

Interaction between classification detail and prediction of community types: implications for predictive modelling of benthic biotopes

Genoveva Gonzalez-Mirelis*, Per Bergström, Mats Lindegarth

Department of Marine Ecology, Tjärnö, University of Gothenburg, 45296 Strömstad, Sweden

ABSTRACT: Ecological classifications lie, explicitly or implicitly, at the foundation of biotic maps, which are in turn an essential tool in ecology and conservation. Techniques are being developed that make efficient use of scarce biological data to create those maps, such as those based on modelling, or predictive mapping. However, little attention has been paid to just how sensitive models are to the pattern they are designed to model. We have examined the predictive accuracy of classification models (Random Forests) relative to the classification detail of community types. The data are from a recently established Marine National Park off the west coast of Sweden where a map of benthic communities has been commissioned. A total of 447 georeferenced, underwater video sequences constituted the sample base for this study. Samples were classified according to increasing values of a similarity threshold, based on faunal composition, in a hierarchical cluster analysis framework. A random forest was fitted at each level of classification detail to predict the class membership of an independent set of samples, based on environmental (terrain and substrate-related) variables. Predictive accuracy was high (multiclass area under the curve, M-AUC = 0.79 to 0.81) across intermediate levels of classification detail (similarity cut off, 40 to 50%). Predictive accuracy was moderate (M-AUC = 0.73 to 0.78) at both ends of the spectrum of classificatory resolutions tested (cut offs: 30%, i.e. very coarse; 60%, very fine). These findings have ramifications in the 'classify first, then predict' approach to predictive mapping at the community level, as they show that calibrating classes before modelling them greatly enhances model reliability.

KEY WORDS: Classification · Benthic communities · Predictive mapping

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Maps showing the spatial distribution of biotic classes (e.g. vegetation types) serve many purposes, including developing hypotheses to explain patterns of biological heterogeneity and supporting spatial planning. Accordingly, these maps are sought after by scientists, decision makers, and conservation practitioners alike. A widespread approach to (predictive) mapping at the community level is: 'classify first, then predict' (adapted from Ferrier & Guisan 2006). In this approach a classification of the biota, i.e. fauna, vegetation, or some predefined set of taxa, is generated up front, based on the degree of resemblance between sampled locations. Subsequently, the relationships

between the labels under this classification and measured environmental covariates, as inducted by a correlative model, are used to classify (i.e. assign labels to) locations for which only the environmental variables are known. By a natural extension of the logic, when full-coverage, high-resolution environmental data layers are available, a prediction can be made for every pixel across a domain of interest and maps can thus be generated on the basis of continuous surfaces of predicted values (e.g. Brzeziecki et al. 1993, Zimmermann & Kienast 1999, Ferrier et al. 2002, Miller & Franklin 2002, Remm & Remm 2009, Riordan & Rundel 2009, Grossmann et al. 2010). The technique is most efficient when the proportion of observed biota is far lower than that unobserved for a given study region. Conse-

*Email: gegm@marecol.gu.se

quently, this approach is rapidly gaining acceptance in remote areas, including in seabed environments (e.g. Holmes et al. 2008, Buhl-Mortensen et al. 2009a), where the cost of direct sampling is often prohibitive and the need for completed biotic maps continues. Clearly, a bottleneck in this process is the development of a comprehensive classification system of regional biodiversity sufficiently unambiguous to be readily incorporated into a model as the target pattern.

As in many other branches of ecology, the development of mapping techniques for benthic environments has lagged behind that for their terrestrial counterparts. Varying degrees of effort and accompanying computational investment have been put into devising classification schemes of benthic communities for mapping purposes, ranging from expert opinion to the use of numerical techniques, or combinations thereof (e.g. Connor et al. 2004, Valentine et al. 2005, Buhl-Mortensen et al. 2009b). Regrettably, arbitrary choices are almost invariably involved and resulting classifications are plagued with subjectivity and may exhibit suboptimal performance (Snelder et al. 2010).

By comparison, predictive mapping of individual species may seem like an appealing alternative. Indeed, distribution modelling of single species has to date been the most common variant of spatial modelling practiced in ecology. Species distribution models, along with their resulting habitat suitability maps, have been used to resolve questions in a wide range of arenas from conservation biology (e.g. Meynard et al. 2009, Müller et al. 2009, Murray et al. 2009) to fundamental sciences such as biogeography (e.g. Hopkins & Burr 2009, Meissner & Darr 2009, Remm & Remm 2009, Titeux et al. 2009), phylogeography (e.g. Richards et al. 2007) and ecology (e.g. Berger & Hildenbrandt 2000, Moore et al. 2009, Vaz et al. 2008, Kearney & Porter 2009, Sundblad et al. 2009, Chatfield et al. 2010, Pleydell & Chretien 2010). The underpinning ecological and evolutionary theories and assumptions have been the major driving force in ecological studies (Guisan & Thuiller 2005).

Notwithstanding the numerous insights that have stemmed from single-species distribution models, a number of pitfalls severely limit their applicability. Relying heavily on niche theory, they fail to establish a distinction between niche and distribution (Pulliam 2000). Also, species interactions, whether intraspecific (but see Berger & Hildenbrandt 2000), or interspecific, are rarely considered. Data requirements of species distribution models pose many exigencies as well, which are not always satisfactorily resolved. These caveats can be easily circumvented by moving from single species to suites of co-occurring species (Zimmermann & Kienast 1999) (i.e. assemblages, communities). The benefits of modelling biological data at the community level include aspects such as (1) increased power to de-

tect shared patterns of response to environmental variables, including species that are rarely recorded or for which only biased field observations exist, and (2) enhanced capacity to synthesize complex data into a form more readily interpretable (Ferrier & Guisan 2006, Riordan & Rundel 2009). Moreover, spatial predictions of community-level models (biotic maps) can be regarded as ecological regionalizations of the landscape, the applications of which include (1) identifying sites likely to respond similarly to human activity or management actions, (2) stratifying sites for survey or monitoring programs and (3) highlighting sites for conservation priority (Snelder et al. 2010).

For the approach to be truly useful, however, it is paramount that these biotic classes, whether they are conceptualized as assemblages, communities, archetypes or any such entity, are indeed influenced by the environmental variables at the analyst's disposal (provided that the spatial scale of both the response and explanatory variables is in conformity). In other words, variation in community composition (at the level of complexity determined by the classification in place) must respond to the environmental gradients that are included in the model. If classes occur throughout the range of habitat gradients, no relationships will be detected. Similarly, if classes are too narrowly defined and linked to unique sets of habitat conditions, or chance events, generalizations may not be possible. Communities derived from classification schemes at intermediate levels of complexity may be more informative, and therefore predictable, than those derived from schemes of either very high or very low complexity. To our knowledge, the interaction between the level of detail, i.e. the resolution of a classification, and the predictive accuracy of (spatially explicit) models has never been formally investigated (but see McBreen et al. 2008, Snelder et al. 2010).

To advance mapping science in the seabed, we explore the effects of varying the classificatory resolution of epibenthic megafaunal communities and its predictability by a suite of environmental variables related to substrate, depth, and topography of the seabed. Additionally, in the present study we (1) advance work toward a classification scheme of epibenthic faunal communities that can be equated to Connor et al.'s (2004) biotopes used as a framework for continuing investigations, and (2) explore ecological relationships and the value of a suite of environmental attributes as potential predictors of biological communities.

GENERAL APPROACH

The approach taken in the present study involved comparing the ability of a model to predict class membership of benthic assemblages under various levels of

classification detail. We suggest that predictive models of community type built around environmental descriptors will have maximal explanatory power when predicting entities that show a collective pattern of response to the environmental gradients considered. Furthermore, these entities will be confined to a limited region of the spectrum of classification resolutions (spanning from very detailed to very coarse).

We investigated this from different standpoints. On one hand, various hierarchically nested classification schemes were applied to field data to look at the performance of classifications based on a single cut-off value for a measure of resemblance between samples. On the other hand, the predictability of a community type was also investigated relative to that of the same class when defined either more or less loosely. Finally, we also looked at a classification based on significant clusters, derived from a series of similarity profile (SIMPROF) tests. The SIMPROF test, developed by Clarke et al. (2008), relies on the contrast between the observed variation in the pair-wise similarities of a set of samples and the variation expected if there were no clear subclasses within it. The outcome of the test can be translated directly into a decision to stop splitting into subclasses.

The work was conducted in 2 steps. First, numerical classification was used to generate the groupings of samples at a number k of cut-off points for compositional similarity, which was in turn calculated solely on the basis of faunal data. Low resolution classifications are characterized by their simplicity and high heterogeneity, and hence have low minimum within-class similarity. As the tolerance for including samples in the same class is decreased, divisions emerge within classes; thus, classes get broken up into even more numerous, smaller classes. In this hierarchical framework, all high-resolution (high minimum within-class similarity) classes are completely contained by classes in higher order classifications.

Following this step, structural modelling was employed to model the relationships between environmental variables and classes of communities at k levels of resolution. A measure of goodness of prediction was then used to compare across levels. Simply, we used the 'classify first, then predict' approach, in an iterative manner, so that another 'classify, then predict' cycle followed, until all levels of classificatory resolutions were exhausted.

Random Forests (Breiman 2001) is used as the modelling technique. The Random Forest (RF) algorithm has been advocated for the analysis of multivariate, complex data, and is gaining favour among ecologists (Prasad et al. 2006, Cutler et al. 2007). RF is a combination of tree predictors. The trees making up the 'forest' are implemented by recursively partitioning the multi-

dimensional space that is defined by the predictor variables into zones that are as homogeneous as possible in terms of the response (Vayssières et al. 2000). This is done through an exhaustive search algorithm that determines which binary division of the explanatory variables best reduces deviance in the response variable (Lawrence & Wright 2001); it then uses this split to divide the data into 2 subsets, thus creating a pair of nodes. For each node, now containing only a subset of the data, the process is repeated and continues until homogeneous terminal nodes are reached in a hierarchical tree. The result is a dichotomous decision tree that permits the classification of new cases for which environmental conditions are known, as each path down the tree, which is determined by a series of classification rules, specifies the conditions that lead to a most probable class.

To overcome the potential instability of single trees, in a RF each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. It also uses a random selection of features to split each node, which yields error rates that compare favorably with other algorithms (Breiman 2001, 2002).

We fitted a RF at each resolution level (k models) using as a target pattern the labels corresponding to the particular classification scheme (Scheme i) in question as a multinomial response. External validation was used to compute a measure of predictive accuracy, which we used to compare between models (i.e. between classifications). Further details are provided in the next section.

DATA AND ANALYSES

Biotic data. The data are from the Koster fjord and archipelago area (Fig. 1) off the west coast of Sweden, where a map of benthic communities has been commissioned to aid the implementation of the recently designated Kosterhavet Marine National Park. The study was circumscribed to the circalittoral offshore zone, an aphotic environment largely dominated by animals, and with only scattered foliose algae (Connor et al. 2004). Data on epibenthic megafauna were obtained from video footage, which had been recorded by means of an underwater remotely operated vehicle (ROV) (additional technical details can be found in Gonzalez-Mirelis et al. 2009). ROV navigation data were time-synchronized with the video signal, enabling the reconstruction of the ROV's path for each of 52 survey sites (Fig. 1), as well as the formalization of a function linking the video footage to the path. The data used in the present study were retrieved from an existing repository of video material for the area. Not sur-

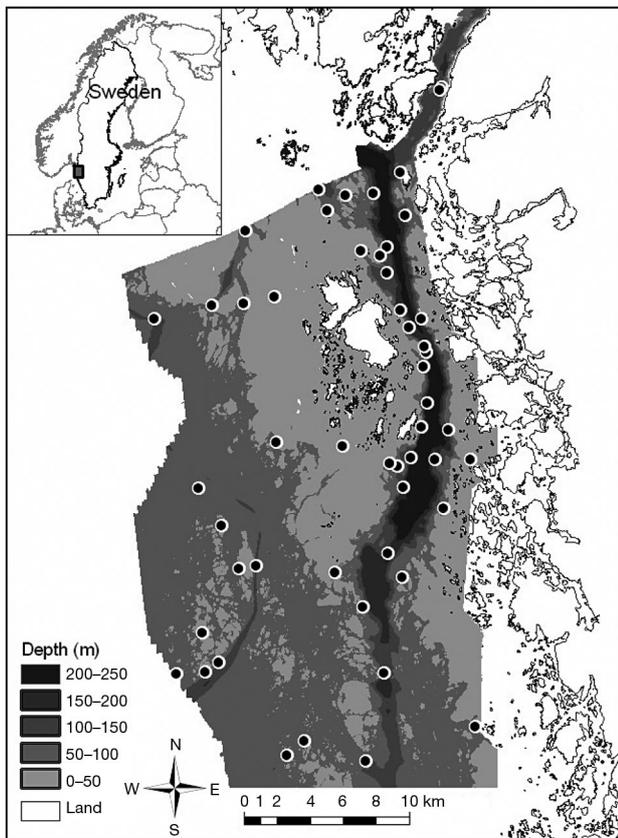


Fig. 1. Study area showing the location of the ROV-surveyed sites (filled circles) and the bathymetry. Inset shows the location of the study area on the west coast of Sweden

prisingly, building a data set suitable for the analyses described here was time-consuming, but in compensation, a large amount of presence/absence data were obtained incurring no extra costs. All ROV surveys had been conducted during the period 2006 to 2008. We used the 30 m isobath as the upper limit for the observations.

In a GIS, a grid lattice was used to clip the survey tracks into sections that could be used as sampling units. Thus, the intersection points between the grid and the survey paths determined the start and end frames of single sampling units. Following recommendations in Gonzalez-Mirelis et al. (2009) based on spatial patterns of epibenthic megafauna, the mesh size (linear scale) chosen for the grid was 15 m. This mesh size defines the spatial scale of analysis, or grain, of this study.

Because the ROV tracks followed disparate surveying conventions, there was variation in the lengths and shapes of the resulting path segments, so additional filtering was exercised. Only grid cells where video duration exceeded 100 s were used to ensure sufficient sampling intensity. Additionally, grid cells were left

out if the track traversed the cell farther than 7 m from its center. This procedure effectively removed sections of video that were liable to identify either of 2 adjacent cells, owing to positional error. In this manner, the uncertainty of the association of the imagery to a particular location in space is rendered insignificant. We refer to these sampling units as cell-samples.

For a grand total of 447 cell-samples, equivalent to approximately 70 h of video recording, faunal data were compiled by identifying all epibenthic megafauna within the relevant clip. Both attached and free-living life-forms were used in this study, but only those pertaining strictly to bottom-dwelling biota were considered. Thus, benthic-pelagic fishes, calcareous sponges, macrophytes and epibiotic fauna (species like *Epitonium* sp. or *Caprellina* sp.) were all excluded from this study. Organisms were identified with as much taxonomic precision as the quality of the imagery permitted. Additionally, the taxon complexes that were used encompassed those taxa that shared similar habitat requirements and were difficult to tell apart (e.g. the 2 species of sponge of the genus *Phakellia*, and the species *Axinella infundibuliformis* were all part of the same complex and were thus considered one variable). During the decision process that led to these settings we considered that as much information as possible should be extracted from the video recording and ensured that only fauna directly associated with the benthic environment were included when using the sampling scale of 15×15 m. Where available, high-definition, still photographic images were used to identify taxa. In all, 178 taxa were identified to species level. A species-by-sites matrix was built, where the rows (sites) corresponded to $n = 447$ cell-samples, and the species (columns) comprised 192 epibenthic megafaunal taxa and taxon complexes.

Environmental descriptors. The environmental (explanatory) data comprised a suite of variables derived from high-resolution multibeam data, including bathymetry and backscatter. The multibeam surveys had been conducted in 2005 at 95 kHz frequency with a Simrad Multibeam EM 1002. We developed a number of proxies for local geophysical processes, alongside others seeking to represent processes operating at larger scales. GIS layers were obtained from the multibeam data sets, and then their values were assessed, by spatial overlay, for the grab of 447 cells for which there were biological data available. Data outside of these cells were not used in this study unless needed for evaluating the neighbourhood of the sampled cells, as was the case in a number of focal variables. Unless otherwise specified, these analyses were carried out in ArcGIS 9.2. The final set of abiotic variables used is described in the following subsections (Table 1).

Depth: Data on depth were obtained directly from the 5 m resolution multibeam bathymetry data. Data were then resampled to the required resolution (15 m) with the mean of all (9) node values encompassed by each grid cell.

Backscatter: Backscatter is the amount of acoustic energy (at 95 kHz for this study) that is scattered back to each beam of the receiver array, referenced to the energy of the transmitted pulse. The amount of backscattered energy, measured in decibels, depends on the interference caused by the interaction of an acoustic wave with a volume of sediment or, in the case of hard rock, the seabed (Gardner et al. 2003). Backscatter intensity is correlated with seafloor texture and therefore indicative of surficial geology (Keeton & Searle 1996). This layer was compiled as a mosaic of 3 separate point feature layers, into which backscatter data had previously been processed, covering the western portion of the study area, the central portion, and the trench areas. The spacing of the nodes at the bottom of the fjord, as well as the inlet known as Säcken (trench data set), was 30 m, so this layer had to be downscaled. A raster was interpolated using the inverse distance weighted (IDW) method. This zone is known to be significantly more homogeneous than the rest of the area so we believe that adding this interpolated layer did not appreciably affect the overall level of detail of the backscatter mosaic. The other 2 layers (West data set and Central data set) were of sufficient resolution (10 m) and they were only resampled (using the mean) to a 15 m resolution.

Slope: Slope was derived from the 15 m bathymetry, using a standard 3×3 running window.

Aspect: Aspect was derived from the 15 m bathymetry, using a standard 3×3 running window.

Topographic complexity: Topographic complexity is a measure of the ruggedness of the seafloor, which is related but not equivalent to relief. It represents the rate of change of the slope, which in turn represents the rate of change in elevation, on a pixel by pixel basis. Topographic complexity was calculated approximately following the method provided by Ardron

(2002). We used the finest bathymetry available (5 m resolution), which did not need be exaggerated, as proposed by Ardron (2002). The layer obtained (slope of the slope) was then reclassified and generalized by calculating density using a search radius of 8 m, which reflects the complexity of the topography of the seafloor at a very small scale. Lastly, this was reclassified into 3 natural-break classes.

Surface area: The surface area refers to the total amount of available surface in the landscape, a measure that is related to seafloor roughness and also to topographic complexity as described above. This descriptor represented the total surface area of the grid cell, in square metres, and was computed by means of the Surface Areas and Ratios from Elevation Grids v. 1.2 ArcView extension (Jenness 2002) in ArcView GIS 3.3. It was calculated by means of the 5 m resolution bathymetry raster and resampled to 15 m resolution by means of bilinear interpolation.

Morphometric features: Morphometric features are the basic landforms, pits, channels, ridges, peaks and passes, as well as planar areas (see Fisher et al. 2004). These were extracted from the 15 m bathymetry in Landserf v. 2.3 (Wood 2009). Initially, 2 scales were tested: 3×3 cells, to represent micro-features, and 25×25 , to represent more local features. The default parameters were used for curvature (0.5) and slope tolerance (2°). The micro-scale features did not seem to have a strong relationship with the biological composition of samples, and only the local features were used. For simplicity, pits were removed by reclassifying them to planar.

Exposure: A proxy for exposure to bottom currents was developed based on aspect, under the assumption that areas of unchanging general orientation would be less disturbed by large scale turbulence and therefore more intensely swept by the current. Classes in this layer represent the extent of the area where aspect remains constant regardless of what the actual aspect is. The 3 classes used went from 1 (low) to 3 (high). Polygons smaller than 500 m^2 were incorporated into the containing polygon.

Table 1. Explanatory variables used to assess benthic community types for modelling. na: not applicable

Variable	Alias	Original resolution	Scale	Type of variable
Depth	depth	5 m	na	Numerical, continuous (range: 30 to 262 m)
Slope	slope	15 m	45 m	Numerical, continuous (range: 0 to 74°)
Aspect	aspect	15 m	45 m	Categorical (levels: N, NE, E, SE, S, SW, W, NW, flat)
Topographic complexity	complex	5 m	15 m	Categorical (levels: low, medium, high)
Morphometric features (local scale)	mfeatloc	15 m	375 m	Categorical (levels: channel, pass, ridge, peak, planar)
Surface area	sarea	5 m	15 m	Numerical, continuous (range: 225 to 909 m^2)
Exposure	exposure	na	$>500 \text{ m}^2$	Categorical (levels: low, medium, high)
Backscatter	bscraw	10 m, 30 m	na	Numerical, continuous (range: 3.2 to -46 dB)

METHODS

Biotic classifications. The required pool of classification schemes was obtained from hierarchical cluster analysis. The clustering procedure was based on pairwise Bray-Curtis similarities for $n = 447$ cell-samples and 192 biotic variables, by means of the group average clustering method. Class memberships were derived at 10 similarity thresholds, increasing from 10 to 90%, in 10 point increments, referred to as Schemes 0 through 9 in decreasing order of resolution so that classes in, for example Scheme 3, contain samples that are at least 70% similar to one another (Fig. 2).

Labels for the clusters (our biotic classes) were composed by assigning characters in alphabetic order, in which a new character was attached at the end of the string for each new resolution level. This way, the number of characters in the string was indicative of the cut-off level, and the characters themselves indicate the classes from which each nominated class had 'descended' (e.g. 'AAA' is the label for the first cluster of the third cut-off level starting from the top [Scheme 7] and it is a subgroup of 'AA', which in turn is a subgroup of 'A'). This labelling system served the purpose of making explicit the relationships within class types across levels of resolution. This feature enabled us to gain further insight into the interaction between classificatory resolution and predictability of

classes within single class types. At this point the biological data (192 variables) were replaced by class memberships according to these 10 classification schemes as well as the 8 environmental variables (see previous section) incorporated as columns.

Out of 447 total, 1 cell-sample constituted a singleton at all levels of similarity. Only 2 organisms were encountered by the ROV at the time of the survey: a sponge, *Suberites* sp., and the crevice brittle star *Ophiopholis aculeata*, which pushed the level of dissimilarity to a point where it did not cluster with any other sample at any cut-off level. This sample was removed from the analysis and $n = 446$ cell-samples remained for modelling.

Predictive modelling. For the modelling phase, a further 89 observations had to be removed owing to missing values of the predictors, which cannot be handled by RF. To compute a measure of model performance based on external validation the data set was split into a training set, comprising two-thirds of the samples, and a testing set, comprising the remaining one-third. The split was quasi-random: some clusters in the high-resolution classifications were so small that they were split informatively to maximize the number of classes represented in both sets. This data split was maintained throughout all models to enable legitimate comparisons. Model performance is defined here simply as predictive accuracy. All RFs were fitted with

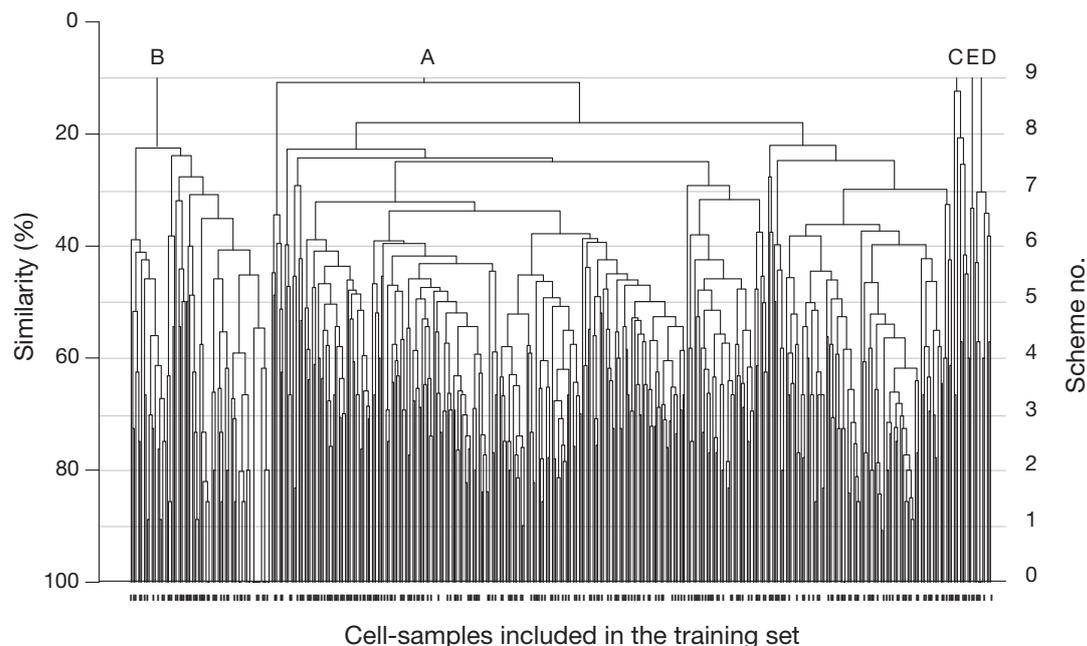


Fig. 2. Analyzed cell-samples based on Bray-Curtis similarity, showing the cut-off levels used to derive the pool of classification schemes. Each instance where a horizontal line (similarity cut-off) crosses a branch of the dendrogram defines one class. The labels of Scheme 9 (A to E) are included at the top of the figure. Cell-samples are not labelled owing to lack of space; however, those that were included in the training set are indicated by means of vertical lines. Details regarding the structure of clusters, particularly at the bottom of the dendrogram, are less noteworthy than the ratio of training to testing samples, which is emphasized by the (few) gaps left between the vertical bars at the bottom. Also, the figure stresses the increase in number of classes as classificatory resolution (similarity) increases

a training set of 235 observations and were tested on a set of 122.

General features of RFs include the following. Each tree in a RF is built with a bootstrap sample of the training set and each node in the tree is in turn based on a randomly chosen subset of the available predictors. By growing each tree to maximum size and selecting the best split only among a random subset at each node, RF maintains prediction strength while inducing diversity among trees (Breiman 2001). Random predictor selection diminishes correlation among trees and keeps the bias low (Prasad et al. 2006). Because a large number of trees are grown, variance is reduced and generalization error (that is, the true error of the population as opposed to the training error only) is limited, which means that no overfitting is possible. This results in overcoming a known problem for single classification trees. Aggregation is done by vote counting.

We fitted a forest of 1000 trees with 4 predictors tried at each split. The selected output was class membership probability, which is derived from the proportion of trees in the forest that voted for a given class.

The modelling pertaining to each classification remains mostly (though not completely) in a 'black box', out of which comes the information central to our analyses: the predictive accuracy of the model when used to classify the cell-samples in the testing set. The metric used for this was the area under the receiver operating characteristic (ROC) curve, or AUC, a measure considered appropriately indicative of general predictiveness (Fawcett 2006). In particular, the multi-class AUC metric (hereafter, M-AUC) developed by Hand & Till (2001) was implemented. This measure was then averaged across all classes to produce an overall measure of model performance, or class predictability, for each classification scheme.

The ROC curve, along with its associated AUC statistic, provides a way to assess the overlap (and hence, the separation) between the 2 probability distributions that can be defined for an object of belonging to a class versus not belonging to it, given a set of measured descriptors for that object. In other words, it is a measure of how well the classification model differentiates between the distributions of the 2 classes. Importantly, the assessment arrived at is not influenced by class distribution (Hand & Till 2001) or misclassification costs (Fawcett 2006). A classification model that does no better than chance is represented by an ROC curve that follows the diagonal of a unit square, made up of all points representing that the probability of an object of, for example Class 1, has an equal probability of being classified as either Class 0 or 1. The better the classification rule, the lower the probability is of a Class 1 object being predicted as Class 0, the incorrect class. Good classification rules yield ROC curves that lie above this

diagonal; the better the classifier, the farther away the curve is from the diagonal. The AUC, then, is a global estimate of the performance of the classification rule, as it is an estimate of the area that lies beneath the ROC curve, such that a classifier that is as good as chance would have an AUC of 0.5 (half of the unit square), and a perfect classifier would have an AUC of 1.

The multiple class extension of this measure (M-AUC) essentially measures the unweighted pairwise comparisons of the probabilities of all classes. In other words, it compares the probability of predicting the true class of an object, not just versus the probability of predicting that the object 'belongs to another class', but also versus the specific probabilities of all other classes. This means that if it is very easy to discriminate between a given class and everything else as a whole, but particularly difficult with respect to one or a few other classes, the classifier will have a rather high AUC value for that class but a somewhat lower M-AUC. In general terms, either AUC or M-AUC for a class can be interpreted as follows: a value of, for example 0.7, means that a randomly drawn sample of that class will be associated with a vector of descriptor values (in our case, environmental ranges) inconsistent with that of a sample of another class 70% of the time (see also Zweig & Campbell 1993).

For completeness (and to shed some light into the 'black box'), variable importance was also included in the output. The importance of the explanatory variables in terms of predicting class membership was assessed by means of the Gini criterion. At every split of a single tree a variable is chosen to form the split, which results in a decrease in the Gini index. The sum of all decreases in the forest due to a given variable and normalized by the number of trees forms this measure of variable importance (Breiman 2002). Lastly, the pair-wise correlations between all predictors were also evaluated.

All analyses were implemented in the statistical package R (R Development Core Team 2008), complemented by the following collection of packages: 'vegan' (Oksanen et al. 2010), 'clustsig' (Whitaker & Christman 2010), 'randomForest' (Liaw & Wiener 2002) and 'ROCR' (Sing et al. 2009). The faunal composition of classes was assessed by means of the SIMPER routine in PRIMER.

RESULTS

Broad classification of benthic communities

The broadest classification, Scheme 9 (within-class, pair-wise similarity > 10%), broke up the data into 5 groups, labelled A to E (Fig. 2). These groups are

described here as ‘families’ of samples, within which classes are related to one another in a top-down, branching framework. Cell-samples in the A family were characterized by various kinds of axinellid sponges, brachiopods *Novocrania anomala* and *Terebratulina retusa*, tube-dwelling polychaetes (such as *Sabella pavonina*, *Placostegus tridentatus*, and other serpulids), cup coral *Caryophyllia smithii* and bryozoans from the Flustridae family, among others; while not a characterizing taxon, this class also encompassed all occurrences of the reef-forming, cold-water coral *Lophelia pertusa*. Typical taxa for the B family included sea feathers *Kophobelemnon stelliferum*, various cerianthids, *Mesothuria intestinalis*, and Norway lobster *Nephrops norvegicus*. Both C and D families were rather small classes. C samples included a different assemblage of sea feathers (*Pennatula phosphorea*, *Funiculina quadrangularis*) and various types of starfish (predominantly *Asterias rubens* and *Crossaster papposus*). Samples classified as D typically featured the great scallop *Pecten maximus* and the edible crab *Cancer pagurus*. Only 2 samples, characterized by the presence of gobiid fishes (*Gobius niger*, *Thorogobius ephippiatus*), made up the last of the broad families (E).

Approximately 80% of the samples belonged to the A family, and 16% to B, with the remainder of the families (C to E) comprising less than 5% of the data. Clearly, the ROV surveys were targeted at a particular kind of habitat (deep, rocky areas), which caused a class imbalance problem about which little can be done. Incidentally, C, D and E families were all engulfed by one single class within the SIMPROF classification scheme. The most detailed classification comprised 434 classes in total, a number that is naturally very close to the total number of samples available from the outset. Out of those, 232 were actually modelled.

Effect of classification on predictive modelling of benthic communities

Mean M-AUC scores ranged from 0.73 to 0.81, peaking at 40 and 50% similarity (Schemes 5 and 6) (Fig. 3). Thus, a pattern of interaction emerges between the classificatory resolution of benthic communities and the predictability of the resulting entities, where as postulated, very broad classifications (in this case, approximately below a 30% similarity threshold) as well as very detailed ones (above 60%) represent a difficult learning task and are not so readily predicted by the classification model (M-AUC = 0.73 to 0.78). Scheme 6 was the single best classification in terms of overall mean M-AUC. Interestingly, the SD of M-AUC of these 2 schemes was also largest, indicating that

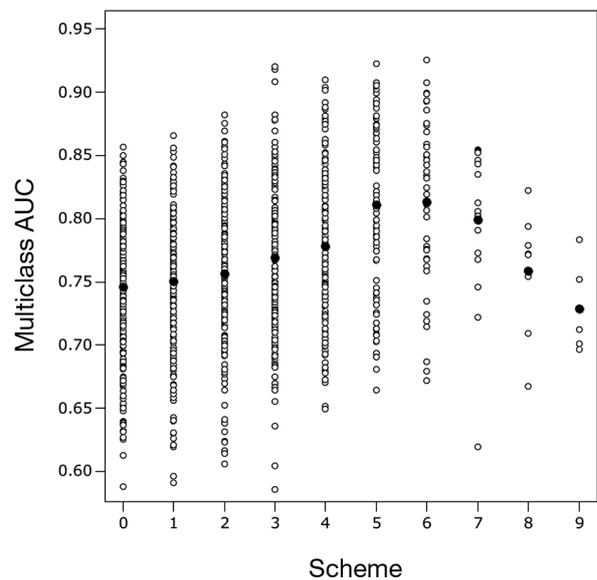


Fig. 3. Multiclass area under the curve (M-AUC) values plotted against classification scheme. Each scheme comprises a number of classes and a single M-AUC value was calculated (open circles) for each class. Mean M-AUC values are represented by solid circles

some classes at these levels were as difficult to simulate as those on the extremes. In all, 26% of the classes modelled had an M-AUC greater than 0.8, which is considered to be excellent (Hosmer & Lemeshow 2000); Scheme 9 was the only classification in which no single class was above this threshold.

The top 2 single-class values (0.92 and 0.93) corresponded to Classes AAACF (Scheme 5), which represents an intermediate depth (30 to 50 m), hard substrate habitat occupied mostly by various serpulids, ascidians (*Ascidia mentula*, *A. virginea*, and *Corella parallelogramma*), as well as the sponge *Pseudosuberites sulphureus*, and AACA (Scheme 6), which was characterized by the presence of the sea anemone *Urticina eques* in association with various sponges, hydroids and serpulids, as well as the shrimp *Atlantopandalus propinquus*. A total of 4 classes in Scheme 3, 4 in Scheme 4, 6 in Scheme 5 and 3 in Scheme 6 scored higher than 0.9 in M-AUC.

By linking the circles in Fig. 3 that constitute one single ‘lineage’ in the dendrogram (Fig. 2), a profile can be built illustrating how predictability varies for a single sample relative to just what other observations are included in the same class. Though there were some slight departures from the general pattern outlined above, all lineages experienced decreasing predictability to some degree with increasing classification detail. Profiles of M-AUC for all cell-samples in the models were explored, but we report here only the most general findings.

We found that some lineages had a very strong peak of M-AUC at some intermediate level, whereas others showed a much more flat profile. To illustrate this we describe 2 cases. The majority of cell-samples in class AAAE are associated with a sharp peak at the level of Scheme 6. At this level, the group is characterized by the presence of *Novocrania anomala*, *Sabella pavonina*, serpulids as a whole, axinellid sponges, *Geodia barretti*, *Placostegus tridentatus* and *Ascidia obliqua*. This indicates that any further subdivisions of this class (i.e. removing observations based on a similarity threshold) will always result in worse predictions. At the next resolution down (Scheme 5) the class is already broken up into as many as 6 subclasses (taxa that seem to be making a difference include the sponge *Mycale lingua*, the serpulid *Hydroides norvegicus*, *Atlantopandalus propinquus* and others). In contrast, Scheme 6's class ABAA (serpulids, *Echinus esculentus*, *Caryophyllia smithii*, flustrids, *Novocrania anomala*, *Ascidia mentula*, *Porania pulvillus*) shows a very gentle drop of AUC at further subdivisions, indicating that in this case, predictability does not deteriorate ostensibly, no matter how much further the class is subdivided.

Occasionally, a secondary peak could be found. This was the case for classes BABA (*Pandalus borealis*, *Nephrops norvegicus* and *Sabella pavonina*, main peak) and BABADABAA (secondary peak), which contained only *Stylatula elegans* and *Pandalus borealis*. Here, predictability increased again at this level even though the class was made up of one single observation.

Mean M-AUC for the SIMPROF classification was 0.80. Based solely on mean M-AUC, this classification ranks third after Schemes 6 and 5.

Effect of classification on variable importance

Depth was the most important variable across all classification schemes, with the notable exception of the broadest one in which raw backscatter was marginally more important than depth (Fig. 4). This means that the 5-class division at the top of the dendrogram is better explained by differences in substrate rather than depth. At the levels of the better-predicted classifications (Schemes 5 and 6), depth also had maximal explanatory power (Gini = 61 and 65). The explanatory power of all variables, and more markedly so in the case of depth, decreased with increasing classification detail.

Depth, slope and backscatter consistently outperformed all other

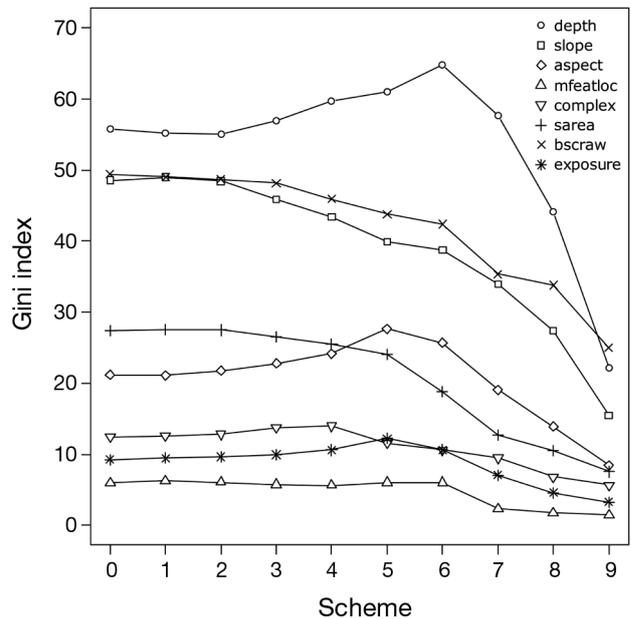


Fig. 4. Variable importance (as Gini index) plotted against classification scheme

predictors. In contrast, the type of local morphological feature, degree of topographic complexity and our measure of exposure to bottom currents were not effective predictors at any level of classification resolution.

The pair-wise correlations between predictor variables showed a strong, yet not surprising, positive correlation between slope and surface area (R = 0.88) and a moderate positive correlation between depth and backscatter (R = 0.66). There was a possibly weak positive relationship between complexity and slope (R = 0.45) (Table 2).

DISCUSSION

In the present study we have examined the discriminability of groups of epibenthic megafaunal taxa at a range of classificatory resolutions (i.e. with varying ecological detail). Even though optimal resolution

Table 2. Pair-wise correlations between predictors used in modelling benthic community types (see Table 1 for definition of predictors)

Predictor	depth	slope	aspect	mfeatloc	complex	sarea	bscraw	exposure
depth	-							
slope	-0.13	-						
aspect	0.16	0.04	-					
mfeatloc	0.10	0.16	-0.07	-				
complex	0.12	0.45	-0.16	0.10	-			
sarea	-0.17	0.88	0.00	0.14	0.42	-		
bscraw	0.66	-0.08	0.07	0.04	0.14	-0.10	-	
exposure	-0.11	0.19	-0.07	0.20	0.05	0.14	-0.26	-

(maximal discriminability) differed between types of communities, overall classifications of biological communities of intermediate resolution were better predicted, and hence easier to discriminate, than either very coarse or very fine classifications, which were in turn only moderately well predicted. Because 85% of all classes modelled were predictable to what is generally considered an acceptable or excellent degree ($M\text{-AUC} > 0.7$), suggests that classification and subsequent prediction of communities constitute an adequate approach to mapping complex assemblages, even though these may not be distinctly organized into discrete communities (*sensu* McCune & Grace's 2002 'abstract community'). However, our results highlight the importance of calibrating ecological classifications before they are used in predictive mapping.

The main challenge of predictive modelling in the context of ecological mapping is to endow models with sufficient predictive power to detect ecologically relevant and informative (benthic) communities. While patterns of species assemblages can be easier to model because of their more uniform response to environmental gradients than those of individual species (Zimmermann & Kienast 1999), the present study points out that how assemblages are defined can have a considerable effect on the detection of distribution patterns.

Discriminative ability was measured as the area under the ROC curve (AUC), with models that were fitted with a training data set and then applied to a test data set. Using this metric as a criterion for selecting the optimal classification level, we concluded that classifications with a minimum within-class Bray-Curtis similarity in the range 40 to 50% performed best. This result is consistent with that of McBreen et al. (2008), who also found an increasing trend in predictability from low-resolution to medium-resolution classifications in a study of benthic assemblages of the southern Irish Sea (although with a very different grain-to-extent ratio). The study by McBreen et al. (2008), however, covered a small range of classificatory resolutions (10 to 30% similarity) and did not involve testing predictions on an independent test data set.

The ability of a model to predict new cases is a measure of the generality of the relationships uncovered by the model and, thus, of the strength of the relationship between the predictors and the levels of the response. We have shown that this relationship is highly dependent on the biological resolution of the target classification, in a way similar to the effect of spatial scale on species–environment relationships (e.g. Thrush et al. 2005). This pattern of interaction between biological resolution and predictability arises through 2 types of mechanisms. The first one relates to the magnitude of the collective response of all the species (or taxa) making up the class to abiotic variables (i.e. the response

curves of all species in the community with respect to the gradients measured). This property depends ultimately on the performance of each individual species relative to all environmental gradients. Secondly, the degree of structure within and between classes also affects class predictability; classes that are highly structured (i.e. homogeneous within and distinct from other classes) are easier to predict than those displaying more variation.

Accordingly, classificatory resolution had a substantial effect on the absolute and relative power of environmental predictors. While the general trend was that variables lost explanatory power as classifications became coarser, the importance of depth in discriminating between communities increased substantially at intermediate resolutions before this final decline. Not only was depth more important than all other descriptors at the levels of Schemes 5 and 6, but also the effect of depth on species composition, in absolute terms, was more evident at this level of classification than any other. At more detailed resolutions (e.g. Schemes 0 to 2) backscatter (substrate) and slope (topography) are equally important, and only slightly less so than depth. However, depth dominates the explanatory power of the model between Schemes 3 and 7. Probably because of this strong effect of depth, schemes at these resolutions are better predicted by the model.

In contrast, at the coarse end of the spectrum, classes responded more strongly to substrate variation than to all other gradients, including depth. Close inspection of the data showed that Class A observations had species typically associated with hard substrates, such as bedrock outcrops and large boulders. In contrast, Class B samples were from areas covered with unconsolidated sediment, mostly fine particles (sand and mud). Both classes spanned the whole range of the depths studied (30 to ~250 m). Classes C, D and E also included observations made predominantly on sediments, but with particle size varying from sandy gravel to gravel and including highly localized areas of shell hash build-up. While the link between substrate type and biological composition is robustly established (Gray 1974, Rhoads 1974, Snelgrove & Butman 1994), little evidence exists to support the possibility of using backscatter data to predict biotic attributes of the benthos directly (but see Brown & Collier 2008). Further work is encouraged to unravel the conditions under which this is possible as enormous benefits could be obtained with this approach, potentially reducing the cost of biological sampling on the seabed. For example, we have shown that the correlation with raw backscatter was highly dependent on the level of detail at which communities are classified in this type of (circalittoral) environment.

This complementary effect of depth- and substrate-related variables at different classificatory resolutions (where one variable explains most of the variation observed and the other explains the remainder of it) is not fully consistent with other studies. For example, Bergen et al. (2001) found a pattern where the coarsest divisions of a dendrogram of benthic communities in southern California were strongly associated with depth, while the finest divisions were better explained by sediment and not by depth. This appears to be contrary to our own results, but there are 2 critical differences. The depth range studied by Bergen et al. (2001) included shallow (infralittoral) sites; also, their focus was on infauna, rather than epifauna. The question of which environmental factors drive the biological composition of seabed-dwelling communities at multiple levels of biological resolution remains therefore largely unanswered, though it seems clear that both depth and attributes of the surficial geology of the seabed play a crucial role, including affecting one another.

Classification and description of benthic communities, biocenoses and biotopes have a long history in Scandinavian waters (e.g. see Olenin & Ducrottoy 2006 for a review of concepts; see Petersen 1913, Molander 1928, Thorson 1957 for early Scandinavian examples). In recent years, the need for broad-scale mapping of European shelf areas driven by developments in national and international policies, such as the Habitats Directive and the Marine Strategy Framework Directive (e.g. Council directive 92/43/EEC, Directive 2008/56/EC), has resulted in remarkably comprehensive schemes for classification of benthic biotopes such as the 'Marine Habitat Classification' for the British Isles (Connor et al. 2004, www.jncc.gov.uk) and the

pan-European EUNIS system (<http://eunis.eea.europa.eu/>). The present study has shown that hierarchical classifications such as those may be hindered by differences in the validity of their various levels, and their usefulness needs to be evaluated instead by the flexibility they afford to managers. In light of this, methods need to be devised that relax the need for subjective criteria when classifying biota (Boesch 1973, Snelder et al. 2010) and, furthermore, account for the variability of the factors driving the structure of the target pattern (Snelder et al. 2010).

Defining biotopes is not a straightforward task. As a general guideline, the observed differences in M-AUC profiles (Figs. 4 & 5) provide evidence against the adequacy of classifications based on a single similarity threshold, in conformity with the SIMPROF routine (Clarke et al. 2008), as well as non-hierarchical methods of classification. Furthermore, the large spread of M-AUC values within schemes indicates that highly structured classes can exist together with much more heterogeneous ones within levels of compositional similarity, and there is no biological reason to presume otherwise. Our analysis can provide a basis to amalgamate hierarchical classification schemes: by considering the varying predictability of observations that are grouped differently, one can infer the cut-off level at which that particular type of community is optimally defined.

The SIMPROF method is very useful for collapsing sets of observations into single classes. The method bears an obvious relationship with our own analyses, but it neglects patterns of response to environmental factors. The examples above (see 'Results') also serve to illustrate this point. Both the sharply peaking profile of AAAE and the flat-topped one of ABAA were 2 of the

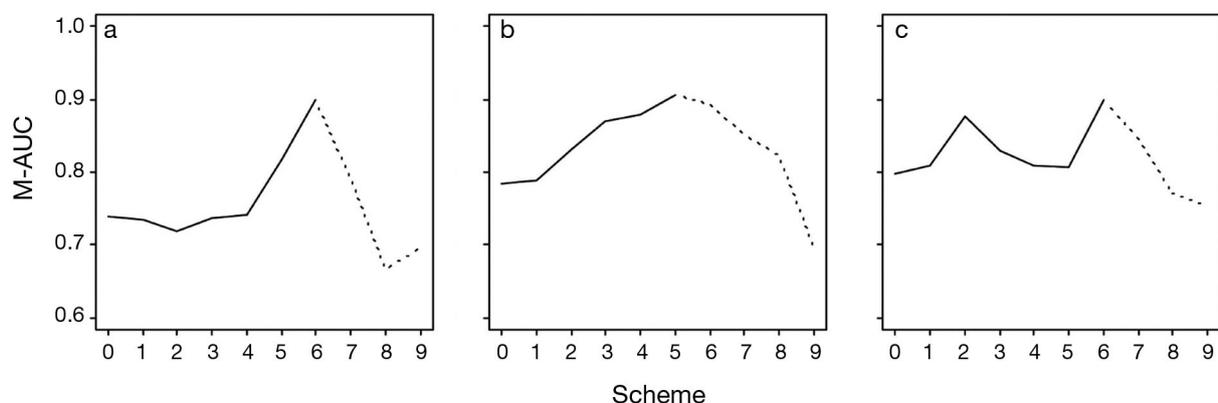


Fig. 5. Multiclass area under the curve (M-AUC) values plotted against classification scheme for 3 separate cell-samples. Note that as one moves to the right in each panel, the relevant sample is incorporated into its corresponding class for that Scheme. (a) A peak in M-AUC is found when the sample is classified at the level of Scheme 6 ($n = 21$) in Class AAAE. (b) M-AUC is maximal when the sample is classified at the level of Scheme 6 ($n = 10$) in Class ABAA, but the decline is much more gentle when incorporated into increasingly smaller classes. (c) Two peaks of M-AUC can be observed for this soft-sediment sample: one at Scheme 6 (BABA, $n = 16$) and Scheme 2 (BABADABAA, $n = 1$)

clusters defined by SIMPROF as significant in our data set. The SIMPROF analysis indicates that further splitting is unwarranted on the basis of faunal composition, but it fails to emphasize the point that this is because the former is so heterogeneous that interfering patterns of response appear, and the latter is so homogeneous that the patterns remain unchanged all the way through. Nevertheless, it is clear that the desired level of ecological detail varies among different mapping and management situations, and aside from discriminative ability, aspects of classificatory complexity are also important to consider in each mapping situation.

There is a tension between the continuous nature of the variation of species performance in response to complex environmental (including temporal) gradients (Boesch 1973) and the discrete nature of (conventional) thematic maps, which is not easily resolved. Despite the difficulties and uncertainty introduced by imperfect classification systems, our results suggest that at some level of resolution these classes are predictable and informative. Thus, such classifications can impose some order onto the otherwise considerable variation in time and space displayed by living organisms, even if we accept that, no matter how fine a classification is used, there will always remain some degree of variation within the classes (Levin 1992). In a marine protected area, adequate knowledge of diversity patterns is important for the development of management plans that aim at their conservation, and repeated or even exhaustive surveying may not be possible. In the process of finding new ways to explore, display and visualize continuous phenomena more readily (e.g. generalized dissimilarity modelling; Ferrier & Guisan 2006), optimizing classificatory procedures should be a top priority for applied ecologists.

Acknowledgements. This research received funding from the European Community's Seventh Framework Programme (FP/2007-2013) under grant agreement no. 217246 made with the joint Baltic Sea research and development programme BONUS, and from the Swedish Environmental Protection Agency from contract 08/391 PREHAB and FORMAS from contract 217-2006-357. The ROV surveys were conducted by T. Lundälv and L. Jonsson, and their willingness to let third parties access and use the data was vital for the completion of this study.

LITERATURE CITED

- Ardron J (2002) A GIS recipe for determining benthic complexity: an indicator of species richness. In: Breman J (ed) Marine geography: GIS for the oceans and seas. ESRI Press, Redlands, CA, p 169–175
- Bergen M, Weisberg SB, Smith RW, Cadien DB and others (2001) Relationship between depth, sediment, latitude, and the structure of benthic infaunal assemblages on the mainland shelf of Southern California. *Mar Biol* 138: 637–647
- Berger U, Hildenbrandt H (2000) A new approach to spatially explicit modelling of forest dynamics: spacing, ageing and neighbourhood competition of mangrove trees. *Ecol Model* 132:287–302
- Boesch DF (1973) Classification and community structure of macrobenthos in the Hampton Roads area, Virginia. *Mar Biol* 21:226–244
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Breiman L (2002) Manual on setting up, using, and understanding random forests, v. 3.1. http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf
- Brown CJ, Collier SJ (2008) Mapping benthic habitat in regions of gradational substrata: an automated approach utilising geophysical, geological, and biological relationships. *Estuar Coast Shelf Sci* 78:203–214
- Brzeziecki B, Kienast F, Wildi O (1993) A simulated map of the potential natural forest vegetation of Switzerland. *J Veg Sci* 4:499–508
- Buhl-Mortensen P, Dolan M, Buhl-Mortensen L (2009a) Prediction of benthic biotopes on a Norwegian offshore bank using a combination of multivariate analysis and GIS classification. *ICES J Mar Sci* 66:2026–2032
- Buhl-Mortensen P, Buhl-Mortensen L, Dolan M, Dannheim J, Kröger K (2009b) Megafaunal diversity associated with marine landscapes of northern Norway: a preliminary assessment. *Norw J Geol* 89:163–171
- Chatfield BS, van Niel KP, Kendrick GA, Harvey ES (2010) Combining environmental gradients to explain and predict the structure of demersal fish distributions. *J Biogeogr* 37:593–605
- Clarke KR, Somerfield PJ, Gorley RN (2008) Testing of null hypotheses in exploratory community analyses: similarity profiles and biota–environment linkage. *J Exp Mar Biol Ecol* 366:56–69
- Connor DW, Allen J, Golding N, Howell K, Lieberknecht L, Northen KO, Reker J (2004) The marine habitat classification for Britain and Ireland, version 04.05. Joint Nature Conservation Committee, Peterborough. www.jncc.defra.gov.uk/MarineHabitatClassification/
- Cutler R, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. *Ecology* 88:2783–2792
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27:861–874
- Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. *J Appl Ecol* 43:393–404
- Ferrier S, Drielsma M, Manion G, Watson G (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in north-east New South Wales. II. Community-level modelling. *Biodivers Conserv* 11: 2309–2338
- Fisher P, Wood J, Cheng T (2004) Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Trans Inst Br Geogr* 29:106–128
- Gardner JV, Dartnella P, Mayer LA, Clarke JEH (2003) Geomorphology, acoustic backscatter, and processes in Santa Monica Bay from multibeam mapping. *Mar Environ Res* 56:15–46
- Gonzalez-Mirelis G, Bergström P, Lundälv T, Jonsson L, Lindgarth M (2009) Mapping the benthos: spatial patterns of seabed-dwelling megafauna in a Swedish Fjord, as derived from opportunistic video data. *Mar Biodivers* 39:291–302
- Gray JS (1974) Animal–sediment relationships. *Oceanogr Mar Biol Annu Rev* 12:223–261
- Grossmann E, Ohmann J, Kagan J, May H, Gregory M (2010) Mapping ecological systems with a Random Forest model:

- tradeoffs between errors and bias. US Geol Surv, GAP Prog Rep, GAP Anal Bull 17:16–22
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 8:993–1009
- Hand DJ, Till RJ (2001) A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach Learn* 45:171–186
- Holmes KW, van Niel KP, Radford B, Kendrick GA, Grove SL (2008) Modelling distribution of marine benthos from hydroacoustics and underwater video. *Cont Shelf Res* 28:1800–1810
- Hopkins RL II, Burr BM (2009) Modeling freshwater fish distributions using multiscale landscape data: a case study of six narrow range endemics. *Ecol Model* 220:2024–2034
- Hosmer D, Lemeshow S (2000) Applied logistic regression, 2nd edn. Wiley-Interscience, New York, NY
- Jenness J (2002) Surface areas and ratios from elevation grid (surfgrids.avx) extension for ArcView 3.x, v. 1.2. Jenness Enterprises. www.jennessent.com/arcview/surface_areas.htm
- Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol Lett* 12:334–350
- Keeton J, Searle R (1996) Analysis of Simrad EM12 multibeam bathymetry and acoustic backscatter data for seafloor mapping, exemplified at the Mid-Atlantic Ridge at 45°N. *Mar Geophys Res* 18:663–688
- Lawrence RL, Wright A (2001) Rule-based classification systems using classification and regression tree (CART) analysis. *Photogramm Eng Remote Sens* 67:1137–1142
- Levin SA (1992) The problem of pattern and scale in ecology: the Robert H. McArthur Award Lecture. *Ecology* 73:943–1967
- Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2:18–22
- McBreen F, Wilson JG, Mackie ASY, Nic Aonghusa C (2008) Seabed mapping in the southern Irish Sea: predicting benthic biological communities based on sediment characteristics. *Hydrobiologia* 606:93–103
- McCune B, Grace JB (2002) Analysis of ecological communities. MjM Software Design, Gleneden Beach, OR
- Meissner K, Darr A (2009) Distribution of *Magelona* species (Polychaeta: Magelonidae) in the German Bight (North Sea): a modeling approach. *Zoosymposia* 2:567–586
- Meynard CN, Howell CA, Quinn JF (2009) Comparing alternative systematic conservation planning strategies against a politically driven conservation plan. *Biodivers Conserv* 18:3061–3083
- Miller J, Franklin J (2002) Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecol Model* 157:227–247
- Molander AR (1928) Animal communities on soft bottom areas in the Gullmar Fjord. *Kristineberg Zool Stn* 1877–1927 No. 2:1–90
- Moore CH, Harvey ES, van Niel KP (2009) Spatial prediction of demersal fish distributions: enhancing our understanding of species–environment relationships. *ICES J Mar Sci* 66:2068–2075
- Müller D, Schröder B, Müller J (2009) Modelling habitat selection of the cryptic hazel grouse *Bonasa bonasia* in a montane forest. *J Ornithol* 150:717–732
- Murray JV, Goldizen AW, O'Leary RA, McAlpine CA, Possingham HP, Choy SL (2009) How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. *J Appl Ecol* 46:842–851
- Oksanen J, Blanchet FG, Kindt R, Legendre P and others (2010) Vegan: community ecology package, R package version 1.17-2. R Foundation for Statistical Computing, Vienna. Available at <http://cran.r-project.org/package=vegan>
- Olenin S, Ducrottoy JP (2006) The concept of biotope in marine ecology and coastal management. *Mar Pollut Bull* 53:20–29
- Petersen CGJ (1913) Valuation of the sea. 2. The animal communities of the sea bottom and their importance for marine zoogeography. *Rep Dan Biol Ann Rev* 16:229–311
- Pleydell DRJ, Chretien S (2010) Mixtures of GAMs for habitat suitability analysis with overdispersed presence/absence data. *Comput Stat Data Anal* 54:1405–1418
- Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9:181–199
- Pulliam HR (2000) On the relationship between niche and distribution. *Ecol Lett* 3:349–361
- R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Remm K, Remm L (2009) Similarity-based large-scale distribution mapping of orchids. *Biodivers Conserv* 18:1629–1647
- Rhoads DC (1974) Organism–sediment relations on the muddy seafloor. *Oceanogr Mar Biol Annu Rev* 12:263–300
- Richards CL, Carstens BC, Knowles LL (2007) Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *J Biogeogr* 34:1833–1845
- Riordan EC, Rundel PW (2009) Modelling the distribution of a threatened habitat: the California sage scrub. *J Biogeogr* 36:2176–2188
- Sing T, Sander O, Beerenwinkel N, Lengauer T (2009) ROCR: visualizing the performance of scoring classifiers, R package version 1.0-4. R Foundation for Statistical Computing, Vienna. Available at <http://cran.r-project.org/package=ROCR>
- Snelder T, Lehmann A, Lamouroux N, Leathwick J, Allenbach K (2010) Effect of classification procedure on the performance of numerically defined ecological regions. *Environ Manag* 45:939–952
- Snelgrove PVR, Butman CA (1994) Animal–sediment relationships revisited: cause versus effects. *Oceanogr Mar Biol Annu Rev* 32:111–177
- Sundblad G, Harma M, Lappalainen A, Urho L (2009) Transferability of predictive fish distribution models in two coastal systems. *Estuar Coast Shelf Sci* 83:90–96
- Thorson G (1957) Bottom communities (sublittoral or shallow shelf). *Geol Soc Am Memoir* 67:461–534
- Thrus SF, Hewitt JE, Herman PMJ, Ysebaert T (2005) Multi-scale analysis of species–environment relationships. *Mar Ecol Prog Ser* 302:13–26
- Titeux N, Maes D, Marmion M, Luoto M, Heikkinen RK (2009) Inclusion of soil data improves the performance of bioclimatic envelope models for insect species distributions in temperate Europe. *J Biogeogr* 36:1459–1473
- Valentine PC, Todd BJ, Kostylev VE (2005) Classification of marine sublittoral habitats, with application to the North-eastern North America region. *Am Fish Soc Symp* 41:183–200
- Vayssières MP, Plant RE, Allen-Diaz BH (2000) Classification trees: an alternative non-parametric approach for predicting species distributions. *J Veg Sci* 11:679–694

- Vaz S, Martin CS, Eastwood PD, Ernande B, Carpentier A, Meaden GJ, Coppin F (2008) Modelling species distributions using regression quantiles. *J Appl Ecol* 45:204–217
- Whitaker D, Christman M (2010) Clustsig: significant cluster analysis, R package. R Foundation for Statistical Computing, Vienna. Available at <http://cran.r-project.org/package=clushsig>
- Wood J (2009). The LandSerf manual, version 1.0, for LandSerf 2.3.1. www.soi.city.ac.uk/~jwo/landserf/landserf230/doc/landserfManual.pdf
- Zimmermann NE, Kienast F (1999) Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *J Veg Sci* 10:469–482
- Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 39:561–577

*Editorial responsibility: Laura Airoidi,
Ravenna, Italy*

*Submitted: December 15, 2010; Accepted: March 14, 2011
Proofs received from author(s): May 30, 2011*