Distribution of benthic macrofaunal communities in the western Baltic Sea with regard to near-bottom environmental parameters. 2. Modelling and prediction

Mayya Gogina, Michael Glockzin, Michael L. Zettler *

Leibniz Institute for Baltic Sea Research, University Rostock, Seestrasse 15, D-18119 Rostock, Germany

ABSTRACT

The detailed analysis of patterns of benthic community distribution related to selected environmental parameters provides a basis for predictive modelling of species distribution. Species-specific models predicting the probability of occurrence relative to environmental and sedimentological characteristics were developed in this study for 29 macrofaunal species common for our study area using a logistic regression modelling approach. This way, a good description of the occurrence of species along gradients of single environmental variables was obtained. Subsequently, we used a technique for a predictive modelling of species distributions in response to abiotic parameters based on single-factor logistic regression models, utilizing AIC and Akaike weights for multimodel inference. Thus, probabilities of occurrence for selected exemplary species (*Arctica islandica*, *Hediste diversicolor*, *Pygospio elegans*, *Tubificoides benedii* and *Scoloplos armiger*) were modelled and mapped. For all species the use of this newly available combination of methods provided fairly accurate results of a distribution prediction. Water depth that represents a type of integral parameter remained the key factor determining the species distribution among the parameters considered within the study scale. This is particularly relevant for species that find their optima habitat here, but also for those as *H. diversicolor* that occur only locally and in comparatively low densities. Total organic content, sorting and, for *S. armiger*, salinity also had noticeable effect in the determination of suitable habitats for benthic macrofauna. The employed technique proved to be appropriate for modelling of the benthic species habitat suitability, at least within comparable spatial scales and variability of environmental factors.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Climate change models assume a drastic change in the food web structure, a shift in species composition towards warm water species and growing benthic deserts on the sea floor as up-coming changes for the Baltic Sea ecosystem (Philippart et al., 2007). Predictive modelling of species distribution can be a valuable tool in management directed towards the sustainable development of the Baltic Sea. Studies on various scales are required to extend our knowledge of habitat change effects.

Response of macrofaunal assemblages to substrate composition, hydrographic parameters and their variation is declared by many different studies (e.g. Sanders, 1968; Rhoads, 1974; O'Brien et al., 2003; Laine, 2003; Perus and Bonsdorff, 2004; Ellis et al., 2006). Particular establishments regarding the dynamics and structure of biotic/environmental interactions are required to evaluate natural and anthropogenic influences and effects on the ecological systems (Pavlikakis and Tsihrintzis, 2000; Glockzin and Zettler, 2008a). An exploratory statistical description of the prevailing ecological structure based on observations is always the indispensible first step (Bourget and Fortin, 1995).

Recently, a number of studies have succeeded in the development of effective statistical models of benthic distribution. Ysebaert et al. (2002) successfully applied logistic regression to derive response surfaces of distributions for 20 common macrobenthic species found in the Schelde estuary in the Netherlands related to salinity, depth, current velocity, and sediment characteristics. Thrush et al. (2003) developed species-specific models for 13 benthic species of New Zealand estuaries that predicted probability of occurrence as well as maximum abundance relative to sediment mud content using logistic regression for distribution modelling and 'factor ceiling' method (Blackburn et al., 1992) for maximum density modelling. Ellis et al. (2006) modelled the distribution of 13 representative macrobenthic species in New Zealand estuarine gradients using logistic regression and classification system based on 'controlling factors' with sediment characteristics, elevation, tidal currents, and wind-wave disturbance employed as predictors. They faced complications to fully test the latter approach due to differences in scales of collected benthic data and of higher level physical variables. Meissner et al. (2008) developed habitat models for *Nephtys* species in the German Bight (North Sea) with median grain size, mud content, depth, and salinity as explanatory variables by application of

* Corresponding author. Leibniz Institute for Baltic Sea Research, Seestrasse 15, D-18119 Rostock, Germany. Fax: +49 381 5197 440.
E-mail address: michael.zettler@io-warnemuende.de (M.L. Zettler).

multivariate adaptive regression spline techniques (MARS). Many researches indicated difficulties due to the complexity of identification of the underlying causal mechanisms controlling species distribution, further extended by the fact that animals modify their physical environment, and many physical parameters co-vary (Ellis et al., 2006).

The present study contributes to the development of statistical models that are able to predict the distribution of benthic macrofaunal species as a function of environmental variables. Models that forecast the behaviour of species distribution versus changes in environmental factors (sensu Legendre and Legendre, 1998) provide an insight into chronic habitat change (regarding these parameters), though they do not provide insight into the acute effects associated with disturbance events (Thrush et al., 2003; Ellis et al., 2006). Model estimations based on a data set consisting of the response variables (e.g. species occurrence or abundance) and on a set of predictor variables (e.g. environmental parameters) can be used to predict the spatial distribution of species in a habitat with known or defined environmental settings (Ysebaert et al., 2002; Ellis et al., 2006; Meissner et al., 2008). Statistical models are able to relate ecological features to environmental factors and, through validation and modification, are able to reveal details in the underlying mechanisms responsible for structure and organization of communities (Austin, 1987; Glockzin and Zettler, 2008a).

Our investigation is focused on a limited area in the south-western Baltic Sea. The pre-work of an exploratory statistical description of the prevailing ecological structure is an essential first step towards modelling, and it was completed for the area of interest beforehand (Gogina et al., 2010-this volume). The identified distinct benthic assemblages have proved to associate with certain spatial regions and specific limits of environmental parameters. The reduction of macrozoobenthic data to presence/absence is forced by the absence of temporal homogeneity and is used here to eliminate the influence of patchiness in spatial distribution that macrofauna exhibits (McArdle and Blackwell, 1989; Legendre et al., 1997; Thrush et al., 2003).

2. Materials and methods

2.1. Study area

The study area is located in the south-western Baltic Sea, between 11.55° to 12.55° E and 54.09° to 54.96° N (Fig. 1). It is bounded by the eastern part of the Mecklenburg Bight and the western region of the Kadetrinne, with its northern and southern limits defined by Danish and German land boundaries. Some geographical details about the area, which is characterised by a relatively high biodiversity of both saline and brackish water species, as well as the analysis of benthic community structure, can be found in Gogina et al., 2010-this volume.

2.2. Data used for model estimation

The study is based upon the data of benthic macrofauna and associated sediment and near-bottom environmental characteristics, sampled at 208 stations (Fig. 1a). For 72 of these stations a full set of the abiotic parameters considered is available. For modelling purposes the species abundance data was reduced to presence/absence.

The description of methods for benthic macrofauna sampling and abiotic factor determination, as well as the selection process for extraction of 29 representative macrobenthic species modelled here, can be found in Gogina et al., 2010-this volume.

2.3. Additional environmental data for predictive modelling

Additional data sets were required to compile the grids of each abiotic descriptor, needed for predictive estimates of species distribution (probability of occurrence) for the whole investigation area. The distribution surfaces obtained for each of the environmental variables considered are presented in Fig. 1.

For the bathymetry a high-resolution digital elevation model (DEM) was created using measured data provided by the Federal Maritime and Hydrographic Agency (BSH) and a regional grid data set from Seifert et al. (2001), covering the Belt Sea region. For more details see the description of DEM design in Meyer et al. (2008). Grid data sets for near-bottom oxygen content and salinity were based on the modelled hydrographical data, averaged for years 1960–2005 with the resolution of 3 nautical miles (Neumann and Schernewski, 2008), covering the whole western Baltic sea area. Grids for sediment parameters like median grain size, sorting, skewness and permeability are derived from the internal database of the Leibniz Institute for Baltic Sea Research Warnemuende (IOW; Bobertz and Harff, 2004), integrating the data of about five decades of marine investigations. The average distance between adjacent sample sites is less than 1 nautical mile. From the IOW database external data on total organic content was also available, however, only for a limited area. Hence, this data was agglomerated together with the observed data used for model estimation to increase the area covered and the density of data points. Nevertheless, only a part of the investigation area could be covered with the compiled grid of this parameter (Fig. 1d). Parameters were interpolated using ordinary kriging with spherical fitted models of semivariograms into a grid with the resolution of about 0.005 decimal degrees (approximately 0.5 km with respect to longitude).

Ysebaert et al. (2002) favoured the usage of modelled estimates of environmental variables over the data measured directly and simultaneously with benthic sampling. The argumentation included the available high spatial resolution and a sort of smoothing caused by simulation, e.g. elimination of outfits. However, taking into account the complexity of the functioning of ecosystems, the uncertainty of simulations may increase the complexity of the interpretation of derived empirical relationships. Utilization of simulated data for the model estimation is forced merely by the necessity and absence of alternatives. The preliminary explicit exploratory analysis of environmental framework should exclusively be based on direct in situ measurements. Therefore, to enable the investigation of autecological relationships we rely our model calibration on directly observed data to the highest extent possible, applying minimum transformations to lessen the reduction of information contained in the data (Gogina et al., 2010-this volume). Yet, the prediction is based partly on modelled data of sufficient resolution available for the study area, thus, allowing the validation of modelling success.

2.4. Statistical analysis and data treatment

2.4.1. Univariate logistic regression

Logistic regression of biotic data reduced to presence/absence was employed to model the probability of occurrence of 29 discriminating species, using the considered environmental factors (water depth, salinity, oxygen concentrations, total organic content, median grain size, sorting, skewness and permeability of sediments) as explanatory variables. These factors are generally assumed to have direct or indirect impact on distribution of macrobenthic species. The logit function in a logistic regression is the special case of a link in a generalised linear model, known as canonical for the binomial distribution. Application of logistic regression methods in modelling species distribution is not new. This method was widely used in plant ecology (e.g. Guisan et al. 1999) and also in aquatic ecology, but to a lesser extent. Thrush et al. (2003) concentrated their investigation on a single environmental factor—sediment mud content, Ysebaert et al. (2002) performed a comprehensive study, using salinity, depth, flow parameters, median grain size and mud content as predictors. The present study represents one of the first applications of this technique to benthic habitats of the Baltic Sea.
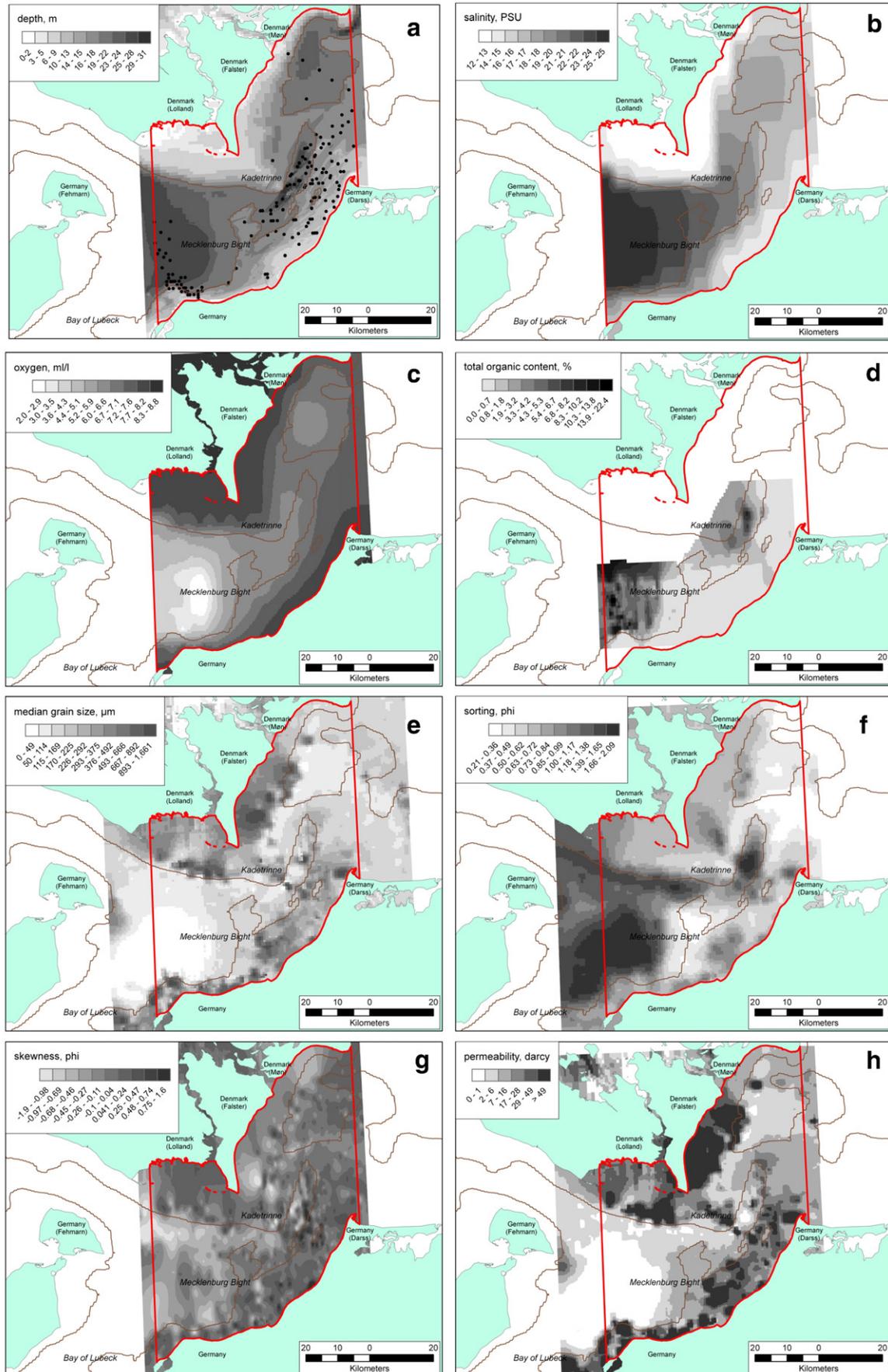
Fig. 1. Distribution surfaces covering the study area (limited by the thick line) generated for each of the environmental variables considered using ordinary kriging. Dots indicate stations sampled for macrofauna. Geographical data ESRI (2003); projection UTM on WGS84.

The logistic regression model relates the probability of observing the species p to one or more predictor variables x (in our study separately to each of the environmental factors) using the logistic link function. The regression model can be written as

$$p(x) = e^z / (1 + e^z) \tag{1}$$

where z is the function of the explanatory variable(s). When this is a first-order polynomial, the response is a logistic, S-shaped function. In the special case of Gaussian logistic model when z is a second-order polynomial, the response will approximate a bell-shaped function. For this particular case Eq. (1) can be written as

$$p(x) = e^{(b_0 + b_1 x + b_2 x^2)} / (1 + e^{(b_0 + b_1 x + b_2 x^2)}) \tag{2}$$

where b_0, b_1, and b_2 are regression parameters. They are estimated by maximum likelihood, assuming a binomially distributed error term (Legendre and Legendre, 1998; Ysebaert et al. 2002; Wisz and Guisan, 2009). When the estimation of z term parameters is based on log-transformed data this can be interpreted as a further extension of the method, aiming to produce an ecologically more plausible response for certain species. This idea was adopted from Thrush et al. (2003, 2005) who found such a transformation to produce the most realistic response to changes in sediment mud content for the occurrence of some of the investigated species.

Thus, for each factor and taxon combination, different functions were used (linear, Gaussian, polynomial) and were based on either raw or log-transformed data. The Wald statistic was used to estimate the model's significance, with a significance level defined at 0.05. The final single-factor model used for each species was the function (of the particular factor) that explained the most variability. The evaluation of the model fit was performed by means of visual control of half-normal plots of residuals and plots of residuals versus predicted values and by considering the percentage of concordant pairs.

2.4.2. Collinearity among predictors

The complexity of prediction of species distribution is caused by the complexity of interactions of various factors. To give some insight in the extent and direction of collinearity among the considered factors, relationships between pairwise combinations of individual predictors are portrayed in Fig. 2 using scatterplot matrix. Additionally, analyses of correlations among environmental parameters and values of Pearson correlation coefficients can be found in Gogina et al., 2010-this volume.

2.4.3. Information-theoretic approach and utilization of the Akaike weights

To combine the results of single-factor models and to draw the multimodel inference we considered the information-theoretic philosophy described in Burnham and Anderson (2004). It relies on the calculation of the Akaike's information criterion (AIC) as a model
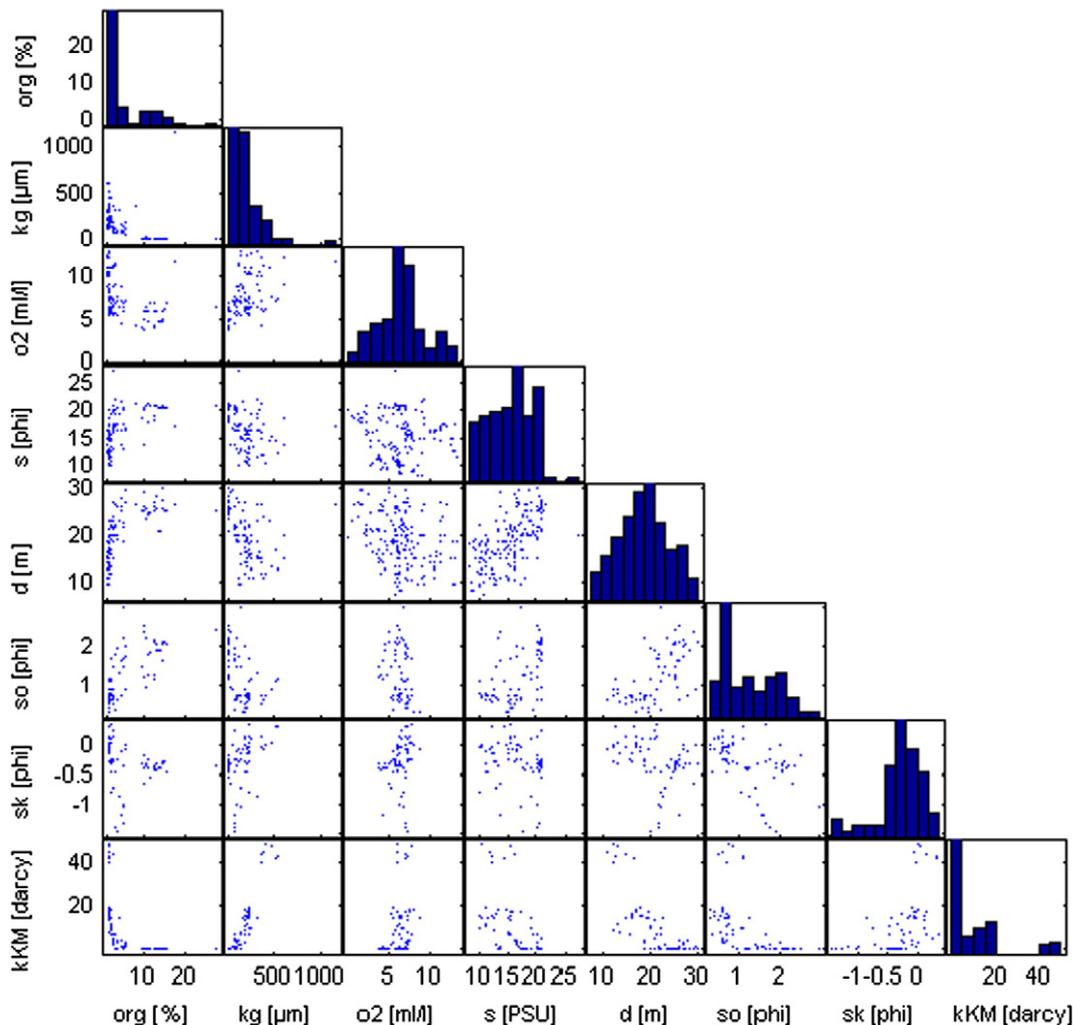

Fig. 2. Scatterplot matrix showing relationships between environmental factors over the data set. Environmental factors notation: d—water depth, s—near-bottom salinity, org—total organic content, o2—near-bottom oxygen concentration, kg—median grain size, so—sorting, sk—skewness, kKM—permeability.

selection tool. AIC is a measure of the relative Kullback–Leibler information loss when the candidate model i is used to approximate the truth j. Generally appropriate small-sample version (designed for $n/K > 40$) of criterion AIC_c is calculated as

$$AIC_c = -2\ln(L) + 2K(K+1)/(n-K-1) \tag{3}$$

where L is a value of the maximised log-likelihood over the estimated parameters given the data and the model, K—number of parameters in model i, and n is the sample sizes. AIC (or AIC_c) are calculated for each of the candidate models, then these values are rescaled to calculate delta AIC (Δ_i) so that the model possessing the lowest AIC value has a Δ_i value of 0:

$$\Delta_i = AIC_i - AIC_{min} \tag{4}$$

where AIC_{min} is the smallest AIC value in the model set. The model with the lowest AIC value is considered to be the best approximating model in the candidate set. The larger the value of Δ_i the less plausible the fitted model i is. Burnham and Anderson (2004) suggest a simple rule of thumb in assessing relative merits of models in a set: models with $\Delta_i \leq 2$ have strong support, models with Δ_i values between 2 and 10 have considerable support and those with $\Delta_i > 10$ have essentially no support. Akaike weights (ω_i) are now calculated for each of the r candidate models:

$$\omega_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum\limits_{r=1}^{R} \exp\left(-\frac{1}{2}\Delta_i\right)}. \tag{5}$$

The ω_i are scaled between 0 and 1, and all Akaike weights sum to one—ω_i values represent the proportion of evidence for a particular model i in the total evidence supporting all of the models (Wisz and Guisan, 2009). A model that possesses the largest ω_i value is the most parsimonious and has most support among the specified candidate models given the data. When more than one model is supported by the data it is possible to calculate a global model that is a weighted average of all the candidate models in the *a priori* defined set. New parameter estimates for each term in the global model can be computed by weighting them by the Akaike weights

$$\hat{\bar{\theta}} = \sum_{i=1}^{R} \omega_i \hat{\theta}_i \tag{6}$$

where is $\hat{\bar{\theta}}$ the model averaged parameter estimate based on all R models, and $\hat{\theta}_i$ is the parameter estimate for a term in a candidate model i with the Akaike weight ω_i. For terms that do not feature in a candidate model but are present in the global model the parameter estimate is taken to be zero. Thus, if the goal is prediction, the point inference can be based on the entire set of models using Akaike weights within the overlapping of single-factor models as weight factors to generally estimate the probability of species occurrence.

2.4.4. Habitat suitability mapping

The method applicability was tested on five selected species (*Arctica islandica*, *Hediste diversicolor*, *Pygospio elegans*, *Tubificoides benedii* and *S. armiger*). Finally, estimates derived for these species were implemented in a geographical information system. As the grid data for total organic content was only available for a limited part of the investigation area (see Section 2.3), two model sets were considered for the final prediction: one with and one without the model of species occurrence probability as a function of total organic content included. The final value assumed for the overlapping areas of two model sets was the value from the model set with all parameters considered; the rest of the area was filled out with the results based on the model set lacking the total organic content factor. Concordance was calculated to provide the estimate of modelling success.

All analysis were carried out using SPSS (SPSS, Inc.), Statistica (StatSoft, Inc., 2007), PRIMER (PRIMER-E, Plymouth Marine Laboratory;
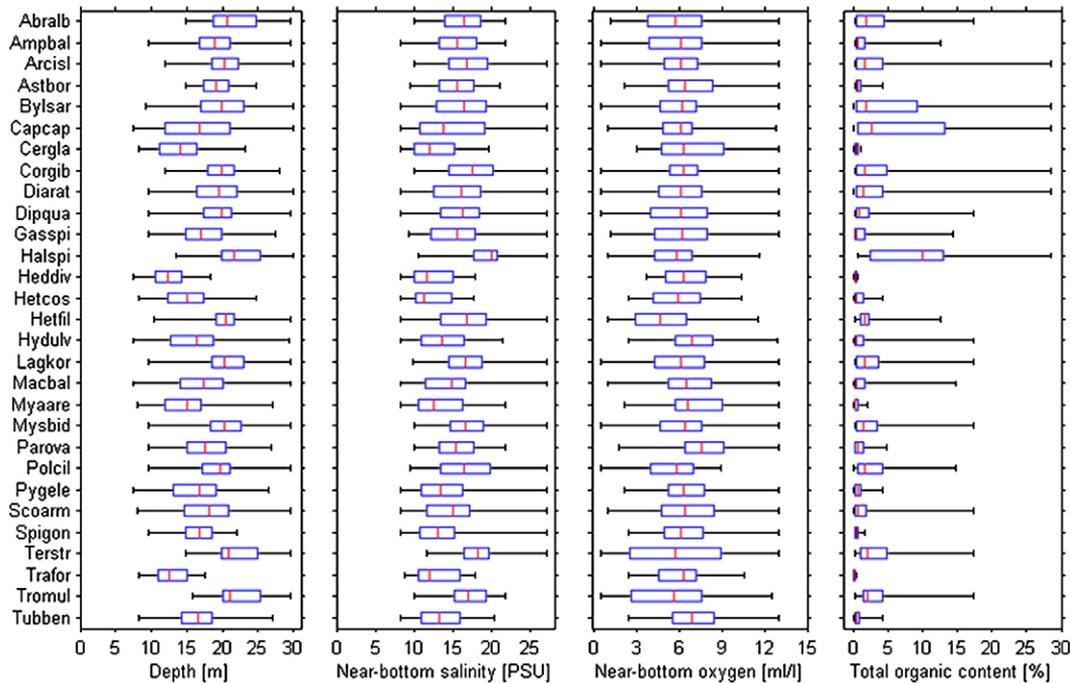

Fig. 3. Box-and-whisker plots for 29 macrobenthic species with respect to depth, near-bottom salinity and oxygen, total organic content. Species are ordered alphabetically, 6-letter codes from top to the bottom referring to *Abra alba*, *Ampharete baltica*, *Arctica islandica*, *Astarte borealis*, *Bylgides sarsi*, *Capitella capitata*, *Cerastoderma glaucum*, *Corbula gibba*, *Diastylis rathkei*, *Dipolydora quadrilobata*, *Gastrosaccus spinifer*, *Halicryptus spinulosus*, *Hediste diversicolor*, *Heterochaeta costata*, *Heteromastus filiformis*, *Hydrobia ulvae*, *Lagis koreni*, *Macoma balthica*, *Mya arenaria*, *Mysella bidentata*, *Parvicardium ovale*, *Polydora ciliata*, *Pygospio elegans*, *Scoloplos armiger*, *Spio goniocephala*, *Terebellides stroemi*, *Travisia forbesii*, *Trochochaeta multisetosa*, and *Tubificoides benedii*. The tops and bottoms of each "box" are the 25th and 75th percentiles of the samples, respectively. The line in the middle of each box is the sample median. The "whiskers" extending to the left and the right of each box represent minimum and maximum of the observations.

Clarke and Warwick, 1994), MATLAB and ArcMap (ESRI Inc., Redlands, USA).

3. Results

3.1. Distribution of the observed species occurrence along considered environmental variables

The observed frequency distribution of 29 macrobenthic species (selected as representative/dominating within the study area as described in Gogina et al., 2009—this volume) along considered environmental variables is presented as box-and-whisker plots in Figs. 3 and 4.

With respect to depth, while the range of species occurrence represented by "whiskers" generally corresponds to the limits defined by cumulative abundance distribution curves (see Gogina et al., 2010-this volume), analysis of "boxes" suggests differences between distributions of cumulative abundance and occurrence curves. E.g. 25th and 75th occurrence percentiles for *Capitella capitata* correspond to ca.12 and 21 m depth, however, the cumulative percentage of its abundance increases between depth values of 20 and 22.5 m from less than 0.2 to more than 0.8, indicating the range of maximum response against depth within the data limits. This is noteworthy, because this sudden increase in cumulative abundance is due to very few locations where the density of these polychaetes is up to two orders of magnitude higher than the average abundance in the region.

Though generally all the dominating species are mesohaline, and the salinity gradient is limited within the study area, the graphs clearly distinguish characteristic species tending to polyhalinity (*Halicryptus spinulosus*, *Terebellides stroemi* and *Trochochaeta multisetosa*) from species preferring the lower salinity within the considered factor range (*Cerastoderma glaucum*, *H. diversicolor*, *Heterochaeta costata*, *Hydrobia ulvae*, *Mya arenaria*, *Spio goniocephala*, *Travisia forbesii*, *T. benedii*). Other species take intermediate positions regarding salinity, with widest range of occurrence observed in our data, e.g. for *Bylgides sarsi* and *Diastylis rathkei*.

The influence of total organic content on the occurrence of species seems to be most evident. Some species indicate an exceptionally narrow range of occurrence at low values of organic content, e.g. *C. glaucum*, *H. diversicolor*, *T. forbesii*, and, to a smaller degree, *Astarte borealis*, *H. costata*, *M. arenaria*, and *S. goniocephala*. Others, such as *B. sarsi*, *C. capitata*, *H. spinulosus*, occur along the whole range of organic content.

3.2. Single-factor response curves

Logistic regression models predicting the probability of species occurrence as a function of each of the abiotic parameters considered were generated for each of the 29 species. Single-factor models are defined here following Eqs. (1) and (2). Table 1, Figs. 5 and 6 give an example of obtained response curves for a single environmental variable, showing the fitted logit curves for the 29 macrobenthic species in relation to water depth and to the total organic content. These are the factors which reveal the most distinct patterns in explaining the variability of benthic fauna distribution. Water depth actually represents a type of integral parameter ('proxy') that combines the effects of various habitat features that are either described by the available data, or not measured (or even not measurable). The response curves obtained were in general agreement with the observed distribution from Figs. 3 and 4, which supports the accuracy of the logistic modelling approach employed. Models predicting the probability of species occurrence relative to water depth show a concordance between 64.1 and 93.5%, relative to the total organic content—between 57.4 and 94.3% (Table 1). The variety of functional forms among the species response to water depth indicates that the occurrence of species is driven by species-specific sensitivity to that factor, with a non-constant rate of change characteristic for all of the species. The probability of occurrence of species such as *C. glaucum*, *M. arenaria*, *H. costata*, *H. diversicolor* and *T. forbesii* was higher at shallow depths. The sharp drop of the curve with increasing depth indicates that some species, e.g. *H. ulvae* and *P. elegans*, are highly sensitive to this parameter and avoid deep regions. Bell-shaped curves with an optimum at intermediate depths, as seen for *S. goniocephala*, *T. benedii* and *Gastrosaccus spinifer*, indicate the decrease of probability of occurrence both at the lower and upper ends of the factor range. Some species, e.g. *Mysella bidentata*,
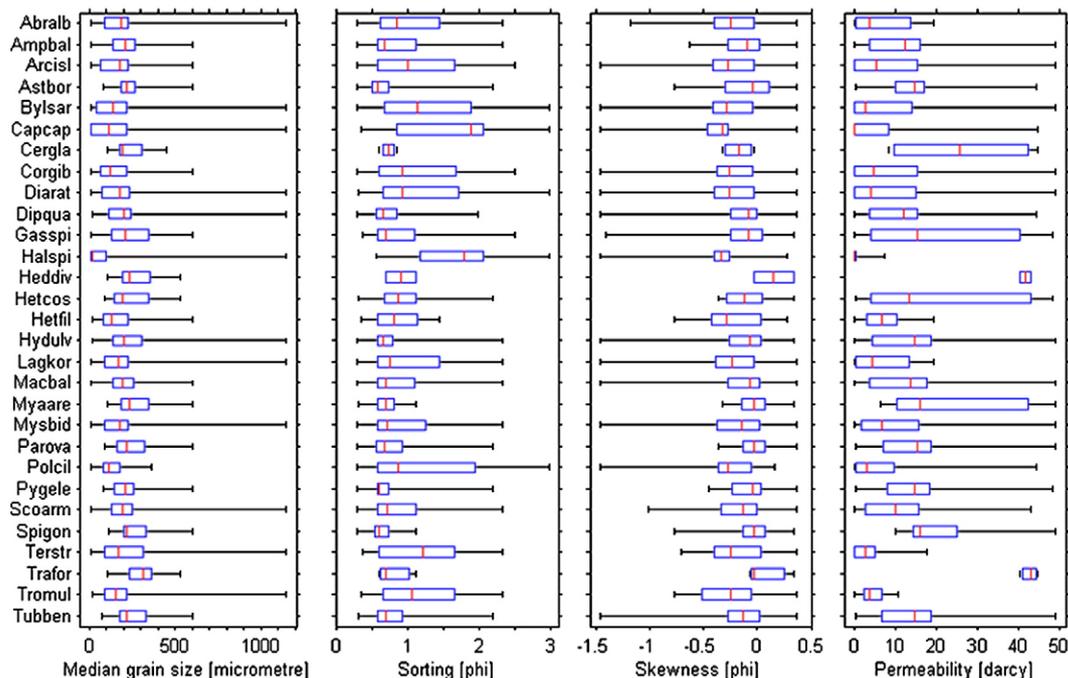

Fig. 4. Box-and-whisker plots for 29 macrobenthic species with respect to median grain size, grain-size parameters sorting and skewness, and permeability of surface sediments. For all the explanation see caption of Fig. 3.

Table 1
Logistic regression models of species occurrence.

Species	Model_d (x — water depth, m)	p(model)	Con %	Model_org (x — total organic content, %)	p(model)	Con %
Abralb	$p(x) = \exp(-4.43 + 0.21{*}x)/(1 + \exp(-4.43 + 0.21{*}x))$	<0.0001	76.0	$p(x) = \exp(-0.84 + 0.48x - 0.04x^2)/$ $(1 + \exp(-0.84 + 0.48x - 0.04x^2))$	0.0046	71.6
Arcisl	$p(x) = \exp(-20.2 + 1.97x - 0.04x^2)/$ $(1 + \exp(-20.2 + 1.97x - 0.04x^2))$	<0.0001	82.1	$p(x) = \exp(-0.62 + 2.32\log(x+1) - 0.75(\log(x+1))^2)/$ $(1 + \exp(-0.62 + 2.32\log(x+1) - 0.75(\log(x+1))^2))$	0.0161	65.5
Astbor	$p(x) = \exp(-42.3 + 4.55x - 0.12x^2)/$ $(1 + \exp(-42.3 + 4.55x - 0.12x^2))$	<0.0001	86.7	$p(x) = \exp(0.35 - 0.52x)/(1 + \exp(0.35 - 0.52x))$	<0.0001	71.0
Cergla	$p(x) = \exp(4.02 - 0.30x)/(1 + \exp(4.02 - 0.30x))$	<0.0001	84.0	$p(x) = \exp(-0.22 - 1.5x)/(1 + \exp(-0.22 - 1.5x))$	<0.0001	59.9
Corgib	$p(x) = \exp(-16.0 + 1.59x - 0.04x^2)/$ $(1 + \exp(-16.0 + 1.59x - 0.04x^2))$	<0.0001	73.9	$p(x) = \exp(-0.84 + 2.1\log(x+1) - 0.64(\log(x+1))^2)/$ $(1 + \exp(-0.84 + 2.1\log(x+1) - 0.64(\log(x+1))^2))$	0.0353	63.4
Macbal	$p(x) = \exp(73.9 - 6.0x + 0.12x^2)/$ $(1 + \exp(73.9 - 6.0x + 0.12x^2))$	<0.0001	92.2	$p(x) = \exp(2.62 - 0.47x)/(1 + \exp(2.62 - 0.47x))$	<0.0001	94.3
Myaare	$p(x) = \exp(8.04 - 0.45x)/(1 + \exp(8.04 - 0.45x))$	<0.0001	89.3	$p(x) = \exp(1.25 - 1.5x)/(1 + \exp(1.25 - 1.5x))$	<0.0001	87.3
Mysbid	$p(x) = \exp(-12.7 + 1.24x - 0.03x^2)/$ $(1 + \exp(-12.7 + 1.24x - 0.03x^2))$	<0.0001	75.7	$p(x) = \exp(-0.91 + 4.14\log(x+1) - 1.53(\log(x+1))^2)/$ $(1 + \exp(-0.91 + 4.14\log(x+1) - 1.53(\log(x+1))^2))$	<0.0001	76.3
Parova	$p(x) = \exp(-5.56 + 0.55x - 0.02x^2)/$ $(1 + \exp(-5.56 + 0.55x - 0.02x^2))$	0.0114	64.1	$p(x) = \exp(0.01 - 0.32x)/(1 + \exp(0.01 - 0.32x))$	<0.0001	68.6
Diarat	$p(x) = \exp(-9.06 + 1.11x - 0.03x^2)/$ $(1 + \exp(-9.06 + 1.11x - 0.03x^2))$	0.0003	80.8	$p(x) = \exp(0.13 + 4.54\log(x+1) - 1.5(\log(x+1))^2)/$ $(1 + \exp(0.13 + 4.54\log(x+1) - 1.5(\log(x+1))^2))$	0.0008	77.8
Hydulv	$p(x) = \exp(8.58 - 0.41x)/(1 + \exp(8.58 - 0.41x))$	<0.0001	89.0	$p(x) = \exp(1.08 - 0.28x)/(1 + \exp(1.08 - 0.28x))$	<0.0001	82.9
Gasspi	$p(x) = \exp(-4.22 + 0.59x - 0.02x^2)/$ $(1 + \exp(-4.22 + 0.59x - 0.02x^2))$	0.0004	69.0	$p(x) = \exp(0.09 - 0.18x)/(1 + \exp(0.09 - 0.18x))$	0.0002	69.5
Hetcos	$p(x) = \exp(2.01 - 0.2x)/(1 + \exp(2.01 - 0.2x))$	<0.0001	77.0	$p(x) = \exp(-1.96 - 0.29x)/(1 + \exp(-1.96 - 0.29x))$	0.0448	69.8
Tubben	$p(x) = \exp(-6.02 + 0.93x - 0.03x^2)/$ $(1 + \exp(-6.02 + 0.93x - 0.03x^2))$	<0.0001	78.5	$p(x) = \exp(1.23 - 0.78x)/(1 + \exp(1.23 - 0.78x))$	<0.0001	81.7
Ampbal	$p(x) = \exp(-11.8 + 1.25x - 0.03x^2)/$ $(1 + \exp(-11.8 + 1.25x - 0.03x^2))$	<0.0001	70.3	$p(x) = \exp(0.19 - 0.17x)/(1 + \exp(0.19 - 0.17x))$	0.0003	61.1
Bylsar	$p(x) = \exp(-3.75 + 0.31x)/(1 + \exp(-3.75 + 0.31x))$	<0.0001	84.1	$p(x) = \exp(0.15 + 1.31x)/(1 + \exp(0.15 + 1.31x))$	<0.0001	83.0
Capcap	$p(x) = \exp(7.05 - 0.76x + 0.02x^2)/$ $(1 + \exp(7.05 - 0.76x + 0.02x^2))$	0.0003	66.2	$p(x) = \exp(-1.41 + 0.16x)/(1 + \exp(-1.41 + 0.16x))$	<0.0001	68.6
Dipqua	$p(x) = \exp(-12.4 + 1.22x - 0.03x^2)/$ $(1 + \exp(-12.4 + 1.22x - 0.03x^2))$	<0.0001	71.2	$p(x) = \exp(0.12 - 0.14x)/(1 + \exp(0.12 - 0.14x))$	0.0013	57.4
Heddiv	$p(x) = \exp(7.5 - 0.56x)/(1 + \exp(7.5 - 0.56x))$	<0.0001	93.5	$p(x) = \exp(-0.38 - 3.62x)/(1 + \exp(-0.38 - 3.62x))$	<0.0001	86.6
Hetfil	$p(x) = \exp(-20.6 + 1.91x - 0.04x^2)/$ $(1 + \exp(-20.6 + 1.91x - 0.04x^2))$	<0.0001	79.4	$p(x) = \exp(1.94 - 0.49\log(x+1) + 0.04(\log(x+1))^2)/$ $(1 + \exp(1.94 - 0.49\log(x+1) + 0.04(\log(x+1))^2))$	0.0005	73.6
Lagkor	$p(x) = \exp(-3.21 + 0.17x)/(1 + \exp(-3.21 + 0.17x))$	<0.0001	72.7	$p(x) = \exp(-1.2 + 4.07\log(x+1) - 1.48(\log(x+1))^2)/$ $(1 + \exp(-1.2 + 4.07\log(x+1) - 1.48(\log(x+1))^2))$	<0.0001	73.7
Polcil	$p(x) = \exp(-8.07 + 0.76x - 0.02x^2)/$ $(1 + \exp(-8.07 + 0.76x - 0.02x^2))$	0.0020	68.5	–	–	–
Pygele	$p(x) = \exp(5.99 - 0.3x)/(1 + \exp(5.99 - 0.3x))$	<0.0001	82.2	$p(x) = \exp(0.69 - 0.5x)/(1 + \exp(0.69 - 0.5x))$	<0.0001	74.4
Scoarm	$p(x) = \exp(3.69 - 0.12x)/(1 + \exp(3.69 - 0.12x))$	0.0003	68.3	$p(x) = \exp(2.18 - 0.28x)/(1 + \exp(2.18 - 0.28x))$	<0.0001	81.8
Spigon	$p(x) = \exp(-13.1 + 1.82x - 0.06x^2)/$ $(1 + \exp(-13.1 + 1.82x - 0.06x^2))$	<0.0001	79.7	$p(x) = \exp(0.91 - 2.49x)/(1 + \exp(0.91 - 2.49x))$	<0.0001	80.1
Terstr	$p(x) = \exp(-5.28 + 0.22x)/(1 + \exp(-5.28 + 0.22x))$	<0.0001	78.1	$p(x) = \exp(-3.14 + 4.06\log(x+1) - 1.29(\log(x+1))^2)/$ $(1 + \exp(-3.14 + 4.06\log(x+1) - 1.29(\log(x+1))^2))$	0.0006	71.7
Trafor	$p(x) = \exp(3.89 - 0.34x)/(1 + \exp(3.89 - 0.34x))$	<0.0001	87.4	$p(x) = \exp(0.97 - 7.78x)/(1 + \exp(0.97 - 7.78x))$	<0.0001	92.3
Tromul	$p(x) = \exp(-5.77 + 0.24x)/(1 + \exp(-5.77 + 0.24x))$	<0.0001	80.5	$p(x) = \exp(-4.49 + 6.32\log(x+1) - 2.06(\log(x+1))^2)/$ $(1 + \exp(-4.49 + 6.32\log(x+1) - 2.06(\log(x+1))^2))$	<0.0001	80.1
Halspi	$p(x) = \exp(-5.79 + 0.23x)/(1 + \exp(-5.79 + 0.23x))$	<0.0001	78.0	$p(x) = \exp(-2.15 + 0.35x)/(1 + \exp(-2.15 + 0.35x))$	<0.0001	91.2

$p(x)$ is the estimated probability that the species occurs. Species abbreviation is as in Fig. 3. p(model) indicates the model significance level, tested using Wald test based on the χ^2-distribution ($p < 0.05$), and Con % is the percentage of concordant pairs.

Dipolydora quadrilobata and *D. rathkei*, showed a clear optimum towards the higher end of the depth range, differing in the position of their optimum and in their tolerance towards the lower end of the factor range. *Abra alba*, *H. spinulosus*, *T. multisetosa*, *T. stroemi* and, to a lesser extent, *B. sarsi* had the highest probability of occurrence at the deepest zones. A broad tolerance for water depth was indicated, for instance for *S. armiger*. The inverse form of the Gaussian response towards that factor shown by *Macoma balthica* and *C. capitata* can be interpreted owning to their cosmopolite behaviour in the context of depth within the studied ranges and dependency on other variables. It may also be regarded as a realistic form of the response function, when considered as a slice of the bimodal response explained by the competition exclusion in the middle of broad tolerance to an environmental gradient.

Most macrobenthic species showed a high probability of occurrence at the lower end of the total organic content range (e.g. *M. balthica*, *M. arenaria*, *D. quadrilobata*) with only two species, *C. capitata* and *H. spinulosus*, having an increasing probability of occurrence with the increase of total organic content. Species such as *M. bidentata*, *A. alba*, *Lagis koreni*, *T. stroemi* showed skewed unimodal curve forms with an optimum tending towards the lower end of the factor range. The models developed for *Heteromastus filiformis*, and to a lesser extent *A. islandica* and *Corbula gibba*, have indicated that the occurrence of these species is not sensitive to a wide range of sediment total organic content.

3.3. Modelling—predicting the probability of species occurrence

The multimodel inference technique was applied to five selected species (*A. islandica*, *H. diversicolor*, *P. elegans*, *T. benedii* and *S. armiger*), exemplarily chosen as representatives for soft and sandy-bottom regions of the investigation area. Maximum densities and frequencies of occurrence for selected species are given in Table 2.

Table 3 provides maximum-likelihood estimates of logistic regression parameters for species response surfaces to each of the abiotic factors. Single-factor models are defined as described in Section 3.2. For *H. diversicolor* models of the factors sorting, skewness and permeability were not supported by a sufficient value of occurrence events (the species was recorded only at 2 stations out of 78 covered with data on corresponding abiotic variables), and thus
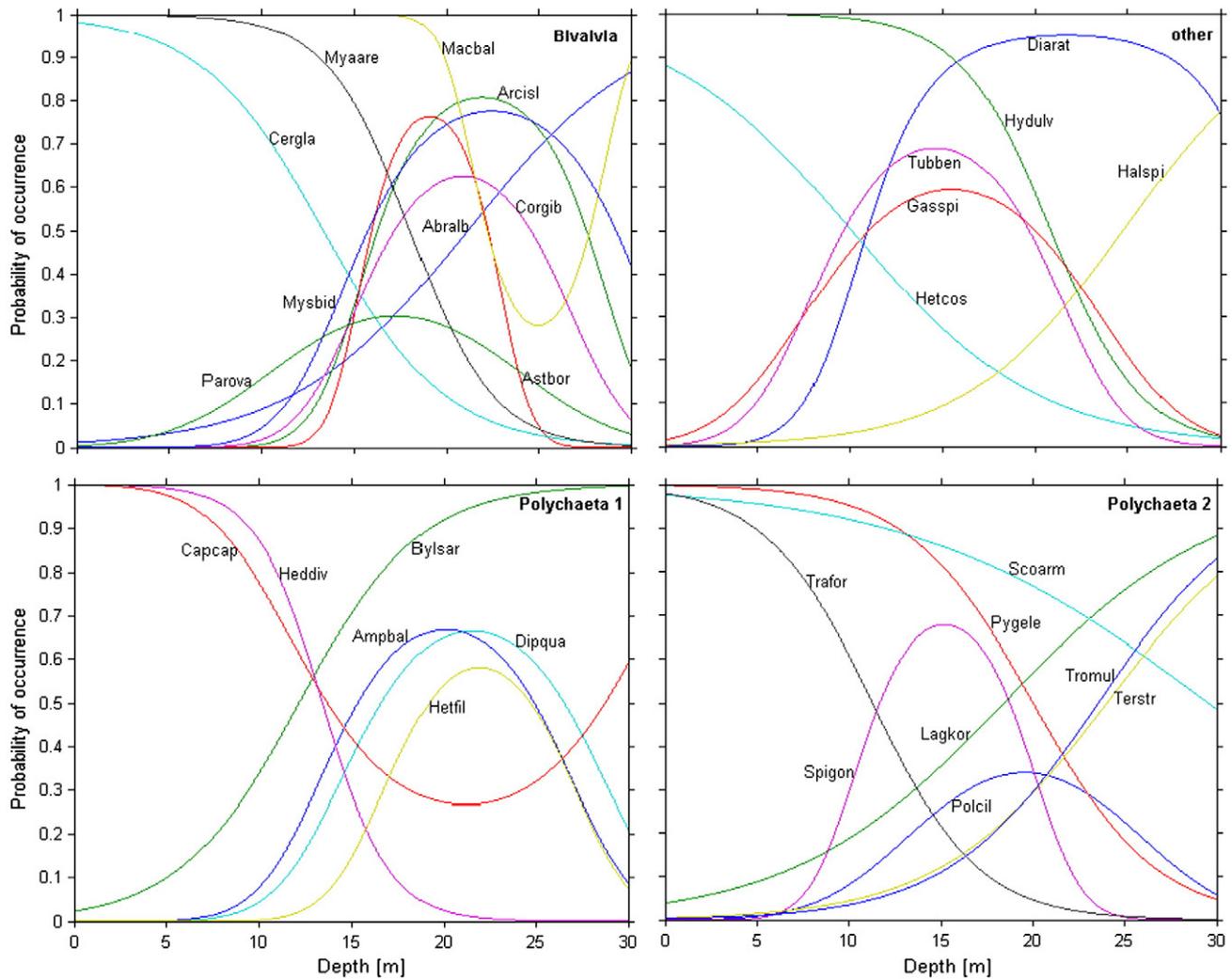
Fig. 5. Logistic regression models predicting probability of occurrence for 29 macrobenthic species relative to water depth. Species abbreviation as in Fig. 3.

models for these factors were not included in the set for model averaging defined for this species.

Akaike weights derived for two model sets for each species—one set including the single-factor model of total organic content used as predictor and one without it—are presented in the lower part of the Table 3. To assess the accuracy of prediction the values of concordance were estimated.

Regarding Akaike weight values, the factor dominating in the explanation of variability of *A. islandica* was depth. Some significant part of distribution variability was also explained by permeability, while all other environmental parameters considered explain a small part of uncertainty, though models with oxygen, median grain size and skewness did not differ significantly from the random chance. The water depth model also provided the greatest strength of evidence for *H. diversicolor*, followed by the model of total organic content, which possessed a considerably lower weight in the model averaging. Factors dominating in the explanation of variability in distribution of *P. elegans* were total organic content and depth. Some significant part of distribution variability was also explained by sorting. When it was not possible to take the total organic content into account salinity also explained a noticeable part of variance. *S. armiger* indicated a well-defined response to salinity, to total organic content, (when it was possible to use this factor as predictor) and to sorting. Among the considered abiotic factors, the variability of distribution of *T. benedii* was almost entirely explained by depth, with some information hidden in the gradient of total organic content.

In Fig. 7 the results of the prediction with application of the described technique are mapped and compared to the observed species abundance data. The visual inspection reveals that higher values of species abundance prevail where higher probabilities of species occurrence are modelled.

For *A. islandica* the preferable conditions for the settlement within an area and considered ranges of environmental factors included high values of depth (approximately >18 m) and salinity as well as wide ranges of total organic content and sedimentological parameters. On the contrary, *H. diversicolor* preferred low-saline regions shallower than 18 m. Therefore, these species seem to present the biological antipodes in our study area. *P. elegans* revealed disinclination to regions with high total organic content and preferred more permeable substrate (see Fig. 7).

4. Discussion

As pointed out by Praca et al. (2008) the use of temporally heterogeneous data confounds the effect of interannual variations in species occurrence and environmental conditions. However, our objectives were to attempt a general description of species habitats and to investigate the predictive abilities of the modelling technique at the selected spatial scale. Models determining the distribution of exemplary macrofaunal species common for the south-western Baltic Sea from changing environmental variables, such as depth, salinity and sediment characteristics, have been successfully developed. These
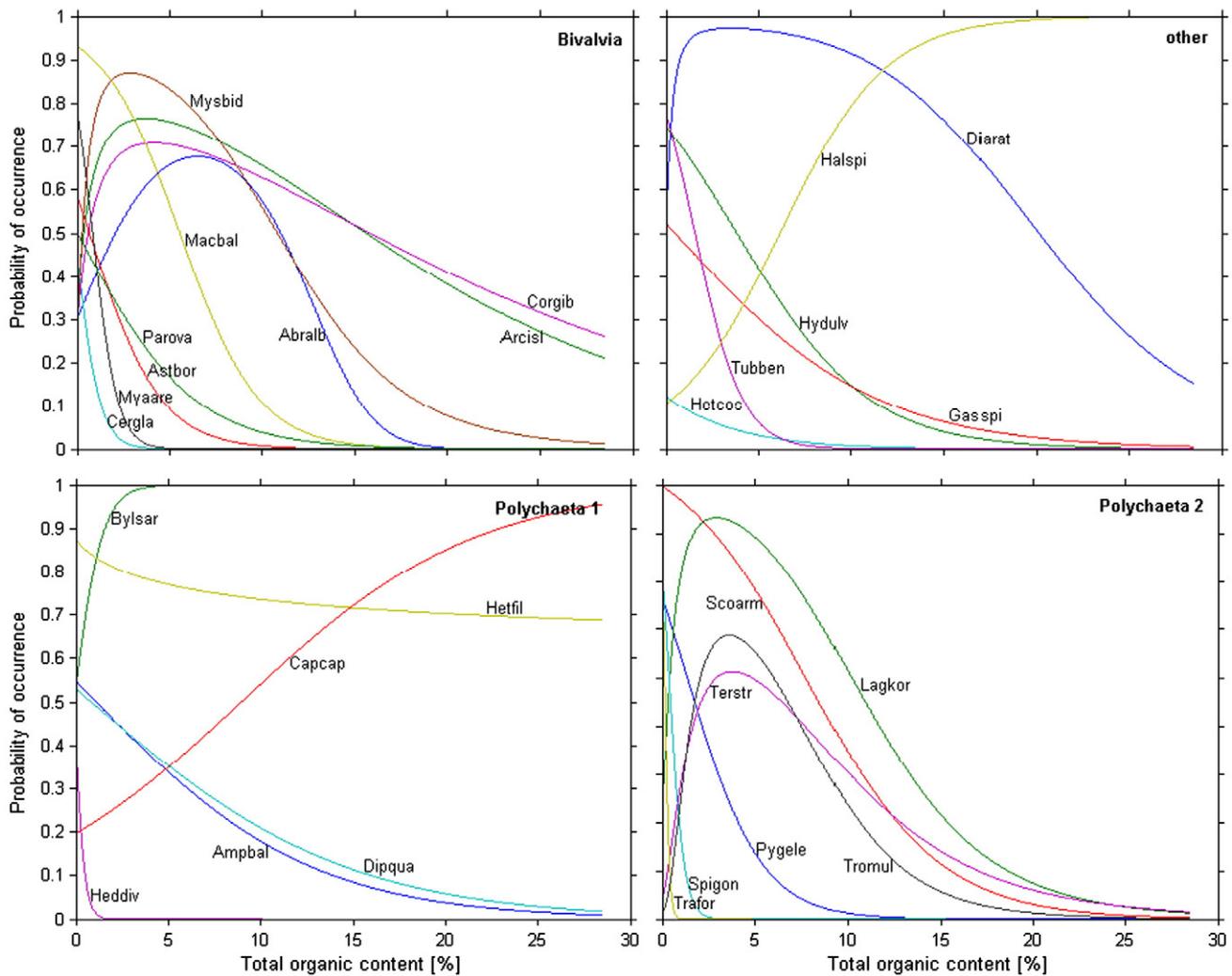
Fig. 6. Logistic regression models predicting probability of occurrence for 29 macrobenthic species relative to sediment total organic content. Species abbreviation as in Fig. 3.

models are considered to be reasonably general, i.e. they are able to resolve the distribution of species over a comparatively large spatial scale (ca. >100 m) as opposed to a metre-by-metre basis. Confirming the postulates of Ysebaert et al. (2002), Thrush et al. (2003), and Ellis et al. (2006), we have found that logistic regression is a useful and relatively transparent approach to predict the response of species occurrence as a function of various environmental conditions.

Beside the widely used generalised linear models (GLM; Guisan et al., 1998; Wisz and Guisan, 2009), other techniques employed in order to unravel the complexity of interactions between distribution and environmental factors are generalised additive models (GAM; Yee and Mitchell, 1991), classification and regression trees (CART; Moore et al. 1991), artificial neural networks (ANN; Fitzgerald and Lees, 1992) and multivariate adaptive regression splines (MARS; Friedman, 1991). The generation of 'potential habitat distribution maps' is stated to be

among the predictive modelling goals (Munoz and Felicisimo, 2004), therefore the convenience of cartographic implementation is crucial. Such properties are mainly assigned to generalised methods as well as to the MARS method, that builds complex regression models by fitting piecewise linear regressions. A successful development of habitat models includes both geographical and ecological discrimination of species. Since only very few species have been studied in detail in terms of their dynamic responses to environmental changes, static distribution modelling often remains the only approach for studying the possible consequences (Woodward and Cramer, 1996). For the regions where the distribution is strongly and directly coupled to physiochemical processes statistical models are also capable of satisfactorily predicting the species distribution (Ellis et al., 2006). This is the case for our study area as was concluded in Gogina et al., 2010-this volume. Among the objectives we had for selecting the relatively simple and intuitive GLM approach was its ability to construct a parsimonious model that strikes a balance between bias and variance; identify the relative importance of the predictor variables; explore and interpret the response of the species to each predictor; estimate the uncertainty associated with parameter estimates; predict the probability of observing the species (rather than predicting binary presence–absence) and explore spatially explicit patterns of uncertainty in predictions. Decisive was its availability for direct interpretation of the results, accompanied by its reputation for providing a competitively high accuracy (e.g. Praca et al., 2008; Wisz and Guisan, 2009). The comparable method that employed logistic regressions together with weighted overlay was also successfully applied and

Table 2
Maximum densities and frequencies of occurrences of selected species within the data set.

Species	Maximum density, ind/m²	Frequency of occurence, %
A. islandica	622	52.4
H. diversicolor	667	20.7
P. elegans	11459	57.2
S. armiger	1684	77.9
T. benedii	1469	44.7

Table 3
The maximum-likelihood estimates of logistic regression parameters for response surfaces of selected species to each of abiotic factors and normalized Akaike's weights derived for the model set including single-factor model of total organic content as predictor and the model set without it.

Parameter estimates		A. islandica	H. diversicolor	P. elegans	S. armiger	T. benedii
d	b_0	−20.1829	7.5037	5.9871	3.6877	−6.0225
	b_1	1.9694	−0.5584	−0.2993	−0.1247	0.9314
	b_2	−0.0449	–	–	–	−0.03179
	p(model)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
s	b_0	−2.8957	3.3046	5.0447	6.6534	5.5642
	b_1	0.1868	−0.3310	−0.3101	−0.3215	−0.3635
	b_2	–	–	–	–	–
	p(model)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
o2	b_0	0.5838	−3.9298	0.0453	4.1259	−1.1231
	b_1	−0.0904	0.7660	0.0183	−1.1573	0.1690
	b_2	–	−0.0520	–	0.1029	–
	p(model)	0.1149[**]	0.0946[**]	0.7473[**]	0.0192	0.0085
org	b_0	−0.6234[*]	−0.3851	0.6887	2.1783	1.2297
	b_1	2.3222[*]	−3.6250	−0.5053	−0.2814	−0.7844
	b_2	−0.7467[*]	–	–	–	–
	p(model)	0.0161	<0.0001	<0.0001	<0.0001	<0.0001
kg	b_0	0.7970	−5.3292	−2.7251	−0.3493	−2.5964
	b_1	−0.0023	0.0207	0.0288	0.0087	0.0177
	b_2	–	−0.00003	−0.00004	–	−0.00002
	p(model)	0.0647[**]	0.0215	<0.0001	<0.0001	<0.0001
so	b_0	1.4313	–	2.2991	2.8781	1.0022
	b_1	−0.4871	–	−2.8793	−2.0383	−1.4740
	b_2	–	–	–	–	–
	p(model)	0.2075	–	<0.0001	<0.0001	0.0008
sk	b_0	0.7319	–	0.1435	0.9031	−0.2341
	b_1	−0.3262	–	3.7708	1.9679	1.6366
	b_2	–	–	–	–	–
	p(model)	0.6379[**]	–	<0.0001	0.0043	0.0285
kKM	b_0	0.5043	–	−2.7986	−1.0221	−2.0592
	b_1	0.1260	–	0.3755	0.5344	0.2287
	b_2	−0.0031	–	−0.0068	−0.0118	−0.0036
	p(model)	0.0300	–	<0.0001	<0.0001	0.0128

Akaike's weights estimated for two model sets

Set	A. islandica Full	No org	H. diversicolor Full	No org	P. elegans Full	No org	S. armiger Full	No org	T. benedeni Full	No org
d	0.7591	0.7669	0.8646	0.9939	0.4431	0.9885	0.0000	0.0000	0.9902	0.9999
s	0.0188	0.0190	0.0003	0.0003	0.0000	0.0001	0.5861	0.9450	0.0001	0.0001
o2	0.0517	0.0522	0.0046	0.0053	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
kg	0.0197	0.0199	0.0004	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
so	0.0411	0.0415			0.0025	0.0056	0.0341	0.0549	0.0000	0.0000
sk	0.0208	0.0210			0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
kKM	0.0786	0.0794			0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
org	0.0101		0.1301		0.5543		0.3798		0.0097	

Concordance of merged model results, %				
82.86	94.59	81.24	75.78	77.75

The environmental factors notation is as follows: org—for total organic content, kg—for median grain size, o2—for oxygen concentration, s—for salinity, d—for water depth, so—for sorting, sk—for skewness, and kKM—for permeability.

  [*] Estimated for log-transformed total organic content variable, that indicated a better model fit.

 [**] No significant relation of species occurrence probability to the factor could be derived.

tested on a similar spatial scale by Glockzin and Zettler (2008b), who modelled habitat suitability maps for exemplary species of the Pomeranian Bay. In their work they emphasised not only the importance of spatial scale, but in-depth knowledge of species autecology used in such studies and models. This is especially important since top-down correlations between the macrobenthos and the environmental variables represented by response models do not always reflect direct cause and effect relationships, since many environmental variables co-vary (van der Wal et al., 2008).

It is essential to differentiate between the questions that the two types of model solutions performed serve to answer. Each single-factor model alone is able to answer the question of how certain species respond to changes of this separate factor, e.g. describe it as euryoecious or stenoecious organism. Thrush et al. (2005) acknowledge that simple models may fit well the purposes of response management, for instance

by defining the sufficiently consistent general pattern of which species prefer muddy, intermediate or sandy sediment types across scales, so that a rank order of species can be developed. Such classification can be used to interpret changes in the distribution and abundance of monitored species or incorporated with other predictions of habitat change, and used in environmental risk assessment. The multimodel inference serves to predict the distribution of species within the common limits of combined environmental factors, e.g. to fill out the lack of information in the locations, where no observational data is available. Thus among the applications of statistical modelling exercises such as the one presented in this paper are: testing the hypotheses about the ranges of species distribution along environmental gradients and benthic stress impacts (for instance, the Pearson and Rosenberg (1978) model that reported a gradual loss of species as the degree of stress increased over space and/or time driven by the replacement pattern
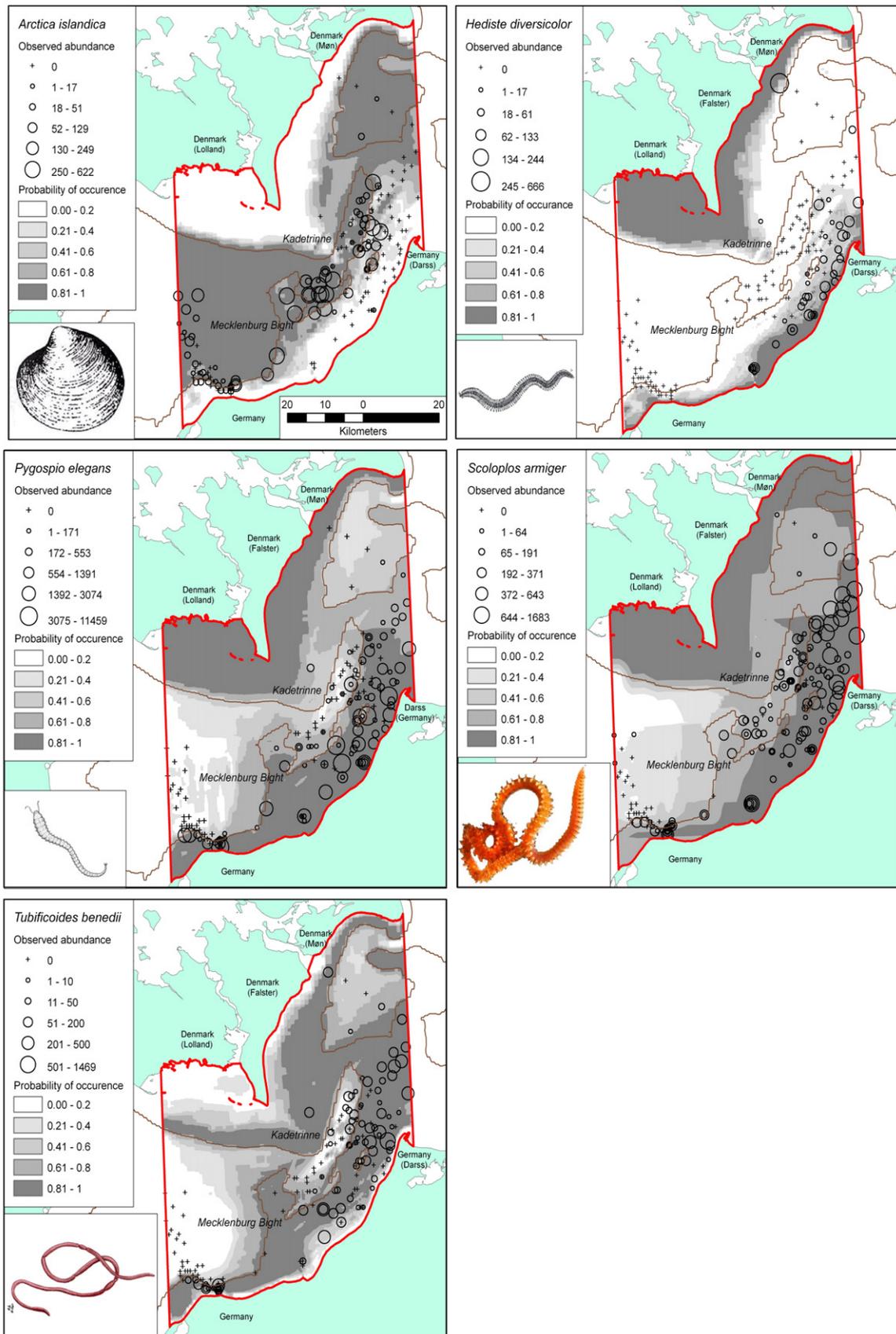
Fig. 7. Modelled probability of species occurrence determined using Akaike's weghts for multimodel inference. Circles indicate the observed distribution with their size corresponding to the value of abundance density (abundance, ind/m^2). The intensity of shading corresponds to increasing probability of occurrence.

defined by different tolerance of species to stress); generalization of habitat suitability maps that predict the specific ecological potential of a habitat (with limitations defined by the data analyzed) which can be considered in marine spatial planning and conservation management (Degraer et al., 2008); and predicting the possible consequences of habitat changes (either natural or antropogenic). When accompanied by other relevant developments and investigations possible use can be found in the comparison the of species' spatial distribution at different scales (e.g. Thrush et al., 2005). The coupling of species ecological functions (filtration rates, bioturbation modes, etc.) with the results of such modelling exercises via biochemical or sediment transport models may help to assess the ecosystem functioning (e.g. Bobertz et al., 2009).

Considerable differences of the importance of various environmental conditions were highlighted by the prediction maps for selected species (Fig. 7). All obtained response surfaces indicated a relatively high percentage of concordance, though a more robust validation of derived models is still to be executed and will unambiguously require external data. The essential autecological features found in literature sources and evidences derived from modelling results are summed up in Table 4.

The discussion above illustrates, that to distinguish between colonization types, prior biological knowledge must support the statistical examination. Bonsdorff (2006) states the "ecological age" of the present Baltic Sea ecosystem to be only about 8000 years, resulting in still ongoing primary succession processes and numerous ecological niches remaining available for immigration. This seems to be evidenced by the fact, that

Table 4

Essential autecological features of exemplary species and its evidences based on modelling results (Fig. 7).

Species	Essential autoecological features	Evidence from the modelling results
A. islandica[a]	Habitat: arctic-boreal bivalve, found in high concentrations at 25–61 m (recorded at 8 to 256 m), in the Baltic Sea found in high saline areas (at depths of 16 to 30 m, eastern distribution limit—Arkona basin). Substrate: firm sediments, medium to fine grain sand, sandy mud, silty sand. Oxygen: resistant to oxygen deficiency; can burrow into substrate and respire anaerobically for up to week. Salinity: found at oceanic salinities, in the lab resists salinities as low as 22 psu. Feeding mode: suspension feeder Notes: occur in dense beds over level bottoms. Extremely long life-span (ca. 80 years in the Baltic and about 405 in the Atlantic). Strong recruitment of the species in the Mecklenburg Bight during the past decades and probably the displacement of recruitment zone from below 20 to 15–20 m depth.	Confirm: wide ranges of organic content and sediment parameters, high values of depth (>18 m) and salinity define suitable habitat. Very unlikely to occur in the shallow coastal zone, influenced by freshwater runoff. Highest densities found at depth between 15 and 20 m, whereas highest probabilities of occurrence coincide with the regions slightly below 20 m. As densities of species are known to correlate negatively with size/age this confirms the displacement of recruitment zone. Not recorded in the region at salinities below 10 psu, highest densities found between 14 and 16 psu (much lower than oceanic salinity), but occurrence increases with salinity. The preference of dense beds reflected in low likelihood of occurrence on coarse sediments and at high permeability values. Oxygen concentration had no strong effect.
H. diversicolor[b]	Habitat: inhabits shallow marine and brackish waters. In the Baltic Sea prefers enclosed bays and lagoons with smooth slopes and absence of strong currents. Substrate: sandy mud but also gravels, clays, even turf. Oxygen: able to survive drastic conditions of hypoxia. Salinity: euryhaline, lowest limit of salinity (determined through its larvae development) of about 5 psu. Feeding mode: carnivore, scavenger, filter feeder and a surface deposit feeder, however also having the ability to live as a suspension feeder Notes: generally endobethic species, able to penetrate the substrate up to depths of about 30 cm.	Confirm: does not find its optima within the study area (greatly higher densities are recorded in the enclosed coastal estuarine regions of the Baltic Sea). Preference of low-saline regions shallower than 18 m within the investigated spatial and environmental limits. Among considered abiotic factors the largest effect size featured water depth, followed by total organic content. Species also responded here noticeably to oxygen, grain size and salinity, but with relatively low weights in the model averaging inference. Its opportunistic abilities seem to allow species to avoid locations where it may be disturbed by other competitive species.
P. elegans[c]	Habitat: mainly lower superlittoral down to depth of 100 m, in the Baltic Sea from ca 5 m down to 20 m. Substrate: fine to medium sands Oxygen: highly sensitive to hypoxia, hardly ever found in areas affected by the phenomena Salinity: from 2 psu to hypersaline pools Feeding mode: deposit and filter feeder Notes: features the penetration depth of 4–6 cm. Evolves opportunistic life strategies: the species is able to rapidly re-colonize defaunated substrates. Regarded as a semi-sedentary species. Avoids strong currents. Negative response to organic enrichment.	Confirm: the model with highest weight in the model averaging was the function of total organic content. Response indicates almost no chance of finding the species in sediments with organic content exceeding 5%. Water depth (containing the effect of currents) indicated slightly smaller but comparable influence on species distribution. Typically favors shallower waters, but the response against increasing water depth is not as rapid as in case of H. diversicolor. Surprisingly, only sorting had non-zero (yet low) model averaging weight among sediment factors, perhaps due to the limiting power of covarying organic content.
S. armiger[d]	Habitat: cosmopolitan species with intertidal as well as in the subtidal occurrence encountered in all zoogeographic regions. In the Baltic Sea, e.g. in the Mecklenburg Bight, species habitat is limited from 5 to ca. 30 m, with highest abundance values found around 10 m depth and absence in purely silt regions. Substrate: muddy sands and mud Oxygen: resistant to hypoxia down to 0.5 ml/l Salinity: cannot survive at salinities lower than 10.5 psu Feeding mode: deposit feeder Notes: mobile non-selective species burrowing freely through sediments, that builds non-persisting tubes in the sediment down to 15 cm.	Confirm: among the considered abiotic factors only salinity, total organic content and, to small extend, sorting explained the distribution pattern of S. armiger in multimodel inference. Bleidorn et al. (2006) revealed that S. armiger represents a species complex and is not a cosmopolitan species that explains the inconsistency between general environmental setting found in literature and its observed and predicted distribution limits within the study area. Thus, there is a negative response along increase of both salinity and organic content in terms of both occurrence and abundance within the region (though observed salinities have the lower limit of 8 psu).
T. benedii[e]	Habitat: ubiquitous marine oligochaete that dominates in coastal areas. It is often typified as 'opportunist' that is adapted to the rapid environmental fluctuations and harsh conditions in estuaries. Substrate: fine organic- enriched sediments Oxygen: resistant to hypoxia Salinity: found in meso- to euhaline waters. Feeding mode: deposit feeder Notes: very successful adaptive strategies in sulfidic benthic environments. Able to penetrate into the substrate up to 10 cm.	Confirm: the prediction map for this species is derived nearly exclusively from water depth as a proxy, with highest probability of its occurrence around 10 to 20 m, and solitary occurrence events below 23 m. In the investigation area this species never seems to settle on the truly mud substrate and does not favor high organic contents as such conditions here coincide with undesirable water depth (that probably comprises a range of other limiting factors for this animal).

[a] Cargnelli et al., 1999; Zetter and Röhner, 2004; Wanamaker et al., 2008.
[b] Nithart et al., 1998; Scaps, 2002; Kristensen, 2001; Henning et al., 2004; Zetter and Röhner, 2004; van der Wal et al. 2008.
[c] Fauchald and Jumars, 1979; Morgan et al., 1999; van der Wal et al. 2008.
[d] Fauchald and Jumars, 1979; Zettler et al., 2000; Bleidorn et al. 2006.
[e] Dubilier et al., 1994; Giere, 2006.

most of the representative species selected for predictive modelling in the study area are named as opportunists.

The presented models should be best applicable for species that find their optima within the observed ranges of environmental factors. *Inter alia*, such models can also reflect the tendencies of an opportunistic species distribution with satisfactory result, e.g. *H. diversicolor* that reaches only low abundances within the study area compared to its average density in more favourable in-shore estuarine conditions. For such species the modelling results reveal the best of available conditions in contrast to their optima. However, it should be mentioned that, as for all statistical methods, the inference about the model selection uncertainty is conditional on both data and the full set of a priori models considered (Burnham and Anderson, 2004). Coudun and Gégout (2006) suggest a general minimum value of 50 occurrences for species to derive acceptable ecological response curves with logistic regression. The available data for *H. diversicolor* slightly drops out of this condition with 43 cases of occurrence within total 208 observations.

A high variance in the species distribution that usually causes an impediment for modelling, is often not a sampling error or random "noise", but rather the mechanistic consequence of shifts between limiting resources or other effects and factors (e.g. intra- and inter-species competition, predation, mortality or dispersal). The abundance of species may be very low under favourable conditions if, for some reason, the number of propagule is very low or species never even reach a given area. This natural phenomena of a species failing to colonize all areas where it could potentially thrive (Huston, 2002) may partly explain some inconsistencies between the predicted high probability of species occurrence and its factual absence according to observations. Under optimal conditions species can reach maximal reproduction rates and maximal abundances, but macrobenthic surveys often show an entirely different reality: species and communities are distributed rather patchily and often the relatively smooth structure of abiotic gradients and other characteristics can increase statistical uncertainty and blur the picture.

5. Conclusions and outlook

For all exemplary species the used technique provided the results of distribution prediction based on environmental data with a fairly satisfactory accuracy. The method combining the parsimony of single-factor logistic regression models with an AIC solution of multimodel inference can therefore be recommended for modelling of the habitat suitability for benthic species, at least on comparable spatial scales and environmental gradients.

It is important to point out that only a fragment of the whole factor network responsible for the macrozoobenthos distribution was covered in this study. Consideration of additional variables may reveal more insightful analysis. Results of the present study may only be interpreted within the context of the momentary state of benthic communities.

In the present study only the probability of benthic species occurrence based on presence/absence data was modelled, allowing the prediction of species distribution derived from environmental data. However, the density of species is a considerably more informative figure. Modelling of benthic species abundance as a function of abiotic parameters is likely to be performed applying 'factor ceiling' (Thrush et al., 2003) or quantile regression (Cade and Noon, 2003) methods.

Acknowledgments

References

Austin, M.P., 1987. Models for the analysis of species' response to environmental gradients. Vegetatio 69, 35–45.

Blackburn, T., Lawton, J.H., Perry, J.N., 1992. A method of estimating the slope of upper bounds of plots on body size and abundance in natural animal assemblages. Oikos 65, 107–112.

Bleidorn, C., Kruse, I., Albrecht, S., Bartolomaeus, T., 2006. Mitochondrial sequence data expose the putative cosmopolitan polychaete *Scoloplos armiger* (Annelida, Orbiniidae) as a species complex. BMC Evol. Biol. 6, 47.

Bobertz, B., Harff, J., 2004. Sediment facies and hydrodynamic setting: a study in the south western Baltic Sea. Ocean. Dyn. 54, 39–48.

Bobertz, B., Harff, J., Bohling, B., 2009. Parameterisation of clastic sediments including benthic structures. J. Mar. Syst. 75 (3–4), 371–381.

Bonsdorff, E., 2006. Zoobenthic diversity-gradients in the Baltic Sea: continuous post-glacial succession in a stressed ecosystem. J. Exp. Mar. Biol. Ecol. 330, 383–391.

Bourget, E., Fortin, M.-J., 1995. A commentary on current approaches in the aquatic sciences. Hydrobiologia 300 (301), 1–16.

Burnham, K.P., Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. 33–2, 261–304.

Cade, B.S., Noon, B.R., 2003. A gentle introduction to quantile regression for ecologists. Front. Ecol. Environ. 1, 412–420.

Cargnelli, L.M., Griesbach, S.J., Packer, D.B., Weissberger, E., 1999. Ocean quahog Arctica islandica, life history and habitat characteristics. NOAA Tech Memo NMFS-NE-148.

Clarke, K.R., Warwick, R.M., 1994. Change in Marine Communities: An Approach to Statistical Analysis and Interpretation. Plymouth Marine Laboratory, Plymouth, UK.

Coudun, Ch., Gégout, J.-C., 2006. The derivation of species response curves with Gaussian logistic regression is sensitive to sampling intensity and curve characteristics. Ecol. Model. 199, 164–175.

Degraer, S., Verfaillie, E., Willems, W., Adriaens, E., Van Lancker, V., Vincx, M., 2008. Habitat suitability as a mapping tool for macrobenthic communities: an example from the Belgian part of the North Sea. Cont. Shelf Res. 28 (3), 369–379.

Dubilier, N., Giere, O., Grieshaber, M.K., 1994. Concomitant effects of sulfide and hypoxia on the aerobic metabolism of 241 the marine oligochaete *Tubificoides benedii*. J. Exp. Zool. 269, 287–297.

Ellis, J., Ysebaert, T., Hume, T., Norkko, A., Bult, T., Herman, P., Thrush, S., Oldman, J., 2006. Predicting macrofaunal species distribution in estuarine gradients using logistic regression and classification systems. Mar. Ecol. Prog. Ser. 316, 69–83.

ESRI, 2003. ESRI Data & Maps 2003. Media Kit. ESRI, Redlands, California, USA.

Fauchald, K., Jumars, P.A., 1979. The diet of worms: a study of polychaete feeding guilds. Oceanogr. Mar. Biol. Annu. Rev. 17, 193–284.

Fitzgerald, R.W., Lees, B.G., 1992. The application of neural networks to the floristic classification of remote sensing and GIS data in complex terrain, in: American Society of Photogrammetry and Remote Sensing (Eds.), Proceedings of the XVII Congress ASPRS, Bethesda, MD, pp. 570–573.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–141.

Giere, O., 2006. Ecology and biology of marine Oligochaeta—an inventory rather than another review. Hydrobiologia 564, 103–116.

Glockzin, M., Zettler, M.L., 2008a. Spatial macrozoobenthic distribution patterns in relation to major environmental factors—a case study from the Pomeranian Bay (southern Baltic Sea). J. Sea Res. 59, 144–161.

Glockzin, M., Zettler, M.L., 2008b. Which factors cause benthic diversity in a low salinity environment? An approach to species response. Book of Abstracts, MARBEF World Congress on Biodiversity, Valencia, Spain, Session 1.10, p. 18.

Gogina, M.A., Zettler, M.L., Glockzin, M., 2010. Distribution of benthic macrofaunal coenoses in the western Baltic Sea with regard to near-bottom environmental parameters. Part 1. Casual analysis. J. Mar. Syst. 79 (1–2), 112–123.

Guisan, A., Theurillat, J.-P., Kienast, F., 1998. Predicting the potential distribution of plant species in an alpine environment. J. Veg. Sci. 9, 65–74.

Guisan, A., Weiss, S.B., Weiss, A.D., 1999. GLM versus CCA spatial modelling of plant species distribution. Plant Ecol. 143, 107–122.

Henning, D., Zauke, G.-P., Stief, P., Vareschi, E., 2004. Integrated effects of marine benthic photosynthetically active microorganisms and a bioturbating polychaete, *Hediste diversicolor* (O. W. Müller), on oxygen distribution at the sediment–water interface: a geostatistical analysis. Proceedings of the Second International Symposium on GIS/Spatial Analyses in Fishery and Aquatic Sciences, Fishery-Aquatic GIS Research Group, Kawagoe-City, Japan.

Huston, M.A., 2002. Critical issues for improving predictions. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, pp. 7–21.

Kristensen, E., 2001. Impact of polychaetes (*Nereis* spp. and *Arenicola marina*) on carbon biogeochemistry in coastal marine sediments. Geochem. Trans. 12.

Laine, A.O., 2003. Distribution of soft-bottom macrofauna in the deep open Baltic Sea in relation to environmental variability. Estuar. Coast. Shelf Sci. 57, 87–97.

Legendre, P., Legendre, L., 1998. Numerical ecology, Second English edition. Developments in Environmental Modelling, vol. 20. Elsevier Science B.V., Amsterdam.

Legendre, P., Thrush, S.F., Cummings, V.J., Dayton, P.K., Grant, J., Hewitt, J.E., Hines, A.H., McArdle, B.H, Pridmore, R.D., Schneider, D.C., Turner, S.J., Whitlatch, R.B., Wilkinson, M.R., 1997. Spatial structure of bivalves in a sandflat: scale and generating processes. J. Exp. Mar. Biol. Ecol. 216, 99–128.

McArdle, B.H., Blackwell, R.G., 1989. Measurement of density variability in the bivalve *Chione stutchburyi* using spatial autocorrelation. Mar. Ecol. Prog. Ser. 52, 245–252.

Meissner, K., Darr, A., Rachor, E., 2008. Development of habitat models for *Nephtys* species (Polychaeta: Nephtyidae) in the German Bight (North Sea). J. Sea Res. 60, 271–286.

Meyer, M., Harff, J., Gogina, M., Barthel, A., 2008. Coastline changes of the Darss-Zingst Peninsula—a modelling approach. J. Mar. Syst. 74, 147–154.

Moore, D.M., Lee, B.G., Davey, S.M., 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. Environ. Manag. 15, 59–71.

Morgan, T.S., Rogers, A.D., Paterson, G.L.J., Hawkins, L.E., Sheader, M., 1999. Evidence for poecilogony in *Pygospio elegans* (Polychaeta: Spionidae). Mar. Ecol. Prog. Ser. 178, 121–132.

Munoz, J., Felicisimo, A., 2004. A comparison between some statistical methods commonly used in predictive modelling. J. Veg. Sci. 15, 285–292.

Neumann, T., Schernewski, G., 2008. Eutrophication in the Baltic Sea and shifts in nitrogen fixation analyzed with a 3D ecosystem model. J. Mar. Syst. 74, 592–602.

Nithart, M., Alliot, E., Salen-Picard, C., 1998. Population dynamics and secondary production of Nereis diversicolor in a north Norfolk Saltmarsh (UK). J. Mar. Biol. Assoc. U. K. 79, 1029–1037.

O'Brien, K., Hänninen, J., Kanerva, T., Metsärinne, L., Vuorinen, I., 2003. Macrobenthic zonation in relation to major environmental factors across the Archipelago Sea, northern Baltic Sea. Boreal Environ. Res. 8, 159–170.

Pavlikakis, G.E., Tsihrintzis, V.A., 2000. Ecosystem management: a review of a new concept and methodology. Water Resour. Manag. 14, 257–283.

Pearson, T.H., Rosenberg, R., 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanogr. Mar. Biol. A. Rev. 16, 229–311.

Perus, J., Bonsdorff, E., 2004. Long-term changes in macrozoobenthos in the Åland archipelago, northern Baltic Sea. J. Sea Res. 52, 45–56.

Philippart, C.J.M., Anadón, R., Danovaro, R., Dippner, J.W., Drinkwater, K.F., Hawkins, S.J., Oguz, T., O'Sullivan, G., Reid, P.C., 2007. Impacts of climate change on European marine and coastal environment. ESF Marine Board Position Paper 9.

Praca, E., Gannier, A., Das, K., Laran, S., 2008. Modelling the habitat suitability of cetaceans: example of the sperm whale in the northwestern Mediterranean Sea. Deep-Sea Res. I. doi:10.1016/j.dsr.2008.11.001.

Rhoads, D.C., 1974. Organism sediment relations on the muddy sea floor. Oceanogr. Mar. Biol. Annu. Rev. 12, 263–300.

Sanders, H.L., 1968. Marine benthic diversity: a comparative study. Am. Nat. 102, 243–282.

Scaps, P., 2002. A review of the biology, ecology and potential use of the common ragworm *Hediste diversicolor* (O.F. Müller) (Annelida: Polychaeta). Hydrobiologia V470, 203–218.

Seifert, T., Tauber, F., Kayser, B., 2001. A high resolution spherical grid topography of the Baltic Sea, Baltic Sea Science Congress, Stockholm, Poster #147, 2nd edition.

StatSoft, Inc., 2007. Electronic Statistics Textbook. StatSoft, Tulsa, OK. http://www.statsoft.com/textbook/stathome.html.

Thrush, S.F., Hewitt, J.E., Norkko, A., Nicholls, P.E., Funnell, G.A., Ellis, J.I., 2003. Habitat change in estuaries: predicting broad-scale responses of intertidal macrofauna to sediment mud content. Mar. Ecol. Prog. Ser. 263, 101–112.

Thrush, S.F., Hewitt, J.E., Herman, P.M.J., Ysebaert, T., 2005. Multi-scale analysis of species–environment relationships. Mar. Ecol. Prog. Ser. 302, 13–26.

van der Wal, D., Herman, P.M.J., Forster, R.M., Ysebaert, T., Rossi, F., Knaeps, E., Plancke, Y.M.G., Ides, S.J., 2008. Distribution and dynamics of intertidal macrobenthos predicted from remote sensing: response to microphytobenthos and environment. Mar. Ecol. Progr. Ser. 367, 57–72.

Wanamaker Jr., A.D., Heinemeier, J., Scourse, J.D., Richardson, C.A., Butler, P.G., Eiríksson, J., Knudsen, K.L., 2008. Very-long lived molluscs confirm 17th century AD tephra-based radiocarbon reservoir ages for north Icelandic shelf waters. Radiocarbon 50, 3.

Wisz, M.S., Guisan, A., 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecol. 9 (8).

Woodward, F.I., Cramer, W., 1996. Plant functional types and climatic changes: introduction. J. Veg. Sci. 7, 306–308.

Yee, T.W., Mitchell, N.D., 1991. Generalized additive models in plant ecology. J. Veg. Sci. 2, 587–602.

Ysebaert, T., Meire, P., Herman, P.M.J., Verbeek, H., 2002. Macrobenthic species response surfaces along estuarine gradients: prediction by logistic regression. Mar. Ecol. Prog. Ser. 225, 79–95.

Zetter, M.L., Röhner, M., 2004. Verbreitung und Entwicklung des Makrozoobenthos der Ostsee zwischen Fehmarnbelt und Usedom—Daten von 1839 bis 2001. Bundesanstalt für Gewässerkunde (Hrsg.), Die Biodiversität in der deutschen Nord- und Ostsee, Band 3. Bericht BfG-1421, Koblenz.

Zettler, M.L., Bönsch, R., Gosselck, F., 2000. Das Makrozoobenthos der Mecklenburger Bucht—rezent und im historischem Vergleich. Meereswiss. Ber. 42, 1–144.