amplification (Guinan et al., 2009; Mohn et al., 2014; Wilson et al., 2007), internal waves and vertical mixing, leading to higher organic matter fluxes to the seafloor (Davies et al., 2009; Dolan et al., 2008; Mohn et al., 2014; Mosquera Giménez et al., 2019; Pearman et al., 2020; Rengstorf et al., 2014). This dynamic phenomenon has been described around Cabo Verde, where the interaction of the island's complex bathymetry with mesoscale eddies (Cardoso et al., 2020) propagates large-amplitude internal waves that increase phytoplankton and zooplankton production above the seamounts' summits (Mohn et al., 2021), eventually resulting in higher food supply to benthic communities.

Terrain variables measured at a broad scale presented the highest relative contribution to the distribution models of the octocoral *Metallogorgia* spp., whereas for the other taxa, the contribution of oceanographical and terrain parameters at finer scales was also evident. The predictive maps of *Metallogorgia* spp. showed a wider distribution compared to the predicted distribution of the other CWCs (see Figure 8, Figure S7), which might result from the higher contribution of such broader-scale processes influencing its distribution. In fact, *Metallogorgia* spp. are commonly observed on seamounts, across a wide bathymetric and spatial distribution (Auscavitch et al., 2020; Lapointe et al., 2020). On the other hand, the influence of variables at finer-to-intermediate scales could explain why there were more limited suitable areas for the other CWCs. For example, the predicted suitable areas of *E. rostrata* were located between 1700 and 1200 m depth on steeper slopes along the pronounced ridges of seamount flanks. However, in these large features, at finer scales, terrain orientation and curvature seem to be driving the distribution of *E. rostrata*. Similar results showcasing the importance of slope, aspect and curvature were observed for *E. rostrata* and other scleractinian corals on the seamounts of New Zealand (Rowden et al., 2017; Tracey et al., 2011), since steeper and elevated seafloor features (i.e. positive curvature), facing the direction of prominent regional currents, ensure more efficient transport and capture of food particles (Dolan et al., 2008; Lecours et al., 2016; Wilson et al., 2007).

Our results suggest that curvature and aspect of large terrain features also influenced the distribution of *Paramuricea* spp. on the seamounts of Cabo Verde. Studies in the Gulf of Mexico, where the distribution of a similar morphospecies of *Paramuricea* spp. as the one in this study was modelled, showed that this taxon was associated with a very restricted suitable habitat driven by a

narrow temperature tolerance (Georgian et al., 2020), bathymetric highs (Georgian et al., 2020) and local current structure (Doughty et al., 2014) and orientation (Liu et al., 2021). In this study, most seamounts presented a small suitable extent for the presence of *Paramuricea* spp., while Cadamosto presented the largest suitable areas. This could be because, according to the available observations for our study region, the suitable environmental characteristics for this taxon involve large surfaces with a convexly shaped curvature, typical of seamounts' summits (Yesson et al., 2011). Cadamosto was the only seamount where the modelling spatial extent covered the summit due to the restricted modelled depth considered to reduce the risk of extrapolating results. This could explain the limited predicted area of this CWC on the other seamounts, given that the summit of the other seamounts considered in our study was not included in the modelling extent because they were outside the modelling depth range considered here.

Mean bottom salinity was revealed to be important for the distribution of the bamboo coral *A. arbuscula* and the scleractinian coral *E. rostrata*, suggesting that some specific water-column properties could be driving the suitable habitat of these CWCs. Other modelling studies have shown the importance of salinity for the distribution of both *A. arbuscula* (Buhl-Mortensen et al., 2015; Yesson et al., 2012) and *E. rostrata* (Anderson, Guinotte, Rowden, Clark, et al., 2016; Anderson, Guinotte, Rowden, Tracey, et al., 2016). Indeed, water physical–chemical boundaries (Davies & Guinotte, 2011; Dullo et al., 2008; Georgian et al., 2020) and, consequently, different water masses (Auscavitch et al., 2020; Buhl-Mortensen et al., 2015; Lapointe et al., 2020; Mosquera Giménez et al., 2019; Puerta et al., 2020, 2022; Quattrini et al., 2017; Taranto et al., 2023) are known to influence the distribution of CWCs. Interestingly, our results showed a disagreement between the values of mean bottom salinity for the maximum probability of presence of the two species, with the salinity peak for *A. arbuscula* around 34.85 psu and over 35.00 psu for *E. rostrata* (see Figure 5). In fact, the distribution of water masses in the region shows the influence of the Antarctic Intermediate Waters (AAIW) around 900 m depth, where a salinity minimum is observed, and a slight salinity increase at 1200 m, indicating the presence of the North Atlantic Deep Waters (NADW) (Mosquera-Giménez et al., 2022). The predicted depth range of these species matches the distribution of these water masses, with the association of *A. arbuscula* with the AAIW and *E. rostrata* with the NADW. This suggests that a narrow salinity envelope could limit the distribution of CWCs to specific areas on the seamounts that also meet the suitable terrain characteristics for each taxon. Furthermore, besides the salinity threshold, and contrary to *E. rostrata* discussed above, the suitable areas for *A. arbuscula* were also associated with flat areas on the seamounts. A similar result was obtained for the SDMs of another bamboo coral *Isidella elongata* on seamounts of the Mallorca Channel (Mediterranean Sea), where the models presented higher suitability on the seamount's flat terraces (Standaert et al., 2023), significantly exposing this species to bottom trawling impacts (González-Irusta et al., 2022).

4.2 | Modelling data-limited regions: Caveats and lessons learned

Models fitted for data-poor regions are often built with a priori intrinsic caveats not only due to reduced number of species records but also due to the use of environmental datasets with coarse resolutions. This is no exception for our study area, where limitations of data availability prevented the inclusion of important known environmental variables for the distribution of CWCs.

Substrate type is an important variable to consider for CWC distributions. However, full-coverage substrate type maps are rarely available for deep-sea regions at appropriate resolutions, which often contributes to low SDMs accuracy (Anderson, Guinotte, Rowden, Clark, et al., 2016; Anderson, Guinotte, Rowden, Tracey, et al., 2016; Bennecke & Metaxas, 2017; Burgos et al., 2020). In this study, substrate type was not included in the final SDMs because detailed substrate type maps or backscatter data were not available for the whole spatial extent of the seamounts. However, it is likely that substrate type has an important influence on the distribution of the modelled CWCs, as previously seen in other investigations (De Clippele et al., 2019; Orejas et al., 2009; Purser et al., 2013; Victorero et al., 2018). When data on substrate type are not available, slope can be used as a proxy since flat terrain indicates areas of higher sediment deposition (Lecours et al., 2016; Wilson et al., 2007) and, therefore, the presence of soft bottoms. This is supported by our results showing that flat areas on the seamounts represent areas with high probability of presence for *A. arbuscula*, and is corroborated by image observations where a higher abundance of this species was associated with muddy and sandy bottoms (Hansteen et al., 2014; Orejas et al., 2022), being consistent with observations of *A. arbuscula* in other CWC habitats in the N Atlantic (Baker et al., 2012; Buhl-Mortensen et al., 2015; Edinger et al., 2011; Lapointe et al., 2020; Morris et al., 2013; Orejas et al., 2017).

In recent years, the inclusion of high-resolution physical hydrodynamic models (with few hundreds of metres of spatial horizontal resolution) in CWC distribution models has improved model accuracy (Mohn et al., 2023; Rengstorf et al., 2014), given that local hydrodynamic processes, such as internal waves (Mosquera Giménez et al., 2019; Pearman et al., 2020; Rengstorf et al., 2013, 2014) and kinetic energy dissipation (Mohn et al., 2023), are important descriptors for the presence of CWCs. In addition, data on chemical water properties, such as oxygen concentration, POC flux, aragonite and calcite, are also important for the distribution of CWCs (Davies & Guinotte, 2011; Tittensor et al., 2009). In the light of climate change, incorporating biogeochemistry data in SDMs is crucial to project future changes in the distribution of CWCs (Morato et al., 2020) and to identify region-specific species tipping points (Puerta et al., 2020). Yet, assessing the influence of these variables through SDMs is often hindered by the low resolution of the available physical oceanographic data (Burgos et al., 2020; Yesson et al., 2012). This was the case for our study since full-coverage data of these parameters were not available at a high spatial resolution.

Despite the lack of substrate type data and a high-resolution hydrodynamic model, using a multi-scale approach to calculate terrain variables in our study was a good compromise to obtain first insights and interpretations on how terrain characteristics, as proxies for ocean circulation and terrain geomorphology, might influence the distribution of CWCs, at finer and broader scales. Therefore, for other studies where such detailed datasets are not available, we recommend considering terrain variables calculated with differing numbers of grid cells in the focal neighbourhood, to better capture the range of environmental conditions where species occur.

In our study, there were some observed disagreements between the two modelling methods (RF and GAM), both in terms of the relative contribution and response to environmental parameters, as well as in terms of the areas predicted by each model. However, despite challenges of modelling the distribution of species with few occurrences, the ensemble model consistently outperformed the other models, showing that this was the most suitable approach to increase model accuracy. Ensemble models often perform better than individual models (Liu et al., 2019), since, in the case of weighted average ensembles, the final predictions are created taking into account the highest-performing individual models. Nonetheless, it is noteworthy that ensemble performance is influenced by several factors, including the methods used to create the ensemble and to validate model predictions (Hao et al., 2019), and fine-tuned individual models can achieve better performances than ensembles (Hao et al., 2020). However, this was not the case in our study since the ensemble presented higher model performance metrics than the individual models considered.

In addition, using an ensemble model helped to identify areas of high/low uncertainty and model disagreement/agreement, which is essential when models are built on limited datasets. We observed that the level of uncertainty at each seamount was proportional to the amount of species data available, where the seamount with more species records (Cadamosto) presented higher confidence in predictions and, conversely, the seamount with fewer or no species records (Boavista) presented higher uncertainty. This result reflects the range of environmental conditions sampled, and not just the absolute number of observations. It is thus important to consider that the good model performance obtained in the ensemble does not necessarily indicate high model accuracy, since models fit with limited data might result in overfitted and inflated performance metrics (Borokini et al., 2023). Despite the high evaluation metrics obtained for all models, a large proportion of the modelled areas in our study are subject to high model uncertainty and large areas of extrapolation, especially in places where no survey data were available. This clearly highlights the need to conduct more surveys on the seamounts of Cabo Verde, following a balanced sampling strategy across a wide range of environmental conditions on all seamounts, in order to obtain a robust dataset to validate and improve model predictions. For example, future sampling efforts should take into account broader depth ranges and also cover flatter areas of the seamounts, since these were the areas shown to be the most contributing to areas of model extrapolation.

The ensemble models presented higher uncertainty in predicted areas of CWC absence (see Figures 8 and 9). Furthermore, the conversion of probability values into a binary classification of presence–absences depends on the threshold used. In our study, we chose a threshold based on maximizing the sum of sensitivity (proportion of presences predicted correctly) and specificity (proportion of absences predicted correctly), since maximizing the sum of these measures provides a balance between true presences and true absences. The choice of an appropriate threshold is important as it significantly affects model outputs and performance metrics (Lawson et al., 2014; Liu et al., 2005, 2011, 2019), being especially important for smaller datasets where the cost of misclassification is higher (Liu et al., 2011). For example, a higher threshold might yield higher-performance metrics, potentially with higher sensitivity due to more predicted areas considered suitable, but with lower specificity since more observed absence areas would be predicted as presences (Lawson et al., 2014; Liu et al., 2005). This means that the choice of the threshold should be made based on the specific objectives of each study. Correctly predicting absences at known areas of species absences (i.e. models with higher specificity) might be beneficial for marine spatial planning purposes, since it allows identification of areas where human activities could be permitted (Wilson et al., 2005). In the case of data-poor models, a model with a good capability of predicting absence areas is also useful to help identify areas for future sampling efforts. Conversely, a model with higher sensitivity is useful for survey planning since it identifies target areas where species are likely to occur (Meller et al., 2014). Additionally, if the precautionary principle is to be applied, predicted areas of species presence are also beneficial for management purposes (Armstrong et al., 2014), giving an indication of where human activities should be limited.

Our study provides a first stepping stone towards identifying suitable areas for VME indicator taxa and where future exploratory research should occur on the unexplored seamounts of Cabo Verde. Using an ensemble model proved to be a good initial approach to address data scarcity in our study. However, given the high uncertainty and large areas of extrapolation of model predictions, it is difficult to rightfully judge model performance and accuracy until additional data are collected. Nonetheless, model predictions showed that all seamounts present suitable areas where VMEs are likely to occur, supporting the expected ecological relevance of the seamounts in the region. Indeed, available image data showed pristine and diverse CWC gardens on Cadamosto Seamount (Orejas et al., 2022; Vinha et al., 2022) and exceptional sponge and crinoid fields and scleractinian coral reefs on Nola Seamount (Hansteen et al., 2014). In addition, three of the seamounts considered in our study – Boavista, Cabo Verde and Nola – are currently included in designated ecologically and biologically significant areas (EBSAs) (UNEP/CBD/COP/DEC/XII/22, 2014), where significant fishing activity occurs (Martins

et al., 2022; Roast et al., 2023). Priority should, therefore, be given to collecting new data in global under-sampled deep-sea regions, involving local researchers and communities.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ddi.13896.

DATA AVAILABILITY STATEMENT

Presence–absence data of CWCs on the seamounts of Cabo Verde are available in PANGAEA (Vinha et al., 2023; https://doi.org/10.1594/PANGAEA.963704) while bathymetry data and R code to calculate terrain parameters are available in Dryad (Vinha et al., 2024; https://doi.org/10.5061/dryad.0vt4b8h5g). Data from the VIKING20x model is available through GEOMAR in Schwarzkopf, 2024 (https://hdl.handle.net/20.500.12085/20248b0a-49fd-4868-90bf-581c61f4b396).

ORCID

Beatriz Vinha https://orcid.org/0000-0001-7193-8387
Ellen Kenchington https://orcid.org/0000-0003-3784-4533

REFERENCES

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Clark, M. R., Mormede, S., Davies, A. J., & Bowden, D. A. (2016). Field validation of habitat suitability models for vulnerable marine ecosystems in the South Pacific Ocean: Implications for the use of broad-scale models in fisheries management. *Ocean and Coastal Management*, 120, 110–126. https://doi.org/10.1016/j.ocecoaman.2015.11.025

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Tracey, D. M., Mackay, K. A., & Clark, M. R. (2016). 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. https://doi.org/10.1016/j.dsr.2016.07.006

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Ardron, J. A., Clark, M. R., Penney, A. J., Hourigan, T. F., Rowden, A. A., Dunstan, P. K., Watling, L., Shank, T. M., Tracey, D. M., Dunn, M. R., & Parker, S. J. (2014). A systematic approach towards the identification and protection of vulnerable marine ecosystems. *Marine Policy*, 49, 146–154. https://doi.org/10.1016/j.marpol.2013.11.017

Armstrong, C. W., Foley, N. S., Kahui, V., & Grehan, A. (2014). Cold water coral reef management from an ecosystem service perspective. *Marine Policy*, 50(Part A), 126–134. https://doi.org/10.1016/J.MARPOL.2014.05.016

Auscavitch, S. R., Lunden, J. J., Barkman, A., Quattrini, A. M., Demopoulos, A. W. J., & Cordes, E. E. (2020). Distribution of deep-water scleractinian and stylasterid corals across abiotic environmental gradients on three seamounts in the Anegada passage. *PeerJ*, 8, e9523. https://doi.org/10.7717/peerj.9523

Baco, A. R., Morgan, B., & Roark, E. B. (2020). Observations of vulnerable marine ecosystems and significant adverse impacts on high seas seamounts of the northwestern Hawaiian ridge and emperor seamount chain. *Marine Policy*, 115, 103834. https://doi.org/10.1016/j.marpol.2020.103834

Baker, K. D., Wareham, V. E., Snelgrove, P. V. R., Haedrich, R. L., Fifield, D. A., Edinger, E. N., & Gilkinson, K. D. (2012). Distributional patterns of deep-sea coral assemblages in three submarine canyons off Newfoundland, Canada. *Marine Ecology Progress Series*, 445, 235–249. https://doi.org/10.3354/meps09448

Bean, W. T., Stafford, R., & Brashares, J. S. (2012). The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography*, 35(3), 250–258. https://doi.org/10.1111/j.1600-0587.2011.06545.x

Beazley, L., Wang, Z., Kenchington, E., Yashayaev, I., Rapp, H. T., Xavier, J. R., Murillo, F. J., Fenton, D., & Fuller, S. (2018). Predicted distribution of the glass sponge Vazella pourtalesi on the Scotian shelf and its persistence in the face of climatic variability. *PLoS One*, 13(10), e0205505. https://doi.org/10.1371/JOURNAL.PONE.0205505

Bennecke, S., & Metaxas, A. (2017). Is substrate composition a suitable predictor for deep-water coral occurrence on fine scales? *Deep Sea Research Part I: Oceanographic Research Papers*, 124, 55–65. https://doi.org/10.1016/j.dsr.2017.04.011

Bergstad, O. A., Gil, M., Høines, Å. S., Sarralde, R., Maletzky, E., Mostarda, E., Singh, L., António, M. A., Ramil, F., Clerkin, P., & Campanis, G. (2019). Megabenthos and benthopelagic fishes on Southeast Atlantic seamounts. *African Journal of Marine Science*, 41, 29–50. https://doi.org/10.2989/1814232X.2019.1571439

Biastoch, A., Schwarzkopf, F. U., Getzlaff, K., Rühs, S., Martin, T., Scheinert, M., Schulzki, T., Handmann, P., Hummels, R., & Böning, C. W. (2021). Regional imprints of changes in the Atlantic meridional overturning circulation in the eddy-rich ocean model VIKING20X. *Ocean Science*, 17(5), 1177–1211. https://doi.org/10.5194/os-17-1177-2021

Borokini, I. T., Nussear, K., Petitpierre, B., Dilts, T. E., & Weisberg, P. J. (2023). Iterative species distribution modeling results in the discovery of novel populations of a rare cold desert perennial. *Endangered Species Research*, 50, 47–62. https://doi.org/10.3354/esr01218

Bouchet, P. J., Miller, D. L., Roberts, J. J., Mannocci, L., Harris, C. M., & Thomas, L. (2020). Dsmextra: Extrapolation assessment tools for

density surface models. *Methods in Ecology and Evolution*, *11*(11), 1464–1469. https://doi.org/10.1111/2041-210X.13469

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, *6*(10), 1210–1218. https://doi.org/10.1111/2041-210X.12403

Bridges, A. E. H., Barnes, D. K. A., Bell, J. B., Ross, R. E., Voges, L., & Howell, K. L. (2023). Filling the data gaps: Transferring models from data-rich to data-poor deep-sea areas to support spatial management. *Journal of Environmental Management*, *345*, 118325. https://doi.org/10.1016/j.jenvman.2023.118325

Bridges, A. E. H., Howell, K. L., Amaro, T., Atkinson, L., Barnes, D. K., Bax, N., Bell, J. B., Bernardino, A. F., Beuck, L., Braga-Henriques, A., & Brandt, A. (2023). Review of the central and South Atlantic shelf and Deep-Sea benthos: Science, policy and management. *Oceanography and Marine Biology: An Annual Review*, *61*, 127–218. https://doi.org/10.1201/9781032426969-4

Buhl-Mortensen, L., Olafsdottir, S. H., Buhl-Mortensen, P., Burgos, J. M., & Ragnarsson, S. A. (2015). Distribution of nine cold-water coral species (Scleractinia and Gorgonacea) in the cold temperate North Atlantic: Effects of bathymetry and hydrography. *Hydrobiologia*, *759*(1), 39–61. https://doi.org/10.1007/s10750-014-2116-x

Buisson, L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, *16*(4), 1145–1157. https://doi.org/10.1111/j.1365-2486.2009.02000.x

Burgos, J. M., Buhl-Mortensen, L., Buhl-Mortensen, P., Ólafsdóttir, S. H., Steingrund, P., Ragnarsson, S. Á., & Skagseth, Ø. (2020). Predicting the distribution of indicator taxa of vulnerable marine ecosystems in the Arctic and sub-arctic waters of the Nordic seas. *Frontiers in Marine Science*, *7*, 131. https://doi.org/10.3389/fmars.2020.00131

Cardoso, C., Caldeira, R. M. A., Relvas, P., & Stegner, A. (2020). Islands as eddy transformation and generation hotspots: Cabo Verde case study. *Progress in Oceanography*, *184*, 102271. https://doi.org/10.1016/J.POCEAN.2020.102271

Chi, X., Dierking, J., Hoving, H. J., Lüskow, F., Denda, A., Christiansen, B., Sommer, U., Hansen, T., & Javidpour, J. (2020). Tackling the jelly web: Trophic ecology of gelatinous zooplankton in oceanic food webs of the eastern tropical Atlantic assessed by stable isotope analysis. *Limnology and Oceanography*, *66*, 289–305. https://doi.org/10.1002/lno.11605

Clark, M. R., Schlacher, T. A., Rowden, A. A., Stocks, K. I., & Consalvey, M. (2012). Science priorities for seamounts: Research links to conservation and management. *PLoS One*, *7*(1), e29232. https://doi.org/10.1371/journal.pone.0029232

Crimmins, S. M., Dobrowski, S. Z., & Mynsberge, A. R. (2013). Evaluating ensemble forecasts of plant species distributions under climate change. *Ecological Modelling*, *266*, 126–130. https://doi.org/10.1016/j.ecolmodel.2013.07.006

Davies, A. J., Duineveld, G. C. A., Lavaleye, M. S. S., Bergman, M. J. N., van Haren, H., & Roberts, J. M. (2009). Downwelling and deep-water bottom currents as food supply mechanisms to the cold-water coral Lophelia pertusa (Scleractinia) at the Mingulay reef complex. *Limnology and Oceanography*, *54*(2), 620–629. https://doi.org/10.4319/lo.2009.54.2.0620

Davies, A. J., & Guinotte, J. M. (2011). Global habitat suitability for framework-forming cold-water corals. *PLoS One*, *6*, e18483.

Davies, J. S., Stewart, H. A., Narayanaswamy, B. E., Jacobs, C., Spicer, J., Golding, N., & Howell, K. L. (2015). Benthic assemblages of the Anton Dohrn seamount (NE Atlantic): Defining Deep-Sea biotopes to support habitat mapping and management efforts with a focus on vulnerable marine ecosystems. *PLoS One*, *10*, e0124815. https://doi.org/10.1371/journal.pone.0124815

De Clippele, L. H., Huvenne, V. A. I., Molodtsova, T. N., & Roberts, J. M. (2019). The diversity and ecological role of non-scleractinian corals (Antipatharia and Alcyonacea) on Scleractinian cold-water coral mounds. *Frontiers in Marine Science*, *6*, 184. https://doi.org/10.3389/fmars.2019.00184

Dolan, M. F. J., Grehan, A. J., Guinan, J. C., & Brown, C. (2008). 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. https://doi.org/10.1016/j.dsr.2008.06.010

Doughty, C. L., Quattrini, A. M., & Cordes, E. E. (2014). Insights into the population dynamics of the deep-sea coral genus Paramuricea in the Gulf of Mexico. *Deep Sea Research Part II: Topical Studies in Oceanography*, *99*, 71–82. https://doi.org/10.1016/j.dsr2.2013.05.023

Downie, A.-L., Piechaud, N., Howell, K., Barrio Froján, C., Sacau, M., & Kenny, A. (2021). Reconstructing baselines: Use of habitat suitability modelling to predict pre-fishing condition of a vulnerable marine ecosystem. *ICES Journal of Marine Science*, *78*(8), 2784–2796. https://doi.org/10.1093/icesjms/fsab154

Dullo, W.-C., Flögel, S., & Rüggeberg, A. (2008). Cold-water coral growth in relation to the hydrography of the Celtic and Nordic European continental margin. *Marine Ecology Progress Series*, *371*, 165–176. https://doi.org/10.3354/meps07623

Edinger, E. N., Sherwood, O. A., Piper, D. J. W., Wareham, V. E., Baker, K. D., Gilkinson, K. D., & Scott, D. B. (2011). Geological features supporting deep-sea coral habitat in Atlantic Canada. *Continental Shelf Research*, *31*(2), S69–S84. https://doi.org/10.1016/J.CSR.2010.07.004

Evans, J. S., & Ram, K. (2021). *Package "spatialEco"*, R CRAN Project [Preprint].

FAO. (2009). *International guidelines for the management of deep-sea fisheries in the high seas*. Food and Agriculture Organization. https://www.fao.org/4/i0816t/i0816t00.htm

Franklin, J. (2010). *Mapping species distributions: Spatial inference and prediction*. Cambridge University Press.

Freeman, E. A., & Moisen, G. (2008). PresenceAbsence: An R package for presence absence analysis. *Journal of Statistical Software*, *23*, 1–31. https://doi.org/10.18637/jss.v023.i11

Georgian, S. E., Anderson, O. F., & Rowden, A. A. (2019). Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. *Fisheries Research*, *211*, 256–274. https://doi.org/10.1016/j.fishres.2018.11.020

Georgian, S. E., Kramer, K., Saunders, M., Shedd, W., Roberts, H., Lewis, C., Fisher, C., & Cordes, E. (2020). Habitat suitability modelling to predict the spatial distribution of cold-water coral communities affected by the Deepwater horizon oil spill. *Journal of Biogeography*, *47*(7), 1455–1466. https://doi.org/10.1111/jbi.13844

González-Irusta, J. M., Cartes, J. E., Punzón, A., Díaz, D., de Sola, L. G., & Serrano, A. (2022). Mapping habitat loss in the deep-sea using current and past presences of *Isidella elongata* (Cnidaria: Alcyonacea). *ICES Journal of Marine Science*, *79*, 1888–1901. https://doi.org/10.1093/ICESJMS/FSAC123

Goode, S. L., Rowden, A. A., Bowden, D. A., & Clark, M. R. (2020). Resilience of seamount benthic communities to trawling disturbance. *Marine Environmental Research*, *161*, 105086. https://doi.org/10.1016/j.marenvres.2020.105086

Grenouillet, G., Buisson, L., Casajus, N., & Lek, S. (2011). Ensemble modelling of species distribution: The effects of geographical and environmental ranges. *Ecography*, *34*(1), 9–17. https://doi.org/10.1111/j.1600-0587.2010.06152.x

Guinan, J., Brown, C., Dolan, M. F. J., & Grehan, A. J. (2009). Ecological niche modelling of the distribution of cold-water coral habitat using underwater remote sensing data. *Ecological Informatics*, *4*, 83–92. https://doi.org/10.1016/j.ecoinf.2009.01.004

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in R*. Cambridge University Press.

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

Hansteen, T. H., Klügel, A., & Kwasnitschka, T. (2014). *Cape Verde Seamounts – Cruise No. M80/3 – December 29, 2009 – February 1, 2010 - Dakar (Senegal) - Las Palmas de Gran Canaria (Spain)*. METEOR-Berichte, M80/3, 42 pp., DFG-Senatskommission für Ozeanographie'. https://doi.org/10.2312/cr_m80_3

Hao, T., Elith, J., Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, 25(5), 839–852. https://doi.org/10.1111/ddi.12892

Hao, T., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography*, 43(4), 549–558. https://doi.org/10.1111/ecog.04890

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29(5), 773–785. https://doi.org/10.1111/j.0906-7590.2006.04700.x

Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, P., Bevan, A., Racine, E. B., Shortridge, A., & Hijmans, M. R. J. (2015). *Package "raster"*, R package, 734, p. 473.

Hogg, O. T., Huvenne, V. A. I., Griffiths, H. J., Dorschel, B., & Linse, K. (2016). Landscape mapping at sub-Antarctic South Georgia provides a protocol for underpinning large-scale marine protected areas. *Scientific Reports*, 6, 1–15. https://doi.org/10.1038/srep33163

Hoving, H.-J. T., Neitzel, P., Hauss, H., Christiansen, S., Kiko, R., Robison, B. H., Silva, P., & Körtzinger, A. (2020). In situ observations show vertical community structure of pelagic fauna in the eastern tropical North Atlantic off Cape Verde. *Scientific Reports*, 10, 1–14. https://doi.org/10.1038/s41598-020-78255-9

Howell, K. L., Bridges, A. E., Graves, K. P., Allcock, L., la Bianca, G., Ventura-Costa, C., Donaldson, S., Downie, A. L., Furey, T., McGrath, F., & Ross, R. (2022). Performance of deep-sea habitat suitability models assessed using independent data, and implications for use in area-based management. *Marine Ecology Progress Series*, 695, 33–51. https://doi.org/10.3354/MEPS14098

Ilich, A. R., Misiuk, B., Lecours, V., & Murawski, S. A. (2023). MultiscaleDTM: An open-source R package for multiscale geomorphometric analysis. *Transactions in GIS*, 27(4), 1164–1204. https://doi.org/10.1111/tgis.13067

Jeliazkov, A., Gavish, Y., Marsh, C. J., Geschke, J., Brummitt, N., Rocchini, D., Haase, P., Kunin, W. E., & Henle, K. (2022). Sampling and modelling rare species: Conceptual guidelines for the neglected majority. *Global Change Biology*, 28(12), 3754–3777. https://doi.org/10.1111/gcb.16114

Jones, M. C., & Cheung, W. W. L. (2015). Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES Journal of Marine Science*, 72(3), 741–752. https://doi.org/10.1093/icesjms/fsu172

Kenchington, E., Beazley, L., Lirette, C., Murillo, F., Guijarro, J., Wareham, V., Gilkinson, K., Koen-Alonso, M., Benoît, H., Bourdages, H., Sainte-Marie, B., Treble, M., & Siferd, T. (2016). *Delineation of coral and sponge significant benthic areas in eastern Canada using kernel density analyses and species distribution models*. DFO Canadian Science Advisory Secretariat. Research document. 2016/093. vi + 178 p.

Kerry, C. R., Exeter, O. M., & Witt, M. J. (2022). Monitoring global fishing activity in proximity to seamounts using automatic identification systems. *Fish and Fisheries*, 23(3), 733–749. https://doi.org/10.1111/faf.12647

Knudby, A., Kenchington, E., & Murillo, F. J. (2013). Modeling the distribution of Geodia sponges and sponge grounds in the Northwest Atlantic. *PLoS One*, 8(12), e82306. https://doi.org/10.1371/journal.pone.0082306

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, 36, 1–13. https://doi.org/10.18637/jss.v036.i11

Kursa, M. B., & Rudnicki, W. R. (2015). *Package "Boruta", Search in* [Preprint].

Kwasnitschka, T., Hansteen, T. H., Ramalho, R. S., Devey, C. W., Klügel, A., Samrock, L. K., & Wartho, J.-A. (2024). Geomorphology and age constraints of seamounts in the Cabo Verde Archipelago, and their relationship to island ages and geodynamic evolution. *Geochemistry, Geophysics, Geosystems*, 25, e2023GC011071. https://doi.org/10.1029/2023GC011071

Langenkämper, D., Zurowietz, M., Schoening, T., & Nattkemper, T. W. (2017). Biigle 2.0-browsing and annotating large marine image collections. *Frontiers in Marine Science*, 4, 83.

Lapointe, A. E., Watling, L., France, S. C., & Auster, P. J. (2020). Megabenthic assemblages in the lower bathyal (700–3000 m) on the New England and corner rise seamounts, Northwest Atlantic. *Deep Sea Research Part I: Oceanographic Research Papers*, 165, 103366. https://doi.org/10.1016/j.dsr.2020.103366

Lawson, C. R., Hodgson, J. A., Wilson, R. J., & Richards, S. A. (2014). Prevalence, thresholds and the performance of presence–absence models. *Methods in Ecology and Evolution*, 5(1), 54–64. https://doi.org/10.1111/2041-210X.12123

Lecours, V., Devillers, R., Schneider, D. C., Lucieer, V. L., Brown, C. J., & Edinger, E. N. (2015). Spatial scale and geographic context in benthic habitat mapping: Review and future directions. *Marine Ecology Progress Series*, 535, 259–284. https://doi.org/10.3354/MEPS11378

Lecours, V., Dolan, M. F. J., Micallef, A., & Lucieer, V. L. (2016). A review of marine geomorphometry, the quantitative study of the seafloor. *Hydrology and Earth System Sciences*, 20(8), 3207–3244. https://doi.org/10.5194/hess-20-3207-2016

Leitner, A. B., Drazen, J. C., & Smith, C. R. (2021). Testing the seamount refuge hypothesis for predators and scavengers in the Western Clarion-Clipperton Zone. *Frontiers in Marine Science*, 8, 705237. https://doi.org/10.3389/fmars.2021.636305

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.

Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28(3), 385–393. https://doi.org/10.1111/j.0906-7590.2005.03957.x

Liu, C., Newell, G., & White, M. (2019). The effect of sample size on the accuracy of species distribution models: Considering both presences and pseudo-absences or background sites. *Ecography*, 42(3), 535–548. https://doi.org/10.1111/ecog.03188

Liu, C., White, M., & Newell, G. (2011). Measuring and comparing the accuracy of species distribution models with presence–absence data. *Ecography*, 34(2), 232–243. https://doi.org/10.1111/j.1600-0587.2010.06354.x

Liu, F., Daewel, U., Samuelsen, A., Brune, S., Hanz, U., Pohlmann, H., Baehr, J., & Schrum, C. (2021). Can environmental conditions at North Atlantic Deep-Sea habitats Be predicted several years ahead? —Taking sponge habitats as an example. *Frontiers in Marine Science*, 0, 1546. https://doi.org/10.3389/FMARS.2021.703297

Lobo, J. M., & Tognelli, M. F. (2011). Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation*, 19(1), 1–7. https://doi.org/10.1016/J.JNC.2010.03.002

Lopes, P. F. M., Verba, J. T., Begossi, A., & Pennino, M. G. (2019). Predicting species distribution from fishers' local ecological knowledge: A new alternative for data-poor management. *Canadian Journal of Fisheries and Aquatic Sciences, 76*(8), 1423–1431. https://doi.org/10.1139/cjfas-2018-0148

Martins, S., Tiwari, M., Rocha, F., Rodrigues, E., Monteiro, R., Araújo, S., Abella, E., de Santos Loureiro, N., Clarke, L. J., & Marco, A. (2022). Evaluating loggerhead sea turtle (*Caretta caretta*) bycatch in the small-scale fisheries of Cabo Verde. *Reviews in Fish Biology and Fisheries, 32*(3), 1001–1015. https://doi.org/10.1007/s11160-022-09718-7

McQuaid, K. A., Attrill, M. J., Clark, M. R., Cobley, A., Glover, A. G., Smith, C. R., & Howell, K. L. (2020). Using habitat classification to assess representativity of a protected area network in a large, data-poor area targeted for deep-sea mining. *Frontiers in Marine Science, 7*, 558860. https://doi.org/10.3389/fmars.2020.558860

McQuaid, K. A., Bridges, A. E. H., Howell, K. L., Gandra, T. B. R., de Souza, V., Currie, J. C., Hogg, O. T., Pearman, T. R. R., Bell, J. B., Atkinson, L. J., Baum, D., Bonetti, J., Carranza, A., Defeo, O., Furey, T., Gasalla, M. A., Golding, N., Hampton, S. L., Horta, S., … Yool, A. (2023). Broad-scale benthic habitat classification of the South Atlantic. *Progress in Oceanography, 214*, 103016. https://doi.org/10.1016/j.pocean.2023.103016

Meller, L., Cabeza, M., Pironon, S., Barbet-Massin, M., Maiorano, L., Georges, D., & Thuiller, W. (2014). Ensemble distribution models in conservation prioritization: From consensus predictions to consensus reserve networks. *Diversity and Distributions, 20*(3), 309–321. https://doi.org/10.1111/ddi.12162

Menezes, G. M., Tariche, O., Pinho, M. R., Sigler, M. F., & Silva, H. M. (2015). Structure and zonation of demersal and deep-water fish assemblages off the Cabo Verde archipelago (northeast-Atlantic) as sampled by baited longlines. *Deep Sea Research Part I: Oceanographic Research Papers, 102*, 118–134.

Mesgaran, M. B., Cousens, R. D., & Webber, B. L. (2014). Here be dragons: A tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity and Distributions, 20*(10), 1147–1159. https://doi.org/10.1111/ddi.12209

Mohn, C., Hansen, J. L. S., Carreiro-Silva, M., Cunningham, S. A., de Froe, E., Dominguez-Carrió, C., Gary, S., Glud, R. N., Göke, C., Johnson, C., Morato, T., Friis Møller, E., Rovelli, L., Schulz, K., Soetaert, K., van der Kaaden, A., & van Oevelen, D. (2023). Tidal to decadal scale hydrodynamics at two contrasting cold-water coral sites in the Northeast Atlantic. *Progress in Oceanography, 214*, 103031. https://doi.org/10.1016/j.pocean.2023.103031

Mohn, C., Rengstorf, A., White, M., Duineveld, G., Mienis, F., Soetaert, K., & Grehan, A. (2014). Linking benthic hydrodynamics and cold-water coral occurrences: A high-resolution model study at three cold-water coral provinces in the NE Atlantic. *Progress in Oceanography, 122*, 92–104. https://doi.org/10.1016/J.POCEAN.2013.12.003

Mohn, C., White, M., Denda, A., Erofeeva, S., Springer, B., Turnewitsch, R., & Christiansen, B. (2021). Dynamics of currents and biological scattering layers around Senghor Seamount, a shallow seamount inside a tropical Northeast Atlantic eddy corridor. *Deep Sea Research Part I: Oceanographic Research Papers, 171*, 103497. https://doi.org/10.1016/j.dsr.2021.103497

Morato, T., Cheung, W. W. L., & Pitcher, T. J. (2006). Vulnerability of seamount fish to fishing: Fuzzy analysis of life-history attributes. *Journal of Fish Biology, 68*(1), 209–221. https://doi.org/10.1111/j.0022-1112.2006.00894.x

Morato, T., González-Irusta, J. M., Dominguez-Carrió, C., Wei, C. L., Davies, A., Sweetman, A. K., Taranto, G. H., Beazley, L., García-Alegre, A., Grehan, A., Laffargue, P., Murillo, F. J., Sacau, M., Vaz, S., Kenchington, E., Arnaud-Haond, S., Callery, O., Chimienti, G., Cordes, E., … Carreiro-Silva, M. (2020). Climate-induced changes in the suitable habitat of cold-water corals and commercially important deep-sea fishes in the North Atlantic. *Global Change Biology, 26*, 2181–2202. https://doi.org/10.1111/gcb.14996

Morris, K. J., Tyler, P. A., Masson, D. G., Huvenne, V. I. A., & Rogers, A. D. (2013). Distribution of cold-water corals in the Whittard canyon, NE Atlantic Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography, 92*, 136–144. https://doi.org/10.1016/J.DSR2.2013.03.036

Mosquera Giménez, Á., Vélez-Belchí, P., Rivera, J., Piñeiro, S., Fajar, N., Caínzos, V., Balbín, R., Jiménez Aparicio, J. A., Dominguez-Carrió, C., Blasco-Ferre, J., Carreiro-Silva, M., Morato, T., Puerta, P., & Orejas, C. (2019). Ocean circulation over North Atlantic underwater features in the path of the Mediterranean outflow water: The Ormonde and Formigas seamounts, and the Gazul mud volcano. *Frontiers in Marine Science, 6*, 702. https://doi.org/10.3389/FMARS.2019.00702/BIBTEX

Mosquera-Giménez, A., Vélez-Belchí, P., Rodríguez, P., & Mouzo, I. (2022). *Physical oceanography off Cabo Verde*. in: Expedition report iMirabilis2 survey https://doi.org/10.5281/zenodo.6352141

Nephin, J., Gregr, E. J., St. Germain, C., Fields, C., & Finney, J. L. (2020). *Development of a species distribution modelling framework and its application to twelve species on Canada's Pacific Coast*, DFO Can. Sci. Advis. Sec. Res. Doc. 2020/004. xii + 107 p. https://www.dfo-mpo.gc.ca/csas-sccs/Publications/ResDocs-DocRech/2020/2020_004-eng.html

Orejas, C., Addamo, A., Alvarez, M., Aparicio, A., Alcoverro, D., Arnaud-Haond, S., Bilan, M., Boavida, J., Cainzos, V., Calderon, R., Cambeiro, P., Castano, M., Fox, A., Gallardo, M., Gori, A., Guitierrez, C., Henry, L. A., Hermida, M., Jimenez, J. A., … Zein, M. (2017). *Cruise summary report - MEDWAVES survey (MEDiterranean out flow WAter and vulnerable EcosystemS)*. Zenodo. https://doi.org/10.5281/zenodo.556516

Orejas, C., Gori, A., Lo Iacono, C., Puig, P., Gili, J. M., & Dale, M. R. T. (2009). Cold-water corals in the cap de Creus canyon, northwestern Mediterranean: Spatial distribution, density and anthropogenic impact. *Marine Ecology Progress Series, 397*, 37–51. https://doi.org/10.3354/meps08314

Orejas, C., Huvenne, V., Sweetman, A. K., Vinha, B., Abella, J. C., Andrade, P., Afonso, A., Antelo, J., Austin-Berry, R., Baltasar, L., Barbosa, N., Barnhill, K. A., Barreiro, A., Bettencourt, R., Blanco, S., Buigues, A., Calado, A., Casal, I., de la Torre, J., … Vélez-Belchí, P. (2022). *Expedition report iMirabilis2 survey*. https://doi.org/10.5281/ZENODO.6352141

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

Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P., & Lees, D. C. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography, 33*(10), 1704–1711. https://doi.org/10.1111/j.1365-2699.2006.01460.x

Puerta, P., Johnson, C., Carreiro-Silva, M., Henry, L.-A., Kenchington, E., Morato, T., Kazanidis, G., Rueda, J. L., Urra, J., Ross, S., Wei, C.-L., González-Irusta, J. M., Arnaud-Haond, S., & Orejas, C. (2020). Influence of water masses on the biodiversity and biogeography of deep-sea benthic ecosystems in the North Atlantic. *Frontiers in Marine Science, 7*, 239. https://doi.org/10.3389/fmars.2020.00239

Puerta, P., Mosquera-Giménez, Á., Reñones, O., Domínguez-Carrió, C., Rueda, J. L., Urra, J., Carreiro-Silva, M., Blasco-Ferre, J., Santana, Y., Gutiérrez-Zárate, C., Vélez-Belchí, P., Rivera, J., Morato, T., & Orejas, C. (2022). Variability of deep-sea megabenthic assemblages along the western pathway of the Mediterranean outflow water. *Deep Sea Research Part I: Oceanographic Research Papers, 185*, 103791. https://doi.org/10.1016/J.DSR.2022.103791

Purser, A., Orejas, C., Gori, A., Tong, R., Unnithan, V., & Thomsen, L. (2013). Local variation in the distribution of benthic megafauna species associated with cold-water coral reefs on the Norwegian margin. *Continental Shelf Research*, 54, 37–51. https://doi.org/10.1016/J.CSR.2012.12.013

Qazi, A. W., Saqib, Z., & Zaman-ul-Haq, M. (2022). Trends in species distribution modelling in context of rare and endemic plants: A systematic review. *Ecological Processes*, 11(1), 40. https://doi.org/10.1186/s13717-022-00384-y

Quattrini, A. M., Gómez, C. E., & Cordes, E. E. (2017). Environmental filtering and neutral processes shape octocoral community assembly in the deep sea. *Oecologia*, 183(1), 221–236. https://doi.org/10.1007/s00442-016-3765-4

Ramirez-Reyes, C., Nazeri, M., Street, G., Jones-Farrand, D. T., Vilella, F. J., & Evans, K. O. (2021). Embracing ensemble species distribution models to inform At-risk species status assessments. *Journal of Fish and Wildlife Management*, 12(1), 98–111. https://doi.org/10.3996/JFWM-20-072

Ramiro-Sánchez, B., González-Irusta, J. M., Henry, L. A., Cleland, J., Yeo, I., Xavier, J. R., Carreiro-Silva, M., Sampaio, Í., Spearman, J., Victorero, L., Messing, C. G., Kazanidis, G., Roberts, J. M., & Murton, B. (2019). 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, 278. https://doi.org/10.3389/fmars.2019.00278

Rengstorf, A. M., Mohn, C., Brown, C., Wisz, M. S., & Grehan, A. J. (2014). 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. https://doi.org/10.1016/J.DSR.2014.07.007

Rengstorf, A. M., Yesson, C., Brown, C., & Grehan, A. J. (2013). High-resolution habitat suitability modelling can improve conservation of vulnerable marine ecosystems in the deep sea. *Journal of Biogeography*, 40(9), 1702–1714. https://doi.org/10.1111/JBI.12123

Roast, M. J., Martins, S., Fernández-Peralta, L., Báez, J. C., Diame, A., March, D., Ouled-Cheikh, J., Marco, A., González-Solís, J., & Cardona, L. (2023). Hidden demographic impacts of fishing and environmental drivers of fecundity in a sea turtle population. *Conservation Biology*, 37(5), e14110. https://doi.org/10.1111/cobi.14110

Robert, K., Jones, D. O. B., Roberts, J. M., & Huvenne, V. A. I. (2016). 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. https://doi.org/10.1016/j.dsr.2016.04.008

Rooper, C. N., Sigler, M. F., Goddard, P., Malecha, P., Towler, R., Williams, K., Wilborn, R., & Zimmermann, M. (2016). Validation and improvement of species distribution models for structure-forming invertebrates in the eastern Bering Sea with an independent survey. *Marine Ecology Progress Series*, 551, 117–130. https://doi.org/10.3354/MEPS11703

Rowden, A. A., Anderson, O. F., Georgian, S. E., Bowden, D. A., Clark, M. R., Pallentin, A., & Miller, A. (2017). High-resolution habitat suitability models for the conservation and management of vulnerable marine ecosystems on the Louisville seamount chain, South Pacific Ocean. *Frontiers in Marine Science*, 4, 335. https://doi.org/10.3389/fmars.2017.00335

Rowden, A. A., Schlacher, T. A., Williams, A., Clark, M. R., Stewart, R., Althaus, F., Bowden, D. A., Consalvey, M., Robinson, W., & Dowdney, J. (2010). A test of the seamount oasis hypothesis: Seamounts support higher epibenthic megafaunal biomass than adjacent slopes. *Marine Ecology*, 31(s1), 95–106. https://doi.org/10.1111/j.1439-0485.2010.00369.x

Santos, R., Medeiros-Leal, W., Novoa-Pabon, A., Silva, H., & Pinho, M. (2021). Demersal fish assemblages on seamounts exploited by fishing in the Azores (NE Atlantic). *Journal of Applied Ichthyology*, 37(2), 198–215. https://doi.org/10.1111/jai.14165

Schumacher, M., Huvenne, V. A. I., Devey, C. W., Arbizu, P. M., Biastoch, A., & Meinecke, S. (2022). The Atlantic Ocean landscape: A basin-wide cluster analysis of the Atlantic near seafloor environment. *Frontiers in Marine Science*, 0, 1411. https://doi.org/10.3389/FMARS.2022.936095

Schwarzkopf, F. (2024). *Supplementary data to Vinha et al. (2024): Ensemble modelling to predict the distribution of Vulnerable Marine Ecosystems indicator taxa on data-limited seamounts of Cabo Verde (NW Africa) [dataset]*. GEOMAR Helmholtz Centre for Ocean Research Kiel [distributor]. https://hdl.handle.net/20.500.12085/20248b0a-49fd-4868-90bf-581c61f4b396

Standaert, W., Puerta, P., Mastrototaro, F., Palomino, D., Aguilar, R., Ramiro-Sánchez, B., Vázquez, J. T., Guillamón, O. S., Marin, P., Blanco, J., & Orejas, C. (2023). Habitat suitability models of a critically endangered cold-water coral, *Isidella elongata*, in the Mallorca Channel. *Thalassas*, 39, 587–600. https://doi.org/10.1007/s41208-023-00531-y

Stenvers, V. I., Hauss, H., Osborn, K. J., Neitzel, P., Merten, V., Scheer, S., Robison, B. H., Freitas, R., & Hoving, H. J. T. (2021). Distribution, associations and role in the biological carbon pump of *Pyrosoma atlanticum* (Tunicata, Thaliacea) off Cabo Verde, NE Atlantic. *Scientific Reports*, 11, 9231. https://doi.org/10.1038/s41598-021-88208-5

Swanborn, D. J. B., Huvenne, V. A. I., Pittman, S. J., Rogers, A. D., Taylor, M. L., & Woodall, L. C. (2023). Mapping, quantifying and comparing seascape heterogeneity of southwest Indian ridge seamounts. *Landscape Ecology*, 38(1), 185–203. https://doi.org/10.1007/s10980-022-01541-6

Taranto, G. H., González-Irusta, J. M., Dominguez-Carrió, C., Pham, C. K., Tempera, F., Ramos, M., Gonçalves, G., Carreiro-Silva, M., & Morato, T. (2023). Spatial distributions, environmental drivers and co-existence patterns of key cold-water corals in the deep sea of the Azores (NE Atlantic). *Deep Sea Research Part I: Oceanographic Research Papers*, 197, 104028. https://doi.org/10.1016/j.dsr.2023.104028

Team, Rs. (2022). *RStudio: Integrated development for R. RStudio, Inc.* http://www.rstudio.com

Thuiller, W., Georges, D., Engler, R., & Breiner, F. (2016). *biomod2: Ensemble platform for species distribution modeling*, R Package Version, 3(3), p. r539.

Tittensor, D. P., Baco, A. R., Brewin, P. E., Clark, M. R., Consalvey, M., Hall-Spencer, J., Rowden, A. A., Schlacher, T., Stocks, K. I., & Rogers, A. D. (2009). Predicting global habitat suitability for stony corals on seamounts. *Journal of Biogeography*, 36(6), 1111–1128. https://doi.org/10.1111/J.1365-2699.2008.02062.X

Tracey, D. M., Rowden, A. A., Mackay, K. A., & Compton, T. (2011). Habitat-forming cold-water corals show affinity for seamounts in the New Zealand region. *Marine Ecology Progress Series*, 430, 1–22. https://doi.org/10.3354/MEPS09164

UNEP/CBD/COP/DEC/XII/22. (2014). *Decision adopted by the conference of the parties to the convention on biological diversity.* https://www.cbd.int/decision/cop/?id=13385

Van Audenhaege, L., Broad, E., Hendry, K. R., & Huvenne, V. A. I. (2021). High-resolution vertical habitat mapping of a Deep-Sea Cliff Offshore Greenland. *Frontiers in Marine Science*, 8, 669372. https://doi.org/10.3389/fmars.2021.669372

van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, 39(6), 542–552. https://doi.org/10.1111/ecog.01509

Victorero, L., Robert, K., Robinson, L. F., Taylor, M. L., & Huvenne, V. A. I. (2018). Species replacement dominates megabenthos beta diversity in a remote seamount setting. *Scientific Reports*, 8, 1–11.

Vinha, B., Hansteen, T. H., Huvenne, V. A. I., & Orejas, C. (2023). *Presence-absence records for four cold-water coral taxa on the seamounts of*

Cabo Verde (NW Africa) [dataset]. PANGAEA, https://doi.org/10.1594/PANGAEA.963704

Vinha, B., Murillo, F. J., Schumacher, M., Hansteen, T., Schwarzkopf, F., Biastoch, A., Kenchington, E., Piraino, S., Orejas, C., & Huvenne, V. (2024). Terrain variables used for ensemble distribution modelling of vulnerable marine ecosystems indicator taxa on data-limited seamounts of Cabo Verde (NW Africa) [Dataset]. Dryad. https://doi.org/10.5061/dryad.0vt4b8h5g

Vinha, B., Simon-Lledó, E., Arantes, R., Aguilar, R., Carreiro-Silva, M., Colaço, A., Piraino, S., Gori, A., Huvenne, V. A. I., & Orejas, C. (2022). *Deep-sea benthic megafauna of Cabo Verde (Eastern Equatorial Atlantic Ocean).* https://zenodo.org/record/6560869

Washburn, T. W., Simon-Lledó, E., Soong, G. Y., & Suzuki, A. (2023). Seamount mining test provides evidence of ecological impacts beyond deposition. *Current Biology, 33*(14), 3065–3071. https://doi.org/10.1016/j.cub.2023.06.032

Williams, A., Althaus, F., Green, M., Maguire, K., Untiedt, C., Mortimer, N., Jackett, C. J., Clark, M. R., Bax, N. J., Pitcher, R., & Schlacher, T. A. (2020). True size matters for conservation: A robust method to determine the size of Deep-Sea coral reefs shows they are typically small on seamounts in the Southwest Pacific Ocean. *Frontiers in Marine Science, 7*, 505850. https://doi.org/10.3389/FMARS.2020.00187/BIBTEX

Wilson, K. A., Westphal, M. I., Possingham, H. P., & Elith, J. (2005). Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation, 122*(1), 99–112. https://doi.org/10.1016/j.biocon.2004.07.004

Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C., & Grehan, A. J. (2007). Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy, 30*(1–2), 3–35. https://doi.org/10.1080/01490410701295962

Wisz, M. S., & Guisan, A. (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecology, 9*(1), 8. https://doi.org/10.1186/1472-6785-9-8

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & NCEAS Predicting Species Distributions Working Group. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions, 14*(5), 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

Wood, S. (2015). *Package "mgcv"*, R package version, 1(29), p. 729.

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

Yesson, C., Taylor, M. L., Tittensor, D. P., Davies, A. J., Guinotte, J., Baco, A., Black, J., Hall-Spencer, J. M., & Rogers, A. D. (2012). Global habitat suitability of cold-water octocorals. *Journal of Biogeography, 39*, 1278–1292. https://doi.org/10.1111/j.1365-2699.2011.02681.x

Zuur, A. F. (2012). *A beginner's guide to generalized additive models with R.* Highland Statistics Limited.

BIOSKETCHES

This manuscript is authored by a multidisciplinary research team working on deep-sea research. The research focus of the authors includes marine ecology, oceanography and geology to investigate how environmental variables affect the distribution and ecosystem functioning of deep-sea benthic habitats, with a particular focus on communities dominated by cold-water corals and deep-sea sponges.

SUPPORTING INFORMATION

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