

CHAPTER 6

HABITAT SUITABILITY MODELLING OF COMMON SPECIES

Bea Merckx, Maaike Steyaert, Ann Vanreusel, Magda Vincx, Jan Vanaverbeke. Habitat suitability modelling of common species. Submitted to Journal of Sea Research.

HABITAT SUITABILITY MODELLING OF COMMON SPECIES

ABSTRACT

Habitat suitability models get increasing attention in conservation management. Rare and specialist species with specific habitat requirements are generally easier to model than common and generalist species. Since habitat requirements of common species are less stringent, these species have been less considered for species level conservation. However, recent research emphasises the importance of common species which appear in high densities to the structure, function and service provision of terrestrial, freshwater and marine ecosystems. Moreover, separating optimal from suboptimal regions for these species may be interesting for other purposes, such as fisheries. Since habitat suitability models are generally based on presence/absence or presence-only data, they are not able to model densities or relative abundances of a given species of interest. However, maps giving an indication of species relative abundances or total densities can be interesting tools for decision makers. Therefore, we constructed habitat suitability maps of marine nematode species including information on species densities. To reach this goal, we used two approaches: 1) the relative abundances of the species are considered to be separate observations of the species. Thus, the number of observations increased with increasing relative abundance; 2) a species was only considered to be present if its relative abundance was higher than a certain threshold (i.e. 1%, 5% and 10%). We show that implementing a threshold on the relative abundances results in most cases in better models which are capable of identifying the habitats where species occur in higher relative abundances.

Keywords

Habitat suitability modelling, relative abundances, Maxent, common species, Nematoda, North Sea

INTRODUCTION

Species with specific habitat requirements are generally easier to model than generalist species (Segurado and Araújo, 2004; Evangelista *et al.*, 2008; Merckx *et al.*, 2011). Identifying suitable habitats for endangered species gets a lot of attention in conservation management. However, recent work emphasises the importance of common species, species which appear frequently in the environment, for ecosystems too (Gaston and Fuller, 2008). If these common species appear in high abundances (commonness), relatively small declines in their relative abundances may result in large declines in individuals and biomass and may affect ecosystem functioning and provisioning of services (Gaston and Fuller, 2008) such as reduced productivity and higher vulnerability to invasions in plant communities (Smith and Knapp, 2003; Smith *et al.*, 2004). In marine benthic environments simulations show that ecosystem functioning, such as organic matter decomposition and the regeneration of nutrients vital for primary productivity, may be seriously impaired when abundant and common macrobenthic species disappear (Solan *et al.*, 2004).

One step in taking targeted protection measures is the understanding and prediction of species requirements to their habitat. Habitat suitability models (HSMs), as the name reveals, give an indication of which habitats are suitable for a species and which are less suitable. Traditional HSMs based on presence/absence or presence-only data may result in a too broad range of habitat characteristics for a species, reflecting both optimal and suboptimal habitats for the species under consideration (Hutchinson, 1957). Narrowing down the suitable habitat to optimal regions with potentially high densities of harvestable species may also be of interest to fisheries (Houziaux *et al.*, 2010).

HSMs are built mainly based on presence/absence or presence-only data. However, this huge data reduction results in neglecting the information about the densities of a species in a sample. These densities may differ strongly according to the habitat, even for common species. Indeed, species will not necessarily thrive in all occupied habitats. In this paper, we incorporate the information on relative abundances of the species by adding presences in areas where the species has been found in high relative abundances or by using minimum thresholds on species relative abundances.

MATERIAL AND METHODS

Data

The research area, with a total surface of about 18 000 km² is situated in the Southern Bight of the North Sea, near the Belgian and the Dutch coast (latitude 51°6'2" - 52°59'19" N; longitude 2°14'39" - 4°30'43" E) (Fig. 6.1). For this area full coverage environmental maps are available.

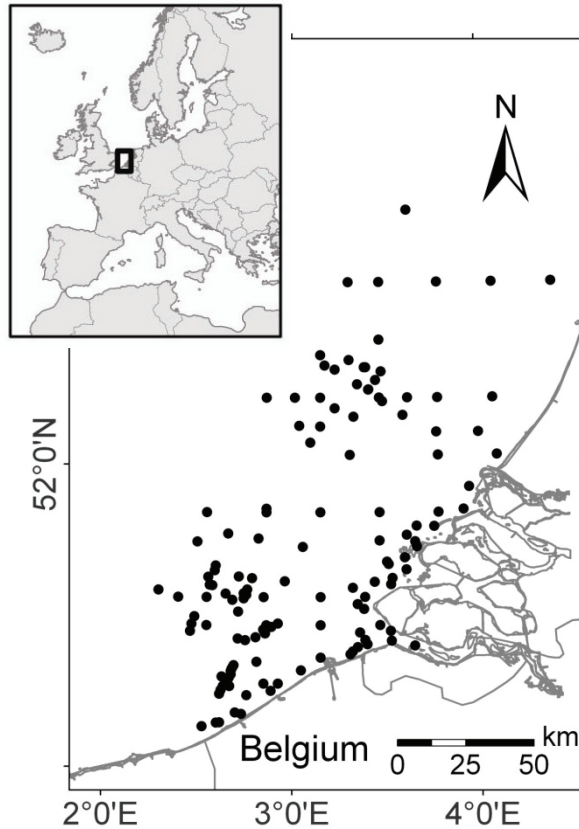


Fig. 6.1. Location of the 140 sampling stations (●).

The performance of the methods applied here, is evaluated by the use of a dataset of free-living marine benthic nematodes from the Southern Bight of the North Sea. Nematodes are usually the dominant taxon within the meiofauna, comprising metazoans passing through a 1 mm mesh sieve but retained on a 38 μ m mesh sieve. These free-living roundworms represent the highest metazoan diversity and densities in many benthic environments (Giere, 2009). The nematode data were retrieved from the MANUELA database. MANUELA is, within the EU Network of Excellence MarBEF, a Research Project focusing on the meiobenthic communities. The MANUELA database contains data of meiobenthic species on a broad European scale (Vandepitte *et al.*, 2009). In this case, the data was restricted to the research area (Fig. 6.1). The total densities of the species in a sample are known for 65% of the stations. Leaving out the data of the 35% remaining stations would seriously reduce the available data. Therefore, the relative abundances, which were known for all the stations, were used. In fact, this means that the potential habitat of a species to reach high relative abundances or dominance in a certain area is modelled. Six nematode species which appear in more than 25% of the sampling stations and which appear at least 5 times in high relative abundances (>10%) were selected from the database (Table 6.1). The six species are found across the entire sampling area. All species, except *Dichromadrea cucullata* and *Onyx perfectus*, are known to appear in high relative abundances near the coastal area (Vincx, 1989a; Vanreusel, 1990), a region characterised by low diversity and low evenness (Merckx *et al.*, 2010), hence with dominance of certain nematode species.

Species	% of the stations where the species is observed
<i>Daptonema tenuispiculum</i> (Ditlevsen, 1918)	27
<i>Dichromadora cucullata</i> Lorenzen, 1973	71
<i>Enoploides spiculohamatus</i> Schulz, 1932	72
<i>Onyx perfectus</i> Cobb, 1891	69
<i>Sabatieria celtica</i> Southern, 1914	63
<i>Sabatieria punctata</i> (Kreis, 1924)	33

Table 6.1. Selected nematode species and the percentage of stations where the species is observed.

The seafloor, the habitat of the nematode species, is not at all homogeneous in this area; it is characterised by sand dunes and a wide range of sediment types, varying from muddy to sandy environments (Lanckneus *et al.*, 2002; Verfaillie *et al.*, 2006). The coastal zone is characterised by a high amount of total suspended matter and chlorophyll *a* in the water column and a high silt-clay fraction in the sediment, especially near the mouth of the Scheldt Estuary and the Eastern side of the Belgian coast (Fig. 6.2) (Fettweis and Van den Eynde, 2003; Eleveld *et al.*, 2004).

The environmental variables were retrieved from maps acquired by remote sensing and maps interpolated from data sampled in the field (Fig.6.2). The first group of maps summarises data on total suspended matter and chlorophyll *a* in the water column (Park *et al.*, 2006). The data is collected by remote sensing by the MERIS spectrometer on board of the Envisat satellite of the ESA. Eighty chlorophyll *a* maps and 90 total suspended matter maps were gathered during the time frame 2003-2005. These maps were reduced to three biologically relevant maps revealing the minimum, maximum and average values. This data reduction technique is often applied in ecological modelling (Loiselle *et al.*, 2008; Cunningham *et al.*, 2009; Echarri *et al.*, 2009). Satellite data are restricted to the water column but are of relevance for seafloor inhabiting organisms as sedimentation of chlorophyll *a* and total suspended matter enrich the bottom organic matter (Druon *et al.*, 2004; Franco *et al.*, 2008). The second group contains maps derived from point sampling at sea. It comprises data on sediment characteristics, such as median grain size and the silt-clay fraction, and bathymetry. These maps were supplied by the Renard Centre of Marine Geology, Ghent University (Verfaillie *et al.*, 2006) and TNO Built Environment and Geosciences-Geological Survey of the Netherlands. The bathymetrical data were provided by the Ministry of the Flemish Community Department of Environment and Infrastructure, Waterways and Marine Affairs Administration and completed with data from the Hydrographic Service of the Royal Netherlands Navy and by the Directorate-General of Public Works and Water Management of the Dutch Ministry of Transport, Public Works and Water Management. The silt-clay fraction and the median grain size are important factors determining the meiobenthic community (Heip *et al.*, 1985; Steyaert *et al.*, 1999;

Vanaverbeke *et al.*, 2002; Merckx *et al.*, 2009). Depth in shallow waters does not directly affect the nematode community, but it modifies effects of other factors such as sea surface temperature, phytoplankton blooms, light penetration, trophic conditions of the benthos and changing water currents. It is clear from Fig. 6.2 that some of the variables have a similar distribution, especially the maps concerning TSM and silt-clay show a strong resemblance. This is not surprising since the silt-clay deposits in front of the Belgian coast can be explained by the combined effect of neap-spring tidal cycles and the presence of TSM in the water column (Fettweis and Van den Eynde, 2003).

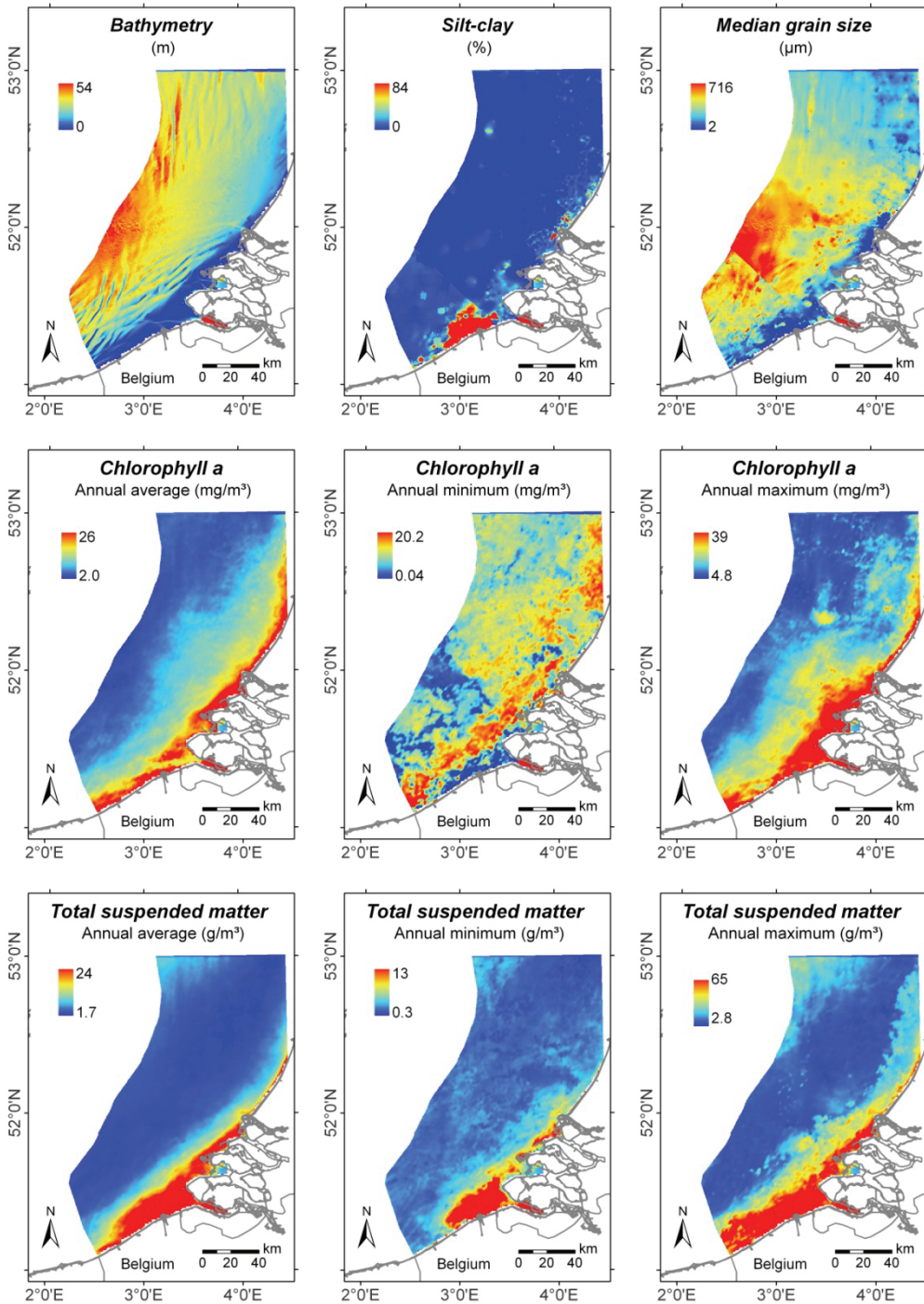


Fig. 6.2. Maps of the environmental variables (sources: see text).

Habitat suitability modelling

Habitat suitability models, attempt to correlate ecological niche elements with species presence and then project this relation into the geographical space to create predictive maps of locations with similar conditions. This approach has been cited by various names such as ‘ecological niche modelling’, ‘species distribution modelling’, ‘habitat suitability modelling’ and ‘bioclimatic envelope modelling’ (Pearson, 2007). We applied a habitat suitability technique (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005) based on presence-only data because absence of nematode species is rarely 100% certain. Generally, only a subsample is analysed and the presence of a species may not be ascertained. When a species is found presence, it is assumed to be present, notwithstanding the fact that sampling artefacts or erroneous determinations may result in false occurrences. Moreover, species may not have occupied their full niche due to unsaturated populations and inter- and intraspecific interactions (Fielding and Bell, 1997). Furthermore nematodes are known to show a patchy distribution (Li *et al.*, 1997; Somerfield *et al.*, 2007; Gingold *et al.*, 2010a). Earlier research pointed out that Maxent is a reliable presence-only modelling technique and it performs well compared to other presence-only modelling techniques (Elith *et al.*, 2006; Hernandez *et al.*, 2006; Ortega-Huerta and Peterson, 2008; Wisz *et al.*, 2008). These good predictive capacities have been attributed to the ℓ_1 -regularisation (see further) which prevents the algorithm from overfitting. Other models often do not apply any form of regularisation, and this can cause the observed difference in predictive performance (Gastón and García-Viñas, 2011). Moreover, Maxent is a generative approach, rather than discriminative. This can be an inherent advantage when the amount of training data is limited (Phillips *et al.*, 2006).

Maxent combines presence-only data with the information of environmental layers using the maximum entropy approach. This algorithm searches the probability distribution which maximises entropy within the constraints of the given data (Phillips *et al.*, 2006). The distribution π assigns a non-negative probability $\pi(x)$ to each point x within the area X , and these probabilities sum to 1. The approximation of π is also a probability distribution, and is denoted as $\hat{\pi}$. The entropy of the set of probabilities $\hat{\pi}(x)$ is defined as $H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x) \cdot \log_e(\hat{\pi}(x))$. H reaches the maximum in the most uncertain situation when a species shows ‘maximum entropy’ and has the same likelihood across the whole region (Shannon, 1948).

The environmental variables or functions thereof are called the ‘features’. These features set limitations to the choice of the probability distribution. The feature types which are used in this study are linear features, quadratic features, product features, threshold features and hinge features. The product features incorporate interactions between predictor variables. Threshold features equal one, once a certain threshold is passed (Phillips *et al.*, 2006). Hinge features, namely the forward hinge feature and the backward hinge feature, are recently introduced features. The forward hinge feature is 0 while the variable is smaller than a threshold h and then increases linearly to 1 at the maximum value of the variable. In a

similar way, the reverse hinge feature is 1 at the minimum value of the variable and drops linearly to 0 at the threshold h (Phillips and Dudík, 2008).

In order to reduce overfitting Maxent applies a penalty term which penalises models with many features (ℓ_1 -regularisation). In this way models with fewer features are favoured. Such models are less likely to overfit (Phillips *et al.*, 2006). However, former research indicated that overfitting is still present (Merckx *et al.*, 2011). Therefore we still performed a backward and forward selection of the environmental variables and a feature selection.

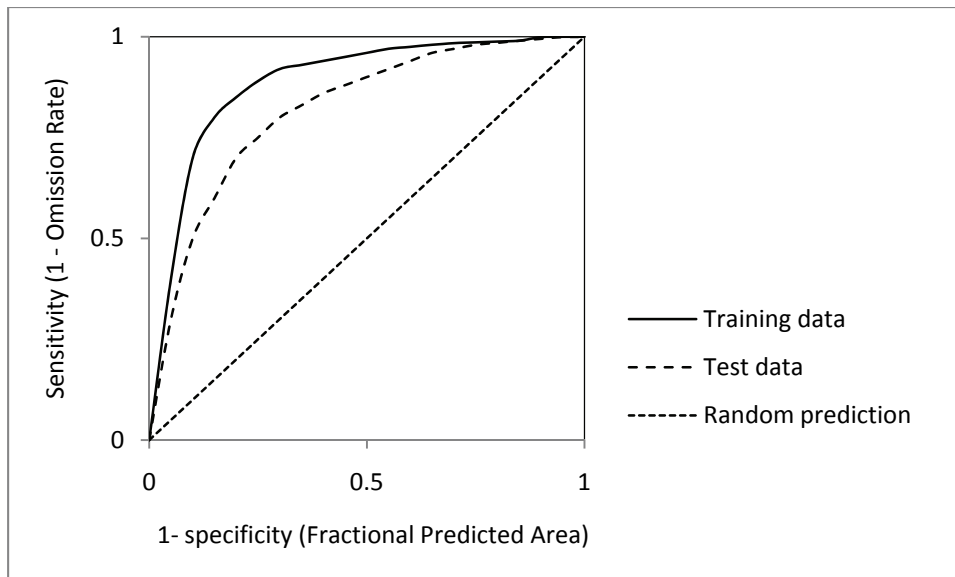


Fig. 6.3. ROC plot for test and training set. The AUC is the area under the ROC plot.

An important quality parameter in habitat suitability modelling is the area under the curve (AUC) (Fig. 6.3). It is a threshold independent measure. For different values of the false positive fraction (1-specificity), the sensitivity values are calculated (true positive fraction). These values are represented in a Receiver Operating Characteristic curve (ROC). The area under this plot equals the AUC. It indicates the overall performance of the model. An AUC of 0.5 indicates that the model predictions do not differ from a random prediction while a maximum value of 1 indicates a perfect model (Fielding and Bell, 1997). However, in presence-only modelling, as is the case for Maxent, there is no absence data available and therefore it is impossible to calculate the false positive fraction. The false positive fraction is the fraction where the species is predicted present, where it is in fact absent. However, this problem is circumvented by distinguishing presence from random, instead of presence from absence. For each AUC-analysis 10 000 pixels are drawn randomly from the study region (Phillips *et al.*, 2006). Then, the true positive fraction is replaced by the 'fractional predicted area', the fraction of the total study area for which the species is predicted to be present. Thus for an ideal model all presences are correctly predicted for an infinitely small predicted area. Consequently, it can be seen that the maximum achievable AUC is less than 1 (Wiley *et al.*, 2003). An AUC of 0.5 still corresponds with a random prediction: predicting x percent of the area as suitable for the species, will result in x percent correctly classified occurrences.

Given the calculation method of the AUC, it is clear that the AUC values of a common or generalist species will rarely be high. Since the species appears on a large fraction of the area, the sensitivity can only be large when a large fraction of the area is predicted. These difficulties of predicting generalist species is not only restricted to presence-only modelling. Evangelista *et al.* (2008) showed that all the modelling techniques they tested, including models based on both presence/absence data and presence-only data, showed difficulties in predicting generalist species. In habitat suitability modelling, data is always reduced to presence-only or to presence/absence data, even if densities are known. Nematode data is mainly obtained from sampling by cores pushed into the sediment. From these cores a subsample of 100 to 200 nematodes is taken and identified. In the database the total nematode densities of the core samples are only known in 65% of the cases. Therefore, the relative abundances of the species were calculated. In this paper, we investigate how the information of the relative abundances can be used to create HSMs which give an indication of the relative abundance of the species. In order to introduce these relative abundances in this presence-only modelling technique, we applied two methodologies: (1) the relative abundances are translated into separate observations of the species in this area (RA). Thus, if the relative abundance of a species is 5 percent, this is translated into 5 observations (Phillips and Pearson, pers. comm.); (2) we constructed 3 presence-only models based on different relative abundance thresholds: a species is considered to be present when it represents at least 1 percent of the local community (T1). The same was done for a threshold of 5 percent (T5) and 10 percent (T10). The performance of these 4 HSMs was then compared with the HSM based on the original presence-only data (PO). Thus, in total for each species 5 models were developed.

Validation of the models

Validation was done in two ways (Fig. 6.4): 1) to find the best model for each of the different datasets, a four-fold cross-validation was applied to 80% of the original data. 2) The remaining 20% of the data was used exclusively at the completion of the analysis to compare the performance of the 5 different habitat suitability models of each species.

The four-fold cross-validation on 80% of the original data is necessary to reduce overfitting. Since overfitting is not completely prevented by the ℓ_1 -regularisation (Merckx *et al.*, 2011) a backward and forward selection for each environmental variable combined with a selection of the best feature combination was done. So, the original 80% of the data was used to create four cross-validation models: 75% of this data was used to create the model and the AUC was calculated for the remaining 25% of the data (i.e. the test data). Each quarter of the data was once used as a test set, thus this resulted in four models. The 4 AUC-values of the test sets were averaged and the model with the highest average AUC is selected. This resulted in 5 final models for each species.

These 5 models were then tested against the remaining unseen 20% of the species data, the test set. Two test statistics are calculated: the AUC and the Spearman rank correlation between the relative densities of the test set and the model output.

For the AUC, the test set was reduced to samples with high relative densities. These HSMs were trained for specific thresholds, thus these thresholds were also applied on the test set. This makes sense, since the calculation of the AUC is based on presence data. Suppose that stations with low relative abundances, not reaching the threshold, are kept in the test set and these stations have low probabilities according to the model, this will be considered as a misclassification (false negative), while it is in fact a correct prediction for this threshold. Thus these samples should be removed from the test set. However, a smaller predicted area will generally entail a larger AUC. Therefore, this statistic is only of secondary importance in the model selection.

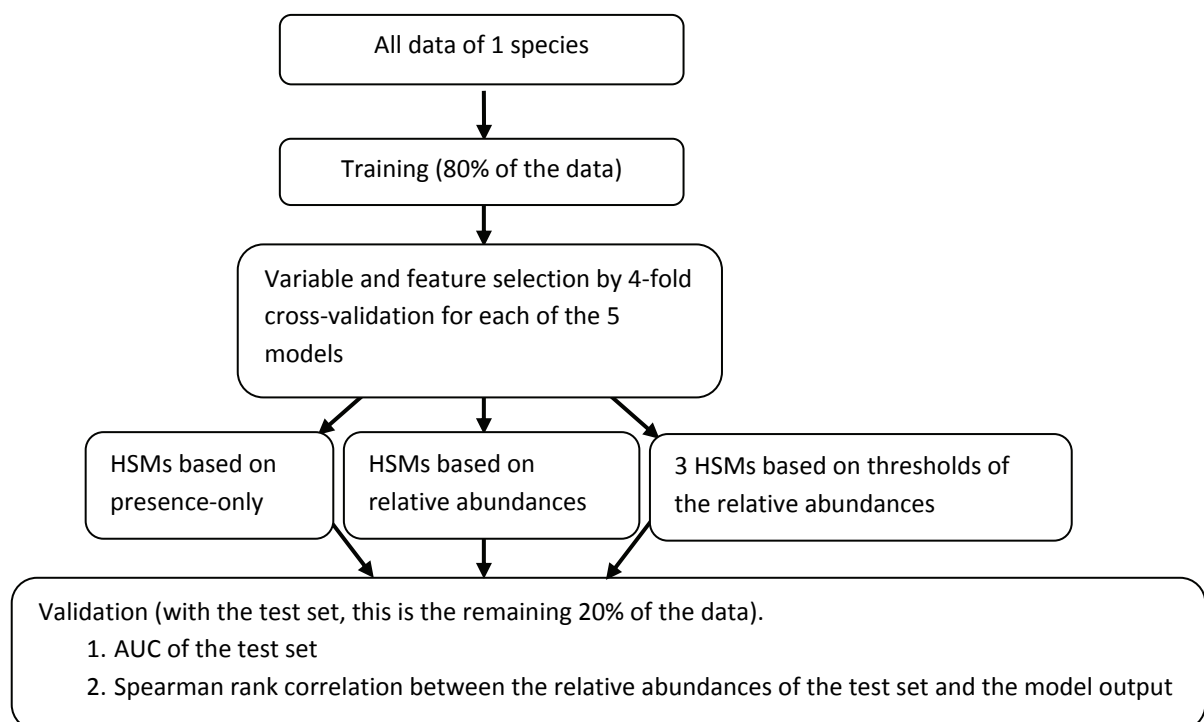


Fig. 6.4. Model validation scheme for 1 species.

More relevant is the Spearman rank correlation between the relative abundances of the test set and the model output. In this case no threshold was applied to the test set. This makes sense since the threshold-based models should predict lower probabilities when lower relative abundances are present and higher probabilities when higher relative abundances are found. Thus, for each species the original test set with all the samples is used to calculate the Spearman rank correlation for each of the 5 models.

Final maps

Based on the results of the final validation, two maps are created for every species: a reference map based on the model for the original presence-only dataset (PO) and a second map for the model showing the highest Spearman rank correlation for the test set.

RESULTS

Model selection

Table 6.2 gives an overview of the correlation between the observed relative abundances of the species in the test set and the output of the selected HSM. It is clear that when all the observations are used to build the model (i.e. in the case of the PO model) the correlation between the relative abundance and the output of the HSM is low and sometimes even negative.

The correlations generally increase with increasing threshold (Table 6.2). Thus it seems that the model is capable in identifying regions with higher relative abundances although less data is furnished to the modelling algorithm. Introducing the relative abundances (RA) as separate observations, results in better correlations than the original PO model in 50% of the cases. When thresholds are applied, the correlation increases five out of six times.

Species	Presence-only	1% threshold	5% threshold	10% threshold	Relative abundance
<i>Daptonema tenuispiculum</i>	-0.02	0.07	0.14	0.31	-0.07
<i>Dichromadora cucullata</i>	0.00	0.03	0.45	0.59*	-0.1
<i>Enoploides spiculohamatus</i>	0.32	0.27	0.68*	0.70*	0.39
<i>Onyx perfectus</i>	0.31	0.40	0.55*	0.39	0.45*
<i>Sabatieria celtica</i>	0.41	0.30	0.28	0.11	0.69*
<i>Sabatieria punctata</i>	0.68*	0.71*	0.61	0.54	0.59

Table 6.2. Spearman rank Correlation coefficients between relative abundances of the test set and the predicted values for the samples of the test set. Values in bold indicate the highest correlation coefficients for the species. Significant correlations ($p < 0.05$) are indicated with an asterisk.

	Presence-only	1% threshold	5% threshold	10% threshold	Relative abundance
<i>Daptonema tenuispiculum</i>	0.86	0.89	0.94	0.92	0.74
<i>Dichromadora cucullata</i>	0.58	0.68	0.51	0.84	0.48
<i>Enoploides spiculohamatus</i>	0.67	0.67	0.95	0.97	0.74
<i>Onyx perfectus</i>	0.66	0.59	0.92	0.92	0.8
<i>Sabatieria celtica</i>	0.79	0.76	0.85	0.9	0.88
<i>Sabatieria punctata</i>	0.94	0.94	0.95	0.95	0.95

Table 6.3. AUC of the test set. Values in bold indicate the highest AUC for the species.

The AUC of the test set also reveals higher values when thresholds are applied (Table 6.3). The AUC of the presence-only models of *Dichromadora cucullata*, *Enoploides spiculohamatus*

and *Onyx perfectus* is too small to be considered as an informative model (Merckx *et al.*, 2011). The presence-only models are depicted in Fig. 6.5 and 6.6, but only to compare the output with the models based on the relative abundances. Moreover, the comparison of the AUC values of the different models should be interpreted cautiously; the interpretation is not as straightforward as for the Spearman rank correlations. The number of samples in the tests set decreases as the threshold increases. It is the purpose of this modelling exercise to restrict the modelled suitable habitats to the actual habitats where the species can thrive at high relative abundances. As mentioned before, a restricted area will often result in an increase in the AUC. Indeed, since there is a reduction in the number of observations, there is an increasing chance that these observations are found within a restricted area, resulting in a small fractional predicted area containing all the observed presences, and thus having a high specificity. The increase in AUC does not necessarily mean a better performance of the model. Thus, in this case the Spearman rank correlation will give a better indication of the performance.

For the four-fold cross-validation it does make sense to use the AUC as a quality parameter, since the models which are compared are constructed with the same dataset, which have obviously the same distribution pattern. Thus if different models are based on the same dataset, a higher AUC will indeed indicate a better overall performance of the model.

	Model	Average Chl <i>a</i>	Maximum Chl <i>a</i>	Minimum Chl <i>a</i>	Median grain size	Water depth	Silt-clay content	Average TSM	Maximum TSM	Minimum TSM
<i>Daptonema tenuispiculum</i>	10%							100% ↗		
<i>Dichromadora cucullata</i>	10%	23% ↘		8% -		35% ↗		30% ↘		
<i>Enoploides spiculohamatus</i>	10%		100% ↗							
<i>Onyx perfectus</i>	5%	19% ↘	6% ↗			54% ↘		17% ↘		
<i>Sabatieria celtica</i>	RA			5% ↗				37% ↘	56% ↘	
<i>Sabatieria punctata</i>	1%				15% ↘				81% ↗	

Table 6.4. Estimate of the relative contributions of the environmental variables to the final models. Only variables contributing more than 5% to the model are shown. Positive (↗), negative (↘) and optimum (↘) correlations are represented.

Final models

The variable contributions of the thresholded models are shown in Table 6.4. Average TSM contributes strongly to the model of *Daptonema tenuispiculum*. *Sabatieria celtica* and *Sabatieria punctata* seem to be strongly influenced by the maximum TSM level. *Enoploides spiculohamatus* shows a strong positive relation with maximum chlorophyll *a*. While species

found in high abundances off-shore such as *Dichromadora cucullata* and *Onyx perfectus* show a relation with water depth. As mentioned in the materials and methods section: the selected variables may represent proxies for other variables. The λ -values describing the thresholded models can be found in Addendum 4.

Fig. 6.5 and 6.6 show the resulting maps. It is clear that the models resulting from the data with the relative abundances thresholds result in narrower distribution patterns. This was to be expected: the number of samples in the model decreased, most probably resulting in smaller ranges of environmental variables.

DISCUSSION

Model selection

According to Table 6.2 the habitat where the species thrives is not favoured by the models based on the occurrence data. This makes sense since this model reflects which habitat is potentially suitable. A single occurrence of a species is considered equally important as high relative abundances of a species. The model does not differentiate between optimal and suboptimal habitats.

Using RA generally increases the correlation between the model output and the relative abundances in the test set. Thus, this methodology may identify regions where the species is found in higher relative abundances, but needs to be examined for each independent case. Applying thresholds seems to have more potential in differentiating between habitats where species occur in high and low relative abundances.

Final maps

In order to analyse if the resulting patterns are also realistic patterns, the habitat preferences of each species were compared with literature sources. It should be noted that some of the data of the cited literature sources (Vincx, 1989a; Vincx *et al.*, 1990; Vanreusel, 1990; Vanaverbeke *et al.*, 2007; Vanaverbeke and Vincx, 2008; Vanreusel, 1991) are actually used for building the models and we are aware that circle reasoning should be avoided when interpreting results. However, it is the first time that the data of 17 different studies are combined in one single analysis and finding the same patterns may reconfirm and strengthen the findings in the literature sources.

Daptonema tenuispiculum is a non-selective deposit feeder (1B) (Wieser, 1953; Vincx and Heip, 1987). The species is often found in stressed environments and can survive in sediments with low oxygen content (Boyd *et al.*, 2000). It is common in the mouth of Western Scheldt (Vincx and Heip, 1987), in unstable sediment due to fluctuating current velocities (Vanreusel, 1990) and dredged material disposal sites (Boyd *et al.*, 2000).

Daptonema tenuispiculum is found in high densities in fine medium sand with a high amount of silt (44.6%) and organic carbon (Vincx, 1989a; Vanreusel, 1990; Schratzberger *et al.*,

2000b). It is dominant at the mouth of the Western Scheldt and at the Belgian coast, except for the Western coast (Vincx, 1989a). It is clear from Fig. 6.5 that the literature sources are confirmed by the model with the 10% threshold: the regions with high TSM-values are highlighted. TSM is a proxy for silt-clay and may indicate elevated levels of organic carbon. The original model with the presence data describes the area where the species is found well (Table 6.3), but it is less clear where high relative abundances can be found. The modified model with the 10% threshold results in more or less the same distribution of the species, but highlights the region where the species is found in high relative abundances.

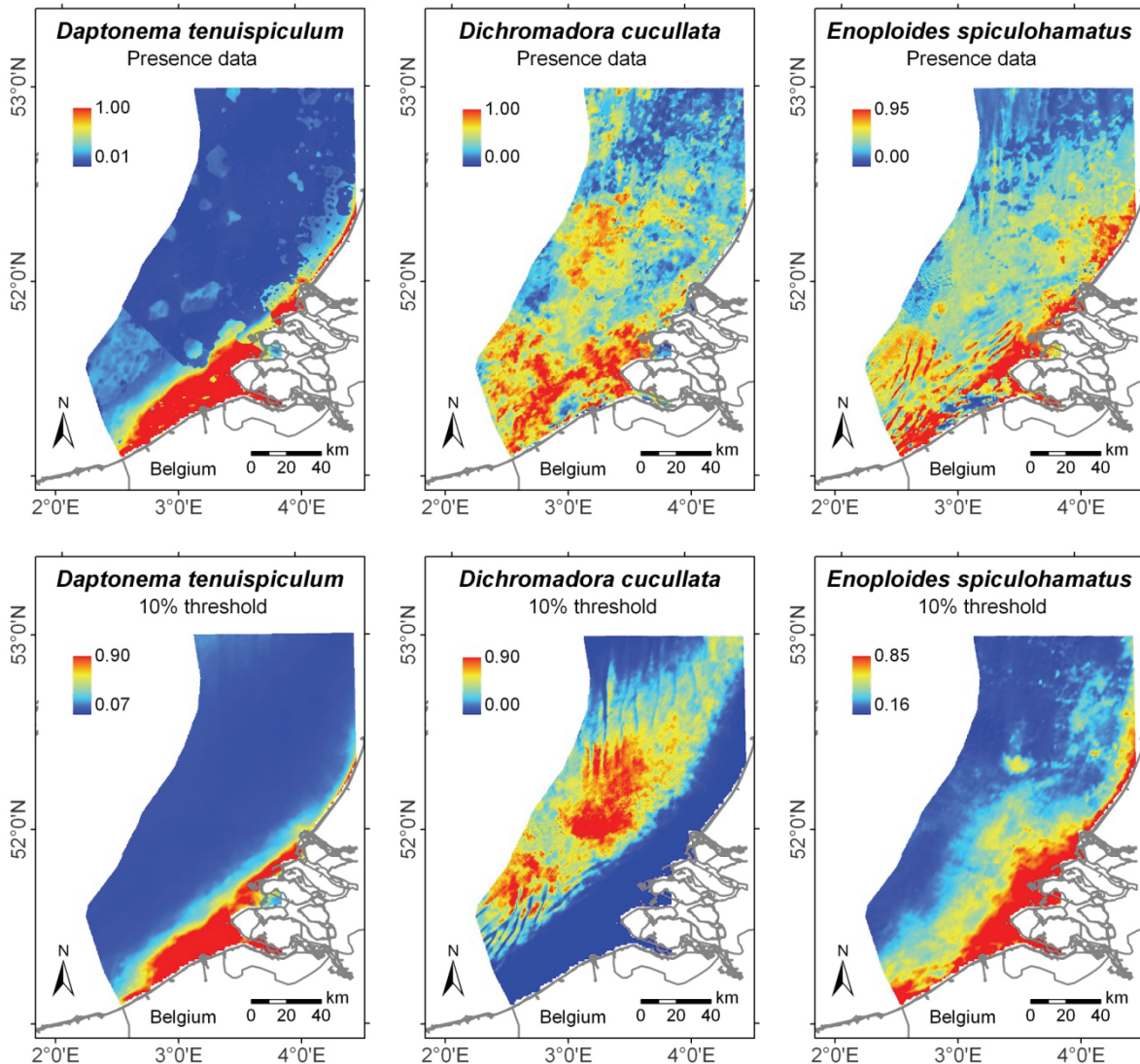


Fig. 6.5. Comparison of the two resulting models using all the presence data and the data incorporating relative abundances for *Daptonema tenuispiculum*, *Dichromadora cucullata* and *Enoploides spiculohamatus*.

Dichromadora cucullata is an epigrowth feeder (2A) (Wieser, 1953; Heip *et al.*, 1983; Vincx, 1989a) and is a common species (Vincx *et al.*, 1990). However it can be found in higher relative abundances in more offshore stations where sediments consist of clean medium sand and little gravel (Vincx, 1989a) and in coarse sand (Lorenzen, 1974). The original model

(Fig. 6.5) confirms the statement that *D. cucullata* is a common species. Limiting the samples to stations where high relative abundances (>10%) were observed restricts the suitable habitats to stations offshore, the region where sediments with higher median grain size prevail.

Enoploides spiculohamatus is a predator 2B (Wieser, 1953; Vincx, 1989a) and a common species (Vincx *et al.*, 1990) which is frequently found in fine medium sand with a low amount of silt (<5%) (Vincx, 1989a) and low chl *a* content in the water column (Vanreusel, 1990), or in coarse sand (Vanaverbeke *et al.*, 2002). It is mainly found along the Dutch coast and the western part of the Belgian coast (Vincx *et al.*, 1990). Both maps display a wide geographical range for this species. Restricting the observations to a threshold of at least 10%, results in a model that shows higher relative abundances along the coast. The original model with presence data clearly leaves out the part of the Belgian coast where a high amount of silt and clay is found (Eastern part of the Belgian coast), which is consistent with the literature sources, while the adjusted model does not exclude this region (Fig. 6.5). This is clearly an artefact of the adjusted model. Moreover, the species seems to occur in regions with high chl *a* content in the water column (Table 6.4), which seems to be in contradiction with the observations of Vanreusel (1990).

Onyx perfectus is a predator (2B) and is found in high relative abundances on the crests of the sand banks (Vincx, 1989a; Vanaverbeke *et al.*, 2007; Vanaverbeke and Vincx, 2008) and stations characterised by medium sand almost without silt (Vincx, 1989a; Vanaverbeke *et al.*, 2002). It is a very rare species in sediments containing more than 5% silt (Vanreusel, 1991). The amount of organic carbon (Vincx, 1989a) and chl *a* (Vanreusel, 1990) can be high. It is also found in high relative abundances in the gullies between the sandbanks of the Belgian Continental Shelf. These gullies are characterised by coarse sediments and high gravel content (Vincx, 1989a). The original model (Fig. 6.6) shows a broad geographical range of the species. The data with the 5% threshold results in a model where the suitability of the habitat is restricted to the sand banks and the gullies in between (Fig. 6.6). And Table 6.4 confirms the positive relation with increased chl *a*, which is consistent with the literature sources.

Sabatieria celtica is a non-selective deposit feeding nematode species (1B) (Wieser, 1953; Vincx, 1989a). It prefers fine to medium sand with a low amount of silt (<5%) (Vincx, 1989a; Soetaert *et al.*, 1995) but can also be found in lower densities in both silty environments and coarse sand (Lorenzen, 1974; Vanreusel, 1990; Vanreusel, 1991). It is mainly found at the Dutch coast and the western part of Belgian coast (Vincx *et al.*, 1990). This is the only model, where the best correlation is found when the relative abundances of the species are introduced as different samples (RA). The modified model (Fig. 6.6) restricts the suitable habitats to the coastal zones and the highest relative abundances are indeed found near the Dutch coast and the Western part of the Belgian coast while for the original model this differentiation is not clear (Fig. 6.6).

Sabatieria punctata is a non-selective deposit feeder (1B) (Wieser, 1953; Vincx and Heip, 1987) which is often found in stressed environments: in dredged material disposal sites (Boyd *et al.*, 2000) and in unstable sediments due to fluctuating current velocities (Vanreusel, 1990). It seems to thrive in fine medium sand with a high amount of silt and organic carbon (Vincx, 1989a; Vanreusel, 1991; Soetaert *et al.*, 1995). It is mainly found at the mouth of the Western Scheldt (Vincx and Heip, 1987) and at the Belgian coast, except for the Western part (Vincx, 1989a). The model selects TSM (Table 6.4) as most important factor which contributes to the presence of silt-clay and organic carbon. Both *Sabatieria punctata* models delineate the coastal zone. There is not so much difference between both models (Fig. 6.6), only at the mouth of the Scheldt estuary there is a small increase in suitability for the modified model, based on data with 1% threshold.

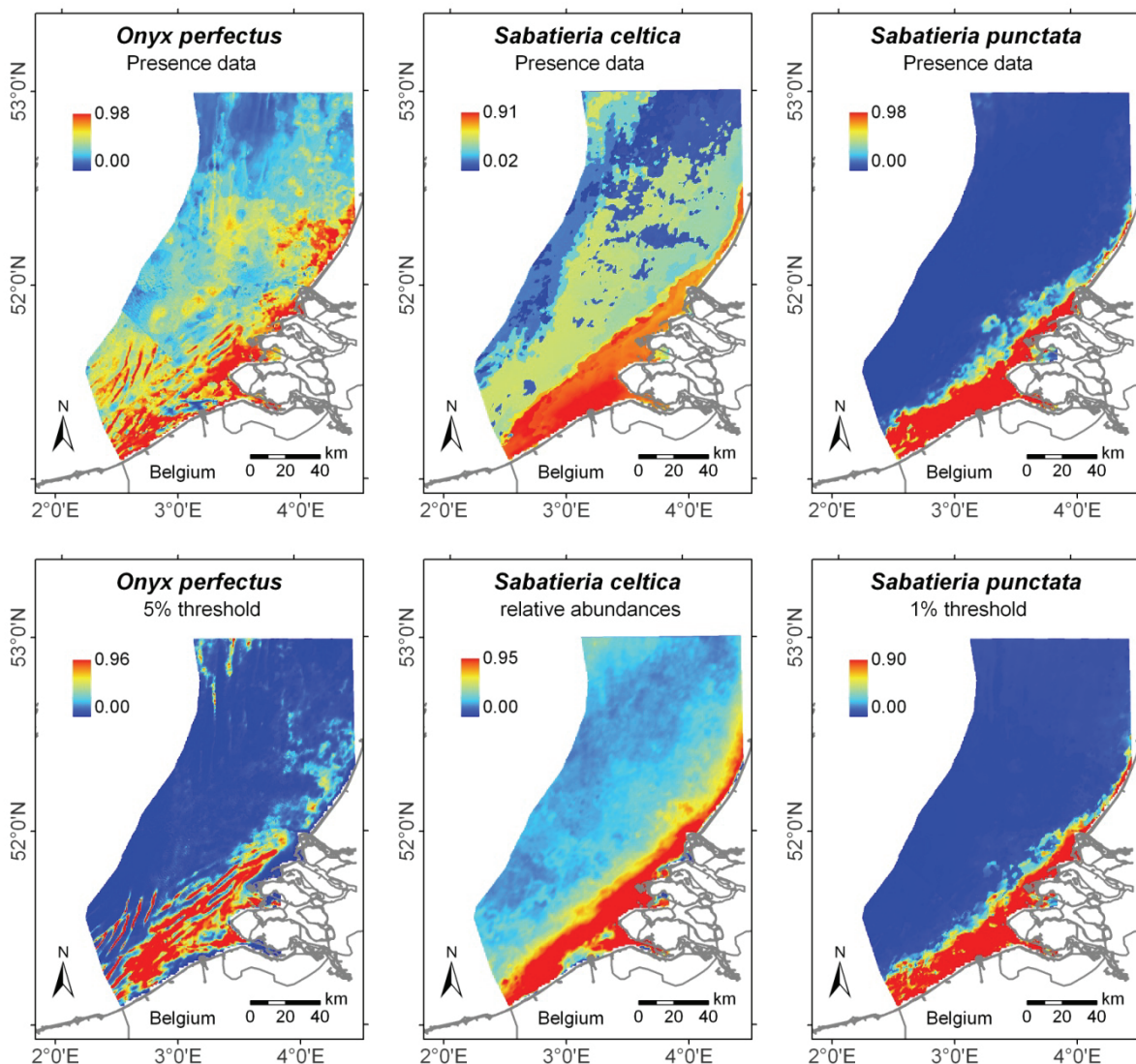


Fig. 6.6. Comparison of the two resulting models using all the presence data and the data incorporating relative abundances for *Onyx perfectus*, *Sabatieria celtica* and *Sabatieria punctata*.

According to literature *Daptonema tenuispiculum* and *Sabatieria punctata* often appear together (Vincx, 1989a; Vincx *et al.*, 1990; Vanaverbeke *et al.*, 2011). The models indeed indicate that the species have the same potential distribution. In addition, our models indicate that high relative abundances of *S. celtica* and *S. punctata* may coincide.

Five out of six nematode species appear in high relative abundances near the coast (Vincx, 1989a; Vanreusel, 1990). The environmental conditions enable species to reach high relative abundances in this region, and thus attributes to the effectiveness of this methodology for these species.

Niche concept

Generalist species tolerate a wide range of environmental conditions, and generally have a diverse diet and a good tolerance for disturbances. All this is important for defining the species' niche. Habitat suitability modelling estimates the species ecological niche. A species' fundamental niche represents a set of all conditions necessary for the survival of a species (Hutchinson, 1957). It is however assumed that every point within the niche has the same probability of persistence of the species, and all points outside the niche have zero probability of survival. This is clearly an oversimplification of reality. There will be optimal and suboptimal conditions in the niche (Hutchinson, 1957). By applying thresholds on the relative abundance, we filter out suboptimal conditions and try to define the fundamental niche for the survival of high relative abundances and dominance of a species. Species appearing across the whole region are generally hard to model since they show no real habitat preference. The relationship between the presence of the species and the environmental variables is therefore not always straightforward. In spite of this, some species may thrive or may be better competitors in certain restricted habitats, but not in others.

In this paper the data was converted to relative abundances, since the total density was not known for a considerable amount of data. In many other cases, absolute densities are known and most probably, this methodology can equally be applied to absolute densities.

CONCLUSIONS

In some cases knowing the potential density or relative abundance of a species in a region may be more important than knowing the suitability of the habitat. In this case it is reasonable to modify the data furnished to the habitat suitability modelling technique in such a way that habitats with high densities or relative abundances are preferentially predicted. The introduction of thresholds seems to be a reliable way to introduce this information into the model. Relating the model to existing knowledge of the species can help in identifying the most reliable model. Thus depending on the purpose of the model, we suggest different approaches in habitat suitability modelling: if the model concerns a rare species, knowing the potential niche may be the main focus of the research. If the species is

common and the species occurs in varying densities, applying thresholds may create opportunities to find the environments where the species can appear in high relative abundances.

ACKNOWLEDGEMENTS

This research is funded by the Fund for Scientific Research (FWO) of the Flemish government (FWO07/ASP/174). The authors wish to thank all the data providers! The environmental data was gathered from different institutes: ESA and MUMM/RBINS are acknowledged for providing and processing MERIS data (chlorophyll and TSM data, www.mumm.ac.be/BELCOLOUR), the Renard Centre of Marine Geology (RCMG, www.rcmg.ugent.be) of Ghent University and the Hydrographic Service of the Royal Netherlands Navy and the Directorate-General of Public Works and Water Management of the Dutch Ministry of Transport, Public Works and Water Management for the oceanographic and sedimentological data. The study was conducted within the framework of the Ghent University BBSea Project (GOA 01600705) and the EU Network of Excellence MarBEF (GOCE-CT-2003-505446).

