Scale-dependent influence of environmental variables on species distribution: a case study on five coastal benthic species in the Baltic Sea

Antonia Nyström Sandman, Sofia A. Wikström, Mats Blomqvist, Hans Kautsky and Martin Isaeus

A number of environmental variables determine the distribution of marine species. For instance, as the depth increases, the benthic primary productivity decreases due to the light attenuation by sea water and turbidity of the water. Salinity, temperature and oxygen concentrations, as well as wave exposure and tidal fluctuations might set the limit for survival and reproduction, and different species have different substrate requirements (cf. Bertness et al. 2001). As marine surveys are costly, detailed knowledge on the distributions of marine organisms is sparse. Therefore accurate maps of underwater habitats are of major importance for efficient management of marine areas, and the possibility to model the distribution of aquatic species or communities is therefore of great interest. Predicted species distribution data is commonly used for conservation planning (Wilson et al. 2005). Statistical modelling is used to relate biological survey data to environmental variables, thereby providing a basis for predictive mapping of species or communities on a local or regional scale (Ferrier and Guisan 2006). Several studies have successfully used available environmental variables to predict the distribution of marine species (cf. Garza-Pérez et al. 2004, Bekkby et al. 2008, 2009, Florin et al. 2009, Mélédé et al. 2010). However, as most of these studies are quite local (the above-mentioned studies range from 12 to 150 km), this raises a number of questions about how variations in scale affect variable selection and performance of models (Araújo and Guisan 2006).

Scale can be viewed in the terms of extent and resolution (Turner et al. 1989, Wiens 1989, Whittaker et al. 2001, Huston 2002, Ricklefs 2004, Guisan and Thuiller 2005, Austin 2007), where extent refers to the area over which a study is carried out, while resolution is the size of the sampling unit at which the data are recorded. It is well acknowledged that environmental variables vary and influence biota at different scales, but models are often fitted without discussion of the effect of this scale dependency on the results (Elith and Leathwick 2009). What is important at one scale may not be as important at another scale (Turner 1989), and although patterns can be found on a large scale, the processes that create them might act on a much finer scale (Levin 1992, Schneider 2001, Huston 2002). Models built on data from large areas may have weak predictive power locally, but locally fitted relationships...
between species and environmental variables might also become weaker with increasing extent (Osborne and Suárez-Seoane 2002). Ecological problems can thus not automatically be addressed by using locally measured variables over larger areas and longer periods of time (Schneider 2001). An increase in extent is often assumed to be accompanied by a decrease in resolution (Wiens 1989, Pearson and Dawson 2003), although a large area not necessarily implies coarse resolution (Guisan and Thuiller 2005). Use of only fine scale data for small areas and coarse scale data for large areas might not always be appropriate, as local topography or other fine scale heterogeneity might modify the impact of coarse scale variables even over large areas (Luoto and Heikkinen 2008, Austin and Van Niel 2011a, b). It is thus interesting to test species–environment relationships across various extents (Hernandez et al. 2006).

It has been suggested that more attention should be given to the relative weight of predictor variables used in models of potential geographic distribution of species (Araújo and Guisan 2006). As environmental variables can act directly or indirectly on species occurrence, abundance and biomass, the use of different geographic extents might give different answers to the same question (Guisan and Thuiller 2005). Austin (1980) defined three types of environmental gradients or variables; resources, direct and indirect. Resources are directly consumed by the species, whereas direct variables have a physiological impact on the species but are not consumed. Indirect variables can be correlated to one or more direct variables but have no physiological impact on the species themselves. Models based on direct or resource variables will be robust, as there is a causal link between the environmental variable and the distribution of a species (Guisan and Zimmermann 2000, Austin 2002), while the true habitat requirements of a species are not fully captured by indirect variables (Randin et al. 2006). As indirect variables co-vary with direct variables depending on geographic context, they are more likely to be location specific. For instance, in different regions the same topographic position can reflect different combinations of resources and direct variables (Guisan and Zimmermann 2000, Austin 2002, Thuiller et al. 2003). Thus, the correlation between indirect variables and species distribution patterns is likely to change with location. In marine species distribution modelling, the use of indirect variables is common practice, as many of the direct variables are highly correlated or difficult to measure. One example is depth, that is often used in marine habitat modelling as a proxy for several direct variables, such as light availability, temperature, salinity or pressure (Elith and Leathwick 2009). Depth also captures changes in wave action and ice scouring, which causes differences in species abundance. On a local scale, depth most often is an important predictor for many macrophyte species. Due to its indirect nature, it is likely that the relationship between depth and species distribution changes when moving from one geographic region to another, or when extending the study area to include a larger region.

There are several studies from the Baltic Sea on variables that influence species distribution, covering various extents from local to regional. Some examples include Rousi et al. (2011, 4 km), Eriksson and Bergström (2005, 10 km), Kautsky and van der Maarel (1990, 25 km), Kūrikki (1996, 100 km coastline), Krause-Jensen et al. (2007a, 300 km), Snickars et al. (2010, 300 km) and Rinne et al. (2011, 400 km). However, none of these have studied how the influence of environmental variables changes with scale. As the studies focus on different response variables and use different environmental variables, it is difficult to make conclusions about the relative importance of scale dependence of environmental variables. Therefore we tested five environmental variables: salinity, depth, wave exposure, substrate and substrate slope, known from previous studies to be important for species distribution in the Baltic Sea (cf. Waern 1952, Remane and Schlieper 1971, Kautsky and van der Maarel 1990, Eriksson and Bergström 2005, Krause-Jensen et al. 2007b).

We analysed the change in relative importance of environmental variables for species abundance with varying extent, using data on the cover of five common benthic species with different characteristics and habitat requirements, from 1731 sites within the photic zone along the Swedish Baltic Sea coast.

More specifically, we wanted to address the following questions: 1) does performance of species distribution models decrease with geographic extent when indirect variables are the most important on a local scale and 2) are variables measured on a fine scale relatively less important predictors of species distributions compared to coarse-scale variables in a larger geographic context.

Study area

The Baltic Sea is one of the largest brackish-water areas in the world. It stretches from the latitudes 54°N to 66°N, resulting in differences from south to north in annual insolation and ice cover (Ehlin 1981). The Baltic Sea has no tides (Ehlin 1981) and the upper limit of perennial vegetation is set by prolonged low-water periods and, especially in the Bothnian Bay and northern part of the Bothnian Sea, by ice-scouring (Kautsky and Kautsky 1989). The salinity changes from 10–12 psu in the southern Baltic Sea, to 1–2 psu in the northernmost part of the Bothnian Bay (Kullenberg 1981). Often, in archipelago areas, and especially when there is a freshwater outlet, salinity gradients from the outer to inner parts may occur. The reduced salinity influences the species composition, and results in a species-poor flora and fauna compared to fully marine environments. The number of marine species is dramatically reduced when entering the Baltic Sea, and decreases northward. Further north, as salinity decreases, the number of freshwater species that can survive increases (Remane and Schlieper 1971, Hållfors et al. 1981, Kautsky 1995a, Nielsen et al. 1995). The well-defined basin and strong gradients in combination with few species, and thus less biotic interactions to consider, makes the Baltic Sea an interesting model system.

Methods

Sampling

In the analysis 1731 diving transects from national and regional environmental monitoring and surveys were used...
The dataset mainly includes phytobenthic plants and algae. Records of associated animals are less well documented, with the exception of the Baltic Sea blue mussel *Mytilus edulis*. Transect data were obtained from the entire Swedish part of the Baltic Sea coast. The dataset was compiled from over 100 different studies performed by 45 divers between the years 2000 and 2009. The data was mainly collected during August (58% of the samples) and September (20%), although some of the samples were from June (4%), July (12%) and October (5%). The same method was used in all studies (Kautsky 1992, 1995b). In short, diving transects were placed perpendicular to the shoreline, from the shore to the deepest occurrence of macroalgae or plants. The divers swam along the transect line from the deepest point towards the shore. Whenever a change was observed of either the type of substrate, species occurring, or their coverage of the substrate a note was made on distance from shoreline (m), depth (m) (obtained from a calibrated divers depth gauge), substrate (% coverage) and species (% coverage). The observations were made in a 6–10 m wide corridor along the transect line. Each observation was made by looking around in an area of at least 10 m² according to the method description. The transect sections described the depth zones with different species composition or dominance. The cover of species and substrate type was estimated based on a 7-grade percentage scale (1, 5, 10, 25, 50, 75 and 100% cover). The substrate was described as rock, boulders, stones, pebbles, sand and soft substratum. For this study, data were subsampled into points every meter along the transect, assigning the information in each segment of the transect to all points within that segment.

**Species**

For the analysis we selected four common macrophytes and one animal species with different life-history and characteristics, all of great ecological interest in the Baltic Sea: the perennial brown alga *Fucus vesiculosus* (including *F. radicans*), the filamentous, annual green alga *Cladophora glomerata*, the filamentous pseudoannual red alga *Ceramium tenuicorne*, the perennial phanerogam *Potamogeton pectinatus*, and the Baltic blue mussel, the Baltic *Mytilus edulis* complex (here referred to as *Mytilus edulis*). The two *Fucus*-species have similar function and are difficult to distinguish from each other when not growing together. Also, especially before 2005, when they were split into two species (Bergström et al. 2005), *Fucus radicans* was recorded as *Fucus vesiculosus*. They are therefore considered as one taxonomic unit in this study and hereafter referred to as *Fucus vesiculosus*. The species selected for the analysis needed to be common and wide-spread enough to produce models in at least most of the analysis iterations. *Fucus vesiculosus* occurred in 41% of the samples, *Cladophora glomerata* in 21%, *Ceramium tenuicorne* in 38%, *Potamogeton pectinatus* in 17% and *Mytilus edulis* in 62% of the samples.

**Environmental variables**

The environmental variables used in the study were depth, slope and percent cover of substrate derived from the field data, together with modelled wave exposure and salinity. Although interactions between the environmental variables most likely are of importance, only the main factors were included in the analysis in order to ensure comparable results.

The depth and slope for each subsampled point along the transect were calculated from the depth change and length of the interval through linear interpolation. The substrate was grouped into the three substrate classes hard (including both rock and boulders), sand and soft. As the chosen species mainly grow on hard substrates (the algae and the mussels) or on sandy or soft substrates (the phanerogames), the substrate types gravel and stones were not

![Figure 1. Map of the Baltic Sea with the location of the transects used in the analysis. The concentric lines indicate the distance steps used in the analysis (from 25 to 1500 km) from one of the southernmost starting points.](image)
included as environmental variables in the analysis. The GIS layer of salinity used in the analysis was retrieved from a large-scale map of surface salinity with 100 m grid resolution. The salinity values were originally derived from an oceanographic model in 200 m grid (Al-Hamdani and Reker 2007), as mean values for the depth interval 0–5 m and the years 2003–2005, but the layer had been further locally refined in the Hanö Bight, the Stockholm archipelago, and the Gävle Bight. In the refined areas additional work was carried out to include better representations of the inner to outer archipelago salinity gradients. Due to the natural gradient of the Baltic Sea, where the salinity decreases with increased latitude (not including the Gulfs of Riga and Finland), salinity is highly correlated with other factors that influence the species distribution, such as insolation, temperature, length of the ice cover period (cf. Kautsky and Kautsky 1995) as well as with concentration of dissolved organic matter (CDOM) (Kratzer et al. 2003). Those factors therefore not included in the analyses. The wave exposure index was calculated in 25 m resolution by a simplified wave model (Isæus 2004). The model integrates the fetch in angular sectors around focal points by grid-based searches for nearby land, and local, mean wind speed from 16 directions. The mean wind speed was calculated for a 10 yr period (1990–2000), using data from 13 wind stations along the coast.

The modelled salinity used in this study is a rather coarse-grained variable with low resolution. The wave exposure is of intermediate resolution, while the depth and the substrate are fine-grained, high resolution variables. Slope is here measured on the scale of meters, although it acts on the scale of decimetres or even less and is therefore considered a coarse-grained variable although the spatial resolution is high.

Data analyses

To evaluate the relative importance of the chosen environmental variables derived from our models, we calculated the average contribution of each factor to the models as a percentage of the total deviance explained. The explained deviance is a measure of how much of the variation in the response variable that can be explained by the predictor variables included in the model.

In order to avoid the problem of spatially dependent samples within transects, we randomly selected only one point per transect for the analysis dataset. The analysis was made by starting in one of the randomly selected points and, within a specific distance from this point (25, 50, 75, 100, 125, 150, 200, 250, 500, 750, 1000, 1250 and 1500 km radius, Fig. 1), model the % cover of each species based on all the chosen environmental variables (full models). We used generalized additive models, GAM (Hastie and Tibshirani 1986, 1990) in the MGCV-package (Wood 2006, 2008) for R (R Development Core Team), with a binomial distribution and k = 2 for the substrate variables and k = 4 for the other predictors, where k sets the upper limit of the degrees of freedom for the smoothing function in the GAM. To keep n constant, 200 points were randomly selected within each distance. For the smallest distances (25 and 50 km), where there were not enough data to use 200 points, we used 120 points.

To get a good spread of the starting points across the gradients of salinity and wave exposure, we randomly selected one starting point per coastal water-body (SMHI 2003). The water-bodies are administrative units where cartographic representations of coastal and sea areas have been grouped based on depth and distance between islands and land areas. The analysis was then run with starting points in all water-bodies where data was present (221 water-bodies). We repeated the analysis five times, with a new randomly selected analysis dataset, using re-sampling with replacement, and a new selection of 221 starting points from the dataset for each repetition. If there were too few points within a specific distance from a starting point, or if the random selection of 200 (120) points did not include any record of the species, no calculations were performed. Therefore, the number of analysis-runs was not equal for all species and for all distances (Table 1), which is also reflected in the confidence intervals in Fig. 2 and 3. The relative contribution of the environmental variables to each model was then calculated using the amount of variation that each variable could create within the model between its maximum and minimum values, expressed in the scale of the linear predictor according to Lehmann et al. (2003). As we used five random selections from the same dataset, we calculated 99% bootstrap confidence intervals for the mean relative contributions of environmental variables for each distance step using the boot-package (Davison and Hinkley 1999, Canty and Ripley 2009) for R (R Development Core Team).

The highest correlation between any two of the chosen environmental variables was found between the hard and the soft substrate (r = −0.43). Auto-correlation in the residuals was tested using spline correlograms in the ncf-package (Bjornstad 2009) for R (R Development Core Team) for the models of three random starting points at four distances (25, 100, 750 and 1500 km) for all species. There was no evidence of autocorrelation in the model residuals.

Table 1. Number of models included in the analysis for each species for all five random selections, with number of field points used (n) within each radius.

<table>
<thead>
<tr>
<th>Species</th>
<th>Radius (km)</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>100</th>
<th>125</th>
<th>150</th>
<th>200</th>
<th>250</th>
<th>500</th>
<th>750</th>
<th>1000</th>
<th>1250</th>
<th>1500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fucus vesiculosus</td>
<td>n</td>
<td>120</td>
<td>120</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Ceramium tenuicorne</td>
<td>n</td>
<td>110</td>
<td>104</td>
<td>504</td>
<td>495</td>
<td>598</td>
<td>750</td>
<td>834</td>
<td>988</td>
<td>1000</td>
<td>1085</td>
<td>1102</td>
<td>1101</td>
<td>1103</td>
</tr>
<tr>
<td>Cladophora glomerata</td>
<td>n</td>
<td>96</td>
<td>464</td>
<td>487</td>
<td>580</td>
<td>710</td>
<td>802</td>
<td>950</td>
<td>1001</td>
<td>1070</td>
<td>1067</td>
<td>1082</td>
<td>1076</td>
<td>1074</td>
</tr>
<tr>
<td>Potamogeton pectinatus</td>
<td>n</td>
<td>116</td>
<td>516</td>
<td>505</td>
<td>611</td>
<td>767</td>
<td>866</td>
<td>1012</td>
<td>1038</td>
<td>1104</td>
<td>1105</td>
<td>1105</td>
<td>1105</td>
<td>1105</td>
</tr>
<tr>
<td>Mytilus edulis</td>
<td>n</td>
<td>102</td>
<td>499</td>
<td>497</td>
<td>604</td>
<td>743</td>
<td>852</td>
<td>974</td>
<td>1001</td>
<td>1094</td>
<td>1103</td>
<td>1105</td>
<td>1105</td>
<td>1105</td>
</tr>
</tbody>
</table>
Results

For all macrophyte species the average total deviance explained by the models was generally quite high, but decreased with increasing extent (Fig. 2). At the local scale it was highest for *Cladophora glomerata* and *Fucus vesiculosus*, while it was low for *Ceramium tenuicorne* at all scales. For *Mytilus edulis* the average total deviance explained by the models increased with extent.

The average contribution of depth decreased with increasing scale for all species (Fig. 3). However, regardless of extent, depth was the most important environmental factor to explain the distribution of all the investigated macrophyte species, except for *Ceramium tenuicorne*, where depth was the most important factor on the local scale only (Fig. 3b). For *Cladophora glomerata*, the relative contribution of depth was around 50% or more of the explained deviance at all scales, ranging from just below 70% at local and regional scales to around 50% for the entire Baltic Sea (Fig. 3a).

The slope was of minor importance for all investigated macrophytes at all scales, but for the blue mussel *Mytilus edulis* it contributed to almost 20% of the explained variation at the most local scale (25 km) (Fig. 3e). Also, the relative contribution of wave exposure decreased with increasing extent for all species except *Potamogeton pectinatus* (Fig. 3d).

For all species, the modelled salinity increased in average contribution when moving from local to Baltic Sea scale. However, for *Cladophora glomerata* and *Potamogeton pectinatus*, modelled salinity was of low importance at all scales (Fig. 3a, d). For *Mytilus edulis*, the modelled salinity contributed to about 10% or less of the total explained variation at 25 and 50 km, but from 250 to 1500 km radius the contribution of the modelled salinity increased to almost 70% (Fig. 3e).

For the filamentous algae *Ceramium tenuicorne* and *Cladophora glomerata* the contribution of substrate increased with extent, while for the perennial species it was fairly constant. For *Mytilus edulis* the importance of the substrate peaked in contribution at intermediate extents (Fig. 3e).

Discussion

Comparisons of species–environmental relationships at different scales may confound the effects of extent and resolution (Collingham et al. 2000). In this study we hold resolution constant, and thus any effect of an increase in extent will be a result of the incorporation of greater spatial heterogeneity (Wiens 1989).

Effect of extent on model performance

The relation between extent and deviance explained by the models differed between the investigated species. Given the same number of points across extents, we expected the explained deviance of the models to decrease with increasing extent due to the use of indirect variables. For all the tested macrophytes the deviance explained decreased as predicted. However, for the blue mussel *Mytilus edulis* the explained deviance increased with extent. The relatively low performance of the *Mytilus edulis* models on the
smaller extents indicates that this species is poorly predicted by the more local variables included in this study, which agrees with previous studies showing that *Mytilus edulis* is mainly limited by salinity (Tedengren and Kautsky 1986, Westerbom et al. 2002). The increase in model performance can be explained by the niche requirements of the species. The larger the extent, the more of the salinity range is incorporated, while locally fitted models only include a small part of the salinity gradient. This nicely illustrates that if the extent does not include a sufficient part species’ niche range, the species response will be truncated (Austin et al. 1990, Guisan and Thuiller 2005, Austin 2007).

The explained deviance of the *Cladophora glomerata* and *Fucus vesiculosus* models was quite high on the local

Figure 3. (a–c) Average % contribution of the explained deviance (± 99% bootstrap confidence intervals) by the different environmental variables for each species at each distance.
scale, while model performance clearly decreased with increasing extent. This was expected, as both species respond strongly to depth, which is an indirect predictor. As the relationship between response variables and indirect variables are likely to change with region (Sundblad et al. 2009), the effect of indirect variables can be expected to be weaker at larger extents (Austin 1985, 2002). Just as Mytilus edulis, on the scale of the entire Baltic Sea, Fucus vesiculosus is also constrained by salinity, but the increase in explanatory power of salinity with extent does not fully compensate for the decrease in importance of depth.

The model performance was generally low for Ceramium tenuicorne. This is well in accordance with the ecology of the species, since it has a wide distribution in both depth and wave exposure range as well as on different substrates. As widespread species probably have local or regional adaptations, resulting in different characteristics and thus different responses, they can also be expected to be less accurately modelled (Stockwell and Peterson 2002).

Our results show that the relation between extent and predictability is complex, as it depends both on the importance of indirect variables and the geographical relationship between extent and niche ranges. Species with narrow niches are expected to be better predicted, as the extent of the study is more likely to capture the entire environmental range (Hernandez et al. 2006, McPherson et al. 2006). The geographical extent of a ‘narrow niche’ depends on the scale of the main structuring variables. For example, when studying rare or endangered insects, amphibians, reptiles, birds and mammals, Hernandez et al. (2006) found that, as distributional extent of a species increased, model prediction accuracy decreased. However, Schneider and Piatt (1986) found that the correlation between murres and puffins and schooling fish increased with increasing measurement distance.

**Relative importance of different environmental variables**

We found that the relative importance of different environmental variables for species distribution changed with extent of the study, and that responses also differed between species. The difference between the species indicates major differences in distributional patterns in relation to the investigated gradients.

The importance of depth decreased with increasing extent for all the investigated species. The decrease in importance of depth is most likely a result of its indirect character. Depth is used mainly as a proxy for light availability, but the light conditions at a certain depth varies along the Baltic Sea north-south gradient, due to differences in solar angle and concentration of dissolved organic matter (CDOM) in the water column (Ehlin 1981, Kautsky and Kautsky 1995, Kratzer et al. 2003). An alternative explanation would be that the variables measured on a fine scale (meters) become less important predictors at over wider geographical scales. However, Austin and Van Niel (2011a) point out that local topography can be important for species distribution modelling even in studies of very large areas. Furthermore, substrate that is also measured on the scale of meters was of fairly constant importance or increased with extent for all macrophytes. Substrate directly influences benthic species distribution, as algal communities are predominantly found on hard substrates, while on soft or sandy substrates, in shallow to intermediate depth, they are replaced by vascular plants vegetation (cf. Kautsky 1995a, Eriksson and Johansson 2003, Krause-Jensen et al. 2007a, b).

While the importance of depth decreased with increasing extent, it is important to note that for the distribution of macrophytes, depth was very important regardless of scale. For example, at a local to regional scale, depth explained about 70% of the explained deviance of Cladophora glomerata, suggesting that a large proportion of the species’ distribution can be predicted based on depth only. Cladophora glomerata has a quite narrow niche in terms of depth, as its natural distribution mainly is restricted to the upper few meters of the photic zone.

For the two filamentous algae, the relative importance of substrate increased with increasing extent. However, as all variables are included in each model, the amount of variation that one environmental variable can create within the model is relative all the other variables included. Therefore, the increase in importance of substrate for the filamentous algae could be an effect of decreasing importance of other variables, rather than an actual increase, suggesting that other factors are more influential on a local scale.

Both wave exposure and slope are documented to influence distribution of benthic marine and brackish water species, acting through various (both direct and indirect) mechanisms. Wave exposure can hamper the settling of propagules and detach organisms from the substrate, but is also correlated with Secchi depth within the archipelago gradient, and removes sediment from hard substrate. Steep slopes have a direct effect on establishment of macrophyte species, but steeper slopes are also less subject to sedimentation. As sedimentation is an important constraint for macroalgal species distribution (Berger et al. 2003, Eriksson and Johansson 2005), the low influence of slope and wave exposure on the macroalgal species was thus unexpected. Our results suggest that these factors are relatively less important predictors for distribution of benthic species in the Baltic Sea, compared to depth, substratum and salinity, although they may be important for some species at fine spatial scales.

Salinity is considered to be the most important factor for species distribution on the scale of the entire Baltic Sea. The importance of salinity increased with increasing extent for all species. The modelled salinity layer used in this study had a relatively coarse resolution, which may contribute to the weak predictive power of salinity at the finest scale. However, the layer was refined in many of the areas where local influence of salinity could be expected. Also with a salinity layer of fine resolution, the salinity range in most coastal areas of the Baltic Sea is too narrow to be a good predictor on a local scale, except for areas with strong inner to outer archipelago gradients where the salinity may determine the occurrence and distribution of species.
Implications for species distribution modelling

The Baltic Sea is a unique system, but the general idea of variation in predictability and variable selection might be valid for other systems as well. In the Baltic Sea, where the system is characterised by a strong gradient, many species are near their distribution limit. Therefore, species distribution models covering large extents within the system would be expected to perform well, as a large portion of the niche is covered. However, if the species strongly respond to indirect predictor variables, the relationships are likely to become weaker with increasing extent.

This has implications for species distribution modelling also from a management perspective. If data is available on a regional or even Baltic Sea level, should the entire region be modelled together, or should it be divided into sub-regions? In modelling smaller areas, fewer data points will be available and there is a risk that data cover a smaller part of the response curve. A larger area will probably include more data, but instead there will be a risk of weaker response due to indirect variables. In the model selection process, the variables that will be included are the ones that are the most important for the entire dataset. If it is not possible to sample the entire gradient, a minimum requirement for data would be to cover the range of environmental variables where changes in species patterns are expected to occur (Rahbek 2005). The critical range is however species specific.

Depth was the most important factor at all scales for three of the four macrophyte species chosen in this study, although the importance of depth did decrease with increasing scale. This stresses the importance of good bathymetry for predictive modelling. It also shows that it is possible to predict the distribution of at least some species on a large extent, given that the bathymetry layer used for prediction is good enough. At least for the Swedish Baltic Sea coast, the lack of substrate maps with reasonable resolution will also negatively influence any attempts to predict species distribution over large areas.

Conclusions

The relationship between spatial extent of the modeled area and the relative importance of predictor variables is complex and depends on the ecology of individual species. One general finding from this study is that indirect variables are likely to become less important at larger spatial extent, which may affect the possibility to predict species distributions over large spatial extent when relying heavily on indirect variables. However, the study also shows that it is possible to predict the distribution of at least some species in fine resolution on a large extent, given that the available environmental layers are of reasonable quality. The persistent importance of depth across scales underlines the importance of not ignoring fine scale predictor variables even though the geographic extent increases.

Acknowledgements – We thank all persons involved in field sampling, data delivery and quality control of data sets. We also thank Göran Sundblad for valuable comments on the statistical analysis and Anna Nikolopoulou at AquaBiota Water Research for the refined salinity layer, which was produced within the MOPODECO project, funded by the Nordic Council of Ministers. Main financial support for collecting the field data was supplied by the Swedish Environmental Protection Agency and County administration boards along the coast.

References


361


Hastie, T. J. and Tibshirani, R. 1990. Generalized additive models. – Chapman and Hall/CRC.

Hastie, T. J. and Tibshirani, R. 1986. Generalized additive models. – Chapman and Hall/CRC.


