Improving predictive mapping of deep-water habitats: Considering multiple model outputs and ensemble techniques

Katleen Robert [a,b,*], Daniel O.B. Jones [a], J. Murray Roberts [c,d,e], Veerle A.I. Huvenne [a]

[a] National Oceanography Centre, Southampton, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK
[b] Ocean and Earth Science, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK
[c] Centre for Marine Biodiversity and Biotechnology, Heriot-Watt University, Edinburgh EH14 4AS, UK
[d] Center for Marine Science, University of North Carolina Wilmington, 601 S College Road, Wilmington, NC 28403-5928, USA
[e] Scottish Association for Marine Science, Scottish Marine Institute, Oban, Argyll, PA37 1QA, UK

ABSTRACT

In the deep sea, biological data are often sparse; hence models capturing relationships between observed fauna and environmental variables (acquired via acoustic mapping techniques) are often used to produce full coverage species assemblage maps. Many statistical modelling techniques are being developed, but there remains a need to determine the most appropriate mapping techniques. Predictive habitat modelling approaches (redundancy analysis, maximum entropy and random forest) were applied to a heterogeneous section of seabed on Rockall Bank, NE Atlantic, for which landscape indices describing the spatial arrangement of habitat patches were calculated. The predictive maps were based on remotely operated vehicle (ROV) imagery transects high-resolution autonomous underwater vehicle (AUV) side-scan backscatter maps. Area under the curve (AUC) and accuracy indicated similar performances for the three models tested, but performance varied by species assemblage, with the transitional species assemblage showing the weakest predictive performances. Spatial predictions of habitat suitability differed between statistical approaches, but niche similarity metrics showed redundancy analysis and random forest predictions to be most similar. As one statistical technique could not be found to outperform the others when all assemblages were considered, ensemble mapping techniques, where the outputs of many models are combined, were applied. They showed higher accuracy than any single model. Different statistical approaches for predictive habitat modelling possess varied strengths and weaknesses and by examining the outputs of a range of modelling techniques and their differences, more robust predictions, with better described variation and areas of uncertainties, can be achieved. As improvements to prediction outputs can be achieved without additional costly data collection, ensemble mapping approaches have clear value for spatial management.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license
(http://creativecommons.org/licenses/by/4.0/).

1. Introduction

As the anthropogenic footprint extends deeper into our oceans, reliable descriptions of the seafloor and the species present are required to devise appropriate management and conservation measures. With very limited areas of seafloor mapped at comparable resolution to terrestrial environments (Sandwell et al., 2006), quantitative spatial information regarding distributions of marine biotic and abiotic components is needed to build benthic habitat maps (Kostylev et al., 2001). Recent advances in acoustic techniques for seafloor mapping (Brown et al., 2011) have made it possible to create detailed geomorphological maps more rapidly.

However, the biological information needed to supplement complete coverage topographic and geological maps has remained limited owing to the time-consuming process of specimen collection and taxonomic identification (Przeslawski et al., 2011).

Full coverage biological sampling is often not an option, and hierarchical approaches involving nested survey designs are often employed. They involve a combination of broader-scale geological map creation based on acoustic data, and detailed ground-truthing biological studies covering smaller spatial extents, often taking the form of imagery transects (Elvenes et al., 2014; Robert et al., 2015). The broader-scale geological maps can be used to define habitat patches allowing the relationships between the spatial arrangement of these patches within the surrounding landscape and their effect on species spatial patterns (Turner and Gardner, 1991) to be examined, modelled and used to make biological predictions across the larger extent covered by the acoustic surveys. The

* Corresponding author.
  E-mail address: katleen.robert@noc.ac.uk (K. Robert).

spatial arrangement of habitat patches can be described using a variety of class and landscape metrics, the former used to describe properties of patches from a single habitat type while the latter are used to characterise all patches present within a landscape (McGarigal et al., 2012). Although such metrics have been shown to help explain species spatial patterns (Teixidó et al., 2002), they have so far rarely been employed for predictive mapping.

In recent years, there have been an increasing number of studies employing a variety of techniques to produce predictive full coverage megabenthic invertebrate habitat maps: maximum entropy (Rengstorf et al., 2012; Ross and Howell, 2012), many types of decision or classification trees (Compton et al., 2013; Gonzalez-Mirelis and Lindegarth, 2012), a variety of multivariate analyses or ordination methods (Buhl-Mortensen et al., 2012; Shumchenia and King, 2010), general additive models, neural networks (Palialexis et al., 2011) and many more. Some of these techniques, such as maximum entropy, are based on records of presence only (with background points), as obtaining reliable absence data can be particularly difficult (Pearce and Boyce, 2006). However, when absence data is available, presence-absence models (such as general linear/additive models or classification trees) can provide more information regarding unsuitable habitats (Brotons et al., 2004; Pearson et al., 2006) and avoid the difficulties associated with selecting appropriate background points (Phillips et al., 2009).

In this study, we used benthic imagery data (photographs and extracted video frames), in addition to acoustic maps to produce predictive maps for megabenthic invertebrate species assemblages, specifically demonstrating the usefulness of class and landscape indices to improve prediction results. The prediction accuracy of three different modelling approaches was investigated for highly heterogeneous sections of seabed mapped at very high resolutions (~0.5 m pixel size): redundancy analysis (RDA) (ter Braak, 1994), maximum entropy (MaxEnt) (Phillips and Dudík, 2008) and random forest (RF) (Breiman, 2001). As these techniques represent three different modelling approaches ('assemble and predict together', RDA; 'assemble first, predict later' using a presence only model, MaxEnt; as well as a presence-absence model, RF) (Ferrier and Guisan, 2006), the aim was to determine which approach and whether a single one may be most appropriate when considered across multiple species assemblages. Based on the results obtained, we also examined whether ensemble maps, which take into account predictions and uncertainties from more than one model (Araújo and New, 2007; Marmion et al., 2009b), could further improve predictions.

2. Materials and methods

2.1. Survey design

As part of the 'UK Marine Environmental Mapping Programme' (MAREMAP; http://www.maremap.ac.uk/index.html) and the 'COmplex Deep-sea Environments: Mapping habitat heterogeneity As Proxy for biodiversity' project (CODEMAP; http://www.codemap.eu/), a section of the western flank of Rockall Bank (200–400 m), Northeast Atlantic, was mapped during the RRS *James Cook* 060 cruise carried out in May–June 2011 (Fig. 1). Over 380 km^2 of ship-based multibeam bathymetry (pixel size of 10 × 10 m), three 12–13 km^2 Autosub6000 autonomous underwater
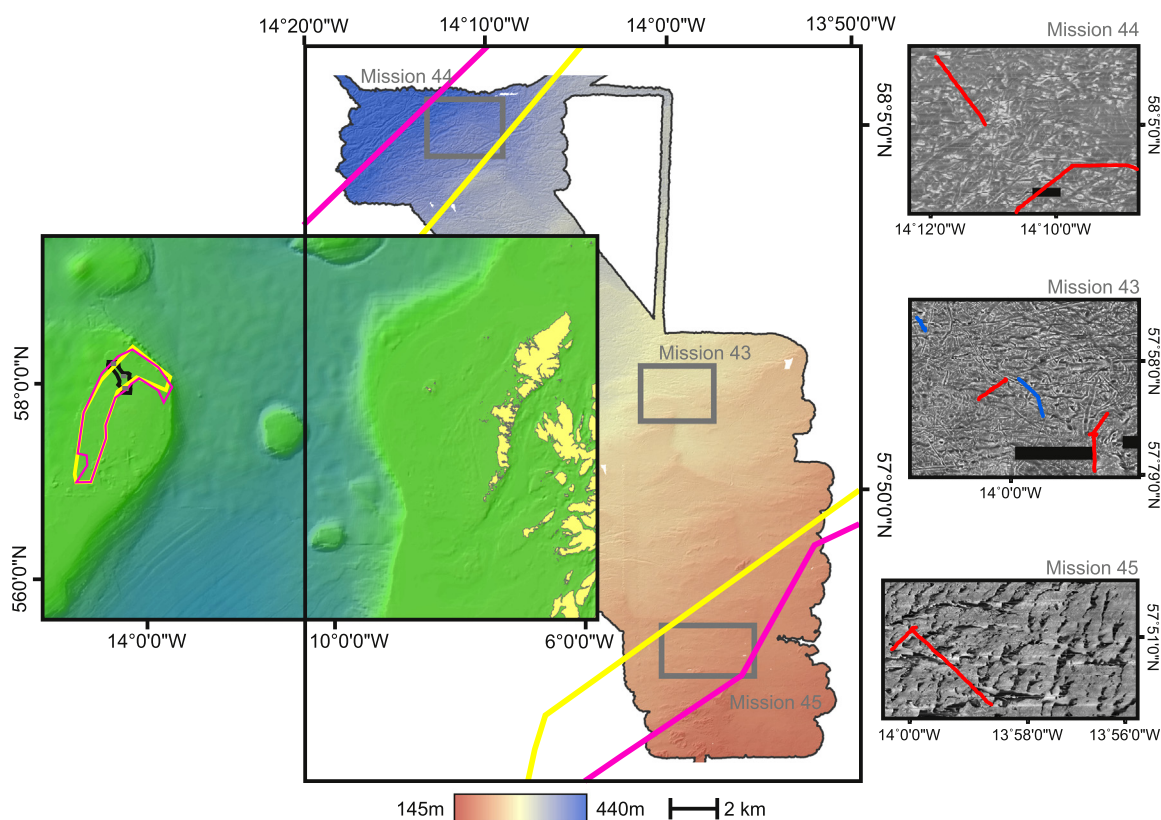
Fig. 1. Map of the surveys carried out on Rockall Bank, Northeast Atlantic. Ship-based bathymetry (black outline) displayed with superimposed outlines of the sidescan sonar data (grey boxes) collected during three autonomous underwater vehicle missions. The remotely operated vehicle imagery transects carried out during the JC-060 cruise are shown in red and the two from JC-073 in blue, superimposed on the autonomous underwater vehicle sidescan sonar maps (high backscatter shown in white). The boundaries of a 2007 fisheries closure area (yellow) and a candidate for 'Special Area of Conservation' (pink) are also displayed. Bathymetry for Northeast Atlantic background from GEBCO (General Bathymetric Chart of the Oceans (IOC IHO and BODC, 2003). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

vehicle (AUV, missions M43, M44 and M45) sidescan sonar surveys (pixel size of 0.5 × 0.5 m) and five *Lynx* remotely operated vehicle (ROV) photographic imagery (2592 × 1944 pixels) transects (1,222 images along ~8 km using a Kongsberg OE14–208 digital stills camera) were collected. During the *JC*-073 cruise carried out in June 2012 as part of the UK Ocean Acidification programme's 'Changing Oceans Expedition', two additional *Holland I* ROV high-resolution (1920 × 1080 pixels) video transects (Insite Mini Zeus camera with direct HDSDI fibre output) were carried out within the more heterogeneous area (M43) surveyed by the AUV. Frames were extracted at a rate representing the distance separating the previously collected digital stills (~1 frame per 5 m, 514 images).

The high resolution AUV surveys were positioned in areas of proposed boundary changes to a conservations zone (Fig. 1). In 2007, a Fisheries Closure was established by the North East Atlantic Fisheries Commission (NEAFC Recommendation IX-2008, EC Regulation No 40/2008) based on reports of cold-water coral occurrence. In 2011, a nearly overlapping, but slightly extended area was put forth as candidate Special Area of Conservation (cSAC) with the main aim of protecting stony and biogenic reefs (JNCC, 2010), a habitat listed under Annex I of the Habitat Directive (92/43/EEC). As such, AUV mapping was conducted in areas outside of the Fisheries Closure, but still inside the cSAC (M44 and M45) as well as inside both protected areas (M43), to identify the status of the seabed habitats. ROV imagery transects were positioned to sample a variety of sediment types within each of the three areas, including areas of high backscatter likely to harbour cold-water corals. To reduce the influence of spatial autocorrelation, images were systematically subsampled into 8 groups in which neighbouring pictures were located at a distance of 40 m (Fig. 2).

All individual organisms larger than 1 cm were counted and identified, using morphospecies when species-level identification could not be achieved. Identification was achieved by consulting image catalogues (Guillaumont et al., 2014; Howell and Davies, 2010; Jones and Gates, 2010; KeyToNature programme, 2015; SERPENT project, 2009; WoRDSS, 2016), species lists compiled from conservation work in the area (Howell et al., 2009; JNCC, 2010) and taxonomic resources (Hayward and Ryland, 1995 ; Mortensen, 1927). Sponges were only described to morphological categories as outlined in Bell and Barnes (2001). Parallel lasers

Table 1.
List of class and landscape variables used to construct predictive maps. Values in italics indicate the size of the moving window used to calculate the metrics. For formulas and descriptions see McGarigal et al. (2012).

Class metrics	
60 m	*150 m*
Sand	
Proportion of Like Adjacencies	Max. Shape Index
Landscape Shape Index	Min. Shape Index
	Proportion of Like Adjacencies
	Mean Patch Area
Mixed Sediments	
Landscape Shape Index	Mean Shape Index
Mean Patch Core Area	
Hard Substrate	
Mean Shape Index	Mean Patch Core Area
Bedrock	
Effective Mesh Size	Min. Patch Area
Patch Density	
Mean Shape Index	
Live Coral	
Total Area	Mean Shape Index
Min. Patch Core Area	Largest Patch Index
Coral Rubble	
Patch Density	Min. Patch Core Area

Landscape Metrics	
60 m	*150 m*
Max. Shape Index	Number of Patches
Largest Patch Index	Mean Core Area Index
	Mean Shape Index

(with 10 cm separation) were mounted on the ROVs to provide a scale on all recorded images. Positioning was achieved using the ROVs' ultra-short baseline (USBL) navigation systems. Only common species, which occurred in at least 10 images, were retained for the analysis, which was carried out with the images as sampling units.

Environmental descriptors were derived from the sidescan backscatter maps (EdgeTech FS2200, 410 kHz). These maps had been classified into sediment interpretation maps (0.5 × 0.5 m pixel size) representing six seabed facies (soft and mixed sediments, hard substratum, exposed bedrock as well as coral stand and rubble) using an unsupervised classification (Robert et al., 2014). From the sediment interpretation maps, class and landscape indices were derived to describe the shape, size, diversity and spatial arrangement (connectivity) of habitat patches (selected for their explanatory power using redundancy analysis and forward selection based on Robert et al. (2014), listed in Table 1). Bathymetry and CTD derived environmental variables were examined, but as they did not significantly improve the models, they were not included and are not discussed further.

Landscape and class metrics were calculated for each pixel of the sediment interpretation maps using moving windows (sizes of 60 × 60 m and 150 × 150 m, see Robert et al. (2014) for description of size choice). Owing to the large number of computations involved, the high performance computer cluster IRIDIS 3 (University of Southampton) was used to run an R script (R Development Core Team, 2011) written for parallel computation. The R package 'SDMTools' was used to compute the metrics and the package 'Snowfall' was used to run the computations in parallel. On smaller datasets, these computations could easily be accomplished on a regular desktop computer (see Appendix A for R code).
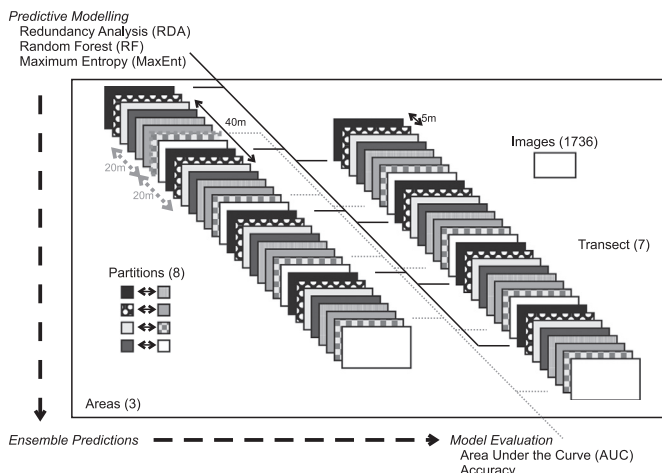
Fig. 2. Schematic of the steps taken. The data were separated into eight partitions where images were located every 40 m. Three statistical approaches were applied separately to each partition and the results were evaluated using the partition whose images were halfway (20 m). For each partition, the three statistical approaches were combined to form ensemble models. These steps were carried out for four species assemblages.

2.2. Predictive modelling

Four species assemblages (A1- *Parastichopus tremulus*, A2- *Munida sarsi* and associated species, A3- *Reteporella* sp. and various sponge spp., and A4- *Lophelia pertusa* and associated species) were identified using K-mean classification, ANOSIM and 'species indicator values' (described in Robert et al., (2014)). Although low numbers of individuals were generally found, the holothurian *Parastichopus tremulus* was most commonly observed in soft sediments. Bryozoan species (mostly *Reteporella* sp.) and various sponge morphotypes were characteristic of hard substratum, while the abundant squat lobster *Munida sarsi* dominated mixed sediment areas. Species associated with the cold-water coral *Lophelia pertusa* included sabellid worms, an unsampled actinarian sp. and many asteroid spp. Analyses were carried out using the R libraries 'vegan', 'randomForest', 'dismo', 'raster' and 'caret'.

2.2.1. Redundancy analysis

A multivariate approach was first considered, and redundancy analysis (RDA) was used to create the first set of full coverage fine scale biological maps. Similarly to Oldeland et al. (2010), we used the estimated coefficients of the linear combination of environmental predictors to position each pixel along the canonical axes. To assign each pixel to a species assemblage, a nearest neighbour classification (k=10) was carried out. The probability of belonging to each of the species assemblages was estimated using the proportion of nearest neighbours.

2.2.2. MaxEnt

Maximum entropy (MaxEnt) predicts an index of relative habitat suitability using presence data compared to randomly selected background points (Phillips and Dudík, 2008) by minimizing the distance between the probability density of species occurrence and the probability density of the covariates as they occur in space (relative entropy) (Elith et al., 2011). The software MaxEnt (version 3.3.3, freely available online http://www.cs.princeton.edu/~schapire/maxent/) was employed with sampling bias grids to select the background points and help account for the transect design. Weighted surfaces based on sampling density were built (with more weight given to areas closer to sampled locations) using a Gaussian kernel estimation (with SD of 500 m) (Clements et al., 2012). Habitat suitability predictions were made separately for each of the four species assemblages. Presence/absence predictions were obtained by setting the threshold level to optimize sensitivity and specificity.

2.2.3. Random forest

Random Forest (RF) is a technique that allows for the building of multiple trees for a dataset, hence the term forest (Breiman, 2001). Each tree is built based on a sub-sample of the data and at each node the data are split based on the best predictor variable, selected out of a smaller number of randomly selected variables. A probability estimate can be obtained based on the number of votes given to each class for a given pixel. Forests were built using a varying number of trees and environmental variables, but a forest containing 1000 trees and considering 15 environmental predictors per node was selected.

2.3. Model evaluation

To minimize spatial autocorrelation between the training and testing datasets, systematic data splitting was carried out. For each of the 8 data partitions, the dataset whose images were located at a distance of 20 m (for example models based on partition 3 were assessed using images in partition 7, Fig. 2) were used to calculate the area under the curve (AUC) of the receiver operating characteristics (ROC) (Fielding and Bell, 1997; Manel et al., 2001). This distance was chosen based on spatial analysis of the data which indicated that spatial autocorrelation became negligible at distances of ~20 m as a result of the high heterogeneity introduced by iceberg ploughmarks in the area (Robert et al., 2014). AUC was calculated for each partition, species assemblage and statistical approach. The test AUC values reported for MaxEnt were based on absences identified within the acquired imagery as opposed to background data. Prediction accuracy (the proportion of correctly assigned presences and absences over total sample size) was also calculated. Full coverage maps were produced for each of the eight partitions. To assess similarities between the predictive maps obtained from each of the statistical approaches, the Hellinger-based niche similarity metric described in Warren et al. (2008) was computed on maps averaged across partitions. This measure can vary from 0 (no overlap) to 1 (identical niches).

2.4. Ensemble predictions

Considering that different models are likely to produce different predictive outputs, but with each containing separate information and areas of uncertainties, the idea of ensemble predictions is to summarise a range of potential outcomes to produce more robust predictions (Araújo and New, 2007). Using the same partitioning of training and test datasets as previously described, for each partition, AUC values for the ensembles were calculated by averaging probability maps from all three models for each species assemblage. Accuracy of the ensemble predictions was calculated by first assigning, for each statistical approach and partition, the species assemblage with the highest predicted probability of occurrence. Subsequently, for each partition, majority voting was carried out based on the species assemblage predicted by each statistical technique. To obtain a visual depiction of prediction confidence, the number of models in agreement at each pixel was also calculated.

3. Results

For the combined *JC*-060 and *JC*-073 datasets, a total of 11,268 individual organisms were observed from 38 morphospecies (present in at least 10 images). For each of the species assemblages considered, AUC values showed all models to perform better than could be expected by chance (Table 2). Based on the eight partitions MaxEnt showed average AUC values ranging from 0.73 (SD=0.05 Assemblage A4; SD=0.02 Assemblage A3) to 0.63 (SD=0.05 Assemblage A2), values of 0.81 (SD=0.02 Assemblage A1) to 0.63 (SD=0.04 Assemblage A2) for RDA and 0.83 (SD=0.02 Assemblage A1) to 0.68 (SD=0.05 Assemblage A2) for the RF classifier. All models had the most difficulties predicting Assemblage A2. Assemblage A1 could be captured by RDA and RF, but predictions using MaxEnt were lower (Mean=0.67, SD=0.02). Overall, RF had the highest AUC values across species assemblages.

The three models showed differences in the maps of habitat suitability for the various species assemblages, but measures of environmental niche indicated similarities between model predictions (Table 3). Generally, RDA and RF showed the most similarities across all species assemblages, and all models tended to select a similar set of environmental descriptors as most important. For RF, the number of patches (150 m), the mean patch size for soft sediments (150 m) and the total area occupied by coral (60 m) were the most important variables. For individual species assemblages, all three environmental descriptors were also the most important for Assemblage A1, while for Assemblage A3, it was the former and for Assemblage A4, the latter. Proportion of like adjacencies for soft sediment (150 m) was also important for

Table 2

Area under the curve (AUC) values observed based on eight partitions for four different species assemblages using three statistical approaches (Maximum Entropy (MaxEnt), Redundancy Analysis (RDA) and Random Forest (RF)) and one ensemble mapping approach. Numbers in bold show highest values obtained for each assemblage and partition. Assemblage A1 - *Parastichopus tremulus* and associated species, Assemblage A2 - *Munida sarsi* and associated species, Assemblage A3 - *Reteporella* sp. and various sponge spp. and Assemblage A4 - *Lophelia pertusa* and associated species.

	Assemblage A1				Assemblage A2				Assemblage A3				Assemblage A4			
	RDA	RF	MaxEnt	Ensemble	RDA	RF	MaxEnt	Ensemble	RDA	RF	MaxEnt	Ensemble	RDA	RF	MaxEnt	Ensemble
1	**0.84**	0.83	0.63	0.83	0.69	0.69	0.68	**0.74**	0.70	**0.81**	0.71	0.75	0.76	0.82	0.81	**0.83**
2	0.80	**0.83**	0.66	0.82	0.63	**0.76**	0.65	0.72	0.74	**0.81**	0.74	0.78	0.59	0.67	**0.71**	0.69
3	0.82	**0.86**	0.66	0.85	0.55	**0.67**	0.63	0.64	0.75	**0.81**	0.71	0.76	0.78	**0.81**	0.76	0.80
4	0.79	**0.84**	0.66	0.82	0.64	0.66	0.61	**0.67**	0.69	**0.73**	0.72	0.72	0.56	**0.67**	0.65	**0.67**
5	0.83	0.79	0.72	**0.84**	0.67	0.69	0.67	**0.72**	0.77	**0.80**	0.75	**0.80**	0.76	**0.78**	0.73	**0.78**
6	0.81	**0.84**	0.69	**0.84**	0.60	**0.73**	0.65	0.69	0.84	**0.85**	0.74	0.83	**0.81**	0.76	0.70	0.77
7	0.83	0.83	0.65	**0.84**	0.63	0.63	0.63	**0.66**	0.79	**0.80**	0.75	0.79	0.72	**0.78**	0.74	0.77
8	0.80	**0.83**	0.67	**0.83**	**0.64**	0.62	0.55	0.59	0.68	**0.75**	0.74	0.74	0.68	0.74	0.74	**0.79**
Mean	0.81	**0.83**	0.67	**0.83**	0.63	**0.68**	0.63	**0.68**	0.74	**0.79**	0.73	0.77	0.71	0.75	0.73	**0.76**
SD	0.02	0.02	0.03	0.01	0.04	0.05	0.04	0.05	0.05	0.04	0.02	0.04	0.09	0.06	0.05	0.06

Table 3

Hellinger-based niche similarities measures (Warren et al., 2008) between habitat suitability predictions for four species assemblages based on three statistical approaches. Number in bold show highest values obtained for each assemblage and survey area. Assemblage A1 - *Parastichopus tremulus* and associated species, Assemblage A2 - *Munida sarsi* and associated species, Assemblage A3 - *Reteporella* sp. and various sponge spp. and Assemblage A4 - *Lophelia pertusa* and associated species.

	Assemblage A1			Assemblage A2			Assemblage A3			Assemblage A4		
	MaxEnt	RDA	RF	MaxEnt	RDA	RF	MaxEnt	RDA	RF	MaxEnt	RDA	RF
M43												
MaxEnt		0.83	0.83		0.81	0.84		0.81	0.81		0.75	0.78
RDA			**0.91**			**0.90**			**0.87**			**0.88**
RF												
M44												
MaxEnt		0.82	0.84		0.74	0.75		0.73	0.77		0.73	0.77
RDA			**0.91**			**0.90**			**0.83**			**0.83**
RF												
M45												
MaxEnt		0.81	0.84		0.82	0.83		0.79	0.82		0.81	0.83
RDA			**0.90**			**0.89**			**0.84**			**0.87**
RF												

Assemblage A2. For MaxEnt, total area covered by coral (60 m) and effective mesh size of bedrock (60 m) were also important for Assemblage A4, while number of patches (150 m) was selected for both Assemblages A2 and A3. For Assemblage A1, it was maximum shape index and proportion of like adjacencies for soft sediment (150 m). Similarly for RDA, the number of patches (150 m) and the proportion of like adjacencies for soft sediment (150 m) were most important, but the mean shape index for corals (150 m) and the effective mesh size for bedrock (60 m) were also valuable.

The areas of variability also differed between models (Fig. 3) and ensemble predictions (Fig. 4 and Appendix B) made by combining all three models exhibited a slightly higher accuracy across species assemblages than could be obtained based on any single model (Table 4). Ensemble predictions showed the highest (or equal) accuracies in every partition for Assemblage A3, and all but one partition for Assemblage A1. For the other two species assemblages, ensemble predictions still showed the highest (or equal) accuracy in half of the partitions. Overall, models completely disagree in less than 10% of the area surveyed while all three agreed in 25.8% of M43, 79.0% of M44 and 42.1% of M45 (Fig. 3).

4. Discussion

By taking advantage of species-environment relationships, abiotic proxies can provide direct applications for the management of natural resources by establishing representations of biotic components via high resolution acoustic survey techniques. The spatial arrangement of habitat patches was successfully included to predict the spatial patterns of four species assemblages across a highly heterogeneous area of seabed. No single approach consistently surpassed the others across species assemblages and although differences occurred between spatial predictions of habitat suitability from the different statistical approaches, ensemble models appeared as a meaningful improvement.

4.1. Model predictions

Of the three models (RDA, RF and MaxEnt) compared in this study, similar AUC values were obtained, but performance varied by species assemblage. As species turnover generally occurs over a gradient, the predictions showed a similar pattern, and overlap between habitat suitability predictions occurred, particularly between Assemblage A3 and A4. This is to be expected as cold-water corals need hard substratum for attachment (Wilson, 1979) and in turn provide hard substratum to a number of species. Assemblage A2 appeared as a transition between the more defined hard substratum and soft sediment associated fauna, and as such prediction performance for this assemblage generally tended to be lower. Across models, areas of highest disagreement tended to occur at the edge of patches and highlighted the difficulty associated in delineating hard boundaries for otherwise continuous gradients of species assemblages. Albeit at a larger scale, higher levels of discrepancies between modelling techniques have been shown to occur at the edge of a species distribution (Grenouillet et al., 2011).
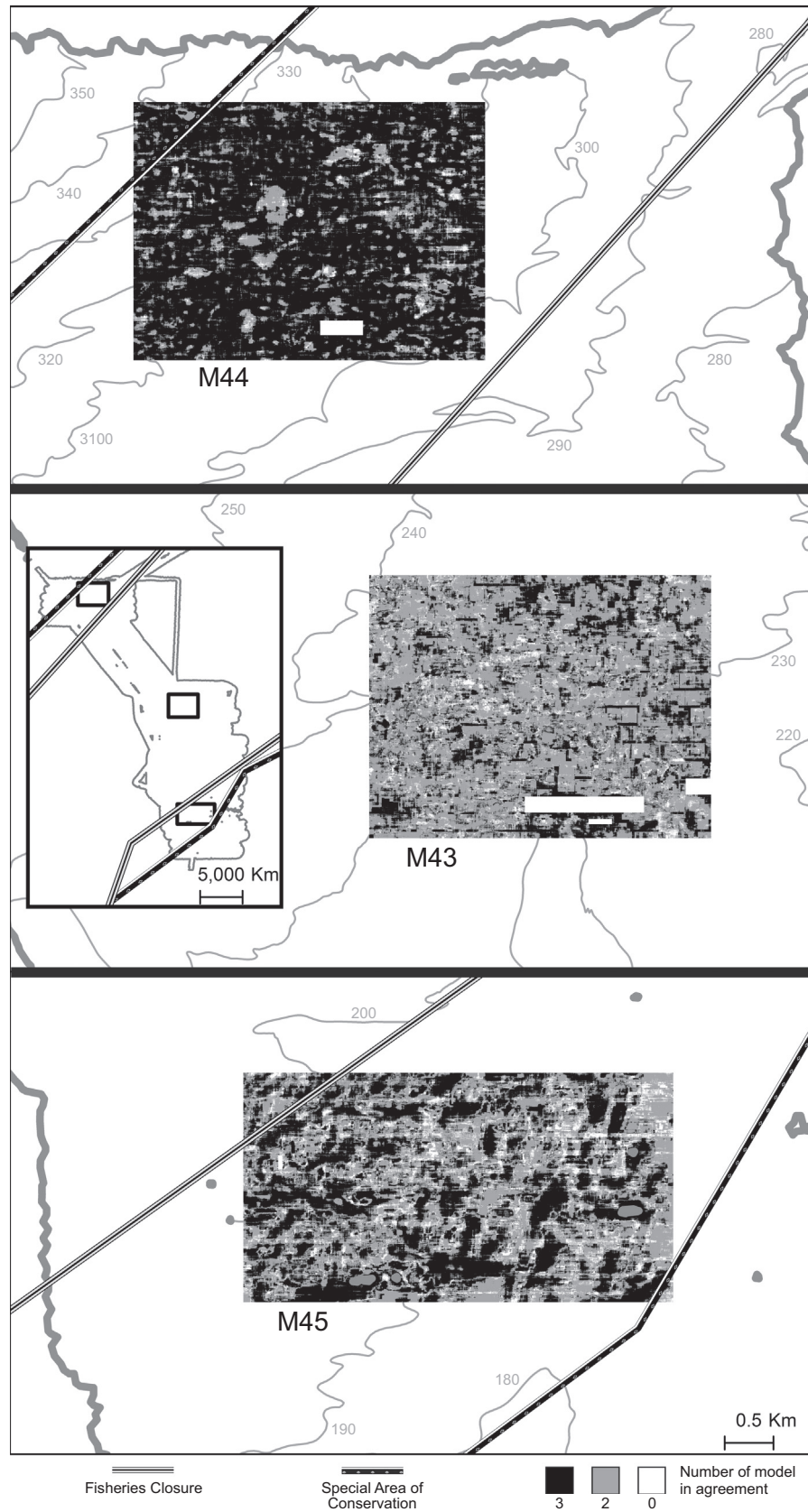
Fig. 3. Maps showing the prediction agreement based on the three models considered for Assemblage A4 (*Lophelia pertusa* and associated species) for three survey areas around two conservation zone boundaries. The maps were based on the habitat suitability averaged across all eight partitions. The inset on the left shows the relative position of the three survey areas M43, M44 and M45 with respect to the boundaries of the conservation zones and the outline of the shipboard-multibeam survey. The white rectangles represent areas for which the acoustic data was corrupted and were not included in the prediction models.
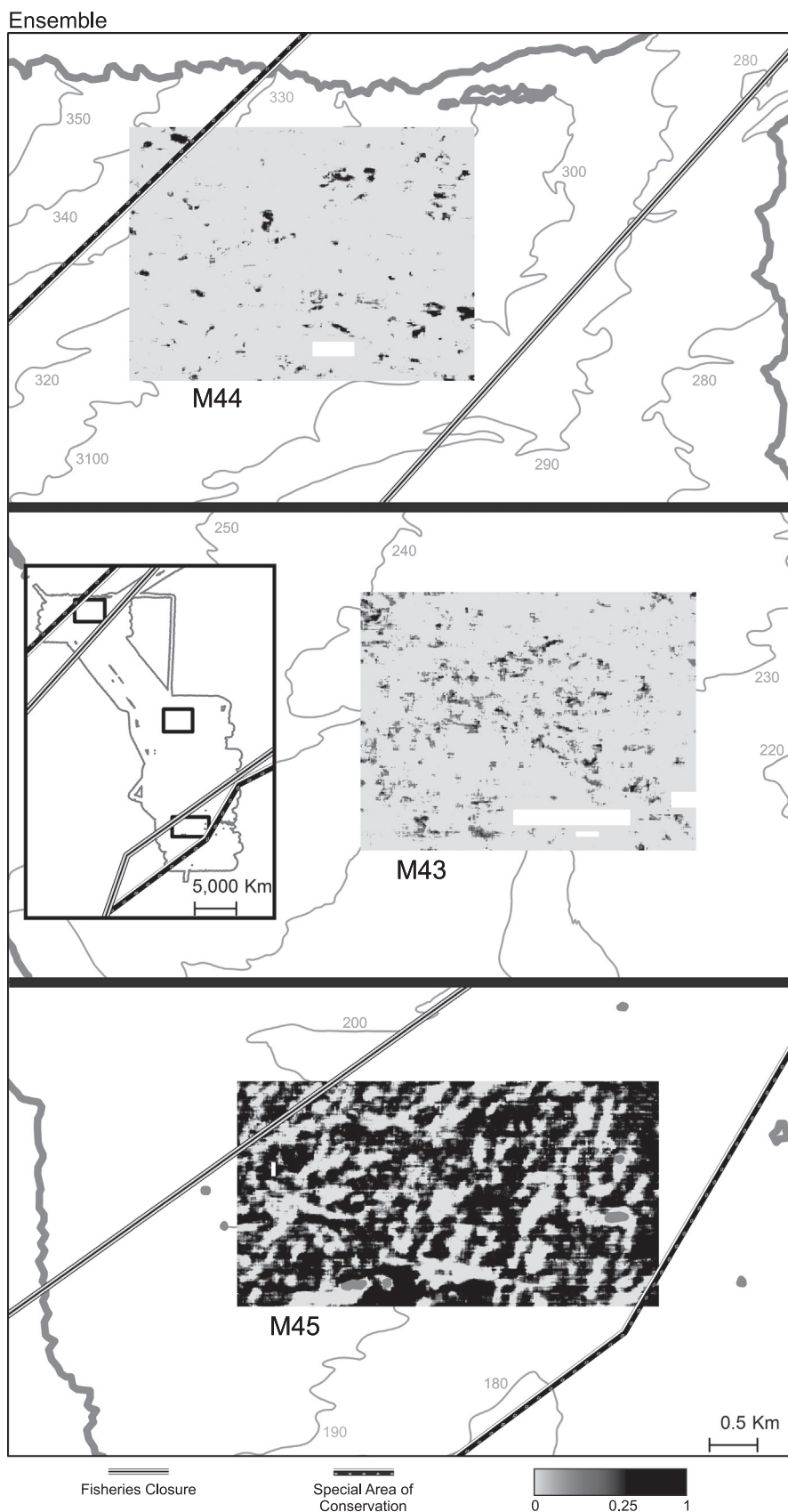
Ensemble

Fig. 4. Maps showing the relative habitat suitability resulting from an ensemble modelling approach for Assemblage A4 (*Lophelia pertusa* and associated species) for three survey areas around two conservation zone boundaries (higher suitability in black). The maps show the habitat suitability averaged across all eight partitions. The inset on the left shows the relative position of the three survey areas M43, M44 and M45 with respect to the boundaries of the conservation zones and the outline of the shipboard-multibeam survey. Maps for the other three groups are presented in Appendix B. The white rectangles represent areas for which the acoustic data was corrupted and were not included in the prediction models.

Table 4
Accuracy values obtained based on eight partitions for four different species assemblages using three statistical approaches (Maximum Entropy (MaxEnt), Redundancy Analysis (RDA) and Random Forest (RF)) and an ensemble mapping approach. Accuracy defined as the proportion of correctly assigned presences and absences over sample size. Number in bold show highest values obtained for each assemblage and partition. Assemblage A1 - *Parastichopus tremulus* and associated species, Assemblage A2 - *Munida sarsi* and associated species, Assemblage A3 - *Reteporella* sp. and various sponge spp. and Assemblage A4 - *Lophelia pertusa* and associated species.

	Assemblage A1				Assemblage A2				Assemblage A3				Assemblage A4			
Partition	RDA	RF	MaxEnt	Ensemble	RDA	RF	MaxEnt	Ensemble	RDA	RF	MaxEnt	Ensemble	RDA	RF	MaxEnt	Ensemble
1	0.76	0.70	0.76	**0.81**	**0.74**	0.73	0.71	0.73	0.76	**0.88**	0.84	**0.88**	0.82	**0.87**	0.83	0.83
2	0.72	0.73	0.76	**0.78**	0.68	**0.77**	0.68	0.75	0.80	0.83	0.82	**0.86**	0.82	0.85	0.82	**0.87**
3	0.70	0.76	0.76	**0.78**	**0.78**	**0.78**	0.70	0.76	0.79	0.85	0.79	**0.88**	0.82	**0.84**	0.82	0.83
4	0.69	0.74	0.72	**0.76**	0.78	0.75	0.61	**0.83**	0.79	0.80	0.81	**0.92**	0.81	0.77	0.82	**0.84**
5	0.72	0.70	**0.76**	0.71	0.71	0.71	**0.76**	0.71	0.83	0.87	0.79	**0.88**	0.82	0.85	**0.85**	0.81
6	0.70	0.73	**0.76**	**0.76**	**0.76**	0.73	0.70	0.73	0.81	0.83	0.82	**0.90**	**0.85**	0.82	0.84	0.78
7	0.75	0.71	0.76	**0.77**	**0.76**	**0.76**	0.71	**0.76**	0.83	**0.87**	0.80	**0.87**	0.79	0.81	0.81	**0.82**
8	0.71	0.73	0.73	**0.76**	0.69	0.72	0.72	**0.74**	0.85	0.85	0.82	**0.86**	0.76	**0.82**	0.80	**0.82**
Mean	0.72	0.72	0.75	**0.77**	0.74	0.74	0.70	**0.75**	0.81	0.85	0.81	**0.88**	0.81	**0.83**	0.82	**0.83**
SD	0.02	0.02	0.02	0.03	0.04	0.02	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.03	0.02	0.02

Assemblage A2 tended to be found in particularly complex areas where a high number of patches, of both hard and soft sediments, appeared. On the other hand, Assemblage A1 was found in areas characterised by few large patches in proximity to other large soft sediment patches. Assemblage A3 or even A4 were generally found in regions of harder substratum, particularly if bedrock was present.

As these three statistical approaches are based on very different modelling strategies, differences in their predictions are to be expected. Presence-absence models generally provide more information about less suitable habitats (if adequate absences are available). As this information is not available to presence only models, overestimation of suitable habitats can occur (Brotons et al., 2004; Pearson et al., 2006). Results can also depend on species characteristics, with generalist species being more difficult to predict accurately, and absence data being more valuable for such species (Brotons et al., 2004; Marmion et al., 2009a). This might be another reason why lower prediction performances were obtained for Assemblage A2. Overall MaxEnt tended to show a lower niche similarity than RDA and RF, which may be due to its different data requirement. In the case of RDA, classification into assemblages was only conducted after predictions of individual species, and as such could be more affected by difficulties associated with predicting rarer species. However, since species are predicted instead of assemblages, it might also be possible to define potentially new assemblages as occurring in areas outside of the originally sampled locations (Ferrier and Guisan, 2006). RF predictions for Assemblage A4 equalled those of the ensemble model. Other studies have found RF to often equal ensemble approaches (Grenouillet et al., 2011; Marmion et al., 2009b), potentially because it already includes a consensus step and might be less affected by species geographical attributes, such as prevalence, range and spatial autocorrelation (Marmion et al., 2009a). On the other hand, Meynard and Quinn (2007) found that although GAM tended to outperform classification trees under many simulated scenarios, the latter were particularly effective at predicting species displaying threshold (on/off) response curves to environmental variables. In the case of Assemblages A3 and A4, a threshold response to the presence of hard substratum could be expected while Assemblages A1 and A2 may be more likely to exhibit more continuous response curves.

Additional factors also complicate predictive habitat modelling, particularly in deeper waters. Precise spatial positioning can be problematic for underwater vehicles, particularly AUVs, owing to difficulties associated with determining the initial position following the descent and correcting for drift (McPhail, 2009). However, as our study site was located at 200–400 m water depth, limited drift would be expected. Accuracy of the ROV's Sonardyne USBL navigation is also expected to be better than 1% of the depth. With the landscape and class metrics calculated at scales of 60 m and 150 m, a small shift in position would have had limited effects on the values of the explanatory variables. The spatial extent for which predictions can be valid is also of importance, as predictions made for areas outside the range of environmental conditions captured by the survey design are problematic (Elith and Leathwick, 2009). The use of transects limited the area surveyed to single narrow lines leaving most of the regions covered acoustically without any biological sampling. Transects are designed to maximise seafloor survey areal coverage for a given bottom time, but also have the disadvantage of causing issues of spatial autocorrelation which need to be taken into account in order to adequately capture predictive ability (Hirzel and Guisan, 2002; Legendre et al., 2002). In our study, this effect was mitigated through a subsampling scheme which increased distances between sample images used for model building.

4.2. Ensemble mapping for conservation

Comparison of the statistical approaches showed differences in predictions, but a single approach did not consistently outperform the others when multiple species assemblages were considered. Instead, our results suggest that taking into account the output of many different models may provide a valuable alternative. Ensembles can be created using an array of approaches (Marmion et al., 2009b), but even the relatively simple approach taken in this study was effective at optimizing different model strengths and increasing accuracy. All three statistical approaches were included in the ensemble mapping of all four species assemblages, but in other cases, the consideration of thresholds for the exclusion of lower performing models may also be valuable. In any case, diversity in the type of approaches selected is needed to increase the likelihood of obtaining better performing ensemble models (Du et al., 2012). Identifying regions of prediction disagreement across models also provides an easy to understand depiction of spatial uncertainties.

On the other hand, under certain circumstances, ensemble mapping may be of less value. One example may be when the main aim is to derive habitat indicators such as extent to monitor the achievement of "Good Environmental Status" as suggested by the Marine Strategy Framework Directive (MSFD) (2008/56/EC) (CEFAS, 2012; OSPAR Intersessional Correspondence Group on the Coordination of Biodiversity Assessment and Monitoring, 2012). In this case, having the most thorough description of an habitat, including across model uncertainties, may not be as important as

having a consistent approach with minimum deviation over time from which to monitor change (Strong, 2015). Employing multiple models can increase the variability as some models may perform less adequately for certain assemblages and make it more difficult to assess the degree of change across surveys. However, this should still not preclude the examination of the data using multiple statistical approaches, as one approach may be more sensitive to a given environmental variable and be able to detect change earlier. Once prediction similarly has been ascertained, the final measure of extent could still rely on one specific technique for consistency.

Cold-water corals can have a strong impact on local diversity and much effort is being made to improve their conservation (Roberts and Hirshfield, 2004), but owing to limited data, spatial planning often must rely only on spatial predictions of habitat suitability. Even so, these maps provide greater insights into their spatial distribution patterns, which helps in understanding their ecology and supports adequate management better than single point observation obtained from limited imagery transects or physical samples. As illustrated by the case of Rockall Bank, different statistical approaches may provide different predictive maps of coral suitability. Predictions of assemblage A4 (mostly composed of the cold-water coral *L. pertusa* and associated filter-feeding species), the least common assemblage, were particularly sensitive to changes in modelling approach. For example, if only random forest had been considered, it would have been tempting to conclude that area M44 was as suitable a conservation area as M45. However, M44 was only found to contain coral rubble in ROV video surveys, likely resulting from past trawling activities. Ensemble models better represented the spatial patterns observed in the video survey as they highlight areas where predictions were consistent across at least two models.

Even though it is the broader-scale patterns in species distributions that may be of interest for management purposes, it is the fine-scale habitat characterisation of the environment, through high-resolution sidescan sonar mapping, that allowed the heterogeneity of the region to be accurately captured and the driving processes identified. The ship-board bathymetry survey carried out during *JC*-60 covered less than 10% of the 4365 km^2 conservation zone and took approximately 2.3 days. Although of much higher resolution than other datasets available for the remainder of this area, compared to the even higher resolutions obtained with the AUV, the ship-board dataset was of limited use in explaining species distribution patterns for the extent covered in this survey (Robert et al., 2014). It is clear that AUV mapping shows great promise for marine management; however there remains a distinct trade-off between the resolution achieved and the extent that can be covered. With current AUV technologies, Autosub6000 can be sent out from a ship to autonomously map an area for $\sim$30 h, covering a distance of $\sim$150 km (the size of the resulting area mapped will vary based on the acquired resolution) (Wynn et al., 2012). In order to map the entirety of the conservation zone to the resolution acquired in this study, $>$200 days would be required. This is well outside the scope of most scientific cruises or conservation projects, but AUVs have been successfully employed to target certain features in other conservations zones such as Haig Fras and the Darwin Mounds (Wynn et al., 2012). The Marine Autonomous and Robotic Systems (MARS) facility is also currently working on developing long-distance AUVs which could be deployed from shore to reach the closer offshore conservation areas with the aims of eventually covering greater extents at high resolutions and instituting repeat long-term monitoring of specific areas without the need for expensive ship-based surveys.

4.3. Conclusion

Predictive habitat maps are of great use for marine management as they represent the best available information to support decision making, but, as they are typically based on a very limited amount of data, they should only serve as general guides until more data become available. The presentation of uncertainty maps should help emphasize this point and can be employed to help select target areas for which further biological sampling will be particularly valuable. Uncritical reliance on a particular statistical method, without comparison with others, may lead to decisions being biased by the chosen method since predictions made from different modelling strategies have been shown to give differing outputs, but whose combination into ensemble models can lead to increased accuracy. Comparison between statistical methods showing one method to outperform the others may not always be extendable to other habitats, species or assemblages, and similarly our results cannot be perfunctorily generalized to all habitats. However, in cases where one statistical approach cannot be identified as performing significantly better, ensemble approaches may provide an elegant alternative. Although this approach can be more involved than other techniques, the additional work requires no further costly sampling or access to specialized equipment and potential increases in prediction performances are clearly of value for spatial planning.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dsr.2016.04.008.

References

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22 (1), 42–47.

Bell, J.J., Barnes, D.K.A., 2001. Sponge morphological diversity: a qualitative predictor of species diversity? Aquat. Conserv.: Mar. Freshw. Ecosyst. 11 (2), 109–121.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Brotons, L., Thuiller, W., Araújo, M.B., Hirzel, A.H., 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27 (4), 437–448.

Brown, C.J., Smith, S.J., Lawton, P., Anderson, J.T., 2011. Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuar. Coast. Shelf Sci. 92 (3), 502–520.

Buhl-Mortensen, L, Mortensen, P.B., Dolan, M., Dannheim, J., Bellec, V., Holte, B.,

2012. Habitat complexity and bottom fauna composition at different scales on the continental shelf and slope of northern Norway. Hydrobiologia 685 (1), 191–219.

Centre for Environment Fisheries and Aquaculture Science (CEFAS), 2012. Proposed UK targets for achieving GES and cost-benefit analysis for the MSFD.

Clements, G.R., Rayan, D.M., Aziz, S.A., Kawanishi, K., Traeholt, C., Magintan, D., Yazi, M.F.A., Tingley, R., 2012. Predicting the distribution of the Asian tapir in Peninsular Malaysia using maximum entropy modeling. Integr. Zool. 7 (4), 400–406.

Compton, T.J., Bowden, D.A., Pitcher, C.R., Hewitt, J.E., Ellis, N., 2013. Biophysical patterns in benthic assemblage composition across contrasting continental margins off New Zealand. J. Biogeogr. 40 (1), 75–89.

Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y., Liu, S., 2012. Multiple classifier system for remote sensing image classification: a review. Sensors 12, 4764–4792.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40 (1), 677–697.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17 (1), 43–57.

Elvenes, S., Dolan, M.F.J., Buhl-Mortensen, P., Bellec, V.K., 2014. An evaluation of compiled single-beam bathymetry data as a basis for regional sediment and biotope mapping. ICES J. Mar. Sci. 71, 867–881.

Ferrier, S., Guisan, A., 2006. Spatial modelling of biodiversity at the community level. J. Appl. Ecol. 43 (3), 393–404.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24 (1), 38–49.

Gonzalez-Mirelis, G., Lindegarth, M., 2012. Predicting the distribution of out-of-reach biotopes with decision trees in a Swedish marine protected area. Ecol. Appl. 22 (8), 2248–2264.

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.

Guillaumont, B., van den Beld, I., Davies, J., Menot, L., 2014. Catalogue des coraux d'eau froide et éponges du Golfe de Gascogne, IFREMER.

Hayward, P.J., Ryland, J.S., 1995. Handbook of the Marine Fauna of North-West Europe. Oxford University Press, Oxford.

Hirzel, A., Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecol. Model. 157 (2–3), 331–341.

Howell, K.L., Davies, J.S., 2010. Deep-sea species image catalogue. ⟨www.marlin.ac.uk/deep-sea-species-image-catalogue/⟩.

Howell, K.L., Davies, J.S., Jacobs, C., Narayanaswamy, B.E., 2009. Broadscale survey of the habitats of Rockall Bank and mapping of Annex I 'reef' habitat. Joint Nature Conservation Committee, Peterborough, UK.

IOC IHO and BODC, 2003. GEBCO digital atlas. Centenary edition of the GEBCO digital atlas, published on CD-ROM on behalf of the Intergovernmental Oceanographic Commission and the International Hydrographic Organization as part of the General Bathymetric Chart of the Oceans, British Oceanographic Data Centre, Liverpool, U.K.

JNCC, 2010. Conservation objectives and advice on operations. Offshore special area of conservation: North West Rockall Bank. Joint Nature Conservation Committee, Peterborough, UK.

Jones, D.O.B., Gates, A.R., 2010. Deep-sea life of Scotland and Norway. Ophiura Publishing.

Key To Nature Programme, 2015. Marine species identification portal. ⟨http://species-identification.org/index.php⟩.

Kostylev, V.E., Todd, B.J., Fader, G.B.J., Courtney, R.C., Cameron, G.D.M., Pickrill, R.A., 2001. Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs. Mar. Ecol. Prog. Ser. 219, 121–137.

Legendre, P., Dale, M.R.T., Fortin, M.-J., Gurevitch, J., Hohn, M., Myers, D., 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25 (5), 601–615.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38 (5), 921–931.

Marmion, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009a. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecol. Model. 220 (24), 3512–3520.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009b. Evaluation of consensus methods in predictive species distribution modelling. Divers. Distrib. 15 (1), 59–69.

McGarigal, K., Cushman, S.A., Ene, E., 2012. FRAGSTATS (4.x): spatial pattern analysis program for categorical and continuous maps. University of Massachusetts, Amherst, USA.

McPhail, S., 2009. Autosub6000: a deep diving long range AUV. J. Bionic Eng. 6 (1), 55–62.

Meynard, C.N., Quinn, J.F., 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. J.

Biogeogr. 34, 1455–1469.

Mortensen, T.H., 1927. Handbook of the Echinoderms of the British Isles. Oxford University Press,, Humphrey Milford, p. 1927.

Oldeland, J., Dorigo, W., Lieckfeld, L., Lucieer, A., Jürgens, N., 2010. Combining vegetation indices, constrained ordination and fuzzy classification for mapping semi-natural vegetation units from hyperspectral imagery. Remote Sens. Environ. 114 (6), 1155–1166.

OSPAR Intersessional Correspondence Group on the Coordination of Biodiversity Assessment and Monitoring, 2012. MSFD advice manual and background document on biodiversity. In: OSPAR Biodiversity Committee (Ed.).

Palialexis, A., Georgakarakos, S., Karakassis, I., Lika, K., Valavanis, V.D., 2011. Prediction of marine species distribution from presence-absence acoustic data: comparing the fitting efficiency and the predictive capacity of conventional and novel distribution models. Hydrobiologia 670 (1), 241–266.

Pearce, J.L., Boyce, M.S., 2006. Modelling distribution and abundance with presence-only data. J. Appl. Ecol. 43 (3), 405–412.

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. J. Biogeogr. 33 (10), 1704–1711.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31 (2), 161–175.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecol. Appl. 19 (1), 181–197.

Przeslawski, R., Currie, D.R., Sorokin, S.J., Ward, T.M., Althaus, F., Williams, A., 2011. Utility of a spatial habitat classification system as a surrogate of marine benthic community structure for the Australian margin. ICES J. Mar. Sci. 68 (9), 1954–1962.

R Development Core Team, 2011. R: A language and environment for statistical computing R Foundation for Statistical Computing.

Rengstorf, A.M., Grehan, A., Yesson, C., Brown, C., 2012. Towards high-resolution habitat suitability modeling of vulnerable marine ecosystems in the deep-sea: resolving terrain attribute dependencies. Mar. Geod. 35 (4), 343–361.

Robert, K., Jones, D.O.B., Huvenne, V.A.I., 2014. Megafaunal distribution and biodiversity in a heterogeneous landscape: the iceberg scoured Rockall Bank, NE Atlantic. Mar. Ecol. Prog. Ser. 501, 67–88.

Robert, K., Jones, D.O.B., Tyler, P.A., Van Rooij, D., Huvenne, V.A.I., 2014. Finding the hotspots within a biodiversity hotspot: fine-scale biological predictions within a submarine canyon using high-resolution acoustic mapping techniques. Mar. Ecol. 36 (4), 1256–1276.

Roberts, S., Hirshfield, M., 2004. Deep-sea corals: out of sight, but no longer out of mind. Front. Ecol. Environ. 2, 123–130.

Ross, R.E., Howell, K.L., 2012. Use of predictive habitat modelling to assess the distribution and extent of the current protection of 'listed' deep-sea habitats. Divers. Distrib. 19, 433–445.

Sandwell, D.T., Smith, W.H.F., Gille, S., Kappel, E., Jayne, S., Soofi, K., Coakley, B., Géli, L., 2006. Bathymetry from space: rationale and requirements for a new, high-resolution altimetric mission. Comptes Rendus Geosci. 338 (14–15), 1049–1062.

SERPENT Project, 2009. Scientific and Environmental ROV Partnership using Existing Industrial Technology (SERPENT): Media Database Archive. In: Jones, D.O.B., Gates, A.R., Curry, R.A., Thomson, M., Pile, A., Benfield, M. (Eds.) ⟨http://archive.serpentproject.com⟩.

Shumchenia, E.J., King, J.W., 2010. Comparison of methods for integrating biological and physical data for marine habitat mapping and classification. Cont. Shelf Res. 30 (16), 1717–1729.

Strong, J.A., 2015. Habitat area as an indicator of Good Environmental Status under the Marine Strategy Framework Directive: the identification of suitable habitat mapping methods with recommendations on best-practice for the reduction of uncertainty. In: Defra Project Code: ME5318 (Ed.).

Teixidó, N., Garrabou, J., Arntz, W.E., 2002. Spatial pattern quantification of Antarctic benthic communities using landscape indices. Mar. Ecol. Prog. Ser. 242, 1–14.

ter Braak, C.J.F., 1994. Canonical community ordination. Part I: basic theory and linear methods. Ecoscience 1, 127–140.

Turner, M.G., Gardner, R.H., 1991. Quantitative Methods in Landscape Ecology: the Analysis and Interpretation of Landscape Heterogeneity. Springer, USA.

Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. Evolution 62 (11), 2868–2883.

Wilson, J.B., 1979. 'Patch' development of the deep-water coral *Lophelia pertusa* on rockall bank. J. Mar. Biol. Assoc. U. K. 59 (1), 165–177.

WoRDSS, 2016. World Register of Deep-Sea species. Glover, A.G., Higgs, N., Horton, T. (Eds.) ⟨http://www.marinespecies.org/deepsea/photogallery.php⟩.

Wynn, R.B., Bett, B.J., Evans, A.J., Griffiths, G., Huvenne, V.A.I., Jones, A.R., Palmer, M.R., Dove, D., Howe, J.A., Boyd, T.J., MAREMAP Partners, 2012. Investigating the feasibility of utilizing AUV and glider technology for mapping and monitoring of the UK MPA network. In: MB0118., D.p. (Ed.). National Oceanography Centre, Southampton, pp. 244.