Original Article

The Baltic Sea scale inventory of benthic faunal communities

Mayya Gogina[1]*, Henrik Nygård[2], Mats Blomqvist[3], Darius Daunys[4], Alf B. Josefson[5], Jonne Kotta[6], Alexey Maximov[7], Jan Warzocha[8], Vadim Yermakov[9], Ulf Gräwe[1], and Michael L. Zettler[1]

[1]*Leibniz Institute for Baltic Sea Research (IOW), Warnemünde, Seestr. 15, 18119 Rostock, Germany*
[2]*Marine Research Centre, Finnish Environment Institute SYKE, PO Box 140, FI-00251 Helsinki, Finland*
[3]*Hafok AB, 179 61 Stenhamra, Sweden*
[4]*Coastal Research and Planning Institute, Klaipeda University, H. Manto Str. 84, LT-92294 Klaipėda, Lithuania*
[5]*Department of Bioscience, Aarhus University, Frederiksborgvej 399, 4000 Roskilde, Denmark*
[6]*Estonian Marine Institute, University of Tartu, Mäealuse 14, 12618 Tallinn, Estonia*
[7]*Zoological Institute Russian Academy of Science, Universitetskaja nab. 1, 199034 St Petersburg, Russia*
[8]*National Marine Fisheries Research Institute, ul. Kołłątaja 1, 81-332 Gdynia, Poland*
[9]*Latvian Institute of Aquatic Ecology, Daugavgrīvas Str. 8, LV-1048 Riga, Latvia*

Corresponding author: tel: +49 381 5197 393; fax: +49 381 5197 211; e-mail: mayya.gogina@io-warnemuende.de

This study provides an inventory of the recent benthic macrofaunal communities in the entire Baltic Sea. The analyses of soft-bottom benthic invertebrate community data based on over 7000 locations in the Baltic Sea suggested the existence of 10 major communities based on species abundances and 17 communities based on species biomasses, respectively. The low-saline northern Baltic, characterized by silty sediments, is dominated by *Monoporeia affinis*, *Marenzelleria* spp., and *Macoma balthica*. Hydrobiidae, *Pygospio elegans*, and *Cerastoderma glaucum* dominate the community in sandy habitats off the Estonian west coast and in the southeastern and southern Baltic Sea. Deep parts of the Gulf of Finland and central Baltic Sea often experience hypoxia, and when oxygen levels in these regions recover, *Bylgides sarsi* was the first species to colonize. The southwestern Baltic Sea, with high salinity, has higher macrofaunal diversity compared with the northern parts. To spatially interpolate the distribution of the major communities, we used the Random Forest method. Substrate data, bathymetric maps, and modelled hydrographical fields were used as predictors. Model predictions were in good agreement with observations, quantified by Cohen's κ of 0.90 for the abundance and 0.89 in the wet weight-based model. Misclassifications were mainly associated with uncommon classes in regions with high spatial variability. Our analysis provides a detailed baseline map of the distribution of benthic communities in the Baltic Sea to be used both in science and management.

Keywords: Baltic Sea, community analysis, large scale, macrozoobenthos, Random Forest, spatial distribution.

Introduction

In the recent decades, there has been a marked increase in benthic habitat mapping, largely inspired by the development and implementation of various research and monitoring programmes, management plans, and legislations, such as the 1992 EU Habitats Directive. This has resulted in an exponential accumulation of data hitherto stored as different non-standardized national datasets. Here for the first time, we bring much of these datasets together and harmonize them to address management issues on the large Baltic Sea scale. Similar large-scale efforts were previously reported for the North Sea in Rees *et al*. (2007), for the Barents Sea in Anisimova *et al*. (2010).

The focus of our study was identification, description, and basin scale mapping of major benthic macrofauna community distribution in the Baltic Sea. A number of works were previously published, describing and mapping the benthic macrofauna communities in different subbasins of the Baltic Sea (e.g. Warzocha, 1995; Laine, 2003; Glockzin and Zettler, 2008; Gogina *et al*., 2010). However,

we identified a lack of joint data analyses covering the entire Baltic Sea region based on new data collected since 2000.

At the beginning of the last century, Petersen (1913) introduced the concept of infaunal communities for the description and quantification of areas with similar fish food. He aggregated the infaunal species in relation to broad geographic areas of relatively uniform sediments, depth, and salinity. Petersen (1913) described various soft sediment benthic macrofauna assemblages in different regions of the Baltic Sea. The scale of the investigated area and its heterogeneity allowed for high resolution of classification detail, and the definition of a number of communities or assemblages. Summaries on the distribution of fauna in the Baltic Sea can be found in works of early investigators such as Hessle (1924), Välikangas (1933), Ekman (1933, 1935), and Segerstråle (1957). Based on the findings outlined by Zenkevitch (1963), Schiewer (2008) listed the most important Baltic Sea coenoses or assemblages of species. These were (i) *Abra alba*-coenosis with *Corbula gibba*, *Arctica islandica*, *Lagis koreni*, *Nephtys* spp., *Diastylis rathkei*, and *Ophiura albida* which predominate in the western Baltic. Towards the east (ii) *Arctica-Astarte* assemblages were increasingly found. In the shallow part of the Baltic Proper (iii) *Macoma baltica* coenosis dominated, whereas in the deeper part (iv) *Macoma calcarea* coenosis (with *Mya truncata* and *Astarte*) occurred. Farther in the eastern soft-bottom areas (v) *Saduria-Monoporeia-Pontoporeia* coenosis with euryhaline species were found, and Gotland Deep, strongly affected by oxygen depletion, was colonized only by (vi) a species-poor hypolimnic community. One of the deepest basins in the Baltic Sea, the Landsort Deep, was characterized by long-term azoic, laminated sediments (Schiewer, 2008). Elmgren (1978) gave a description of the structure and dynamics of the Baltic Sea benthic communities from the Bothnian Bay down to the Baltic Proper and summarized habitats as "an almost uniquely simple benthic ecosystem". Carman and Cederwall (2001) provided an analysis of mean biomass data for the main subbasins of the Baltic Sea and attributed dramatic changes of benthic communities during the late 1960s to oxygen depletion, resulting in extinction of macrofauna in the deepest parts of the Baltic Proper.

Based on the updated data compiled in the HELCOM checklist, 2035 macrozoobenthic species are currently known to live in the Baltic Sea (HELCOM, 2013a). The species composition and diversity in the Baltic Sea are heavily shaped by the large salinity gradient ranging from 2 in the Bothnian Bay to over 20 in the Danish Straits. Zettler *et al.* (2014) reported a continuous decline in overall species numbers with decreasing salinity levels due to decreasing number of species with marine affinity. In contrast, freshwater species were most common in the northern parts and some low-saline coastal subregions.

In this study, we have collated and standardized sample data from all countries around the Baltic Sea to determine ecologically significant macrofaunal groups shaped by the pronounced environmental heterogeneity unique to the Baltic region. This inventory is to our knowledge the most extensive one to date from the Baltic Sea area. Our overall aim was to delineate major communities (*sensu* Petersen) inhabiting the Baltic Sea and to provide their distribution within the entire Baltic Sea area.

The study provides basic information on macrofauna distribution in the Baltic Sea area that can be used in management such as marine spatial planning and fishery management.

Methods
Joint dataset
The standardized dataset comprised over 1000 taxa from over 7000 locations (17 000 visit events) mostly sampled in period 2000–2013

for species abundance and partly for biomass (Figure 1). Overall, 1176 taxonomic units (accepted in the World register of Marine Species WoRMS) were included in the final dataset. Over 70% of the samples were collected from May to July and therefore, the dataset is biased towards spring and early summer.

GIS databases were utilized for collating the distribution of benthic macrofauna covering the entire Baltic Sea basin. Most data (77%) were collected with a van Veen grab (covering a bottom area of 0.1 m^2), 1–3 replicates per sampling were taken, sieved on a 1 mm mesh, preserved in 4% buffered formaldehyde–seawater solution, and identified in the laboratory to the lowest taxonomic level possible. Table 1 provides more details including specifications and the area covered by each laboratory. All species abundance and biomass data were recalculated to the area of 1 m^2. Biomass records were available for 75% of the samples. Wet weight (including calcareous structures, without tubes) were only available for 2943 (39%) locations and ash-free dry weight for 2241 (29.7%) locations. Only wet weight were used for further analysis. Where only dry or ash-free dry weight data were available, in-house (IOW) biomass data conversion factors were used to calculate wet weight. The conversion factors are available upon request.

For the analyses, we used a 5 × 5 km grid (matching one-quarter of the European Environment Agency EEA 10 km reference grid size). Inner coastal fjords, estuaries, and lagoon regions were omitted from the analysis due to higher macrofauna variability, and because these areas contained a large part of freshwater organisms not representative for the broadscale picture of the entire Baltic. Average taxonomic unit abundances were calculated to account for all samples within one grid cell. Biomass records were obtained for 2268 of 5 × 5 km grid cells. Such aggregation of data neglects fine-scale spatial differences in the community structure and ignores temporal variation within the study period in the favour of capturing the most important large-scale patterns.

Community analysis
Cluster analysis (group-average linking) and nMDS (not presented here for brevity) were conducted to reveal similarities between stations. For multivariate analyses, we used taxa determined to the species level, where possible. Difficult taxa such as oligochaetes and chironomids were not identified to lower taxonomic level. Some characteristic Baltic Sea taxa were grouped to genus or even higher taxonomic level (*Diastylis*, *Harmothoe*, *Exogone*, Hydrobiidae, *Idotea*, *Nephtys*, *Ophiura*, *Phoronis*, *Polydora*, Diptera) to overcome taxonomic inconsistencies. For abundance-based analysis, only taxa that occurred at more than ten sampling events and contributed to the top 90% of the cumulative abundance per sampling event were retained. This reduction was done for better detection of the underlying community similarities and resulted in 272 taxa. For biomass analysis, taxonomic units were selected based on two criteria: either among top five species with highest biomass recorded at any sampling event or contributed to 80% of the cumulative total biomass. It resulted in the selection of 381 taxa. Abundance and biomass data matrices can be found in the Supplementary material.

All abundance data were fourth-root transformed, and biomass data were log(x + 1)-transformed to account for the biasing effect of few high and low data values. Zero-adjusted Bray–Curtis similarities were calculated between stations. Clarke *et al.* (2006) suggested solving the problem of denuded samples in data (i.e. no common species among two sites) by using a dummy variable being equal to 1 for all objects allowing for the coefficient to be defined in these cases, thus avoiding bias. We used the dissimilarity matrix to perform a
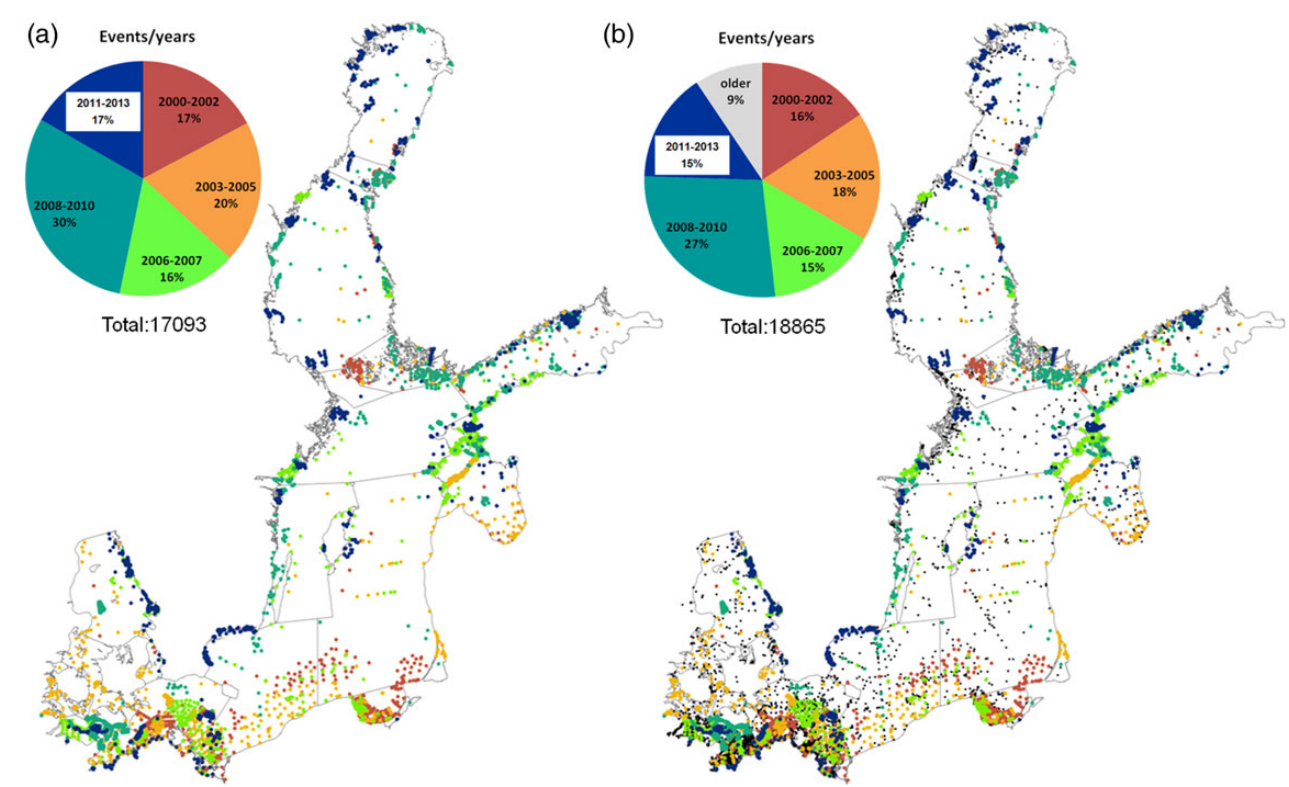
Figure 1. Temporal variability within the joint dataset: (a) only data collected starting with the year 2000 is shown; (b) with data collected before 2000 included. Data from different periods are shown in different colours. Considering the older data (that contributes to only 9% of the overall sampling events and are up to 98% collected after 1990 which considerably increases spatial resolution and coverage especially in the northern and central regions). This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.

Table 1. Sampling equipment used for different datasets included and availability of biomass.

Source	Sampler	Sampler area (m²)	Mesh size (mm)	*n* sampling events	Biomass
DK	Haps	0.0143	1	191	To species
	van Veen	0.1	1	313	
EST	Ekman-Birge	0.021	0.25	2153	Dry weight, to species
	Box	0.04	0.25	19	
	Lenz	0.021	0.25	627	
	Petersen	0.044	0.25	344	
	van Veen	0.1	0.25	427	
DE	van Veen	0.1	1	1370	To species
LT	van Veen	0.1	0.5	313	Fresh weight
LV	van Veen	0.1	0.5	195	–
PL	Reineck	0.0225	1	48	To species
	van Veen	0.1	1	298	
SE	SM	0.1	1	21	To species
	van Veen	0.1	1	6738	
FI	Box	0.04	1	57	To species
	Ekman-Birge	0.0225 – 0.1445	1	247	
	van Veen	0.1	1	1425	
RU	van Veen	0.025	0.4	25	To species
Literature	van Veen mainly	0.1	1	1536	Partly, to species
ICES	van Veen	0.1	1	42	–

hierarchical cluster analysis (group-average). Community clusters were determined by selecting a distance where stations were grouped in well-defined clusters (arbitrary based on expert judgement). The classification results were tested with a one-way PERMANOVA based on the Bray–Curtis similarity matrix using 999 random permutations (Anderson, 2001). See Anderson and Walsh (2013) for the warning on usage of this test for unbalanced designs where heterogeneity of dispersions is known to occur. These analyses were conducted in PRIMER v6 (Clarke and Warwick, 2001).

Finally, we applied Similarity percentage analysis (SIMPER, Clarke and Warwick, 2001) and Indicator value analysis (IndVal

proposed by Dufrêne and Legendre, 1997) on the abundance and biomass data matrix to identify the main taxa which were responsible for differences in community structure. The IndVal method is implemented in R package "labdsv" (Roberts, 2015). IndVal is a measure of association between a taxon and a cluster of stations, calculated as the product of specificity (mean abundance or biomass of a given taxon within a cluster compared with the other clusters) and fidelity (taxon occurrence at stations belonging to a cluster). The maximum IndVal of 100% is reached when a given taxon is observed at all stations of only one cluster. Indicator species in contrast to species associations are indicative of a particular cluster of sites, i.e. good indicator species should be at the same time ecologically restricted to a target site group and frequent within it (De Cáceres et al., 2012). Statistical significances of indicator taxa were tested by 9999 random permutations of the samples and an *a priory* defined 0.05 threshold level was used.

Random Forests for spatial interpolation

Recently, a number of different modelling techniques have become increasingly used for benthic ecological studies (Wei et al., 2010; Collin et al., 2011; Reiss et al., 2011; Šiaulys and Bučas, 2012; Bučas et al., 2013). This study used the Random Forest algorithm developed by Breiman (2001) to generate full-coverage community maps. This method is implemented in the R package "randomForest" (Liaw and Wiener, 2002), a machine learning algorithm comprising an ensemble of randomly constructed decision trees (500 classification trees predicting class membership). A brief description of how the algorithm works can be found in Diesing et al. (2014). The model runs efficiently on large datasets, and can handle thousands of input variables without variable deletion. Moreover, the algorithm avoids the problem of overfitting. Another advantage of the Random Forest algorithm is that there is no need for additional cross validation or a separate test set to reach unbiased estimate of the test set error. This is internally estimated by the OOB (out-of-bag) error estimates, using a different bootstrap sample from the original data (one-third), during the construction of the tree. OOB data are also used to get estimates of variable importance. After building each tree, proximities are computed for each pair of cases. If two cases occupy the same terminal node, their proximity is increased by one. At the end of the run, proximities are normalized by the number of trees. Proximities are used to replace missing data, locate outliers, and to create low-dimensional views of the data. For this case study, it was essential that the algorithm balanced the error in class population unbalanced datasets. Due to the differing data sources used in this study, communities towards the entrance of the Baltic Sea are represented by considerably smaller number of sampling events and grid cells. Thus, no random stratified sampling design is possible to be achieved without a loss of spatial resolution.

Elith et al. (2010) showed that results of such machine learning algorithms should be treated with caution if extrapolated outside of the training area. However, we assume that the spatial cover of our data is sufficient and that we are not going beyond interpolation.

To assess the accuracy of the spatial prediction, we derived measures of classification accuracy from an error matrix. These measures were overall OOB estimate, error rate, purity, and representation. Additionally, we applied Cohen's (1968) κ coefficient of agreement calculated using "psych" R library (Revelle, 2015). This statistic measures inter-rater agreement typically used for categorical data to compensate for the chance agreement and is thought to be

a more robust measure than simple per cent agreement calculation in this kind of study (Diesing et al., 2014).

Environmental variables

To support the spatial modelling, we used environmental variables, which represent direct or indirect linkage to the distribution of benthic macrofauna. We limited ourselves to datasets that were available for the entire Baltic Sea. We identified five major environmental predictors: bathymetric maps, near-bottom hydrographic fields (temperature, salinity, and flow speed), sediment type maps, and particulate organic carbon content of surface sediments.

We tested bathymetry datasets originating from the EU-BALANCE project (Al-Hamdani and Reker, 2007) and from Seifert et al. (2001). The latter showed to be more useful for our purposes, and was used to calculate the slope and indices of northness and eastness, describing the orientation of the slopes (based on first derivate of the bathymetry).

Hydrographic near-bottom fields of temperature, salinity, and flow speed were extracted from the model simulations of Gräwe et al. (2015), covering the entire North Sea and Baltic Sea for the period 1997–2015. The numerical model has a spatial resolution of 1 nm (∼1.8 km), data were provided as either monthly or annual means, minimum or maximum values. Additionally, we used monthly or annual standard deviation to quantify the natural variability. To account for extremes, monthly 5 and 95% percentiles were used. From the original dataset, only the period from 2000 to 2014 was used. The averaging of the hydrographical variables helped to reduce the variance in the resulting fields.

The only available data on seabed sediments covering the whole Baltic Sea region were produced by the EU-BALANCE project. The dataset was categorized in five categorical sediment classes: 1, bedrock; 2, complex sediments; 3, sand; 4, hard clay; 5, mud and clay. As these substrate data are available at a 200 × 200 m resolution, for model training, we down-sampled the data to the final 5 × 5 km grid. This was done by choosing the sediment class with highest coverage within the 5 km grid. Note that this strategy can lead to a mismatch between the dominating sediment class (within a 5 km grid cell) and the actual sediment type sampled for macrofauna. However, we believe that this error caused by the local heterogeneity occurs only at a small number of occasions and the entire Baltic picture aimed at should not be essentially biased.

Full coverage map of particulate organic carbon content in surface sediments with a resolution of 5 km was published by Leipe et al. (2011). All other variables were resampled before the prediction to match the sediment type data resolution to avoid information loss.

Estimate of total biomass distribution

To estimate the distribution of total benthos wet weight biomass over the Baltic Sea basins based on the available data, we utilized ordinary kriging on the log-transformed data averaged per 5 × 5 km grid cell. The resulting raster was then back-transformed to display biomass in gramme per square metre.

Results
Communities based on abundance data

Based on the hierarchal cluster analysis of abundance data, we identified ten major communities inhabiting the Baltic Sea (Figure 2). We placed a cut off at 23% of Bray–Curtis similarity, clusters with a minimum number of ten cells considered as major
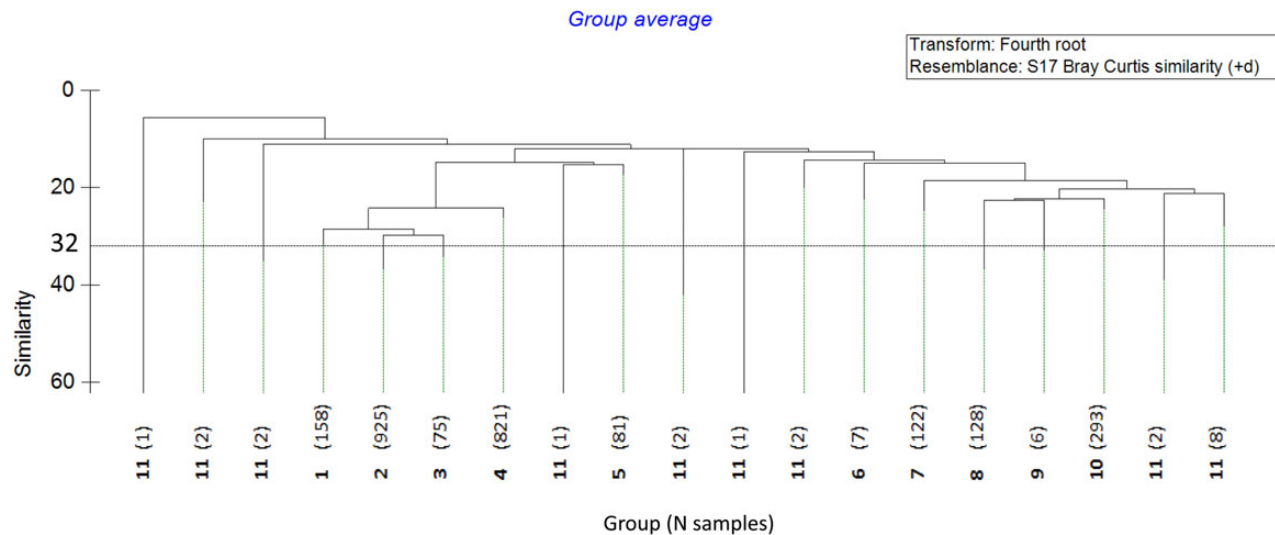
Figure 2. Dendrogram resulting from cluster analysis of benthic samples based on species abundance data. The horizontal dotted line represents the cut-off at 32% similarity. Altogether 10 tight clusters (communities) were detected, with remaining samples coded as 11 ("other" variable communities). This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.

communities. It should be noted that we have chosen an arbitrary level of classification to provide an entire Baltic picture. We have tested various arbitrary cut-off similarity levels and decided on the reported one based on expert judgement. The higher the chosen similarity level, the larger the number of delineated groups (communities). Our aim was to determine a reasonable number of communities that can be well visualized and distinguished. Community structure differed significantly between the delineated communities (PERMANOVA $F_{10,2626} = 175.32$, $p = 0.001$). The results of the indicator species and SIMPER analyses revealed that certain species contributed to observed pattern and helped to determine group distinctions. The characteristics of major communities and characteristic species are listed in Table 2 and the observed distribution aggregated to the 5 km grid cell scale is shown in Figure 3. Grid cells classified outside the major groups are treated as a separate class of "diverse/patchy/variable" community and will be referred to as "other" throughout the text. Distribution of the oxygen minimum zone was used to mask out regions with frequent absence of macrofauna (grey areas in Figure 3). A full-coverage map of the model prediction is presented in Figure 4.

Community 1 was the dominating type in the northern area. Characteristic species in community 1 were *Monoporeia affinis*, *Marenzelleria* spp., and *Macoma balthica*. Their abundance are well correlated with the occurrence of muddy to sandy sediments. There are some interspersed patches of community 2 found in sandy areas characterized by Hydrobiidae, *Pygospio elegans,* and *Cerastoderma glaucum* species. This community was mainly found in sandy habitats off Estonian west coast and in the southwestern Baltic Sea. Figure 3 shows group 4 in the Gulf of Finland and central Baltic down to Arkona Basin. Group 4 is strongly linked to the oxygen conditions in the deeper basins. *Beggiatoa*-periphytons occur frequently in these areas during periods of hypoxia or low oxygen content. After increase in oxygen levels, the motile *Bylgides sarsi* is the one to arrive first.

Communities in the southern part are more diverse than the benthic communities of the Northern Baltic Sea. The Bay of

Mecklenburg is mostly inhabited by sandy community 2 in shallower areas and community 3 (*Diastylis* sp., *C. gibba*, *Dipolydora quadrilobata*, *A. islandica*, *Aricidea suecica*, *Abra alba*) in the deeper part. Communities 5 (characterized by *Amphiura* sp., *Abra nitida*, *Galathowenia oculata*, *Ennucula tenuis*, *Thyasira flexuosa*, *Nucula nitidosa*, *Diplocirrus glaucus*), 7 (*Phoronis* sp., *Tellina fabula*, *Thracia phaseolina*, *Ophelia borealis*, *Spiophanes bombyx*, *Branchiostoma lanceolatum*, *Spio arndti*) and 8 (*Tellina tenuis*, *Ensis directus*, *Haustorius arenarius*, *Lamprops fasciatus*) are mostly found in the Kattegat region. Community 9, probably a degraded habitat (with *L. koreni*, *Cerastoderma edule*, *Polydora* sp., *Halicryptus spinulosus*), is typical for the Flensburg Fjord. The occurrence of community 10 (characterized by *Echinocyamus pusillus*, *Harmothoe* sp., *Bittium reticulatum*, Oligochaeta, *Alitta virens*, *Turritella communis*, *Asterias rubens*) is well restricted to the Great Belt region.

Communities based on biomass data
The hierarchal cluster analysis of biomass data identified 17 major communities in the Baltic Sea. Results are shown in Figure 5 (cut-off at 28%, clusters with minimum number of ten cells were considered as major communities, see Table 3 and Figure 6 for characteristic species). The significance of differences in species composition of 17 distinct communities was confirmed by PERMANOVA ($F_{17,2250} = 91.255$, $p = 0.001$). Figure 7 shows full-coverage map resulting from the model prediction.

Biomass-based community 1 covers most of the northern areas. It appears on both muddy and sandy sediments and has *Saduria entomon*, *Ma. balthica*, *M. affinis*, *H. spinulosus*, *Pontoporeia femorata* as characteristic species. Community 2 is characterized by *Hediste diversicolor*, *Mya arenaria*, Hydrobiidae, and *C. glaucum* and can be classified as the third largest community, based on predicted spatial coverage. Community 2 mostly appears on sandy sediments and can be found in patches in the Archipelago Sea, in shallower parts of the Gulf of Riga, and on the edge of the Eastern Gotland Basin. However, the preferred habitats occur in the Pomeranian Bay, the Rügen-Falster-Plate, and in the northwestern part of the Sound. The second largest coverage was community 5,

Table 2. Characteristics of major communities delineated based on abundance.

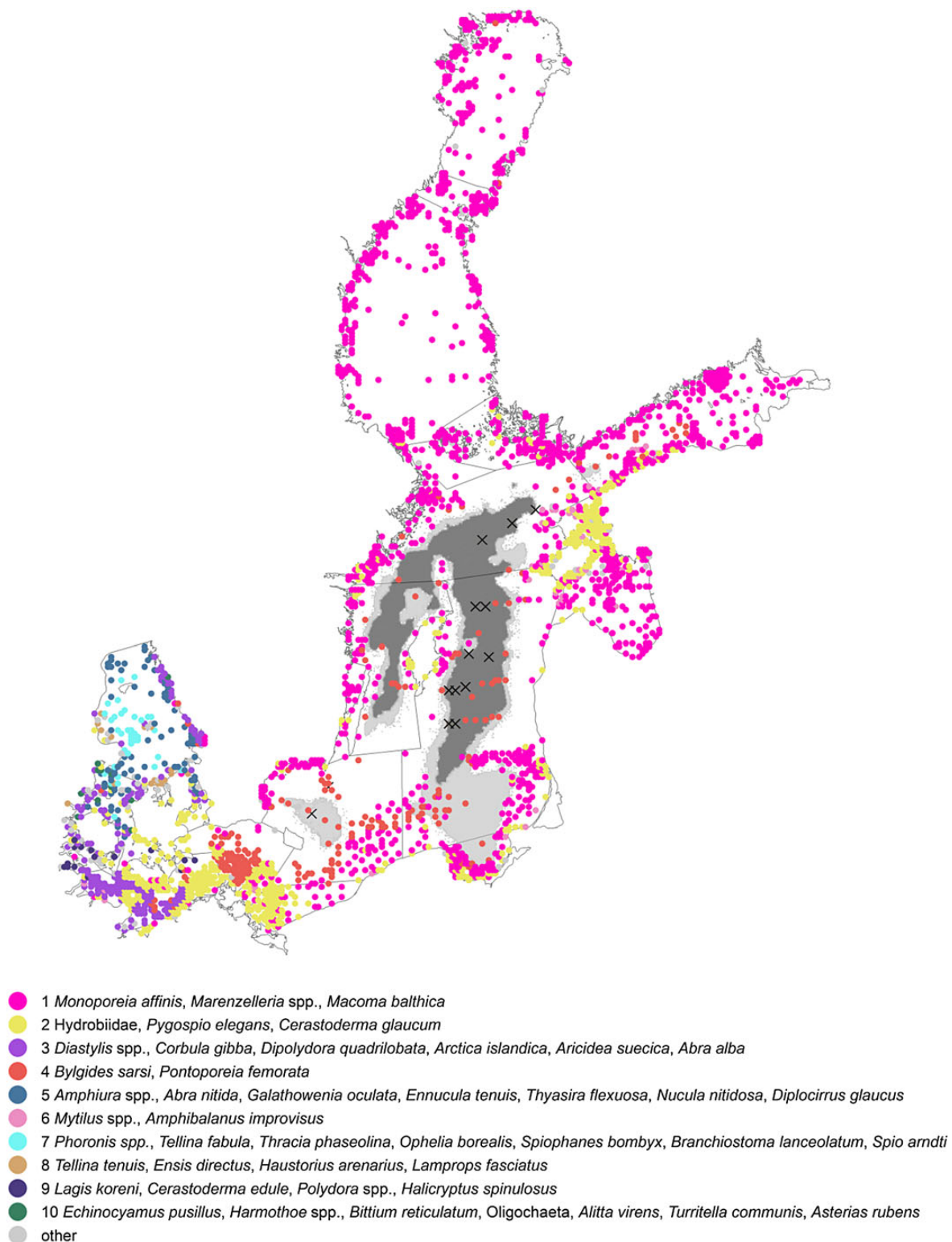| Community	Key species	n cells	n visits	Mean n species/ station (overall)	Mean abundance/ station (ind. m^{-2})	Mean biomass/ station (g m^{-2})	Predicted area km$^2 \times 10^3$	Depth	Av. salinity (min – max)	Sediment class (balance)
1	*Monoporeia affinis, Marenzelleria, Macoma balthica*	1415	9977	5 (338)	1345	83	245.26	5 – 65	6.1 (0.9 – 19.4)	Mud to sand
2	Hydrobiidae, *Pygospio elegans, Cerastoderma glaucum*	557	3827	14 (463)	4797	193	28.59	<30	9.5 (0.1 – 27.7)	Sand
3	*Diastylis, Corbula gibba, Dipolydora quadrilobata, Arctica islandica, Aricidea suecica, Abra alba*	208	1434	38 (543)	3288	358	12.03	10 – 25	20.7 (13.5 – 30.8)	Sand
4	*Bylgides sarsi, Pontoporeia femorata*	205	855	7 (202)	485	54	74.97	20 – 120	12 (2.4 – 24.9)	Mud
5	*Amphiura, Abra nitida, Galathowenia oculata, Ennucula tenuis, Thyasira flexuosa, Nucula nitidosa, Diplocirrus glaucus*	84	600	44 (580)	1903	204	14.33	3 – 40	28.3 (19.1 – 34.7)	Mud
6	*Mytilus* spp., *Amphibalanus improvisus*	39	77	7 (143)	2986	20	0.27	2 – 70	7.6 (4.1 – 21.4)	Hard bottom complex
7	*Phoronis, Tellina fabula, Thracia phaseolina, Ophelia borealis, Spiophanes bombyx, Branchiostoma lanceolatum, Spio arndti*	38	127	29 (391)	1508	466	5.58	8 – 40	30 (25.4 – 34.1)	Sand
8	*Tellina tenuis, Ensis directus, Haustorius arenarius, Lamprops fasciatus*	17	22	21 (159)	1542	44	0.81	<25	26.4 (20.1 – 32.8)	Sand
9	*Lagis koreni, Cerastoderma edule, Polydora, Halicryptus spinulosus*	12	45	11 (105)	1385	41	0.22	<20	18.9 (13 – 22.3)	Sand
10	*Echinocyamus pusillus, Harmothoe, Bittium reticulatum, Oligochaeta, Alitta virens, Turritella communis, Asterias rubens*	11	12	29 (147)	1719	16	0.20	2 – 20	24.2 (19.4 – 28.1)	Gravelly
Other		70		(394)			1.55			
No fauna	–	50	291				11.59		11.3 (5.9 – 15.4)	Mud

Figure 3. Ten major communities with characteristic species according to hierarchal clustering (based on point data aggregated to 5 km grid cells), SIMPER, and Indicator species analysis based on abundance data. Crosses indicate sampled cells with no benthic infauna recorded (mainly dataset PL). Light grey and dark grey areas mask out the deep water hypoxic and anoxic oxygen conditions, correspondingly (winters 2014 and 2015, after Aranda's monitoring cruise, SYKE)—when oxygen depletion takes place, this areas usually lack macrofauna.

which is characterized by *B. sarsi* and corresponds to abundance-based community 4 (i.e. macrobenthic community attributed to the area when conditions are oxic).

The accuracy estimates for the abundance-based classes are summarized in Table 4. The overall accuracy achieved by the spatial model prediction based on abundance data derived communities

Figure 4. Full coverage map of abundance-based communities distribution predicted with Random Forest. See Figure 3 for colour legend, Table 2 for community characteristics, and Table 4 for accuracy estimates. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.

was 93%. However, attempts failed to predict community 10 as well as community assemblages named "other", where the representation was zero for both of these two classes. A slightly better, but still low, representation is achieved for communities 6 and 9 (see Table 2 and Figures 3 and 4 for respective details on characteristic species, observed and predicted distribution).

Figure 5. Dendrogram resulting from cluster analysis of benthic samples (5 km grid cells) based on species biomass data. Communities were arbitrary split at 28% similarity. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.
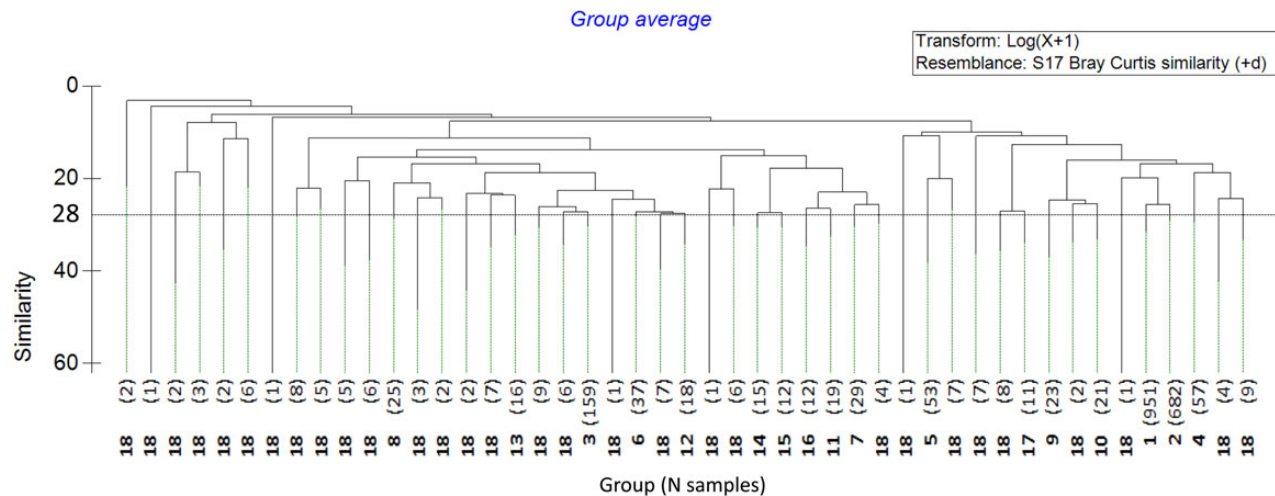
The accuracy achieved for the model based on biomass data was 92%. Table 5 shows the confusion matrix for the Random Forest biomass run-based classes (classes 1–17 are recognized as communities 1–17). Community 18 included all "other" pooled variable communities. Based on our available predictors, the model predicted community assemblages 1–3, 5, 9, 11, 13–16 quite well (see Table 3 and Figures 6 and 7 for respective details on characteristic species, observed and predicted distribution).

Estimated distribution of total wet weight biomass

Interpolation of total wet weight biomass over the Baltic Sea basins derived using ordinary kriging are shown in Figure 8. Areas where biomass data were lacking produced interpolation artefacts: for example, the very low values at the shallow parts of the Eastern Gotland Basin at the west coast off Latvia.

Discussion and conclusions
Community analyses

The present study describes the major benthic communities inhabiting the Baltic Sea in terms of both abundance and biomass.

Community structure is often used to assess changes along environmental gradients. Abundance is often used as the response variable in community data analyses. Biomass, however, is generally used to indicate the functionality of a species within a community, as it is strongly correlated with metabolism (Saint-Germain *et al.*, 2007). Biomass is also a decisive metric in recently developed biotope classifications (such as HELCOM HUB Underwater Biotope and Habitat classification system, Schiele *et al.*, 2015), and is especially relevant for the implementation of Marine Strategy Framework Directive (MSFD) and other legislative directives.

The question of using either abundance (count) or biomass as a response variable always emerges in community analyses. Generally, this dilemma is an understudied question. Which biodiversity "currency" is more relevant for different aspects of ecosystem functioning, and what is the relative importance of the two in relation to the effect they have in ecosystem processes (see Norkko *et al.*, 2013)? Communities themselves are not abundances or biomasses, but

both, and a statistical workflow should be developed in the future to combine these two approaches.

The whole northern area of the Baltic Sea is dominated by abundance-based community 1, characterized by *M. affinis, Marenzelleria* spp.*,* and *Ma. balthica*. This is expected as all the three species are typical for the Baltic Sea according to previous studies (Laine, 2003; Perus and Bonsdorff, 2004). *Monoporeia affinis* is a limnic species, with highest densities observed for salinities between 5 and 9, and only in rare occasions, it occur at stations with salinity values above 10 (e.g. Gogina and Zettler, 2010). *Marenzelleria* spp. are among the most successful non-native benthic species in the Baltic Sea (Maximov *et al.*, 2015). After first appearance in 1985, they quickly colonized the entire sea, now occupying a dominant position in the zoobenthos. Although the distribution of *Ma. balthica* is limited by salinity of ∼3 (Väinölä and Varvio, 1989), community 1, with *Ma. balthica* as a typical constituent species, reaches the whole Bothnian Bay and eastern part of Gulf of Finland. Some sources (e.g. Elmgren, 1978) imply that in contrast to most parts of the Baltic Sea where *Ma. balthica* is most common, bivalves are absent and meiofauna plays the most important role in the Bothnian Bay and the easternmost Gulf of Finland. Since we focused on soft-bottom macrofauna communities, the similarity analysis suggests that these areas still belong to that most typical community 1 cluster. Thus, the similarity of community structure despite the absence of bivalve species is still overall very high. However, if a finer level of classification was considered, these areas would indeed to be allocated into a separate class. Laine (2003) analysed large-scale distribution of soft-bottom macrofauna in the open Baltic Sea based on biomass data collected in 1996/1997, thereby representing the momentary state in distribution for that period. He showed that zoobenthos in the Bothnian Bay and central part of the Bothnian Sea is dominated by the isopod *Saduria entomon* and the amphipods Monoporeia/*Pontoporeia*. This is confirmed by our data, with *Ma. balthica* as another characteristic species of community 1 (both abundance and biomass based) increasing in importance where salinity is higher (above 4.5).

Standardized taxonomic units in abundance/biomass matrices are essential to ensure ecologically meaningful interpretation of clustering of stations. Unresolved taxonomic units are a problem that was

Table 3. Characteristics of major communities delineated based on biomass.

Community	Key species	n cells	n visits	Mean n species/station (overall)	Mean abundance/ station (ind. m^{-2})	Mean biomass/ station (g m^{-2})	Predicted area km^2 × 10^3	Depth	Av. salinity (min – max)	Sediment class (balance)
1	*Saduria entomon, Macoma balthica, Monoporeia affinis, Halicryptus spinulosus, Pontoporeia femorata*	951	8549	5 (255)	1452	111	244.2	2 – 86	9.1 (1.8 – 25.5)	Mud to sand
2	*Hediste diversicolor, Mya arenaria,* Hydrobiidae, *Cerastoderma glaucum*	682	4327	13 (448)	4108	180	43.3	<30	20.1 (14.3 – 28.1)	Sand
3	*Diastylis, Astarte, Mya truncata, Dendrodoa grossularia, Abra alba, Lagis koreni, Trochochaeta multisetosa, Lineus ruber*	159	1156	41 (481)	3383	414	10.0	10 – 46	5.8 (2.4 – 9.6)	Mud
4	*Marenzelleria,* Oligochaeta, Diptera, *Cyanophthalma obscura, Pisidium* sp., *Sphaerium corneum*	57	269	4 (79)	660	22	1.4	<75	11.3 (6.2 – 20.9)	Mud
5	*Bylgides sarsi*	53	286	2 (34)	34	1	62.0	48 – 158	20.9 (15.5 – 30.6)	Mud
6	*Capitella*	37	56	18 (203)	1150	25	1.2	5 – 19	28.8 (21.7 – 33.6)	Sand
7	*Nucula nitidosa, Rhodine gracilior, Owenia fusiformis, Virgularia mirabilis, Thyasira flexuosa*	29	151	43 (413)	2399	430	4.4	12 – 24	24.1 (19.4 – 30.8)	Mud
8	*Tellina tenuis, Ophelia borealis, Haustorius arenarius, Ensis*	25	28	26 (212)	3017	67	0.7	<16	7.3 (4.1 – 14.9)	Sand
9	*Amphibalanus improvisus,* NS *Idotea*	23	50	8 (60)	2788	356	0.1	<58	20.7 (14 – 27.2)	Gravelly to sand
10	*Mytilus, Asterias rubens, Carcinus maenas, Parvicardium pinnulatum, Balanus crenatus, Corbula gibba, Alitta succinea*	21	59	38 (276)	3979	466	0.5	<21	27.7 (20.9 – 31.9)	Gravelly to sand
11	*Phoronis, Lanice conchilega, Arctica islandica, Tellina fabula, Echinocardium cordatum, Nucula sulcata, Macoma calcarea*	19	21	43 (212)	2733	1041	2.4	<27	22.7 (17.9 – 28.6)	Sand
12	*Cerastoderma edule, Nassarius, Pygospio elegans, Scoloplos armiger, Aricidea cerrutii, Spio goniocephala*	18	157	17 (239)	909	27	0.1	<24	27.4 (19.7 – 33.2)	Sand
13	*Alitta virens, Ampelisca, Dipolydora coeca, Phyllodoce groenlandica, Nephtys, Spiophanes bombyx, Aporrhais pespelecani*	16	24	34 (226)	2585	110	1.8	<17	33.2 (28 – 34.7)	Sand to mud
14	*Brissopsis lyrifera, Amphiura, Thracia convexa, Scoletoma fragilis, Hyala vitrea, Lipobranchius jeffreysii*	15	182	49 (364)	1662	141	5.4	13 – 49	26.9 (23.2 – 34.7)	Mud
15	*Diplocirrus glaucus, Pholoe baltica, Abra nitida, Nemertea, Eudorella emarginata, Parvicardium minimum*	12	242	42 (358)	1847	144	0.3	<68	29 (18.8 – 31.1)	Mud
16	*Branchiostoma lanceolatum, Dosinia lupinus, Euspira nitida, Modiolus modiolus, Chamelea striatula*	12	102	25 (256)	1045	806	0.7	9 – 18	5.6 (3.2 – 9.8)	Sand
17	Odonata, *Theodoxus fluviatilis, Radix balthica,* Coleoptera, Gammaridae, *Bithynia tentaculata, Asellus aquaticus*	11	33	5 (26)	2072	34	0.3	5 – 90	20.1 (0.5 – 34.5)	Gravelly to sand
18	other	128	255	18 (557)	1412	89	10.1	3 – 37		

● 1 *Saduria entomon, Macoma balthica, Monoporeia affinis, Halicryptus spinulosus, Pontoporeia femorata*
● 2 *Hediste diversicolor, Mya arenaria, Hydrobiidae, Cerastoderma glaucum*
● 3 *Diastylis, Astarte, Mya truncata, Dendrodoa grossularia, Abra alba, Lagis koreni, Trochochaeta multisetosa, Lineus ruber*
● 4 *Marenzelleria* spp., *Oligochaeta, Diptera, Cyanophthalma obscura, Pisidium* sp., *Sphaerium corneum*
● 5 *Bylgides sarsi*
● 6 *Capitella* spp.
● 7 *Nucula nitidosa, Rhodine gracilior, Owenia fusiformis, Virgularia mirabilis, Thyasira flexuosa*
● 8 *Tellina tenuis, Ophelia borealis, Haustorius arenarius, Ensis* spp.
● 9 *Amphibalanus improvisus, Idotea* spp.
● 10 *Mytilus* spp., *Asterias rubens, Carcinus maenas, Parvicardium pinnulatum, Balanus crenatus, Corbula gibba, Alitta succinea*
● 11 *Phoronis* spp., *Lanice conchilega, Arctica islandica, Tellina fabula, Echinocardium cordatum, Nucula sulcata, Macoma calcarea*
● 12 *Cerastoderma edule, Nassarius* spp., *Pygospio elegans, Scoloplos armiger, Aricidea cerrutii, Spio goniocephala*
● 13 *Alitta virens, Ampelisca* spp., *Dipolydora coeca, Phyllodoce groenlandica, Nephtys* spp., *Spiophanes bombyx, Aporrhais pespelecani*
● 14 *Brissopsis lyrifera, Amphiura* spp., *Thracia convexa, Scoletoma fragilis, Hyala vitrea, Lipobranchius jeffreysii*
● 15 *Diplocirrus glaucus, Pholoe baltica, Abra nitida, Nemertea, Eudorella emarginata, Parvicardium minimum*
● 16 *Branchiostoma lanceolatum, Dosinia lupinus, Euspira nitida, Modiolus modiolus, Chamelea striatula*
● 17 *Odonata, Theodoxus fluviatilis, Radix balthica, Coleoptera, Gammaridae, Bithynia tentaculata, Asellus aquaticus*
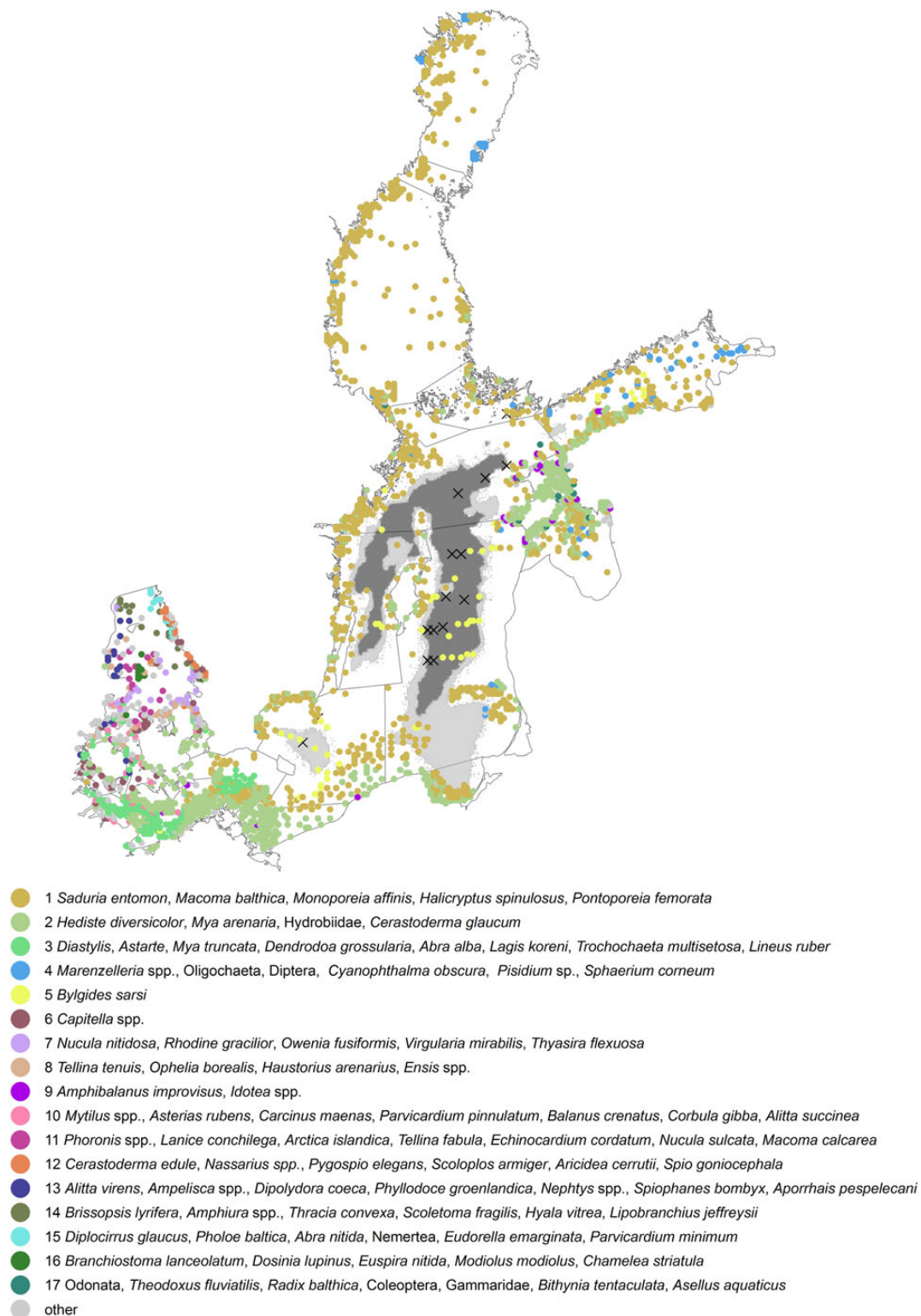● other

Figure 6. Seventeen major communities with characteristic species according to hierarchal clustering (based on point data aggregated to 5 km grid cells), SIMPER, and Indicator species analysis based on biomass data. Crosses indicate sampled cells with no benthic infauna recorded (mainly Polish dataset). Light grey and dark grey areas mask out the deep water hypoxic and anoxic oxygen conditions, correspondingly (winters 2014 and 2015, after Aranda's monitoring cruise, SYKE)—when oxygen depletion takes place these areas usually lack macrofauna. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.
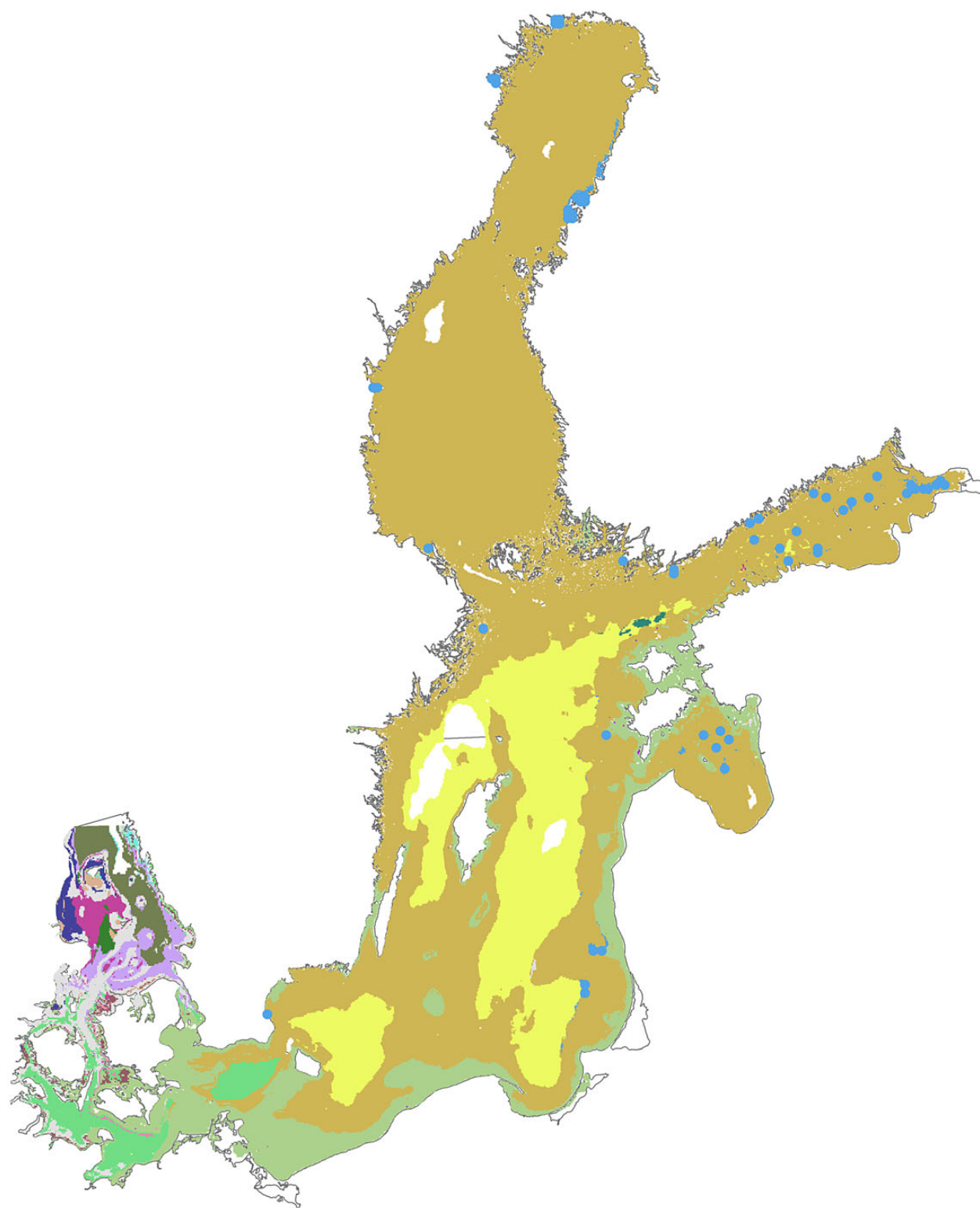
Figure 7. Full coverage map of biomass-based communities distribution predicted with Random Forest. Community 4 is shown by points as Random Forest model failed to predict it with available predictor variables. See Figure 6 for colour legend, Table 3 for community characteristics, and Table 5 for accuracy estimates. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.

solved by grouping difficult taxa to higher taxonomic levels. An exception and a crucial shortcoming of our study was the presence of both Chironomidae and Diptera in data matrices used for clustering,

although it would be correct to aggregate all the corresponding records under Diptera. In the Finnish data, most chironomids are identified to species level, whereas in Estonian data, both individuals

Table 4. Confusion matrix for the Random Forest of abundance-based classes (classes 1–10 are communites 1–10, class 11—"other" communities, class 12—cells with only no macrofauna records, respectively).

	1	2	3	4	5	6	7	8	9	10	11	12	class.error
1	1033	59	1	22	1	3	0	0	0	0	0	0	0.0769
2	75	328	22	7	1	4	0	0	1	1	1	0	0.2545
3	0	16	155	1	9	0	1	3	0	1	0	0	0.1667
4	41	6	8	146	1	0	0	0	0	0	0	4	0.2913
5	0	0	10	0	63	0	6	0	0	1	0	0	0.2125
6	20	12	1	1	0	1	0	0	0	0	0	0	0.9714
7	0	1	2	0	10	0	21	0	0	0	2	0	0.4167
8	0	0	5	0	1	0	1	3	0	0	2	0	0.75
9	0	4	6	0	0	0	0	0	2	0	0	0	0.8333
10	0	2	3	0	0	0	1	0	0	0	0	0	1
11	12	6	16	2	4	0	4	2	0	0	0	0	1
12	2	0	0	6	0	0	0	0	0	0	0	8	0.5

OOB estimate of error rate: 19.8%

Cohen's κ coefficient: 0.90

Note Random Forest fails to predict classes 6, 9, 10, and 11 based on available predictors.

Model note: 2703cellsAvAbu_7_SpatJoin_for_Sed_kl_check: factor(community) ∼ iowtopo_bs + tempulf + speedUlf + saltmean_L + factor(factor(sedkor), levels = c("1", "2", "3", "4", "5")) + saltstd3.

identified to family and to order level could be documented for the same station. As a result of such split, the Diptera as taxonomic unit was underrepresented in the relative abundance and was not picked as characteristic by the indicator species analysis for abundance-based communities. However, Diptera (and particularly the family Chironomidae) are relevant as characteristic species in communities 1 and 2 (where they have especially high subdominance in the shallow northern regions along the coast). The treatment of Diptera as one uniform taxonomic unit would result in the shift of branching between the stations of these two communities.

It is important to highlight that our community analysis results are not directly comparable with HELCOM HUB classification of biotopes. The HELCOM HUB developed particularly for the Baltic Sea is a hierarchical classification structured into six levels, whereby the first three habitat levels are defined by abiotic factors such as the presence of photic zone and substrate type. These habitat levels then constrain levels 4–6 that describe biotope. The finest biotope level 6 accounts only for biomass of dominating taxa, predominantly bivalves (HELCOM, 2013b). In the southwestern German Baltic EEZ, Schiele *et al.* (2015) show that molluscs dominate the biomass. Our data show that the most northeastern regions, where salinities are very low, are dominated by Amphipoda and Isopoda in Bothnian Bay, and Polychaeta (*Marenzelleria* spp.), Oligochaeta, and Chironomidae in the Gulf of Finland. In contrast to 328 biotopes defined by HUB, our communities are delineated based solely and entirely on biological data, taking into account the abundance or biomass of all the species present. We consider changing community composition as a continuum without necessarily ascribing a biological boarder to change of sediment type or photic zone on a map.

The dataset

To provide a solid statistical estimation, we included data with different sampling and laboratory methods, or stations not necessarily fully representative for the respective subareas. As already pointed out by Carman and Cederwall (2001), this approach can sometimes lead to under- or overestimates of total biomass for different subregions. Differences in gear and mesh size within the joint dataset (Table 1) can cause concerns about the validity of the classification. For example, no standard 1 mm sieved data are available for the

Russian part of the Gulf of Finland and for Estonian waters. Instead of excluding these regions, we chose to include the data based on 0.25 or 0.4 mm sieves. The risk of misleading estimate due to this methodological difference while comparing between sites was reduced by (i) considering only macrobenthic species, and (ii) the use of data transformation before the community analysis to reduce variance inflation. The total biomass values depend mainly on a few large species. The findings of Elmgren (1978) supported the point that within regions of relatively high macrofauna biomass, meiofauna are likely to have less importance for the total biomass.

Densities of individual species vary greatly seasonally but also between years. Due to these fluctuations, only rough estimates (with significant errors bars) can be made for the benthos biomass of the whole Baltic Sea and its individual subregions. For some regions and depth strata, the total macrofauna wet weight estimated in this study is in good agreement with Carman and Cederwall (2001), e.g. for the zone 70–120 m in Bothnian Bay and the Quark region. Our estimate for the 30–70 m strata of the same region is at least one order of magnitude higher than reported by Carman and Cederwall. Similar numbers hold for the Baltic Proper. Additionally, total biomass calculated for the deepest part of the region calculated in this study was two orders of magnitude larger (Table 6).

Spatial modelling and mapping

Many recent case studies that report spatial mapping of benthic fauna were focused either on some distinct regions (Glockzin and Zettler, 2008; Gogina *et al.*, 2010) or relating the distribution of biodiversity or single species to three physical predictors. These are sediment properties and two main environmental variables for brackish-water systems, salinity and oxygen supply (Ojaveer *et al.*, 2010; Zettler *et al.*, 2014). The presented analysis was using averages of all the available sampling events for faunal abundance and biomass over a 5 × 5 km grid which includes the integration of the temporal variation over the available data (Figure 1). The use of large grid cells also provides a smoothing effect on certain scales of the predicted major benthic macrofaunal communities over the Baltic Sea. This can be seen as a trade-off between capturing the most important large-scale patterns in their distribution while omitting the fine-scale structures phenomena.
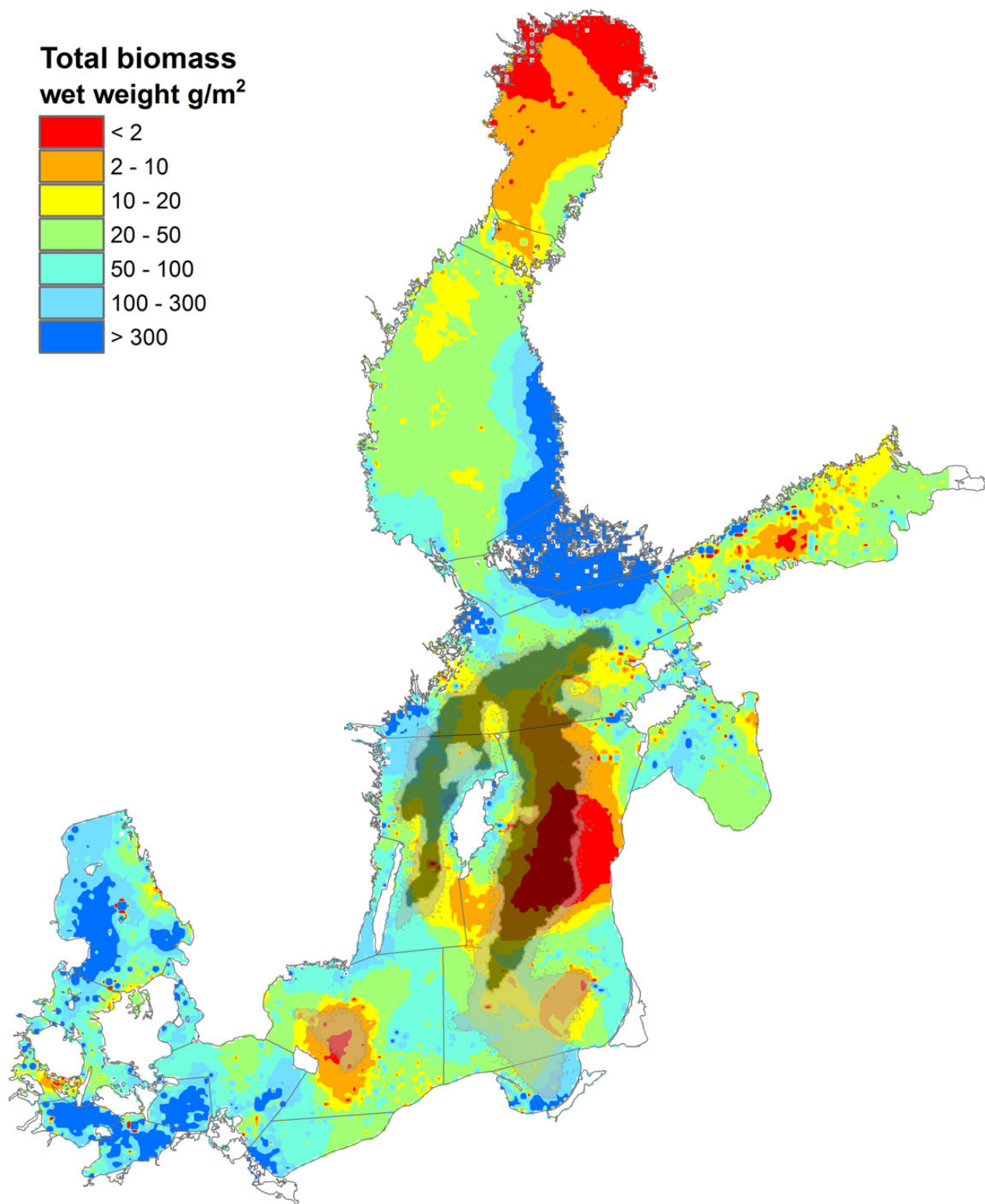
Figure 8. Distribution of interpolated total wet weight biomass, derived using ordinary kriging interpolation of available biomass data averaged per 5 × 5 km grid cell. Transparent light grey and dark grey areas mask out the deep water hypoxic and anoxic oxygen conditions. Note that at the areas where biomass data are lacking interpolation artefacts are evident, for instance, values at the shallow parts of the Eastern Gotland Basin at the west coast off Latvia are presumably too low. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.

To generate full coverage prediction maps based on classified point data, we have used the Random Forest method, chosen due to its best predictive performance achieved in most of the comparative studies (Collin *et al.*, 2011; Reiss *et al.*, 2011; Šiaulys and Bučas, 2012). Despite an overall good performance, this method has limitations. For example, *Marenzelleria* community 4

Table 5. Confusion matrix for the Random Forest of biomass-based classes (classes 1–17 are communites 1–17, class 18—"other", respectively).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	class.error
1	747	44	4	3	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0697
2	80	434	7	0	0	3	0	0	0	0	0	1	0	0	0	0	0	4	0.1796
3	0	5	128	0	0	4	0	0	0	4	0	0	0	0	0	0	0	16	0.1847
4	35	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.878
5	16	0	1	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3208
6	0	6	5	0	0	4	1	3	0	1	1	0	0	0	1	0	0	4	0.8462
7	0	1	2	0	0	0	16	0	0	0	1	1	2	2	0	0	0	4	0.4483
8	0	0	0	0	0	3	1	0	0	0	0	0	2	0	0	1	0	9	1
9	5	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10	0	3	7	0	0	2	0	0	0	0	0	0	0	0	0	0	0	5	1
11	0	0	1	0	0	1	3	0	0	0	7	0	2	0	0	0	0	4	0.6111
12	0	2	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	3	0.875
13	0	1	0	0	0	1	1	0	0	0	2	0	7	0	0	0	0	2	0.5
14	0	0	0	0	0	0	2	0	0	0	0	0	0	10	0	0	0	3	0.3333
15	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	0	3	0.875
16	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	7	0	2	0.4167
17	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
18	8	19	18	0	0	2	9	3	0	1	2	1	2	1	0	1	0	33	0.67

OOB estimate of error rate: 23.1%
Cohen's κ coefficient: 0.89

Note that Random Forest fails to predict classes 4, 6, 8–10, 12, 17, and 18 based on available predictors.
Model note: FWHcl381slog_2_sedcr4: factor(cHclFW381s) ∼ iowtopo_bs + tempulf + speedUlf + saltmean_L + factor(factor(sedcr), levels = c("1", "2", "3", "4", "5")) + saltstd3 + speedst3 + tempstd3.

Table 6. Storage of macrofauna wet weight in tonnes $\times 10^4$ in different subareas (HELCOM division) and depth strata of the Baltic Sea (derived using ordinary kriging interpolation of total biomass averaged per 5 km grid cell) and for selected regions comparison to numbers for these subareas and depth strata published in Carman and Cederwall (2001).

HELCOM subbasin	0–30 m	30–70 m	70–120 m	>120 m	Sum
1. Bothnian Bay	11.2	6.9	2.6		20.8
2. The Quark	7.4	0.2			7.6
Sum for 1 and 2	18.6	7.1	2.6		28.4
Factor	3.2	12.5	1.1		3.3
3. Bothnian Sea	226.3	165.2	78.4	14.2	484.1
4. Aland Sea	14.0	15.0	6.0	7.2	42.2
5. Archipelago Sea	1118.0	11.5			1129.4
Sum for 3, 4, and 5	1358.2	191.6	84.5	21.4	1655.7
Factor	9.4	1.3	2.2	2.1	4.9
6. Gulf of Finland	75.3	30.9	7.5		113.7
Factor	0.7	0.3	6.2		0.5
7. Arkona Basin	150.6	48.9			199.5
8. Bay of Mecklenburg	125.2				125.2
9. Bornholm Basin	93.7	97.3	11.8		202.8
10. Eastern Gotland Basin	54.0	76.0	45.2	5.8	180.9
11. Kiel Bay	148.9				148.9
12. Northern Baltic Proper	179.1	130.1	51.5	13.7	374.5
13. Western Gotland Basin	52.0	43.0	46.7	7.0	148.7
Sum for 7–13	803.4	395.4	155.2	26.5	1380.5
Factor	0.8	1.1	14.1	132.5	1.0
Great Belt	71.9	1.9			73.7
Gulf of Gdańsk	19.7	22.4	15.7		57.8
Gulf of Riga	77.4	34.7			112.1
Kattegat	336.5	59.5	2.4		398.4
Little Belt	20.2	0.1			20.4
The Sound	28.0				28.0
Overall sum	2809.4	743.6	267.8	47.9	3868.7

That is, *Factor* shown in italics is the ratio between current storage and corresponding estimate reported in Carman and Cederwall (2001).

in the biomass-based classification is not well captured by the Random Forest model (see error rate in Table 5). Important predictors that drive and explain the distribution patterns of this community, such as oxygen concentrations, are not included in this study.

Moreover, we did not investigate interspecific interactions. The size of modelled classes strongly influences the resulting model, favouring the performance for the largest class at the cost of the smaller one. This is in agreement with the findings of Diesing *et al.* (2014) that

misclassifications are mainly associated with uncommon (rarely sampled) classes. Apart from an unbalanced design and lack of available well-suited predictors, the low model skill for such class as "other" in our study is caused by very high natural variability. Rather than classifying models as correct or incorrect, they should be viewed as useful starting points for different applications (Reiss *et al.*, 2011). The presented community maps are general overviews and many local details including temporal variability of benthic fauna are masked out in our representation. Due to a coarse approach however, these maps must be used with caution.

The relevance of the presented maps and their temporal validity is also limited by the regional variability of benthic communities. In some regions, the community structure is stable over time, whereas in other areas, higher fluctuations or even regime shifts due to species invasions are recorded [e.g. Gogina *et al.* (2014) in the southwestern Baltic Sea; Maximov (2011), Maximov *et al.* (2014) for the eastern Gulf of Finland; Kauppi *et al.* (2015) for the major part of the Baltic Sea area]. The variable environment of the Baltic Sea (especially oxygen and temperature conditions) is another reason of temporal changes in benthic communities (Laine, 2003; Rousi *et al.*, 2013).

Comparison with historical data

In his work from the beginning of the last century, Petersen (1913) mapped the benthic macrofauna communities of the western Baltic Sea and Swedish west coast, delineating a *Brissopsis lyrifera* community in the Kattegat and an *Echinocardium cordatum* community along the Danish coast southwest of Læsø. *Macoma* communities are mostly reported further south to the Great Belt and the Sound down to Rügen. In our study, these areas match our biomass-based community 14 (*B. lyrifera, Amphiura* spp.*, Thracia convexa, Scoletoma fragilis, Hyala vitrea, Lipobranchius jeffreysii*), community 11 (*Phoronis* spp.*, Lanice conchilega, A. islandica, Te. fabula, Ec. cordatum, Nucula sulcata, Ma. calcarea*) and community 2 (*He. diversicolor, My. arenaria, Hydrobiidae, C. glaucum*), correspondingly. As for abundance-based communities, these were community 5 (*Amphiura* spp., *Ab. nitida, G. oculata, E. tenuis, T. flexuosa, N. nitidosa, D. glaucus*), community 7 (*Phoronis* spp., *T. fabula, Th. phaseolina, O. borealis, S. bombyx, B. lanceolatum, Sp. arndti*), and community 2 (*Hydrobiidae, P. elegans, C. glaucum*), respectively. In Petersen (1918), an *Ab. alba* community occurs around the Island of Funen down to deepest parts of the Femarnbelt and the Bay of Mecklenburg. This roughly overlapped with our biomass-based community 3 (*Diastylis, Astarte, My. truncata, Dendrodoa grossularia, Ab. alba, L. koreni, Trochochaeta multisetosa, Lineus ruber*) and community 6 (*Capitella* spp.), possibly indicating the latter to be a degraded variation of the former one. Naturally, communities' distribution largely deviates, and it is important to outline that this not only mirrors distributional changes over time, or depiction of different characteristic species. This highlights the large transitional nature of areas that cannot be captured by statistical methods employed in this study.

Comparing the penetration of some marine and brackish-water animals into the Baltic Sea reported in Zenkevitch (1963), our study shows that many species overlap (e.g. *Gammarus* spp.). Others indicate different distribution limits according to the recent data. For example, distribution limits for *Mya* penetrate more to the north, whereas *Priapulus* does not occur beyond the 55.5°N. Zenkevitch reports only 25 known polychaete species in the southern part of the Baltic Sea. In contrast, our data list over 90 polychaete species, which are found in the Bay of Mecklenburg alone. Moreover, by comparing our aggregated dataset with the data published by Jarvekulg (1979), we show the more northerly penetration of the Gastropoda *Alderia modesta* in the Bothnian Sea. Bivalvia *Ma. calcarea* has its easternmost record on the edge of Arkona Basin, i.e. absent in the inner parts, according to our data. These inconsistencies might be caused by shifts in the species distribution during the last 30 years. Moreover, our larger set of collected field data helped to detect species, which were overlooked in former times. Finally, the differences in taxonomic expertise can be responsible for the deviations.

Conclusions and outlook

The present standardized dataset represents an important baseline for the present state of the benthic macrofauna in the Baltic Sea. It can therefore be used for multiple analyses of benthic fauna distribution and its features and has been a unique opportunity to collate all benthic macrofauna data from the Baltic Sea region to finally create Baltic Sea benthic communities maps, targeted at the community level.

The results strive towards a Baltic wide picture, and of course when going more in to details, more groups and species would appear important. The work will continue with further refinement of presented results and description of finer community units in selected regions.

A different approach, in contrast to the presented study for spatially explicit mapping, is to model key species distributions separately, similar to Schiele *et al.* (2015), where communities can be derived by overlaying the expected averaged abundance and biomass and estimating the expected dominance. Moreover, smaller grid cell sizes might be considered (1 km, 1 nm, 2 km as in HELCOM), as well as 10 km (EEA reference grid) or even an adaptive grid. An estimation of the spatial variability within a grid cell should be addressed in future studies and should include temporal changes in local community structure. This study considered abundance and biomass-based communities separately, although incorporation of these two metrics into a single meaningful metric remains an open question.

This paper provides a baseline for future studies for better recognition of links between benthic fauna and ecosystem functioning. As an outlook based on the derived maps, the spatial distribution of specific macrozoobenthos functions (e.g. biodeposition and bioturbation) can be estimated. This will help to better understand the possible effects of natural or human-induced changes in macrofauna communities in the hosting ecosystem.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript. Additionally, the data outputs are made accessible through http://gis.ices.dk/sfdev/ and http://geo.ices.dk/geonetwork/srv/en/ under "Benthic faunal communities in the Baltic Sea".

Acknowledgements

References

Al-Hamdani, Z., and Reker, J. 2007. Towards marine landscapes in the Baltic Sea. BALANCE Interim Report #10.

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology, 26, 32–46.

Anderson, M. J., and Walsh, D. C. I. 2013. What null hypothesis are you testing? PERMANOVA, ANOSIM and the Mantel test in the face of heterogeneous dispersions. Ecological Monographs 83: 557–574.

Anisimova, N. A., Jørgensen, L. L., Lyubin, P. A., and Manushin, I. E. (Eds) 2010. Mapping and monitoring of benthos in the Barents Sea and Svalbard waters: results from the joint Russian–Norwegian benthic programme 2006–2008. IMR-PINRO Joint Report Series 1-2010. 114 pp.

Breiman, L. 2001. Random Forests. Machine Learning 45: 5–32.

Bučas, M., Bergstrom, U., Downie, A. L., Sundblad, G., Gullstrom, M., von Numers, M., Šiaulys, A., *et al.* 2013. Empirical modelling of benthic species distribution, abundance, and diversity in the Baltic Sea: evaluating the scope for predictive mapping using different modelling approaches. ICES Journal of Marine Science, 70: 1233–1243.

Carman, R., and Cederwall, H. 2001. Sediments and macrofauna in the Baltic Sea—characteristics, nutrient contents and distribution. *In* A Systems Analysis of the Baltic Sea, pp. 289–327. Ed. by F. Wulff, L. Rahm, and P. Larsson. Springer-Verlag, Berlin. 455 pp.

Clarke, R. K., Somerfield, P. J., and Chapman, M. G., 2006. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. Journal of Experimental Marine Biology and Ecology 330: 55–80.

Clarke, K. R., and Warwick, R. M. 2001. Change in Marine Communities: an Approach to Statistical Analysis and Interpretation, 2nd ed. PRIMER-E, Plymouth. 172 pp.

Collin, A., Archambault, P., and Long, B. 2011. Predicting Species Diversity of Benthic Communities within Turbid Nearshore Using Full-Waveform Bathymetric LiDAR and Machine Learners. PLoS ONE 6: e21265.

Cohen, J. 1968. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70: 213–220.

De Cáceres, M., Legendre, P., Wiser, S. K., and Brotons, L. 2012. Using species combinations in indicator value analyses. Methods in Ecology and Evolution, 3: 973–982.

Diesing, M., Green, S. L., Stephens, D., Lark, R. M., Stewart, H. A., and Dove, D. 2014. Mapping seabed sediments: comparison of manual, geostatistical, object-based image analysis and machine learning approaches. Continental Shelf Research, 84: 107–119.

Dufrêne, M., and Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs, 67: 345–366.

Ekman, S. 1933. Die biologische Geschichte der Nord- und Ostsee. *In* Die Tierwelt der Nord- und Ostsee, 23(1.b). Ed. by G. Grimpe, and E. Wagler. Akademische Verlagsgesellschaft, Leipzig. 40 pp.

Ekman, S. 1935. Tiergeographie des Meeres. Akademische Verlagsgesellschaft, Leipzig. 542 pp.

Elith, J., Kearney, M., and Phillips, S. 2010. The art of modelling range-shifting species. Methods in Ecology and Evolution, 1: 330–342.

Elmgren, R. 1978. Structure and dynamics of Baltic benthos communities, with particular reference to the relationship between macro- and meiofauna. Kieler Meeresforsch Sonderh, 4: 1–22.

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

Gogina, M. A., Darr, A., and Zettler, M. L. 2014. Approach to assess consequences of hypoxia disturbance events for benthic ecosystem functioning. Journal of Marine Systems, 129: 203–213.

Gogina, M. A., Glockzin, M., and Zettler, M. L. 2010. Distribution of benthic macrofaunal communities in the western Baltic Sea with regard to near-bottom environmental parameters. 1. Causal analysis. Journal of Marine Systems, 79: 112–123.

Gogina, M. A., and Zettler, M. L. 2010. Diversity and distribution of benthic macrofauna in the Baltic Sea. Data inventory and its use for species distribution modelling and prediction. Journal of Sea Research, 64: 313–321.

Gräwe, U., Holtermann, P., Klingbeil, K., and Burchard, H. 2015. Advantages of vertically adaptive coordinates in numerical models of stratified shelf seas. Ocean Modelling, 92: 56–68.

HELCOM, 2013a. HELCOM Red List of Baltic Sea species in danger of becoming extinct. Baltic Sea Environmental Proceedings, No. 140.

HELCOM. 2013b. HELCOM HUB models of stratified shelf seas. Ocean Mode biotope and habitat classification. Baltic Sea Environmental Proceedings, No. 139.

Hessle, C. 1924. Bottenbonitoringar i inre Östersjön. Meddelanden fran Kungliga Lantbruksstyrelsen, 250: 1–52.

Jarvekulg, A. 1979. Benthic Fauna of the Eastern Part of the Baltic Sea: Composition and Distribution Ecology. Valgus, Tallinn. 382 pp. (in Russian).

Kauppi, L., Norkko, A., and Norkko, J. 2015. Large-scale species invasion into a low-diversity system: spatial and temporal distribution of the invasive polychaetes *Marenzelleria* spp. in the Baltic Sea. Biological Invasions, 17: 2055–2074.

Laine, A. O. 2003. Distribution of soft-bottom macrofauna in the deep open Baltic Sea in relation to environmental variability. Estuarine, Coastal and Shelf Science, 57: 87–97.

Leipe, T., Tauber, F., Vallius, H., Virtasalo, J., Uścinowicz, S., Kowalski, N., Hille, S., *et al.* 2011. Particulate organic carbon (POC) in surface sediments of the Baltic Sea. Geo-Marine Letters, 31: 175–188.

Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. R News, 2: 18–22.

Maximov, A. A. 2011. Large-scale invasion of *Marenzelleria* spp. (Polychaeta; Spionidae) in the eastern Gulf of Finland, Baltic Sea. Russian Journal of Biological Invasions, 2: 11–19.

Maximov, A. A., Bonsdorff, E., Eremina, T., Kauppi, L., Norkko, A., and Norkko, J. 2015. Context-dependent consequences of *Marenzelleria* spp. (Spionidae: Polychaeta) invasion for nutrient cycling in the Northern Baltic Sea. Oceanologia, 57: 342–348.

Maximov, A. A., Eremina, T. R., Lange, E. K., Litvinchuk, L. F., and Maximova, O. B. 2014. Regime shift in the ecosystem of the eastern Gulf of Finland caused by the invasion of the polychaete *Marenzelleria arctia*. Oceanology, 54: 46–53.

Norkko, A., Villna, A., Norkko, J., Valanko, S., and Pilditch, C. 2013. Size matters: implications of the loss of large individuals for ecosystem function. Scientific Reports, 3: 2646.

Ojaveer, H., Jaanus, A., Mackenzie, B. R., Martin, G., Olenin, S., Telesh, I., Zettler, M. L., *et al.* 2010. Status and change of biodiversity in the Baltic Sea. PLoS ONE, 5: e12467.

Perus, J., and Bonsdorff, E. 2004. Long-term changes in macrozoobenthos in the Åland archipelago, northern Baltic Sea. Journal of Sea Research, 52: 45–56.

Petersen, C. G. J. 1913. Valuation of the sea II. The animal communities of the sea bottom and their importance for marine zoogeography. Reports of the Danish Biological Station, 21. 44 pp.

Petersen, C. G. J. 1918. The sea bottom and its production of fish food. A survey of the work done in connection with valuation of the Danish waters from 1883–1917. Reports of the Danish Biological Station, 26. 62 pp.

Rees, H. L., Eggleton, J. D., Rachor, E., and Vanden Berghe, E. (Eds) 2007. Structure and dynamics of the North Sea benthos. ICES Cooperative Research Report, 288. 258 pp.

Reiss, H., Cunze, S., König, K., Neumann, H., and Kröncke, I. 2011. Species distribution modelling of marine benthos; a North Sea case study. Marine Ecology Progress Series, 442: 71–86.

Revelle, W. 2015. psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, IL, USA. http://CRAN.R-project.org/package=psychVersion=1.5.4.

Roberts, D. W. 2015. labdsv: Ordination and Multivariate Analysis for Ecology. R package version 1.7-0. http://CRAN.R-project.org/package=labdsv.

Rousi, H., Laine, A. O., Peltonen, H., Kangas, P., Andersin, A-B., Rissanen, J., Sandberg-Kilpi, E., *et al.* 2013. Long-term changes in coastal zoobenthos in the northern Baltic Sea: the role of abiotic environmental factors. ICES Journal of Marine Science, 70: 440–451.

Saint-Germain, M., Buddle, C. M., Larrivee, M., Mercado, A., Motchula, T., Reichert, E., Sackett, T. E., *et al.* 2007. Should biomass be considered more frequently as a currency in terrestrial arthropod community analyses? Journal of Applied Ecology, 44: 330–339.

Schiele, K. S., Darr, A., Zettler, M. L., Friedland, R., Tauber, F., von Weber, M., and Voss, J. 2015. Biotope map of the German Baltic Sea. Marine Pollution Bulletin, 96: 127–135.

Schiewer (Ed.) 2008. Ecology of Baltic Coastal Waters. Ecological Studies 19. Springer-Verlag, Berlin Heidelberg 430 pp.

Segerstråle, S. G. 1957. Baltic Sea. *In* Treatise on Marine Ecology and Paleoecology, pp. 751–800. Ed. by J. W. Hedgpeth. Geological Society of America, Memoirs 67.

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

Šiaulys, A., and Bučas, M. 2012. Species distribution modelling of benthic invertebrates in the south-eastern Baltic Sea. Baltica 25: 163–170.

Väinölä, R., and Varvio, S-L. 1989. Biosystematics of *Macoma balthica* in northwestern Europe. *In* Reproduction, Genetics and Distributions of Marine Organisms: 23rd European Marine Biology Symposium, School of Biological Sciences, University of Wales, Swansea, 5–9 September 1988. International Symposium Series, pp. 309–316. Ed. by J. S. Ryland, and P. A. Tyler. Olsen & Olsen, Fredensborg. 469 pp.

Välikangas, I. 1933. Über die Biologie der Ostsee als Brackwassergebiet. Stuttgart. 419 pp.

Warzocha, J. 1995. Classification and structure of macrofaunal communities in the southern Baltic. Archive of Fishery and Marine Research, 42: 225–237.

Wei, C. L., Rowe, G. T., Escobar-Briones, E., Boetius, A., and Soltwedel, T. 2010. Global patterns and predictions of seafloor biomass using Random Forests. PLoS ONE, 5: e15323.

Zenkevitch, L. 1963. Biology of the Seas of the U.S.S.R. Allen & Unwin, London. 953 pp.

Zettler, M. L., Karlsson, A., Kontula, T., Gruszka, P., Laine, A., Herkül, K., Schiele, K., *et al.* 2014. Biodiversity gradient in the Baltic Sea: a comprehensive inventory of macrozoobenthos data. Helgoland Marine Research, 68: 49–57.

Handling editor: Michel Kaiser