Marine Pollution Bulletin 104 (2016) 129–138

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Reducing spatial variation in environmental assessment of marine benthic fauna

Kjell Leonardsson [a,*], Mats Blomqvist [b], Rutger Rosenberg [c,d]

[a] Department of Wildlife, Fish, and Environmental Studies, Swedish University of Agricultural Sciences, SE 901 83, Umeå, Sweden
[b] Hafok AB, SE 179 61, Stenhamra, Sweden
[c] Marine Monitoring AB, Strandvägen 9, SE 453 30, Lysekil, Sweden
[d] Department of Biology and Environmental Sciences — Kristineberg, University of Gothenburg, SE 451 78, Fiskebäckskil, Sweden

ARTICLE INFO

Article history:
Received 21 September 2015
Received in revised form 28 January 2016
Accepted 29 January 2016
Available online 5 February 2016

Keywords:
Benthic quality index
BQI
Depth
Marine Strategy Framework Directive
Salinity
Sediment
Water Framework Directive

ABSTRACT

The Benthic Quality Index, BQI, is widely used for benthic quality assessment. Here, we investigated if spatial variation in the BQI can be reduced by accounting for the environmental factors instead of having different boundaries for different salinity regimes between status classes in the EU Water Framework Directive and Marine Strategy Framework Directive. For this purpose we tested salinity, sediment structure, and depth in a regression model to test their contribution to variations in BQI. The spatial variation in BQI was better explained by depth than by salinity or sediment structure. The proposed assessment method uses the residuals from the regression model between BQI and depth. With this method the variance in BQI between samples was reduced by 50% to 75% in the majority of situations. A method to establish the boundary between *good* and *moderate* status and how to derive EQR-values according to the WFD is presented.

1. Introduction

We live in a changing world and there is a great need to scientifically analyse the temporal and spatial changes we observe and to evaluate how recorded changes affect the ecosystem. Changes in the ecosystem can be beneficial to humans when nutrients have a top-down effect and lead to increased fish catches without significant side effects. In contrast, excess nutrient input can have serious negative effects on ecosystem services and oxygen conditions. It is therefore essential to assess, in a scientific way with accuracy and precision, how ecosystems can cope with various pressures and how resilience can be maintained.

In this study we focus on methods for the assessment of temporal and spatial changes in benthic communities in Swedish marine waters, more specifically in the Skagerrak, Kattegat and the Sound between Denmark and Sweden. Benthic communities are ideal for such assessment studies as the animals are rather stationary and the community structure changes in a rather predictable way to various environmental pressures such as organic enrichment, hypoxia, metal pollution and physical disturbance (Pearson and Rosenberg, 1978; Heip, 1995; Josefson et al., 2009). Along with increased anthropogenic stress, benthic faunal diversity declines and the proportion of tolerant species increases. Similarly, a low or variable salinity reduces the number of benthic species, which has been demonstrated for estuaries and brackish water areas. Elliott and Quintino (2007) described this as the "estuarine quality paradox" and stated that estuarine fauna have features similar to that in anthropogenically-stressed areas, and this makes it difficult to detect anthropogenically-induced stress and separate it from the effect of low and variable salinity.

In year 2000, the European Union introduced the Water Framework Directive (WFD), which was followed by the Marine Strategic Framework Directive (MSFD) in the year 2008, and today both directives provide for the assessment of ecosystem quality. The Benthic Quality Index (BQI) was initially introduced for assessing the ecological status according to WFD in Swedish waters (Rosenberg et al., 2004). BQI was calculated based on individual species sensitivity values, species dominance and number of species. BQI was further evaluated for use in benthic quality assessment by Leonardsson et al. (2009). BQI has proven to be useful for benthic quality assessment in Scandinavian waters (e.g. Perus et al., 2007) and its response to various pressures has been compared and evaluated in relation to Danish and Norwegian indices (Josefson et al., 2009). BQI has also been used for quality assessment in the Mediterranean (Labrune et al., 2006; Dimitriou et al., 2012; Karakassis et al., 2013) and other European waters (Grémare et al., 2009). Recently BQI was also introduced as an indicator for *good* environmental status of open sea and coastal waters for several descriptors within the framework of the MSFD implementation in Sweden (national regulation HVMFS2012:18).

* Corresponding author.
E-mail address: Kjell.Leonardsson@slu.se (K. Leonardsson).

http://dx.doi.org/10.1016/j.marpolbul.2016.01.050

By time, lots of new information about faunal distributions have been added and new assessments have been made. Evidence has also been recorded that sometimes and at certain times a strong and occasional recruitment of a particular species can reduce the sensitivity value considerably with additional negative effects for also other species in the same sample. Such events with strong recruitment of species considered as sensitive by expert judgements will flaw the quality assessment. This drawback has been pointed out by Labrune et al. (2006) and Grémare et al. (2009). Therefore, a change in the calculation of the sensitivity values, which have the greatest impact on BQI of all factors, was proposed and statistically evaluated by Leonardsson et al. (2015). They suggested using the observed number of species in each sample instead of the ES50-value as used earlier (ES50 is the estimated number of species among 50 individuals as interpolated from the rarefaction method (Hurlbert, 1971). Despite this improvement there is still a considerable spatial variation in the BQI-values, even within a single waterbody. The existing method to deal with some of this variation in Swedish west coast waters has been to apply different WFD status class boundaries in shallow and deeper waters (Rosenberg et al., 2004; Leonardsson et al., 2009). The halocline had been shown to have a significant impact on the structure of the benthic community composition on the Swedish west coast (Rosenberg and Möller, 1979). Depth-related boundaries between *High, Good, Moderate, Poor* and *Bad* in the WFD were therefore separated by the depth between the influence of brackish water of Baltic origin and oceanic water and was set at 20 m depth, i.e. the deeper distribution of the halocline in the Skagerrak-Kattegat area. An improvement of the index to explicitly account for this type of impact of, e.g. the salinity, would reduce the amount of sampling effort needed for accurate boundary setting and assessment.

Variation due to environmental factors is in part reduced by the characterization of water bodies into types and associated type specific status class boundaries as specified in the WFD. Types are two dimensional compromises between practical aspects related to the number of types and demands on narrow ranges in environmental factors for reduction of uncertainty in different indicators. Thus, there may still be significant variations in environmental factors within each type, a variation that is especially important for benthic biota dependent on factors related to depths. The effects of this variation can be reduced by stratification of types into subtypes, e.g. based on depths as in the current Swedish assessment system or salinity as in Muxika et al. (2007), or even stratification of water bodies into ecotopes based on salinity and depth (intertidal/subtidal) as in van Loon et al. (2015). As a consequence, optimizations of monitoring programmes have to take subtypes into account, which requires enough data from each subtype. For countries with a long and topographically complex coast as few subtypes as possible would be desired. An alternative has been to include environmental factors when setting reference values for separate metrics within the indicator formula as in the Danish DKI_{v2} (Carstensen et al., 2014), or the British IQI_{vIV} (Phillips et al., 2014).

Here we investigate an alternative approach where we remove as much as possible of the spatial variation in BQI by means of a regression model that includes the main factor(s) that contribute to the spatial variation. Having a regression model that successfully takes the environmental variables into account will simplify the sampling and the assessment since there is no need to divide the types into subtypes. With this approach the final assessment will not be based on the index values per se, but on the residuals from the regression model. This way of dealing with the BQI makes it transparent how the different environmental factors contribute in the model to improve the precision of the assessment. The aim is to reduce as much as possible of the variation, which means that there is also a need to analyse the contribution of each of the components of BQI to the uncertainty. One of the components, the sensitivity factor, has been analysed separately (Leonardsson et al., 2015), which leaves the species number and the abundance factor (see Eq. (1) below) for analyses in this paper. The environmental variables considered to be of relevance for the regression model were salinity, depth, and sediment characteristics. In this paper we evaluate how BQI is related to depth, salinity and sediment structure in Swedish marine waters.

2. Material and methods

2.1. Study area

The distribution of benthic data used in the analyses in this publication originates from Swedish coastal and open waters, more specifically from the Skagerrak, Kattegat and the Sound between Denmark and Sweden (Fig. 1). All samples are from the years 1965 through 2013, and encompass 855 stations from depths between 4 and 153 m. All samples were obtained by a 0.1 m^2 Smith–McIntyre grab and the samples were sieved on 1 mm meshes. Stations from areas with known impact were marked as impacted and excluded from most of the analyses. Impacted areas were occasionally or frequently exposed to hypoxia, physical disturbance or toxic substances. Fishing pressure was not included in our analysis because of lack of suitable data for this pressure and its effects on benthic communities.

2.2. Sediment characteristics

Each station was given a sediment class based on a transformation of static marine geological maps to seafloor surface sediments made by Hallberg et al. (2010). The underlying data for the marine geological maps are of different quality since investigation methods have developed over time resulting in better quality in Kattegat and the Sound compared to northern and outer parts of Skagerrak. The sediment classes are not always matching sediment information associated to each benthic sample, e.g. free text sediment description or sample volume. Since objective ways of classifying sediments from each benthic sample were lacking, e.g. data of grain size or organic content, the surface sediment map was the only way to achieve a sediment classification for all stations.

2.3. Salinity

Median salinity from nearby depths and areas during the period April–June and five years prior to benthic sampling was used as a salinity value for each benthic sample. Measured salinity data downloaded from SMHI (www.smhi.se), ICES (www.ices.dk) and AU (dce.au.dk) was primarily used. If such data were lacking for a sampling occasion, modelled salinity data from SMHI (vattenweb.smhi.se) was used, or as a last option from the EUSeaMap project (jncc.defra.gov.uk/page-5020). Selecting the median salinity during the five year period prior to benthic sampling was done to overcome differences in hydrographic sampling efforts between areas and years. The months April through June were chosen since they generally represent the lowest salinity values during the year. Different species and life stages might be affected by different values (min, max, median, range, etc.), but the choice of the median was because it is less sensitive to the number of values available.

2.4. Statistical analyses

To reduce the uncertainty in the BQI, the underlying sources contributing to the uncertainty need to be identified and explained. One way of doing this is by finding reliable and explicit relationships between the independent variables causing, direct or indirect, the variation in the index. These independent variables need to have measurable values at all possible sites in order to apply to the adjusted index anywhere in the region. Here we use a regression approach to find the environmental variables that will reduce the variation of the index substantially. Three variables were included in these analyses: depth, salinity and sediment type. All these variables have the potential to directly or indirectly affect the benthic community structure.

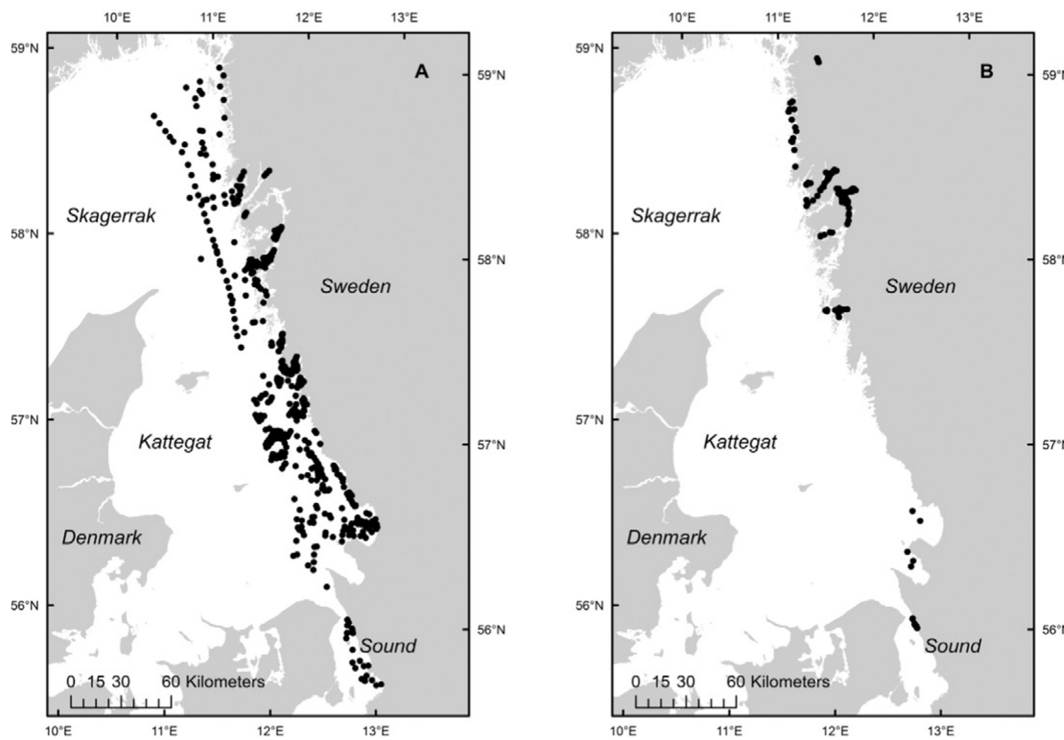

Fig. 1. Distribution of benthic stations in the Skagerrak, Kattegat and the Sound analysed in this study. Left map (A) shows stations without known disturbance and right map (B) shows stations that, at some time, have been disturbed.

Before applying the regression models, the uncertainty in the BQI was analysed in order to modify or remove factors that contribute most to the variation in the index values. This part of the analysis was performed by visual inspection of scatter plots with each of the three components: the sensitivity factor, the species number factor, and the abundance factor; and combinations of them plotted against the three environmental variables. The species sensitivity factor has already been analysed by Leonardsson et al. (2015), but the graphical illustration of the contribution of this factor was included in this paper for comparison with the other factors.

In analyses of the models, the first step was to find, when possible, a proper regression model for each of the variables separately. The medians of all the data per station were used in the analyses to avoid higher weights of data from frequently sampled sites. The visual inspection of graphs with the index plotted against the variables gave information of which type of regression model to apply, e.g. linear, quadratic, logistic, piecewise linear, etc. For comparison of the model fits, the Akaike's information criterion (AIC) from each analysis was used. Thereafter, the residuals from each analysis were plotted against each of the variables not in the model, e.g. the residuals from the model with depth were plotted against salinity. From these residual plots we checked by visual inspection if there was a possibility to explain the deviation among the residuals by means of the other variables. Such possibility should appear as linear or non-linear trends against the other variables. We preferred this manual approach rather than using an automatized stepwise regression since with our data the stepwise regression tended to produce over-fitting with relationships between the variables and the index that were lacking support in the literature, in terms of a biological explanation. One particular pattern that was significantly predicted by a stepwise regression was a dip in the BQI between 50 and 80 m depth. Since the sediments as well as the salinities were the same as at depths below 40 m it was the inclusion of a nonlinear transformation of depth that produced the better fit. Our interpretation is that the dip in the BQI at these depths may be a consequence of bottom trawling rather than a true depth effect, and hence a consequence of over-fitting since we had no data to include bottom trawling as a covariate. If this is the case, then the statistically best model produced by stepwise regression may make it difficult to detect trawling effects on the benthic community in the future assessment. The types of regression models applied were piecewise linear regressions for depth and salinity, and linear categorical regression for the sediment type. The piecewise linear regressions gave better fits to the data than the corresponding logistic regressions, which was the alternative approach to the seemingly sigmoid responses. Each regression analysis was followed by inspection of the residuals to check for outliers and normal distribution of the residuals. All statistical analyses were performed in Mathematica 10.0 (Wolfram Research Inc., 2014).

From the relationships between the index and its components and the variables in Fig. 2, it was clear that there were conspicuous relationships between the different components of the index and the environmental variables, except for the abundance factor ($N/(N+5)$) in the index (Eq. (1)). Based on these findings, we propose a new and updated formula for BQI–BQI_{2015}, that is the same as the original BQI formula (Rosenberg et al., 2004) but the sensitivity values in the sensitivity factor are based on species richness instead of ES50 (Eq. (2)). The reason for changing from ES50 to species richness in the calculations is that Leonardsson et al. (2015) showed that ES50 can be "unfairly" small when there is a strong dominance of one species in a sample. This can lead to erroneous sensitivity values. Despite the fact that the formula is the same as the original formula we have chosen to call it BQI_{2015} to emphasize that the sensitivity values are calculated in a 'new' way. The regression analyses were performed with ($BQI_{2009,}$ Leonardsson et al., 2009) and without the abundance factor (BQI_{2015}) but, in both cases, with the new sensitivity values based on number of species instead of ES50, as to find the index composition with the lowest uncertainty.

$$BQI_{2009} = \sum_{i=1}^{S_{classified}} \left(\frac{N_t}{N_{classified}} * Sensitivity\ value_t \right) * log_{10}(S+1) * N/(N+5) \tag{1}$$

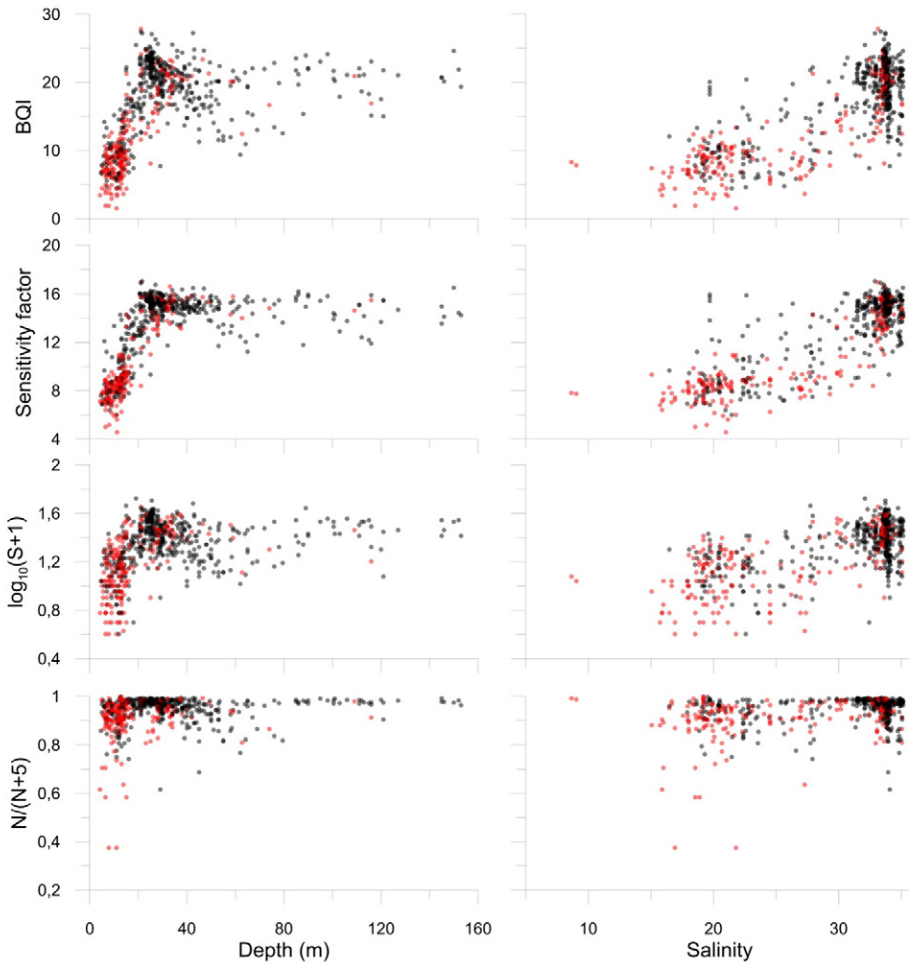

Fig. 2. Relations between water depth and salinity for the different factors in BQI. BQI was calculated according to Eq. (1), i.e. with abundance factor. Red dots are for coarse sediment and sand; black dots represent soft mud. A single dot represents the median from all grab samples at a specific sampling station. Overlaps between dots produce darker colours. Grabs from environmentally impacted stations were excluded.

$$BQI_{2015} = \sum\nolimits_{i=1}^{S_{classified}} \left(\frac{N_t}{N_{classified}} * Sensitivity\,value_t \right) * log_{10}(S+1) \qquad (2)$$

The *Sensitivity values* (Eqs. (1) and (2)) used in this paper were derived by Leonardsson et al. (2015), *i* is taxa number, *S* is the total number of taxa in the sample, $S_{classified}$ is the number of taxa with sensitivity values, N_i is the number of individuals of taxa *i* in the sample and *N* the total number of individuals in the sample (0.1 m^2). Taxa not given a sensitivity value are excluded from the sensitivity factor but included in the total number of species when calculating BQI.

The next step was to establish a method for assessment using the residuals from the regression model based on BQI_{2015} and the relevant environmental variable(s). Since there was a clear shift in BQI from low values at shallow less saline areas towards higher values at deeper and high saline areas there is a need to transform the residuals to avoid situations with larger minimum residuals at shallow bottoms than at deep bottoms, which would be the case when there is no fauna, i.e. BQI = 0. To ensure the same minimum residual independent on depth the transformation function was derived analytically from the regression model by setting BQI = 0 and solving for the residual.

In this paper we treat the boundary for MSFD *good* environmental status in a similar way as the WFD boundary between *good* and *moderate* ecological status despite their theoretical differences (e.g. Van Hoey et al., 2010). To define the WFD boundary between *good* and *moderate* (*GM*) ecological status and the MSFD boundary for open sea *good* environmental status (hereafter together named *GM*-boundary) the transformed residuals from the "undisturbed" areas (Fig. 1A) was used to calculate the lower one-tailed 95% confidence limit based on averages of single samples collected during the same year from five sites per water type to have at least some spatial representativeness of the assessment. The resampling was nested in such a way that in the first step a water type was selected at random to have a balanced design for the spatial dimension, even if there were data from a varying number of years from the different water types. Thereafter one of the available sampling years were selected at random, followed by a random selection of five sampling sites, and finally a single random sample was selected from each of these five sampling sites. This procedure was repeated with replacement 60,000 times from the "undisturbed" data set (baseline data) to reach the precision of one decimal. The underlying assumption in using this approach is that when assessing the environment quality with new data, the mean residual from the regression model will be compared against the *GM*-boundary, which will be the same as testing if the new data is significantly different from the baseline data, independent on the spatial origin of the new data. Since the relationship between BQI and depth turned out to be more or less independent on water type, no effort was made to derive separate GM's for each water type. In the near future, when more data are available from the water bodies, the relevant spatial unit in the first step of the bootstrap should be a water body rather than a water type since it is the water body that needs to be assessed in the WFD.

For transformation of the numerical results to the EQR scale for the WFD, two more steps were taken. First, to avoid the need of adjusting the transformation over time by updating the maximum value of the residuals an upper limit for the residuals (r_{max}) is applied. That is, r_{max} is defined as the absolute value of r_{min}, which in turn is the residual originating from BQI = 0. The EQR-value is then given by Eq. (3). This transformation ensures the EQR-values to be in the range 0–1.

$$EQR = \frac{Mean\,residual + Abs(r_{min})}{2Abs(r_{min})}, \tag{3}$$

where Abs(r_{min}) is the absolute value of the residual obtained when BQI = 0 and mean residual is the mean of the residuals obtained from the regression model when applied to BQI-values from the waterbody to be assessed.

3. Results

3.1. Evaluation of the abundance factor

Using the BQI-formula described in Eq. (1) (including the abundance factor) and comparing the relationship to depth as well as salinity, clearly suggests that the impact of the abundance factor is of minor importance. There was a clear visual relationship between the different components of the index and the variables tested except for the abundance factor (N / (N + 5)) (Fig. 2). Both the sensitivity factor and the species number factor, $\log_{10}(S + 1)$, contributed more to the relationship between BQI and the environmental variables than the abundance factor. This has been proven both for salinity and depth. This implies that the use of the BQI-formula presented in Eq. (2) (that lacks the abundance factor) is the most appropriate way forward.

3.2. Relationships between BQI_{2015}, depth, salinity and sediment type

All three factors of BQI show an increase with increasing depth up to at least 20 m depth (Fig. 2), although the depth dependence is less conspicuous for the abundance factor than for the sensitivity and the species richness factors. BQI and the sensitivity factor had a dip for a number of samples from around 60 to 100 m depth. The relationship between the three factors and salinity was more variable with less clear patterns. Comparing BQI to substrate characteristics showed that low index values were generally associated with coarse sediment (Fig. 3). Sandy stations showed a greater variation in BQI_{2015}, whereas finer sediments were associated with greater index values. Comparison with medians of BQI_{2009} displayed similar results.

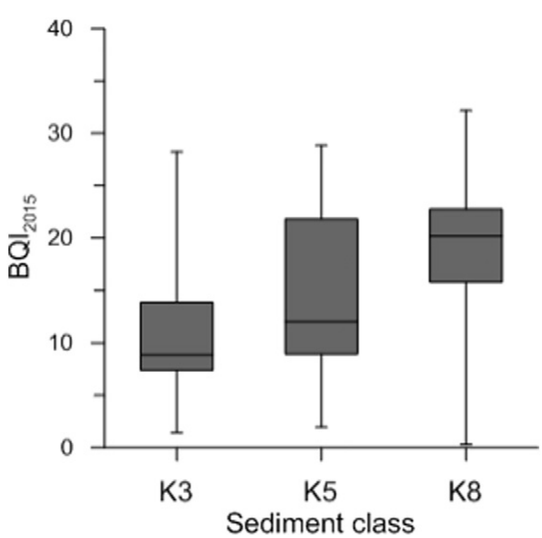

Fig. 3. Medians of BQI_{2015} (without the abundance factor) per station vs. sediment characteristics. K3 is sand, coarse sand, gravel, shell gravel and pebbles, K5 is fine sand, and K8 is soft mud. The horizontal lines show the median, the box the 25th and 75th percentiles and whiskers the full range of data. Sediment classifications are from Hallberg et al. (2010).

3.3. Depth as a proxy for salinity and sediment type

The results implicate that depth can be used as a proxy for both salinity and sediment type since both can be related to the depth. Median salinity distributions in relation to depth show a strong increase in salinity from about 4 m down to 20 to 30 m depth with a fairly stable salinity around 34 units deeper than 30 m (Fig. 4). Coarse sediment is predominantly found in shallow waters, fine sand in exposed areas around the halocline and soft mud dominates in deeper waters. Coarse sediments were associated with low median BQI_{2015} and mud with generally higher indices (Fig. 3). The distribution of sediment characteristics in relation to depth in the Skagerrak, Kattegat and Sound is shown in Fig. 4.

3.4. Regression models to explain spatial variation in the index

The nonlinear relationships between BQI and depth and between BQI and salinity, shown in Fig. 2, may be accounted for by either logistic regression models or piecewise linear regression models. In the comparison between these two types the piecewise regression produced the best fit, i.e. lowest AIC. For the sediment, a categorical variable could be added to the model. The single variable piecewise linear regression model resulted in a better fit using depth as an independent variable compared to using salinity (Table 1). Both depth and salinity generated better fit than using the sediment type as the independent variable. These results were similar for the BQI with and without the abundance factor, but the exclusion of the abundance factor in BQI improved the model fit significantly.

The fit of the best model, piecewise linear regression of BQI_{2015} as a function of depth, was described by a constant level from 4 m to 11.2 m water depth, an increasing trend from 11.2 to 21.7 m, and remaining constant below 21.7 m depth. The residuals from this model did not show any trend in relation to salinity (Fig. 5A). For the three sediment types the residuals centred close to zero (Fig. 5B). Consequently, the inspection of the residuals gave no hint on how to improve the regression model by including salinity or the sediment in the model. For this reason, the model with depth as the only independent variable was selected as the final model for BQI_{2015}; see Table 2 for the model parameters and Fig. 6 for the fit to the data.

The next step in the model-evaluation was to investigate if the residuals from the model centred around zero independently of the typology. The samples originated from nine different water types, including two open sea areas. The model seems applicable in the entire region, although some adjustments of the boundaries may be needed for some of the water types when applied to the WFD and the MSFD (Fig. 7).

There is one complication introduced by the regression model that needs to be solved. According to the regression model the predicted BQI-values for samples from deep areas was greater than that for samples from shallow areas (Fig. 6). Still the BQI_{2015} could be zero both at shallow and deep bottoms, which means that the minimum value of the residuals could differ between shallow and deep bottoms. Thus, a transformation to obtain the same range for the residuals, independent of depth, is necessary.

The transformation needed is given by the regression equation and when using the residual range for the shallow bottoms as the target the following transformation is required for each of the three segments in the regression model:

Depth < 11.2 m; no transformation,
Depth 11.2–21.7 m; negative residuals are multiplied by $8.52 / (-4.85 + 1.19 * \text{depth})$,
Depth ≥21.7 m; negative residuals are multiplied by $8.52 / (-4.85 + 1.19 * 21.7) = 0.405$,

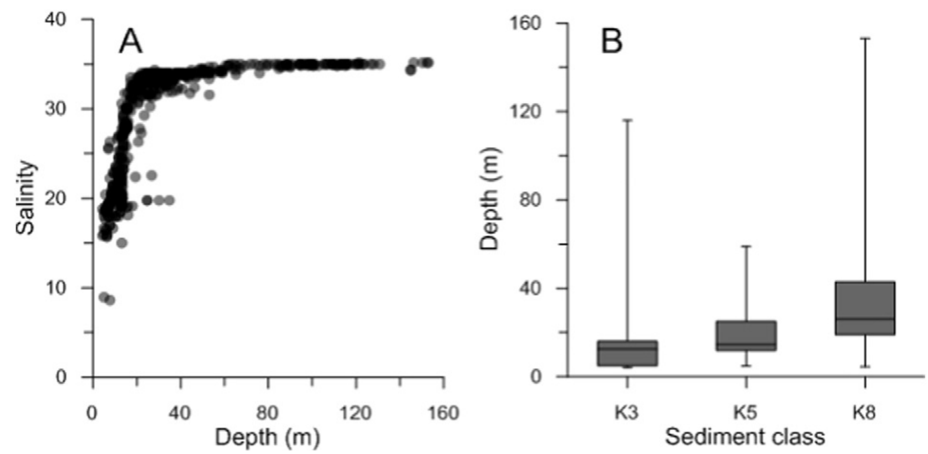

Fig. 4. Median salinity (A) and boxplot of sediment classification (B) at unaffected stations (Fig. 1A) in the Skagerrak, Kattegat and Sound vs. depth. Salinity data are from SMHI, AU, ICES and measurements at similar depths where benthic fauna was collected. Overlaps between dots produce darker colours. K3 is sand, coarse sand, gravel, shell gravel and pebbles, K5 is fine sand, and K8 is soft mud. The horizontal lines show the median, the box the 25th and 75th percentiles and whiskers the full range of data. Sediment classifications are from Hallberg et al. (2010).

where the numbers originate from the regression parameters in Table 2. In order to avoid adjusting the calculation of WFD EQR-values over time, by updating the maximum value of the residuals, an upper limit for the residuals was also applied as the absolute value of the residuals of shallow bottoms (i.e. 8.52). Untransformed and transformed residuals from all samples, including samples from "disturbed" areas are shown in Fig. 8.

3.5. Comparison of uncertainty with and without depth adjustment

To obtain a measure of the improvement by the depth adjustment, we compared the depth transformed variances with those where the predicted value from the regression was added to the residuals for which the comparison was made. The reason to add the predicted value for the already transformed residuals was to remove the effect of the actual depth transformation. A bootstrap was made with 100,000 random draws of five grabs from the entire dataset. For each grab we calculated the depth adjusted residual (r_a) and the depth adjusted residual to which the predicted BQI_{2015} was added (r_b). The ratio of the variances of r_a and r_b was calculated. A ratio of 0.5 means that the depth adjustment reduced the variance by 50%. Out of the 100,000 variance ratios, 84% were below 1.0, 62% were below 0.5, and 42% were below 0.25. That is, for 42% of the variance ratios the variance was reduced by more than a factor four. This means that the depth adjusted values will reduce the uncertainty considerably in the assessment compared to unadjusted values. Another major improvement of the depth adjustment is that we no longer need to have two sets of WFD boundaries, above and below 20 m depth, within each water body in our study area. This will facilitate the design of monitoring programmes and assessment of status significantly.

Table 1
Model fit of the different models in terms of AIC and ΔAIC (difference to best model fit). BQI_{2009} denotes BQI with the abundance factor described in Leonardsson et al. (2009), and BQI_{2015} denotes original BQI without the abundance factor, described in Rosenberg et al. (2004) but with the new way of calculating sensitivity value described in Leonardsson et al. (2015).

Model	AIC	ΔAIC
$Y = BQI_{2015}$(depth)	3173.3	
$Y = BQI_{2009}$(depth)	3297.1	123.7
$Y = BQI_{2015}$ (salinity)	3350.4	177.0
$Y = BQI_{2009}$ (salinity)	3453.0	279.6
$Y = BQI_{2015}$ (sediment)	3919.2	745.8
$Y = BQI_{2009}$ (sediment)	3937.0	763.7

3.6. Boundary setting and assessment

The approach here by using a baseline data without data from disturbed areas based on expert opinion restrict our possibilities in setting other WFD and MSFD status class boundaries than that between *good* and *moderate* ecological status and the *good* environmental status (*GM*-boundary). Nevertheless this *GM-boundary* needs to be derived for the specific conditions required for the status assessment, which in the Swedish system will be the same for the WFD and MFSD. The Swedish BQI-approach calls for at least five samples from each assessment unit to be assessed to allow at least some spatial representativeness. This means that the *GM*-boundary should account for this minimum number of samples. The data used in the regression analyses were considered to represent baseline conditions in the meaning *good* or *high* ecological status and good environmental status, and therefore the *GM*-boundary should correspond to the limit below which the mean residual is significantly lower compared to be expected from the baseline data. That is, if the mean of the five transformed residuals used in the assessment is below the 5% significance limit of the baseline data then the status should be classified as less than *good*. Finding the *GM*-boundary therefore reduces to finding the lower one-tailed 95% confidence limit of the mean transformed residuals based on five samples bootstrapped from the available combinations of waterbody/type and year from the baseline data. The residuals used to set the boundary should not be based on the transformed residuals originating from the median BQI_{2015}-values from each station, but rather on data from single grabs and just one sample from each station.

The mean transformed residuals derived from the bootstrap were normally distributed around zero and the lower one-tailed 95% confidence limit was calculated to −2.5 BQI_{2015} units (Fig. 9). A large number of draws, about 60 000, were needed to reach a precision of one decimal.

The remaining 5% below the boundary means that there is a 5% risk that the environment is classified as less than *good* status by chance even if the true status is *good*. When performing the assessment, the mean residuals from all samples from an assessment unit will be compared to this boundary. If the mean residual is above the boundary, the environment is classified as at least *good* and if below, the environment is classified as below *good* and measures to improve quality has to be taken.

A final step needed to conform to the WFD is to transform the result to have EQR-values between 0 and 1, where 1 denotes reference conditions. The method for this is described in the methods section and when the parameters from the regression model is known as in this case, the application of Eq. (3) yields EQR = (mean residual + 8.52) / 17.04. From

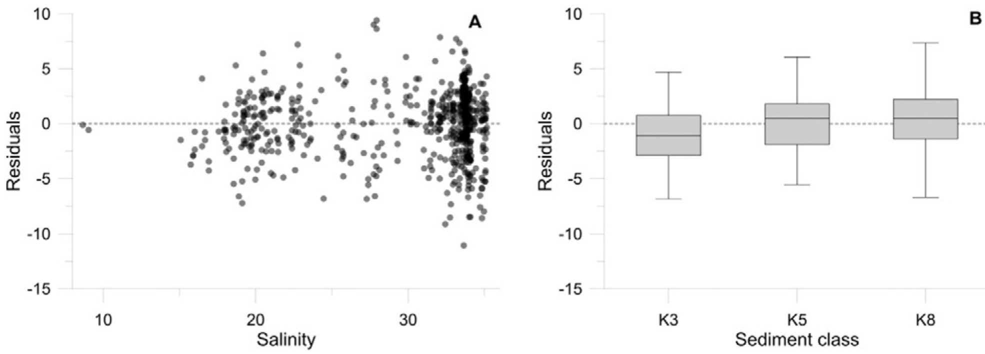

Fig. 5. Residuals from the best fit model against salinity (A) and sediment classes (B) for benthic samples (n = 645) in the Skagerrak, Kattegat and Sound where environmentally impacted stations were excluded. Overlaps between dots produce darker colours. In B the horizontal lines show the median, the box the 25th and 75th percentiles and whiskers the full range of data. No relationships between residuals and salinity or sediment classes can be seen in the graphs.

this the boundary between *good* and *moderate* ecological status is found at $EQR = (-2.5 + 8.52) / 17.04 = 0.35$. If we use the same principle upwards (boundary between *high* and *good* ecological status), one-tailed 95% confidence limit, we obtain $EQR = 0.65$ which means that the EQR-range for *good* ecological status is 0.3 while the range for *high* status is 0.35. The range for *good* conditions covered 90% of the data from the baseline conditions due to the normal distribution of the residuals centering around $EQR = 0.5$.

4. Discussion

About a century ago, one of the pioneers in benthic ecology, the Dane C.G.J. Petersen (Petersen, 1913) described the benthic community composition in the Skagerrak, Kattegat and the Sound. At that time the input of nutrients and pollutants was significantly lower compared to today. He found that the composition, particularly of the dominant species, could be similar over large areas and was related to depth, salinity and sediment composition. These rather uniform distribution patterns have changed since then, and mainly eutrophication and hypoxia have affected benthic communities negatively, particularly in enclosed and salinity-stratified areas (Rosenberg et al., 1990). In Fig. 1 we have identified some of those areas, which were excluded from parts of our analyses. Also trawling has an impact in many areas of the seabed outside the coast, but there is no information about the intensity to relate this to particular areas and times of benthic sampling. However, today there is a great concern to analyse spatial and temporal changes and trends in the marine ecosystem, and new and improved scientific methods are presented and applied.

The EU WFD and MSFD provides for status assessment of spatial units, water bodies and regions/subregions respectively. One assessment unit can only have one status irrespective of the presence of different spatial subareas, such as depth zones, habitats or ecotopes etc., characterized by different environmental variables. These subareas within assessment units can possibly have different levels of index values resulting in a large spatial variation of index values within an assessment unit. Splitting up the environment into subtypes is one way to deal with this variation, but when large and diverse coastal areas have to be assessed this approach requires additional resources for sampling compared to if the variation can be explained by means of a statistical model.

In this paper we suggest a method to reduce spatial variation in a benthic quality index to improve the assessment of status according to the EU WFD and MSFD without the need of stratification into habitats/subtypes. In the benthic marine environment depth, salinity and sediment characteristics are all important structuring factors contributing to the spatial variation in the benthic communities (Gray and Elliott, 2009). Despite the known influence of these three factors our analyses somewhat surprisingly resulted in depth alone as the factor explaining most of the variation. Using a regression model to account for the depth variation in BQI_{2015} would therefore remove much of the spatial variation and hence reduce the uncertainties in the assessments. According to the results of the regression analyses the remaining variation could not be further reduced by using salinity or sediment characteristics as independent variables. While an improvement of the assessment based on depth adjustment alone was acceptable, it was also clear that depth rather than salinity showed to be the best explanatory variable. With salinity as an explanatory variable it could be argued that the underlying process could be differences in salinity tolerances between species. The biological explanation for depth having a better explanation in the model than salinity might be that the depth covaries with the salinity down to 20 to 30 m depth, and that the depth in a certain range could be associated with food availability.

The production and distribution of food is one of the primary factors affecting the distribution and biomass of benthic communities (Pearson and Rosenberg, 1987). Ott (1992) developed general models describing the coupling between the pelagic and the benthic systems regarding food transport and showed that it is related to depth. In oligotrophic

Table 2
Parameter estimates for the best fit model, $Y = BQI_{2015}(\text{depth})$. The regression model is a piecewise linear regression with two breakpoints (BP), for depths $<BP_{Lower}$ then $BQI_{2015} = A$, for depths $\geq BP_{Upper}$ then $BQI_{2015} = B$, and for depths between BP_{Lower} and BP_{Upper} then $BQI_{2015} = -4.85 + 1.19 * \text{depth}$. The regression equation was derived from the four parameters.

Parameter	Estimate	Standard error	t-Statistic	P-value
A	8.52	0.30	28.6	<0.001
B	21.0	0.16	135.6	<0.001
BP_{Lower}	11.2	0.38	27.0	<0.001
BP_{Upper}	21.7	0.43	48.6	<0.001

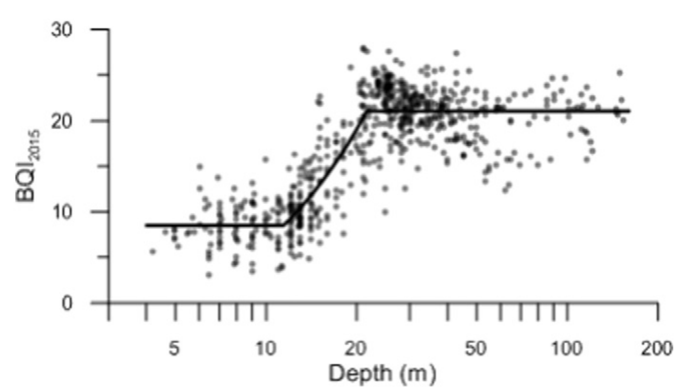

Fig. 6. Relation between depth and station median BQI_{2015} for benthic samples (n = 645) in the Skagerrak, Kattegat and Sound where environmentally impacted stations were excluded. Overlaps between dots produce darker colours. The regression line of the piecewise regression model is denoted by the solid line.

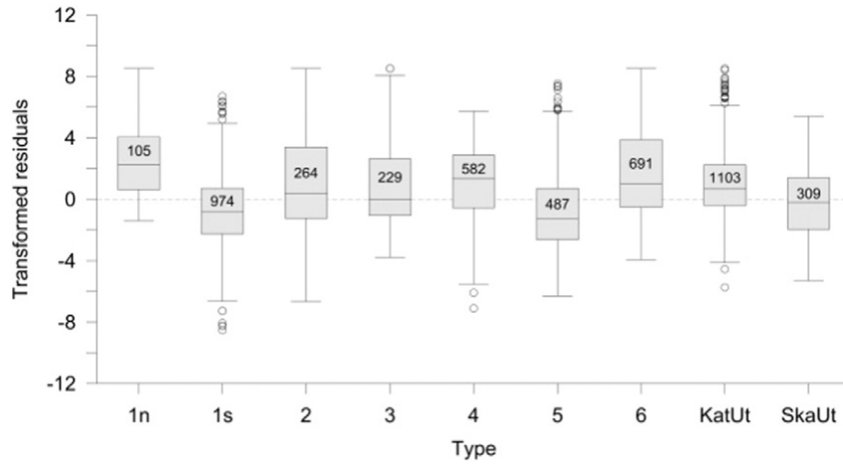

Fig. 7. Box plots of the residuals from the best-fit model for baseline data in each of the water types at the Swedish west coast where the samples originated. The horizontal lines show the median, the box the 25th and 75th percentiles and whiskers are 1.5 times the height of the boxes. Numbers within boxes denote number of samples. Type 1 represents inner coastal waters (northern and southern part), type 2 fjords, type 3 Skagerrak outer coastal waters, type 4 Kattegat outer coastal waters, type 5 south Halland and north Sound coastal waters and type 6 Sound coastal waters. KatUt represents Kattegat open sea and SkaUt Skagerrak open sea outside the WFD coastal zone but within Swedish EEZ and MSFD assessment area.

systems, most of the organic material will be mineralised in the water column, whereas in eutrophic systems, as in the present study, a greater proportion will accumulate at the seabed. The transport of organic matter through lateral advection can be 2 to 8 times greater as lateral velocities generally are higher than vertical transport rates (Graf, 1992). In soft mud at accumulation areas such as below the halocline, the number of functional groups in the benthos is greater compared to in shallow waters (Pearson, 2001). As an example, Rosenberg et al. (2000) found a significant correlation between depth and the faunal variables abundance and biomass in the northern Kattegat and attributed this to advection of organic material. Thus, the comparatively more stable temperatures and salinities in deeper waters and a rich food supply provide generally better conditions for the benthic fauna than in shallow waters where the temporal variation of temperature and salinity is greater. In conclusion, depth may not have a direct effect of the benthic faunal structure, but indirectly via food production as well as a co-variation with salinity and sediment structure motivate its presence in the regression model. Both salinity and sediment classification is also associated with considerable variation and uncertainty which also can explain why depth alone gives the best model fit. High quality sediment classification can objectively be made by measurement of sediment parameters during sampling in the field. In contrast, salinity is more problematic to deal with due to temporal variation and difficulties in selection of a relevant measure (min, max, median, range, etc).

In other parts of the sea, other environmental variables than depth alone might be important for benthos possibly resulting in different models for different sea areas. However the principle of using the residuals after adjustment for environmental variability in assessment of status can still be of use irrespective of the environmental variables used in the model.

One advantage in having a baseline dataset that fulfils the criteria of *good* or *high* ecological status (WFD), without the need to distinguish between the two, and *good* environmental status (MSFD), is that new data can be tested if it is significantly below the *GM*-boundary set by means of the baseline data. We assessed the *GM*-boundary by bootstrapping mean transformed residuals from single samples from each of five sites and consequently the test should also be based on data from single samples from each of five sites. When data comes from more than five sites in a water body, a bootstrap can be made to create a frequency distribution of mean-residuals to allow an assessment of the uncertainty in the status classification, in addition to the 5% risk of making a type I error by using data from five samples in the assessment. These types of tests also open up for the possibility of defining rules for updating the baseline data.

We have here demonstrated a method for the derivation of the *GM*-boundary by using data from the entire study area. The residuals from a majority of water types centred around zero which means that the GM-boundary needs no further adjustment for these types. In the near future, when more data with better spatial coverage will be available, adjustments of the *GM*-boundaries may be needed for those water types where the residuals deviates from the zero-centering, see e.g. Fig. 7.

Long term climatic trends are likely to affect the faunal communities and eventually such impacts need to be reflected in the assessment boundaries since a return to the original state may no longer be possible (Duarte et al., 2009). The work with the assessment needs to be adaptive in the sense that the indicators are improved as new knowledge is

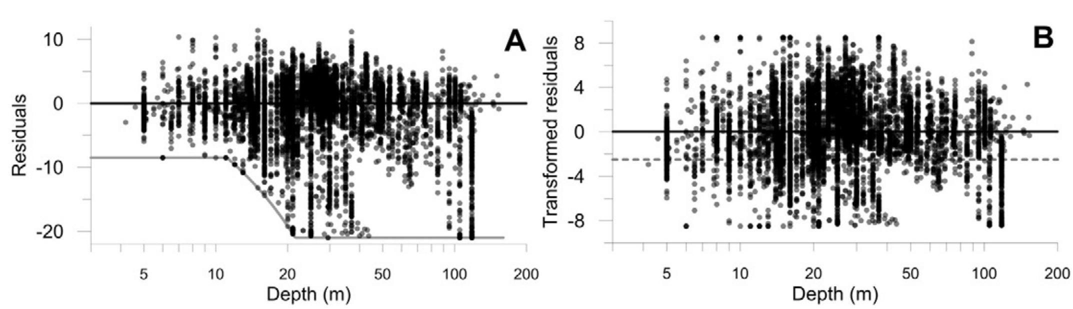

Fig. 8. Residuals (A) and transformed residuals (B) of BQI_{2015} plotted against depth with the model-fit presented as the solid black line and the dashed black line denotes the GM-boundary. The lower grey line in A represents the residuals when BQI_{2015} is zero. Each dot represents one sample from the whole study area, including samples from "disturbed" areas. Overlaps between dots produce darker colours.

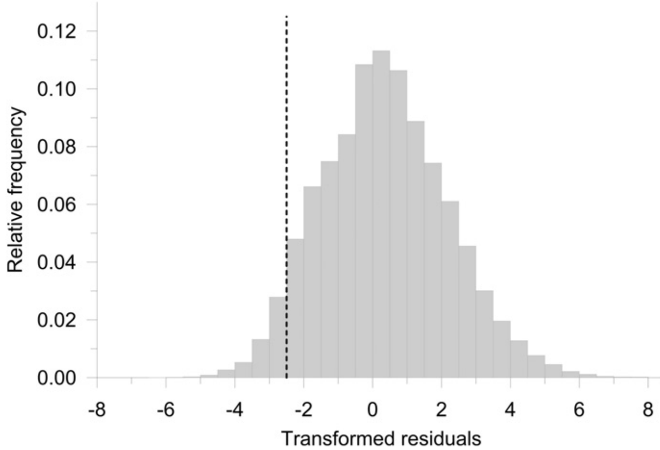

Fig. 9. Histogram of 60,000 bootstrapped means of five transformed residuals after adjusting BQI_{2015} for depth, using data from areas without apparent anthropogenic disturbance. The bootstrapping was hierarchical to balance the spatial variation, with random sampling with replacement in the sequence water type, sampling year, five sites, and finally one sample from each of the five sites. The dashed vertical line denotes the 5th percentile, which is here suggested as the GM-boundary.

obtained but also by the need to adjust the boundaries for the ecological status over time. There are two alternatives available to update the baseline data. One alternative is to use a statistical test with some specified criteria that needs to be fulfilled in order for new data to enter the baseline data. Another alternative is to define geographical areas that serve as baseline areas from which all new data are entering the baseline data. The first alternative is appealing in the sense that there are some distinct criteria to follow. However, there is a risk that criteria specified in advance introduce a trend in the baseline data and this trend will propagate into a corresponding trend in the boundaries. For example, if one uses a traditional test to check if a new dataset is not significantly different from the baseline data, there will be a declining trend if the majority of the new data is in the lower half of the baseline data. Accepting the alternative with baseline areas does not introduce these types of trends, but it opens up for having an improved coverage of the natural variation over time. It also captures long term trends in the environment as could be the case with the climate change (e.g. Altieri and Gedan, 2015), and with the more short term change due to invasion of new species (e.g. Olenin et al., 2014). The drawback of using specified geographical areas as baseline areas is that it is hard to find areas without any type of anthropogenic disturbance, and a certain degree of expert opinion will most likely be needed to specify these areas, as was done with the baseline data we used here.

The assessment method we describe is built on a statistical model which makes it easy to explore and improve the accuracy of the assessment index when more data is available. In that respect it deviates from most of the existing methods used to assess the environmental status in marine benthic fauna. The most common approach in the assessment of the marine environment has been the use of multimetric indices (e.g. M-AMBI, Muxika et al., 2007 Danish DKI_{v2}, Carstensen et al., 2014; British IQI_{vIV}, Phillips et al., 2014; Dutch BEQI2, van Loon et al., 2015). The different components in a multimetric index can be adjusted for environmental variation by means of covariates. A more common approach is to normalize each component by dividing by its maximum value, which in turn may have been adjusted for the environment, e.g. by using salinity. Another solution would be to use a statistical model directly in the assessment, either by means of full structural modelling by adding a latent variable (e.g. Chiu et al., 2013) or by using the residuals as the latent variable. We used the latter approach, which resulted in a large reduction in assessment uncertainty. An advantage in using a statistical model and a latent variable is that the statistical model can be updated when new data are available, without the need to change the index per se. Increased knowledge based on future research can therefore be used to improve the statistical model, which will benefit the accuracy of the assessment of the environment.

Acknowledgment

This work was partly funded by the Swedish Environmental Protection Agency (Dnr 10/179) and Swedish Agency for Marine and Water Management through contract for the research programme WATERS. Data used in this article were to a large extent collected within Swedish National and regional monitoring programmes as well as within research projects on benthic soft bottom invertebrates along the Swedish coast.

References

Altieri, A.H., Gedan, K.B., 2015. Climate change and dead zones. Glob. Chang. Biol. 21, 1395–1406. http://dx.doi.org/10.1111/gcb.122754.

Carstensen, J., Krause-Jensen, D., Josefson, A., 2014. Development and Testing of Tools for Intercalibration of Phytoplankton, Macrovegetation and Benthic Fauna in Danish Coastal Areas. Scientific Report from DCE – Danish Centre for Environment and Energy No. 93. Aarhus University, DCE – Danish Centre for Environment and Energy (85 pp.). (http://dce2.au.dk/pub/SR93.pdf).

Chiu, G.S., Wu, M.A., Lu, L., 2013. Model-based assessment of estuary ecosystem health using the latent health factor index, with application to the Richibucto estuary. PLoS One 8 (6), e65697. http://dx.doi.org/10.1371/journal.pone.0065697.

Dimitriou, P.D., Apostolaki, E.T., Papageorgiou, N., Reizopoulou, S., Simboura, N., Arvanitidis, C., Karakassis, I., 2012. Meta-analysis of a large data set with Water framework directive indicators and calibration of a benthic quality index at the family level. Ecol. Indic. 20, 101–107.

Duarte, C.M., Conley, D.J., Carstensen, J., Sánchez-Camacho, M., 2009. Return to Neverland: shifting baselines affect eutrophication restoration targets. Estuar. Coasts 32, 29–36. http://dx.doi.org/10.1007/s12237-008-9111-2.

Elliott, M., Quintino, V., 2007. The estuarine quality paradox, environmental homeostatis and the difficulty of detecting anthropogenic stress in naturally stressed areas. Mar. Pollut. Bull. 54, 640–645.

Graf, G., 1992. Benthic-pelagic coupling: a benthic view. Oceanogr. Mar. Biol. Annu. Rev. 30, 149–190.

Gray, J.S., Elliott, M., 2009. Ecology of Marine Sediments. From Science to Management. Oxford University Press, Oxford (225 pp.).

Grémare, A., Labrune, C., Vanden Berghe, E., Amouroux, J.M., Bachelet, G., Zettler, M.L., Vanaverbeke, J., Fleischer, D., Bigot, L., Maire, O., Deflandre, B., Craeymeersch, J., Degraer, S., Dounas, C., Duineveld, G., Heip, C., Herrmann, M., Hummel, H., Karakassis, I., Kedra, M., Kendall, M., Kingston, P., Laudien, J., Occhipinti-Ambrogi, A., Rachor, E., Sarda, R., Speybroeck, J., Van Hoey, G., Vincx, M., Whomersley, P., Willems, W., Wlodarska-Kowalczuk, M., Zenetos, A., 2009. Comparison of the performances of two biotic indices based on the MacroBen database. Mar. Ecol. Prog. Ser. 382, 297–311.

Hallberg, O., Nyberg, J., Elhammer, J., Erlandsson, C., 2010. Ytsubstratklassning Av Maringeologisk Information. SGU-Rapport: Dnr 08–1565/2009, Rapport Maringeologi Nr 2010:1 (SGU: 26 pages, In Swedish with summary in English, http://resource.sgu.se/produkter/sgurapp/s1006-rapport.pdf).

Heip, C., 1995. Eutrophication and zoobenthos dynamics. Ophelia 41, 113–136.

Hurlbert, S.H., 1971. Nonconcept of species diversity: critique and alternative parameters. Ecology 52 (4), 577–586.

Josefson, A.B., Blomqvist, M., Hansen, J.L.S., Rosenberg, R., Rygg, B., 2009. Assessment of marine quality change in gradients of disturbance: comparison of different Scandinavian multi-metric indices. Mar. Pollut. Bull. 58, 1263–1277.

Karakassis, I., Dimitriou, P.D., Papageorgiou, N., Apostolaki, E.T., Lampadariou, N., Black, K.D., 2013. Methodological considerations on the coastal and transitional benthic indicators proposed for the Water Framework Directive. Ecol. Indic. 29, 26–33.

Labrune, C., Amouroux, J.M., Sarda, R., Dutrieux, E., Thorin, S., Rosenberg, R., Grémare, A., 2006. Characterization of the ecological quality of the coastal Gulf of Lions (NW Mediterranean). A comparative approach based on three biotic indices. Mar. Pollut. Bull. 52, 43–47.

Leonardsson, K., Blomqvist, M., Rosenberg, R., 2009. Theoretical and practical aspects on benthic quality assessment according to the EU-Water framework directive — examples from Swedish waters. Mar. Pollut. Bull. 59, 1286–1296.

Leonardsson, K., Blomqvist, M., Magnusson, M., Rosenberg, R., 2015. Calculation of species sensitivity values and their precision in marine benthic faunal quality indices. Mar. Pollut. Bull. 93, 94–102.

Muxika, I., Borja, Á., Bald, J., 2007. Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water framework directive. Mar. Pollut. Bull. 55, 16–29.

Olenin, S., Narščius, A., Minchin, D., David, M., Galil, B., Gollasch, S., Marchini, A., Occhipinti-Ambrogi, A., Ojaveer, H., Zaiko, A., 2014. Making non-indigenous species information systems practical for management and useful for research: an aquatic perspective. Biol. Conserv. 173, 98–107. http://dx.doi.org/10.1016/j.biocon.2013.07.040.

Ott, J.A., 1992. The Adriatic Benthos: Problems and Perspectives. In: Colombo, G., Ferrari, I., Ceccherelli, V.U., Rossi, R. (Eds.), Marine Eutrophication and Population DynamicsProceedings of the 25th EMBS. Olsen and Olsen, Fredensborg, pp. 367–378.

Pearson, T.H., 2001. Functional group ecology in soft-sediment marine benthos: the role of bioturbation. Oceanogr. Mar. Biol. Annu. Rev. 39, 233–267.

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

Pearson, T.H., Rosenberg, R., 1987. Feast and Famine: Structuring Factors in Marine Benthic Communities. In: Gee, J.H.R., Giller, P.S. (Eds.), 27th Symposium of the British Ecological Society Aberystwyth 1986. Blackwell Scientific Publications, Oxford, pp. 373–395.

Perus, J., Bonsdorff, E., Bäck, S., Lax, H.-G., Villnäs, A., Westberg, V., 2007. Zoobenthos as indicators of ecological status in coastal brackish waters: a comparative study from the Baltic Sea. Ambio 36, 250–256.

Petersen, C.G.J., 1913. Valuation of the sea. 2. The animal communities of the sea bottom and their importance for marine zoogeography. Rep. Dan. Biol. Stat. 21, 1–44.

Phillips, G.R., Anwar, A., Brooks, L., Martina, L.J., Miles, A.C., Prior, A., 2014. Infaunal Quality Index: Water Framework Directive Classification Scheme for Marine Benthic Invertebrates. Report: SC080016. ISBN: 978-1-84911-319-9.

Rosenberg, R., Möller, P., 1979. Salinity stratified benthic macrofaunal communities and long-term monitoring along the west coast of Sweden. J. Exp. Mar. Biol. Ecol. 37, 175–203.

Rosenberg, R., Elmgren, R., Fleischer, S., Jonsson, P., Persson, G., Dahlin, H., 1990. Marine eutrophication case studies in Sweden. Ambio 19, 102–108.

Rosenberg, R., Nilsson, H.C., Hellman, B., Agrenius, S., 2000. Depth correlated benthic faunal quantity and infaunal burrow structures on the slopes of a marine depression. Estuar. Coast. Shelf Sci. 50, 843–853.

Rosenberg, R., Blomqvist, M., Nilsson, C.H., Cederwall, H., Dimming, A., 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water framework directive. Mar. Pollut. Bull. 49, 728–739.

Van Hoey, G., Borja, A., Birchenough, S., Buhl-Mortensen, L., Degraer, S., Fleischer, D., Kerckhof, F., Magni, P., Muxica, I., Reiss, H., Scröder, A., Zettler, M.L., 2010. The use of benthic indicators in Europe: from the Water framework directive to the marine strategy framework directive. Mar. Pollut. Bull. 60, 2187–2196.

Van Loon, W.M.G.M., Boon, A.R., Gittenberger, A., Walvoort, D.J.J., Lavaleye, M., 2015. Application of the benthic ecosystem quality index 2 to benthos in Dutch transitional and coastal waters. J. Sea Res. 103, 1–13. http://dx.doi.org/10.1016/j.seares.2015.05.002.

Wolfram Research, Inc., 2014. Mathematica, Ver. 10.0 (Champaign, IL).