Ecological Indicators 89 (2018) 667–679

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Article

A regional benthic fauna assessment method for the Southern North Sea using Margalef diversity and reference value modelling

Willem M.G.M. van Loon[a,*], Dennis J.J. Walvoort[b], Gert van Hoey[c], Christina Vina-Herbon[d], Abigayil Blandon[d], Roland Pesch[e], Petra Schmitt[e], Jörg Scholle[e], Karin Heyer[f], Marc Lavaleye[g], Graham Phillips[h], Gerard C.A. Duineveld[g], Mats Blomqvist[i]

[a] Rijkswaterstaat, Ministry of Infrastructure and the Environment, Zuiderwagenplein 2, 8224 AD Lelystad, The Netherlands
[b] Wageningen Environmental Research, Wageningen University & Research, Droevendaalsesteeg 3, P.O. Box 47, 6700 AA, Wageningen, The Netherlands
[c] Flanders Research Institute for Agriculture, Fisheries and Food, Department Aquatic Environment and Quality, Ankerstraat 1, 8400 Oostende, Belgium
[d] JNCC (Joint Nature Conservation Committe), Monkstone House, City Road, Peterborough, PE1 1JY United Kingdom
[e] BIOCONSULT Schuchardt & Scholle GbR, Reeder-Bischoff-Str. 54, 28757 Bremen, Germany
[f] Dr. Karin Heyer, Consultancy, Krähenhorst 9, 22587 Hamburg, Germany
[g] NIOZ Royal Netherlands Institute for Sea Research and Utrecht University, P.O. Box 59, 1790 AB, Den Burg, The Netherlands
[h] Environment Agency (EA), Kingfisher House, Goldhay Way, Orton Goldhay, Peterborough PE2 5ZR, United Kingdom
[i] Hafok AB, SE 179 61, Stenhamra, Sweden

ARTICLE INFO

Keywords:
OSPAR
MSFD
Marine benthos
Index
MMI
Multi-metric index
Benthic assessment method
Reference value estimation
Model
Margalef diversity
Species richness
AMBI
ITI
Shannon index
PIE
SN
SNA
Fishing pressure
Organic enrichment
Sedimentation

ABSTRACT

The aims of this study are to develop an optimized method for regional benthic fauna assessment of the Southern North Sea which (a) is sensitive and precise (quantified as the slope and the R^2 value of the pressure-impact relationships, respectively) for the anthropogenic pressures bottom fishing and organic enrichment, (b) is suitable for estimating and modelling reference values, (c) is transparent, (d) can be efficiently applied using dedicated software; and to apply this method to benthic data from the Southern North Sea.

Margalef diversity appeared to be the best performing benthic index regarding these criteria, even better than several Multi-Metric Indices (MMIs) containing e.g. AMBI (AZTI Marine Biotic Index) and ITI (Infaunal Trophic Index). Therefore, this relatively simple and very practical index, including a new reference value estimation and modelling method, and BENMMI software were selected as a common OSPAR (Oslo Paris convention) method for the benthic fauna assessment of the Southern North Sea. This method was applied to benthic fauna data from the Southern North Sea collected during the period 2010–2015. The results in general show lower normalized Margalef values in coastal areas, and higher normalized Margalef values in deeper offshore areas.

The following benthic indices were compared in this study: species richness, Margalef diversity, SNA index, Shannon index, PIE index, AMBI, ITI. For each assessment area, the least disturbed benthic dataset was selected as an adjacent 6 year period with, on average, the highest Margalef diversity values. For these datasets, the reference values were primarily set as the 99th percentile values of the respective indices. This procedure results in the highest stable reference values that are not outliers. In addition, a variable percentile method was developed, in which the percentile value is adjusted to the average bottom fishing pressure (according to data from the International Council for the Exploration of the Sea, ICES) in the period 2009–2013. The adjusted percentile values were set by expert judgement, at 75th (low fishing pressure), 95th (medium fishing pressure) and 99th (high fishing pressure) percentile. The estimated reference values for Margalef diversity correlate quite well with the median depth of the assessment areas using a sigmoid model (pseudo-$R^2 = 0.86$). This relationship between depth and Margalef diversity was used to estimate reference values in case an assessment area had insufficient benthic data

For testing the effects of bottom fishing pressure, normalized index values (NIV; index value divided by reference value) were used. The rationale for using NIVs is the assumption that, although a certain level of bottom fishing pressure will have a larger absolute effect on more biodiverse benthic communities in deeper waters than on more robust and less biodiverse coastal benthic communities, the relative effects (tested using NIVs) are comparable. A clear exponentially decreasing relationship ($R^2 = 0.26$–0.27, $p < 0.00001$) was found between both bottom surface and subsurface fishing activity (penetration depth < 2 cm and > 2 cm, respectively) and normalized Margalef diversity values, with an asymptotic normalized Margalef value of 0.45 at a

* Corresponding author.
E-mail address: willem.van.loon@rws.nl (W.M.G.M. van Loon).

http://dx.doi.org/10.1016/j.ecolind.2017.09.029
Received 31 December 2016; Received in revised form 13 September 2017; Accepted 14 September 2017
1470-160X/

subsurface fishing activity > 2.3 sweeps/year. This asymptotic value is predominantly found in coastal waters, and probably shows that the naturally more robust coastal benthic communities have been transformed into resilient benthic communities, which rapidly recover from increasing fishing pressure.

1. Introduction

There is a need for the European Marine Strategy Framework Directive (MSFD) to develop indices able to detect differences and changes in benthic fauna condition in relation to anthropogenic pressures. A wide variety of benthic fauna multi-metric indices (MMIs) exist, mainly developed for the European Water Framework Directive (WFD), in particular the m-AMBI (Muxika et al., 2007), the BAT (Teixeira et al., 2009), the BEQI2 (Van Loon et al., 2015), the BQI (Rosenberg et al., 2004), the IQI (Borja et al., 2007), the DKI (Borja et al., 2007) and the SN index (Rygg, 2006). In these MMIs, the benthic indices species richness, Margalef diversity (Margalef, 1958), SN, Shannon index (Shannon and Weaver, 1963), Simpson measure of concentration (λ, Peet, 1974), Probability of Interspecies Encounter (1-λ, Hurlbert, 1971), Total Abundance (N) and AMBI (Borja et al., 2000) have been used one or more times. These indices all showed a useful sensitivity to anthropogenic pressures in a broad European study (Borja et al., 2011), and were therefore selected for testing in this study. In addition, the Infaunal Trophic Index (ITI) was selected because it is a relevant biological trait index that has yielded useful results (Maurer et al., 1999).

For the assessment of the condition of macrobenthic communities in a marine area, it is common practice to assign the benthic data to specific benthic habitat types. This approach is essential for improving the comparability of the benthic data, by reducing the benthic variation due to differences in habitat types and anthropogenic pressures (Van Hoey et al., 2013). A currently available and widely applicable habitat classification is the EUNIS system. EUNIS level 3, which discriminates Sand, Mud, Coarse and Mixed as sediment types, is often applied (Long, 2006). In addition, depth is an important environmental variable which can be used as a proxy for several natural pressures (including salinity, flow and wave pressure and light limitation) on the benthic habitats and the associated benthic communities (Leonardsson et al., 2016, Armonies et al., 2014). According to Kröncke et al. (2011), clearly discernible benthic communities can be found in the depth ranges < 50 m, 50–100 m, and > 100 m. In the Southern North Sea depths of < 50 m occur most frequently, and in the Northern North Sea deeper areas (> 100 m) are found. Coastal areas are usually defined by depths of < 20–25 m.

Benthic habitats and associated benthic communities are submitted to natural as well as anthropogenic pressures. The latter pressures act on top of the natural pressures, and induce a certain amount of degradation of the benthic communities, leading to decreases of benthic index values compared to reference values. In order to be able to clearly discriminate between natural and anthropogenic pressures, useful estimates of reference values, as well as quantitative pressure data and pressure-impact relationships, are needed.

The main environmental variables driving the natural variation are salinity, currents, wave and storm activity in tidal areas (Van Loon et al., 2015), primary production, light limitation and temperature. Some of these variables are influenced by climate driven processes such as the North Atlantic Oscillation (Kröncke et al., 2011). Higher natural pressures in the coastal zone result in lower reference index values than in the deeper offshore areas, where natural pressures are lower (Leonardsson et al., 2016; Armonies et al., 2014; Van Denderen et al., 2015). Because of higher natural stress levels in coastal areas the local benthic communities are probably more robust by nature (Duineveld et al., 1991). Common anthropogenic pressures in the North Sea are fishing (Collie et al., 2000; De Groot 1984; Kaiser et al., 2006; Tillin et al., 2006; Rumohr and Kujawski, 2000), organic enrichment and eutrophication (especially in the coastal zone, Borja et al., 2011), dumping of harbour sludge, sand extraction (in the coastal zone), and pollution e.g. Total Organic Hydrocarbons and heavy metals (in the coastal zone and near oil platforms; Olsgard and Gray, 1995; Hiscock et al., 2005). Fishing is a broadly occurring pressure in the Southern North Sea. In the coastal zone, organic enrichment is the second most widely occurring pressure. These often more or less chronic anthropogenic pressures have led to a certain degree of degradation of the benthic condition in the past. This degradation can be assessed by dividing the current benthic index assessment values with the estimated reference values, resulting in normalized index values (NIV; analogous to ecological quality ratios (EQR)). It is also known that restoration of coastal benthic communities after the cessation of fishing pressure proceeds slowly, and can take more than 8–10 years (Bergman et al., 2015; Coates et al., 2016; Lambert et al., 2014).

An important step in a benthic assessment method is the estimation of reference values under natural conditions. There are several methods to estimate these reference values: (OSPAR, 2012; Borja and Tunberg, 2011): (a) collection or use of a reference benthic dataset from a pristine or high quality area within the same ecoregion (such as the Southern North Sea). The high quality of this area must be demonstrated using a set of appropriate pressure data (fishing, organic matter, oxygen, suspended matter, heavy metals, etc.), (b) use of historical data from the least impacted areas or period, i.e. with the lowest level of anthropogenic disturbance (see OSPAR, 2012), (c) modelling and (d) expert judgement. Although method (a) is in principle preferable, it appears that in the Southern North Sea, which is broadly and chronically impacted by fishing activity (ICES 2016), pristine or high quality areas are in general not available any more (Kröncke et al., 2011; Kröncke and Bergfeld, 2003). Therefore, a combination of methods b, c and d was used in this study, similarly as in the WFD MMIs for marine benthos (Van Hoey et al., 2013; Van Loon et al., 2015). A commonly used WFD method is to estimate reference values as 95th or 99th index percentile values of a sufficiently large set of index values obtained for the least disturbed (baseline) period (Van Hoey et al., 2013; Van Loon et al., 2015). In addition, the variable percentile method of Hering et al. (2006) is a relevant method in which the height of the index percentile value used is adjusted to the known amount of anthropogenic pressure.

The aims of this study are to develop an optimized regional benthic fauna assessment method for the Southern North Sea which (a) is sensitive and precise (quantified as the slope and the R^2 value of the pressure-impact relationships, respectively) with regard to the anthropogenic pressures fishing and organic enrichment, (b) is suitable to estimate and model reference values, (c) is transparent, (d) can be efficiently applied using dedicated software; and to apply this method to benthic data from the Southern North Sea.

2. Methods

2.1. Benthic and pressure data

The Southern North Sea (SNS) is the central part of the Greater North Sea (Fig. 1). Benthic fauna data (species-abundance) were delivered by Belgium, Germany, the Netherlands and the United Kingdom for the Southern North Sea (SNS) area outside the 1 M Water Framework Directive (WFD) zone, and for three UK areas just above the Northern SNS borderline (Farnes East, North East of Farnes Deep, Swallow Sand; Fig. 2). Each national area, e.g. the German Dogger bank, is given an area code, e.g. DE_DoggerBank. In addition, each national area is divided in the relevant EUNIS 3 habitats (Long, 2006),

Fig. 1. OSPAR (Oslo Paris convention) marine regions in the North East Atlantic. Region I: Arctic waters, region II: Greater North Sea, region III: Celtic Sea, region IV: Bay of Biscay and Iberian Coast, region V: Wider Atlantic (source: www.ospar.org). The location of the Southern North Sea (SNS) in region II is indicated.

i.e. sand, mud, coarse, mixed. This results in an area-habitat code, e.g. DE_DoggerBank-Sand. Depth data (median values) were collected for each assessment area.

Samples in the SNS area was in most cases taken with a grab sampler (Germany and Belgium: Van Veen grab 0.1 m^2; United Kingdom: Van Veen, Hamon, mini Hamon or Day grab, 0.1 m^2) or with a boxcorer (Netherlands, 0.078 m^2). The samples were sieved over 1 mm in the field to remove the sediment and to retain the benthic macrofauna specimens. The sampling months were respectively: Belgium, September-October; Germany July-October; United Kingdom, March, April, May and November; Netherlands; March-April (before the formation of juvenile specimens). Macrofauna was identified to the species level whenever possible, but in some cases specimens could only be assigned to a higher taxonomic level (ISO, 2014). For each species or lowest taxon level the number of individuals was registered. The above dataset contains no data from areas with anthropogenic stressors such as dredge disposal, sand extraction or wind-farms. Bottom fishing is the only main anthropogenic disturbance operational in the sampled areas.

ICES bottom fishing activity data (Vessel Monitoring System-VMS data in GIS format) were obtained from the ICES website (ICES, 2016). These data were available for the period 2009–2013 (vessels > 12 m) and consist of the annual average fishing activity per ICES grid cell of $0.05 \times 0.05^\circ$ with respect to surface fishing with a bottom penetration < 2 cm (swept area ratio, unit number of sweeps/year) and subsurface fishing with a bottom penetration > 2 cm (fraction (0–1) of the surface activity, unit sweeps/year) (Eigaard et al., 2015). Note that the subsurface fishing activity is defined by ICES as a fraction (0–1) of the surface fishing activity. The swept area ratios were calculated by ICES using the VMS data points which fishing vessels produce once every two hours. Annual average bottom surface and subsurface fishing activities were calculated for each assessment area for the period 2009–2013. The relationships between respectively bottom surface and subsurface fishing activity, and between median depth and subsurface fishing activity, were investigated. The sampling dates were checked whether they fell within the first or second half of a year. Since this distribution was practically equal, fishing data from the same year as the year of benthic sampling were used.

To compensate for the lack of benthic data from areas with other stressors than fishing, the following three sets of species-abundances data were added to this study: a) data from the Swedish Saltkälle fjord which discharged organic matter via the Gullmar fjord into the Northern North Sea for the period 1964–1978, b) data from in- and near a German sediment dumping site near Helgoland in the SNS for the period 2005–2015, and c) data and corresponding sediment concentrations of Titanium dioxide from the Norwegian Jossing fjord which enters the Northern North Sea, for the period 1983–1995.

2.2. Index/MMI calculation

The benthic indices tested are listed in Table 1. Normalized index values (NIV) were calculated as:

$$NIV = [Index(ass.val.) - Index(bad\ val.)]/[Index(ref.val.) - Index(bad\ val.)] \quad (1)$$

in which ass.val. is the assessment value, ref. val. is the reference value, bad val. is the bad value. For all benthic indices the bad (lowest possible) value is 0, with the exception of AMBI, for which a bad value of 6 is used. Therefore, only index reference values have to be estimated.

2.3. Reference value estimation

A sufficiently large set of index values, with sufficient spatial and temporal coverage, is necessary to obtain the highest stable reference values. Therefore, for each assessment area the following data demands and corresponding quality code were formulated for estimating reference values: 6 data years and a minimum of 10 samples per year were assigned quality code 3; 3–5 data years and a minimum of 10 samples per year quality code 2; and less than 3 data years and a minimum of 10 samples per year a quality code 1. Note that only estimated reference values with a quality code 3 or 2 were used to analyze relationships of reference values with median depth. For UK assessment areas with reference values having a quality code 1, model values of the depth-reference Margalef model were used, if the depth fell within the calibrated part of the model (−10 to −50 m).

Initial estimates of reference values for the SNS assessment areas were made on basis of the 99th percentile value of the least disturbed data of a period of preferably 6 years (Van Hoey et al., 2013; Borja et al., 2011, Van Loon et al., 2015Borja et al., 2011, Van Loon et al., 2015). The least disturbed data were defined as the ones with on average the highest Margalef index values. The percentile value was subsequently modified according to Hering et al. (2006), by making the percentile value dependent on the average subsurface fishing activity (FA_{ss}) in the period 2009–2013, for which ICES data are available (ICES, 2016). This was done by assigning a higher percentile value to an area with higher fishing pressure. The choice of combinations of FA_{ss} and percentile values was based on expert judgement, and was as follows: $FA_{ss} < 0.1$ subsurface sweeps/year, 75 percentile; FA_{ss} 0.1–0.5, 95 percentile; $FA_{ss} > 0.5$, 99 percentile.

Next, the index reference values estimated using the variable percentile method (with quality code 3 or 2) were related to the median

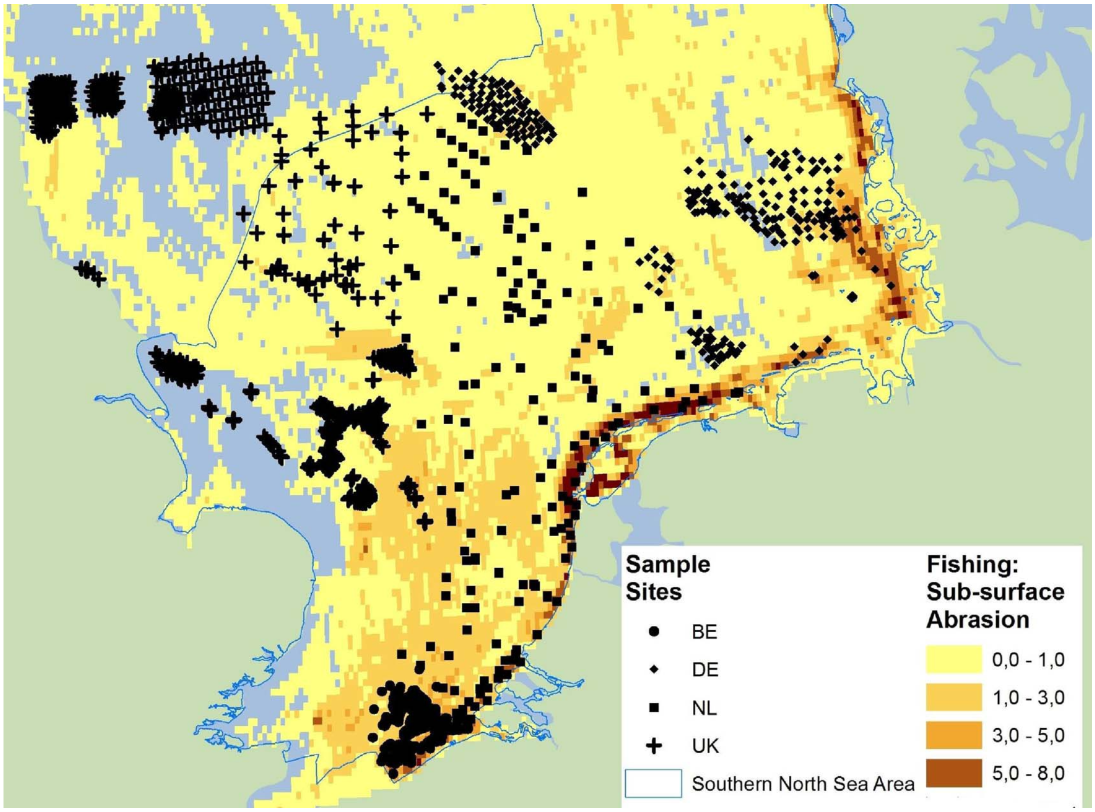

Fig. 2. The Southern North Sea (SNS) and the spatial distribution of the benthic samples used in this study (source: www.ospar.org). For illustration, the average subsurface fishing activity (bottom penetration > 2 cm, in sweeps/year) in 2013 is shown.

depth of the assessment areas. For Margalef diversity and species richness, sigmoid curve fitting was used. A generalized logistic function was programmed in R to model a sigmoid curve through the data points (Richards, 1959; https://en.wikipedia.org/wiki/Generalised_logistic_function, Walvoort and van Loon, 2017b). The model formula used is:

$$Y = A + (K - A)/(1 + \exp(-B * (x - M))) \quad (2)$$

In this model Y is the index reference value; A the lower asymptote; K the upper asymptote; x is the median depth; M is the median depth where the index reference value (Y) reaches half of its range; and B is related to the slope of the function. The precision of the model (explained variance) was calculated as 100*(1-(residual sum of squares/total sum of squares)). The average absolute deviation of Margalef percentile reference values from the depth-reference value model was calculated for the quality codes 3, 2 and 1, respectively. For the other benthic indices, linear regression was used to relate the index reference values and median depth. The correlation coefficients of the respective curve fits were used as one of the index evaluation criteria.

2.4. Pressure testing of indices/MMIs

The selected indices (Table 1) and multi-metric indices (S + AMBI, S + ITI, D + AMBI, D + ITI) were evaluated for their anthropogenic pressure sensitivity using the following criteria and quantification

Table 1
Benthic indices tested in this study. These indices were also used to construct and test multi-metric indices.

Benthic index	Formula	Reference and comments
Species richness		Taxa must preferably be identified at the species level.
Margalef diversity	D = (S − 1)/lnN	Margalef, 1958. S = species richness and N = total abundance. Note that for samples with N = 0, Margalef diversity is set to 0.
SNA	SNA = ln(S)/ln(ln(N + 1) + 1)	S = species richness and N = total abundance. SNA is a new improved version of the SN index (Rygg, 2006), which avoids a numerical problem at S and N = 2 or 3 (Walvoort and van Loon, 2017c).
Shannon index	See reference	Shannon and Weaver, 1963
Probability of Interspecies Encounter	PIE = 1 − λ	https://en.wikipedia.org/wiki/Diversity_index λ is the Gini-Simpson index.
AMBI	See reference	Borja et al., 2000. The Borja species sensitivity list of November 2014 was used. For a list of 100 species in the Dutch part of the North Sea, optimized AMBI species sensitivity values were used (Gittenberger and van Loon, 2013).
Infaunal Trophic Index (ITI)	$ITI = 100 - \frac{100}{3}\left(\frac{N_2 + 2N_3 + 3N_4}{N_1 + N_2 + N_3 + N_4}\right)$	Maurer et al., 1999. A biological trait index (Word, 1979), which classifies species to into one of four feeding classes: filter feeders (class 1), interface feeders (facultative filter and deposit feeders, class 2), deposit feeders (class 3) and subsurface deposit feeders (class 4). The index value ranges from 100% (only filter feeders) to 0% (only subsurface deposit feeders). N_n is the number of individuals within the respective feeding group. Species feeding habits were primarily collected from WoRMS http://www.marinespecies.org/), and supplemented by M. Lavaleye.

methods.

For bottom fishing pressure, the sensitivity (slope) and precision (pseudo-R^2) of the fishing activity-benthic impact relationships were determined as follows. The fishing surface and subsurface activity data and normalized sample index values were aggregated and tested at the assessment area-year level. Normalized index values (Eq. (1)) were used, because it was assumed that a specific fishing activity value will have a larger absolute effect on the benthic biodiversity in deeper and more sensitive benthic communities compared to more robust and less biodiverse coastal benthic communities, but that comparable relative (normalized) effects on benthic biodiversity occur in all areas. Selected benthic data and ICES fishing pressure data (period 2009–2013) were analysed using exponential curve fitting, which gave the best results (Formula 4). The pseudo-R^2 value, and the slope of the exponential curve at a subsurface fishing activity of 1 sweep/year, were calculated as measures for the precision and sensitivity of the index tested, respectively.

For organic enrichment, the precision of the relationship between recovery time after cessation of organic matter discharge and benthic condition was tested for all indices using benthic data from the Swedish Saltkälle fjord for the period 1966–1978. For more experimental details see Rosenberg (1973). The sampling time was used as a proxy for the decreasing organic matter pressure after the cessation of the pulp mill discharges in 1966. Data from the station L12, which has an intermediate spatial position in the fjord and did not become completely azoic in the past, were analysed (Rosenberg, 1973). For each benthic index, the most appropriate linear or sigmoid curve fit model (Eq. (2)) was used, and the R^2 or pseudo-R^2 value of the curve fit was used as a measure of the precision of this relationship.

For the median depth-reference value model (see paragraph 2.3), the precision of this model was determined for all indices using the pseudo-R^2 value for a sigmoid model or the R^2 value for a linear model, depending on which model is the most suitable.

For the transparency of an index an expert judgement rank score (11-1) was used.

For sediment dumping and sedimentation, the sensitivity and precision of Margalef for this pressure was tested using benthic fauna data from a German dumping site near Helgoland, and the results are presented graphically in a bar chart with 95% confidence intervals. The samples were collected in this area for the period 2005–2015, and the distance of each sample from the dumping site was registered.

For the contaminant Titanium dioxide, data from the Norwegian Jossing fjord, from the period 1983–1995, were investigated. For more experimental details see Olsgard and Hasle (1993). Linear curve fits between the titanium dioxide concentration (weight%) and the respective benthic index scores were made, and the precision (R^2) of these relationships was determined.

The single benthic indices were scored for each criterion using a normalized index score (0–1), or if this was not possible using a ranking score (11–1 points; then normalized to 0–1), and the total score per index/MMI was calculated (Table 3). For MMIs composed of two indices, the combined performance per criterion was estimated as follows: for sensitivity the average index sensitivity was calculated; for precision (R^2 or pseudo-R^2 values) the square root of the quadratic sum of the index R^2 values was calculated; for transparency the lowest single index score was used which is considered to be the limiting factor. By exception, for the calculation of the combined precision (R^2) for organic enrichment, for AMBI an increase of +0.02 and for ITI an increase of +0.04 was used, because the sum of the R^2 values exceeded 1 for some of the MMIs including AMBI and ITI.

The criteria scores were given equal weight, with the exception of the precision for organic enrichment. The latter criterion was given a weight factor of 0.5, in view of the limited dataset and the more localized occurrence compared to the broadly present fishing pressure.

MMI's composed of three indices were preliminary tested, but are not reported as they did not improve the MMI performance.

2.5. BENMMI software

Benthos data analysis software, called BENMMI, was developed for OSPAR/MSFD application to test benthic indices and MMIs using species-abundance and pressure data, and to assess areas-habitats (Walvoort and van Loon, 2017a; Walvoort and van Loon, 2017b). This software essentially performs the following steps: (a) validation and cleanup of benthic input data, (b) standardization of species names using WoRMS (World Register of Marine Species, http://www.marinespecies.org/), (c) index calculation, (d) MMI optimization (using up to 3 benthic indices and pressure data), (e) calculation of index and MMI values for samples and assessment areas, (f) calculation of confidence intervals and sampling power functions, (g) production of index correlation plots, multi-dimensional scaling (MDS) plot and (S-1)/ln(N) (Margalef) plots for quality control, and (g) production of a detailed analysis report and result files. Step (a) involves the checking of the input fields of the BENMMI data format, and the removal of mobile epifaunal species (Decapoda, Mysida) which are sampled unreliably due to their mobility.

Step (d) MMI optimization is performed by BENMMI using benthic and pressure data and Multiple Linear Regression using the following model:

$$NIV\ (MMI) = b_0 + w1 * NIV(index1) + w2 * NIV(index2) + w3 * NIV(index3) \quad (3)$$

in which NIV is the Normalized Index Value (range 0–1), b_0 is the intercept with the Y-axis, w1, w2 and w3 are weight factors which range from 0 to 1, and the sum of w1 + w2 + w3 = 1. For a single metric index, ordinary least squares is used for optimization. For a bimetric index, Brent's method (combination of golden section search and successive parabolic interpolation) is used. For a trimetric index the Downhill simplex method by Nelder & Mead is used (www.nr.com). For each index or optimized MMI, the intercept (b_0), slope (sensitivity), R^2_adjusted (precision) and significance (p) are reported. The Akaike Information Criterion (AIC) was not used because all models had the same complexity in terms of number of parameters (two) and pressures (one).

In addition to multi-linear regression, BENMMI can perform single index exponential curve fitting (user setting) using the following formula:

$$Y = b_0 + b_1 * \exp(-b_2 * X) \quad (4)$$

in which Y is the normalized index value, b_0 is the asymptotic normalized index value at higher pressures, $b_0 + b_1$ is the intercept with the Y-axis, b_2 is the sensitivity of the relationship, and X is the pressure. The sensitivity per index was estimated as the slope of the curve fit at a subsurface fishing activity of 1 sweep/year, using the formula: slope = $b_1 * b_2 * \exp(b_2 * X)$, in which X = 1. Pseudo-R^2 values were calculated to characterize the approximate amount of variation explained by the model as follows:

$$\text{Pseudo-}R^2 = 1 - SS_{res}/SS_{tot} \quad (5)$$

In which $SS_{res} = \Sigma_i (y_i - \hat{y}_i)^2$ is the residual sum of squares, $SS_{tot} = \Sigma_i (y_i - \bar{y})^2$ is the total sum of squares, y_i is the observed (normalized) index value, $\hat{y}_i$ is the predicted (modelled and normalized) index value, $\bar{y}$ is the mean (normalized) index value.

2.6. Assessment method

The optimized benthic index was applied to all assessment areas for the period 2010–2015, because this period is synchronized with the assessment period of the OSPAR BH3 habitat damage indicator

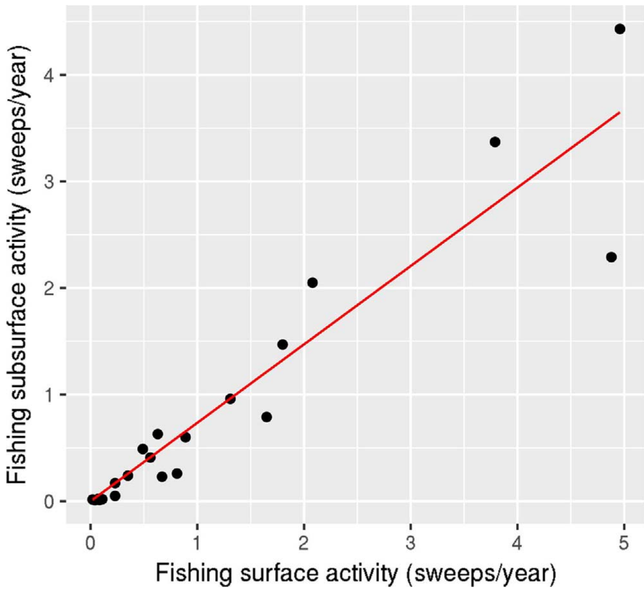

Fig. 3. Relationship between subsurface fishing activity (bottom penetration < 2 cm) and surface fishing activity (bottom penetration > 2 cm) on the x and y-axis, respectively. This figure shows that surface and subsurface fishing activity are strongly linearly related (slope 0.74, $R^2 = 0.93$, $p < 0.00001$).

(OSPAR, 2017). For each assessment area, first the annual average normalized index values, and their 90% confidence intervals, were calculated. Then the period average for 2010–2015 was calculated as the average of the annual averages. These are the final results reported for the OSPAR assessment of the Southern North Sea.

The following quality codes were assigned to the assessment values: 3+ data years and > 30 samples, quality code 3; 2 data years and > 20 samples, quality code 2; 1 data year and > 10 samples, quality code 1.

3. Results

3.1. Benthic and fishing pressure data

The spatial distribution of all analysed samples in this study, available for the period 1994–2015, is shown in Fig. 2. It appears that the spatial coverage of the sampling locations for Belgium, Germany and the Netherlands is good. The spatial coverage for the UK is fairly good, only the Southern part of the UK area is not covered. Danish benthic data were not available for the SNS at the time of this study. The median depths of the assessment area ranges from 11 m in the shallow coastal Belgium North Sea to 77 m in the deep offshore UK Farnes East sand area in the Northern North Sea (Table 1).

A detailed characterization of the benthic samples, and their habitats and fishing pressures, is given in Supplement 1. The total number of samples available for this project was 4833. Some average sample index values are: total abundance, 156; species richness, 23; Margalef diversity, 4.7.

There appears to be a strong linear relationship between the bottom surface and subsurface fishing activity (Fig. 3, $R^2 = 0.93$, p-value < 0.00001). The subsurface fishing activity is on average 74% of the surface fishing activity. There is a decline in bottom fishing activity with increasing depth, with very low levels below 50 m depth in a few UK areas north of the SNS border (Figs. 2 and 4).

3.2. Reference value estimation and modelling

When using data with quality code 3 and 2, the relationship between Margalef diversity reference value and median depth was modelled by means of a sigmoid curve (Fig. 5). This model shows three

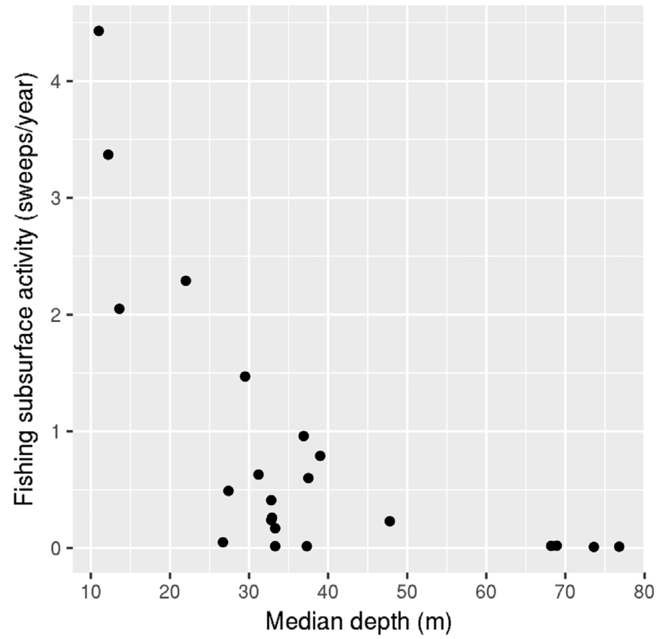

Fig. 4. Relationship between median depth and subsurface fishing activity with bottom penetration > 2 cm. This figure shows that at increasing median depth the fishing activity strongly decreases. The relation between depth and surface fishing activity (i.e. bottom penetration < 2 cm) is similar (not shown).

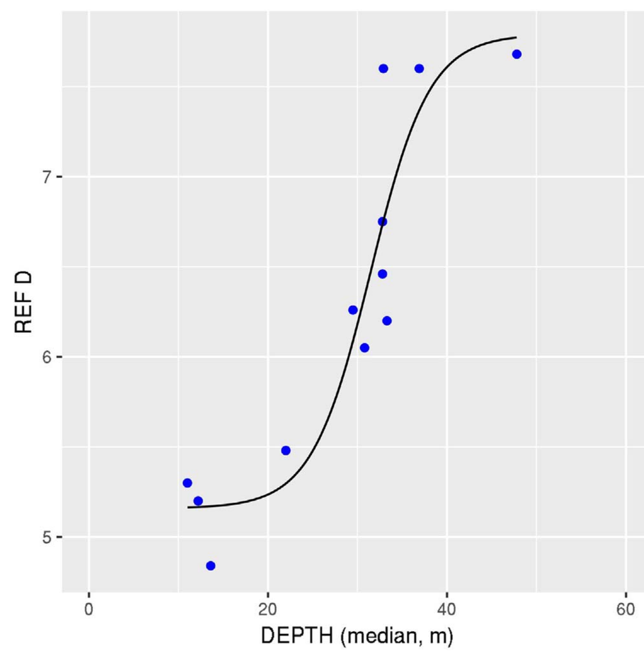

Fig. 5. Relationship between median depth and Margalef reference values for the assessment areas with data quality code 3 and 2. The sigmoid model parameters are: pseudo-$R^2 = 0.86$; A = 5.159; K = 7.79; B = 0.3054; M = 31.49.

linear segments: (a) a stable linear part at lower median depth (< 20m), (b) a rapid increase in the median depth range 20–40 m, and (c) a stable linear part at a median depth > 40 m. The depth-reference value model for Margalef has a pseudo-R^2 of 86%, which is regarded as a good model precision. These result shows that the variable percentile method is effective in producing sufficiently reliable and useful estimates of index reference values for the EUNIS 3 habitats assessed in this study within the calibrated range of 10–50 m depth (Table 2). Note that the start of the curve jump at approximately 20 m median depth can be considered as the borderline between the coastal and offshore area.

Table 2
Estimated index reference values per assessment area using the least disturbed or available benthic data. Supporting information presented per assessment area (national area combined with EUNIS 3 habitat) is: the median depth, the average subsurface fishing activity, the data years used for the reference value estimation, the number of samples used, the quality code of the least disturbed benthic dataset, the percentile value used for reference value estimation, the Margalef percentile reference value, the Margalef model reference value, the absolute difference between the Margalef percentile and model reference values, and the percentile based reference values of the other indices (species richness, SNA, Shannon index, PIE, AMBI and ITI).

National area	Habitat	Median depth (m)	Average subsurface fishing activity (sweeps/year)	Data years reference estimation	Nr data years	Nr benthos samples	Quality code reference data	Percentile value for reference estimation	D reference value (percentile method)
BE_NorthSea	Coarse	22	2.29	2007–2012	6	610	3	99	5.48
BE_NorthSea	Sand	11	4.43	2007–2012	6	136	3	99	5.3
DE_BorkumReefGround	Sand	26.7	0.05	2012	1	38	1	75	3.77
DE_Coastal	sand	13.6	2.05	2009–2014	6	155	3	99	4.84
DE_Doggerbank	sand	37.5	0.6	2000	1	106	1	99	7.82
DE_ElbeUrstromValley	Mud	33.3	0.016	2005–2010	6	22	2	75	6.2
DE_ElbeUrstromValley	sand	33.3	0.17	2012	1	118	1	95	7.21
DE_SyltOuterReef	Coarse	32.8	0.41	2011–2014	4	20	2	95	6.46
DE_SyltOuterReef	sand	32.8	0.24	2011–2014	4	104	2	95	6.75
NL_CoastalZone	Sand	12.2	3.37	1994–1999	6	299	3	99	5.2
NL_DoggerBank	Sand	32.9	0.26	2005–2010	6	103	3	95	7.6
NL_FrysianFront	Sand	36.9	0.96	1999–2004	6	364	3	99	7.6
NL_Offshore	Sand	29.5	1.47	1999–2004	6	180	3	99	6.26
NL_OysterBanks	Mud	47.8	0.23	1997–2002	6	448	3	95	7.68
UK_DoggerBank	Sand	30.8	na	98,03,05,08	4	171	2	95	5.38
UK_FarnesEast	Mixed	68.2	0.019	2012	1	41	1	75	11.1
UK_FarnesEast	Sand	76.8	0.012	2012	1	30	1	75	8.03
UK_HoldernessOffshore	Mixed	37.3	0.016	2012	1	26	1	75	6.66
UK_MarkhamsTriangle	Coarse	39	0.79	2012	1	26	1	99	6.83
UK_NEFarnesDeep	Mixed	68.9	0.021	2012	1	21	1	99	10.5
UK_NNorfolkSandbanks	Coarse	27.4	0.49	2012	1	54	1	95	6.37
UK_NNorfolkSandbanks	Sand	31.2	0.63	2012	1	153	1	99	5.12
UK_SwallowSand	Sand	73.6	0.01	2012	1	67	1	75	7.31

National area	D reference value (model)	Delta percentile – model (absolute)	D reference value (final)	S reference value	SNA reference value	H' reference value	PIE reference value	AMBI reference value	ITI reference value
BE_NorthSea	5.30	0.18	5.48	26.8	1.91	3.93	0.974	0.392	57
BE_NorthSea	5.16	0.14	5.30	37.7	1.81	3.92	0.936	0.333	90.4
DE_BorkumReefGround	5.65	1.88	5.65	18					
DE_Coastal	5.17	0.33	4.84	29.4	1.8	3.54	1	0.363	78.7
DE_Doggerbank	7.43	0.39	7.43	43.4					
DE_ElbeUrstromValley	6.83	0.63	6.20	30.9	2.01	4.32	0.955	0.629	81.8
DE_ElbeUrstromValley	6.83	0.38	6.83	43					
DE_SyltOuterReef	6.73	0.27	6.46	35.3	1.98	4.25	0.937	0.127	91.2
DE_SyltOuterReef	6.73	0.02	6.75	39.8	1.98	4.24	0.936	0.494	76.2
NL_CoastalZone	5.17	0.03	5.20	29.8	1.8	3.58	0.933	0.33	83.9
NL_DoggerBank	6.75	0.85	7.60	42	2.05	4.59	0.948	0.539	85.9
NL_FrysianFront	7.37	0.23	7.60	40	2.04	4.44	0.952	0.677	86.8
NL_Offshore	6.09	0.17	6.26	31.8	1.95	4.2	0.967	0.668	79.7
NL_OysterBanks	7.77	0.09	7.68	37	2.14	4.69	0.967	1.09	85.7
UK_DoggerBank	6.34	0.96	5.38	27	2.02	4.22	0.96	0.88	76.9
UK_FarnesEast	na		11.10	61					
UK_FarnesEast	na		8.03	39.75					
UK_HoldernessOffshore	7.41	0.75	7.41	30.75					
UK_MarkhamsTriangle	7.55	0.72	7.55	32.5					
UK_NEFarnesDeep	na		10.50	50.6					
UK_NNorfolkSandbanks	5.75	0.62	6.37	34.35					
UK_NNorfolkSandbanks	6.42	1.30	5.12	21.84					
UK_SwallowSand	na		7.31	32.5					

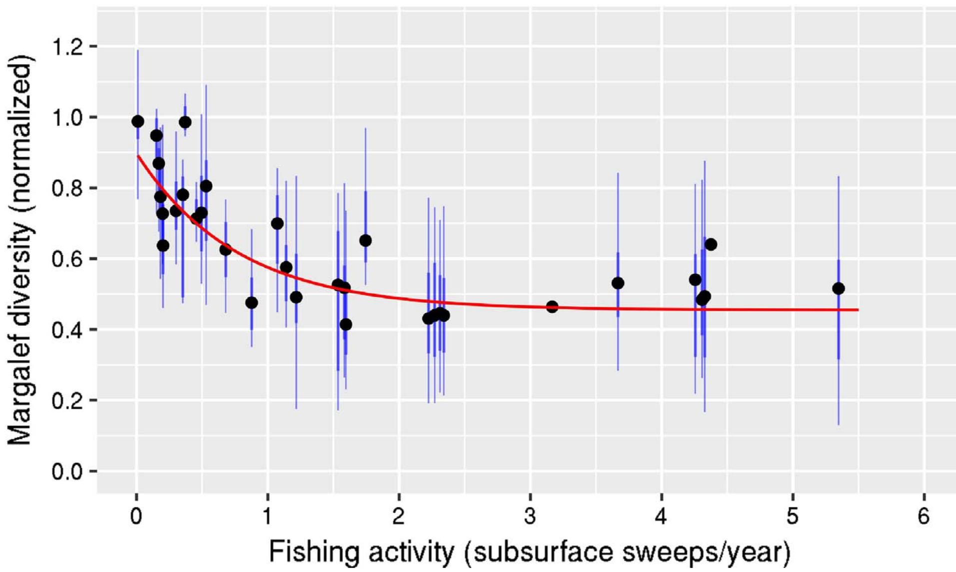

Fig. 6. Relationship between normalized Margalef diversity and bottom subsurface fishing activity. Both variables have been aggregated at the assessment area-year level. Median Margalef values per assessment area-year are shown, and their 50 and 90% confidence intervals (thicker and thin lines), respectively. The figures shows an exponentially decreasing relationship, an asymptotic value at a normalized Margalef value of 0.45, and an intercept of 0.90 at zero fishing pressure. The curve fit parameters are: pseudo-R^2 = 0.26; $p < 0.00001$; b0 = 0.45; b1 = 0.44; b2 = −1.3.

For each of the 3 data quality classes, the average absolute deviation was calculated between the percentile reference values for Margalef diversity and those predicted by the corresponding depth-Margalef model. This yielded an average deviation of 0.25 for quality code 3 (6 data years), of 0.47 for quality code 2 (3–5 data years), and of 0.86 for quality code 1 (1–2 data years) (Table 2). The use of data from a continuous period of 6 years appeared to result in the most precise reference values, as the percentile reference values only deviate on average 4% from the value predicted by the model. This deviation is on average 7% for quality code 2, and 14% for quality code 1. In view of these results, reference values with quality code 1 are considered as unreliable in this study. Therefore, only data sets with reference values with quality code 3 and 2 (for the assessment areas Belgium, North Sea Coarse and Sand; Germany, Coast, Sylt Outer Reef Coarse and Sand; Netherlands, Coast, Offshore, Frisian Front, Oyster Banks and Dogger Bank) were used for the fishing activity-index response testing (Table 2).

The relationship between median depth and reference values was also investigated for the other indices. The sigmoid relationship between median depth and the reference values for species richness (pseudo-R^2 = 0.48, figure not shown) is clearly less strong compared to Margalef reference values (pseudo-R^2 = 0.86, Fig. 5,). A strong linear relation was found between the median depth and the SNA reference value (R^2 = 0.95; figure not shown). A linear relation with good precision (R^2 = 0.85) was also found for the Shannon index. For AMBI a linear relation with reasonable precision (R^2 = 0.45) was found, showing an increase of the AMBI reference value at greater depth. Linear curve fitting of relationships between reference values of ITI and PIE and depth yielded low R^2 values of 0.02 and 0.001, respectively.

The final reference values for Margalef diversity, estimated using the variable percentile method or the depth model, are presented in Table 2. The final reference values for the other indices, estimated using the variable percentile method, are also presented in Table 2.

3.3. Pressure testing of indices/MMIs

Comparison between bottom surface and subsurface fishing activity and normalized Margalef values showed that the precision of the curve for subsurface fishing (pseudo-R^2 = 0.26, Fig. 6) is comparable with the surface curve (pseudo-R^2 = 0.27; Supplement 3; curve fit parameters: b0 = 0.46; b1 = 0.53; b2 = −0.99). Subsurface fishing activity (penetration depth > 2 cm) is known to have a higher impact and damaging effect on infauna than surface fishing activity (bottom penetration < 2 cm), which primarily impacts epifauna (Eigaard et al., 2015). Therefore the subsurface fishing activity data were primarily used for the index optimization reported hereafter.

The fitting of an exponential curve between subsurface fishing activity using individual ICES grid cell values, and normalized Margalef diversity of individual samples in the period 2009–2013 of the 10 selected assessment areas, yielded a pseudo-R^2 = 0.23 and $p < 0.000001$ (figure not shown). However, it appeared that by averaging of the ICES grid cell values, and by averaging the sample normalized Margalef values at the assessment area and year level, a similar curve fit with an improved precision was obtained (pseudo-R^2 = 0.26) (Fig. 6). In addition, Fig. 6 gives a clearer picture of the exponential decrease of the quality of the macrobenthic community. Furthermore, it was found that several benthic samples lie at the border of ICES grid cells, and averaging of ICES cells within an assessment area avoids these border errors. Therefore, ICES grid cell values and normalized sample Margalef values were thereafter averaged at the assessment area-year level. Fig. 6 also shows that above a subsurface fishing activity level of 2.3 sweeps/year (95% of the model asymptotic value) resilient benthic communities have been formed with a stabilized normalized Margalef value of around 0.45. These benthic communities are not further degraded at increasing fishing pressure in the range of 2.3–5.5 sweeps/year.

Sigmoid relationships between the stressors organic enrichment plus associated oxygen depletion in the Swedish Saltkalle fjord, and Margalef diversity (Fig. 7, pseudo-R^2 = 0.87), species richness (pseudo-R^2 = 0.90), and ITI (pseudo-R^2 = 0.90) all have a good precision. The precision (pseudo-R^2) of similar curves for the other indices is: SNA 0.75, H 0.75, AMBI 0.51, PIE 0.28. Note that it took about eleven years for the benthic community to restore to a presumably normal benthic condition (Fig. 7).

The relation between distance from a German sediment dumping site and the Margalef diversity value of samples taken around this dumping site is shown in Fig. 8. The results show that after 3–12 months of dumping, the benthic condition within the dumping site (DS) is still significantly lower than within the 1 km zone (A 1 km), as is shown by the 95% confidence intervals in Fig. 8. In addition, both the benthic condition of both the dumping site (DS) and the 1 km zone (A 1 km) are significantly lower than of the reference area (R), as is shown by the 95% confidence intervals.

The fitting of linear relationships between sediment concentrations of Titanium dioxide and values of the different indices yielded the following correlation coefficients: R^2 = 0.82 for Margalef diversity (Fig. 9), 0.82 for SNA, 0.76 for species richness, 0.50 for H, 0.35 for AMBI, 0.21 for PIE, and 0.03 for ITI. Margalef diversity, SNA and

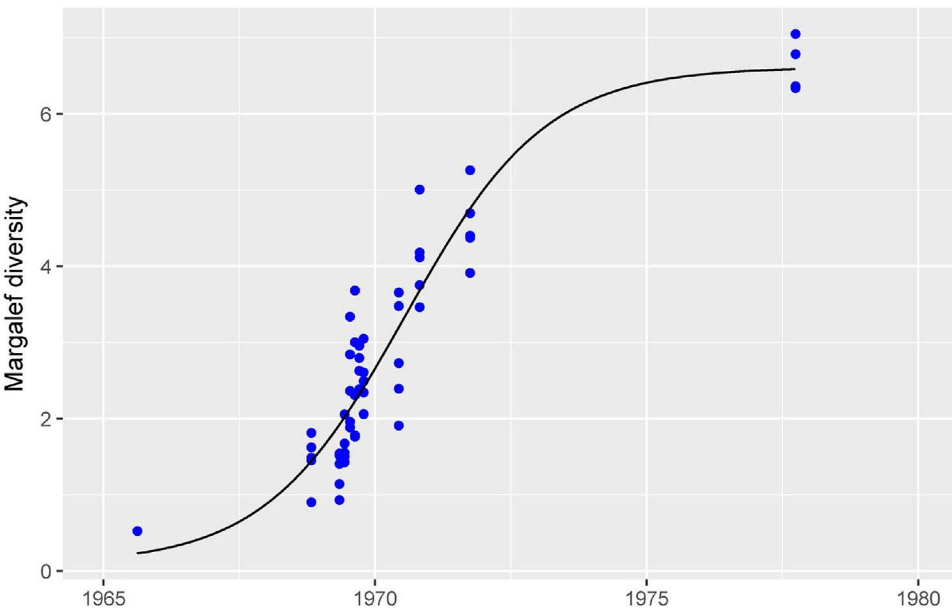

Fig. 7. Restoration of benthic fauna condition as indicated by Margalef diversity, after cessation of organic enrichment due to pulp mill effluent discharge in the Swedish Saltkälle fjord in 1966. The curve fit parameters are: pseudo-R^2 = 0.87; A = 0.0908; K = 6.61; B = 0.00212; M = 202.

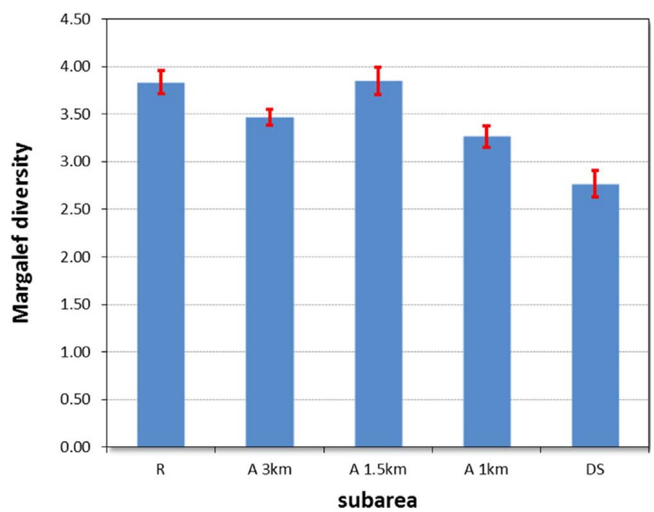

Fig. 8. Relationship of Margalef diversity with dumping and sedimentation pressure (3–12 months after dumping) in and nearby a German sediment dumping site near Helgoland in the Southern North Sea. R = Reference site (approx. 10 km from dumping site); DS = Dumping Site, A 3 km = sampling area approx. 3 km from dumping site; A 1.5 km = sampling area approx. 1.5 km from dumping site; A 1 km = sampling area approx. 1 km from dumping site. The error bars show the 95% confidence intervals.

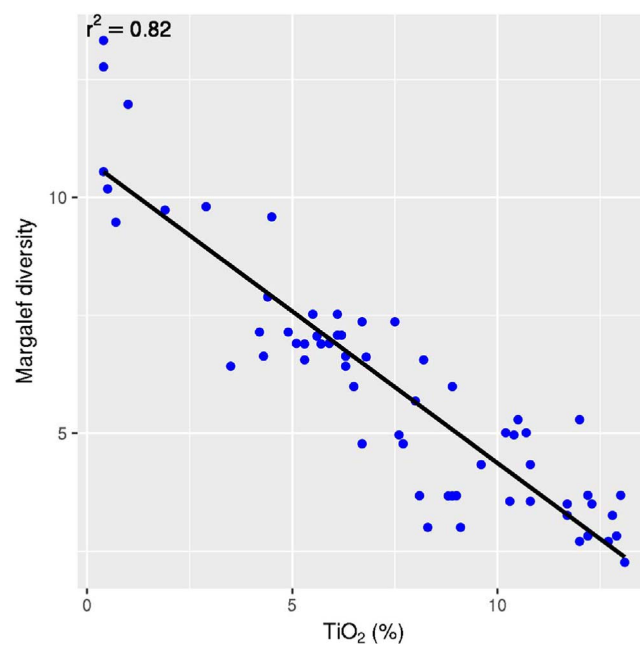

Fig. 9. Relationship of Margalef diversity with Titanium dioxide sediment concentrations (weight%; R^2 = 0.82) in the Norwegian Jossing fjord.

species richness show the most precise response to this type of contamination. Note that the ITI is remarkably insensitive to Titanium dioxide contamination.

The quantitative results for the evaluation of the 7 indices and 4 MMIs are summarized in Table 3. These results show that the total index score of Margalef diversity is the highest (3.99). Clearly, Margalef is the best performing index. Species richness scores second best, having a 5% lower score (3.79) than Margalef, which is a significant difference. The SNA index also shows a good single index score (3.02), in particular regarding the precision (R^2) of the depth-reference value model. The PIE shows the lowest single index score (1.14). Among the multi-metric indices Margalef + ITI (score 3.58) and Margalef + AMBI (score 3.40) score the highest, but not as high as Margalef alone. Margalef diversity performs particularly well in terms of the precision of the reference value model (Table 3). This is a valuable advantage compared to the relatively poor performance of the reference value model for species richness, and was co-decisive in the index evaluation in favour of the Margalef index (Table 3). This higher precision of the Margalef reference value model is probably caused by the correction of Margalef for (limited) differences in sampling device types, and associated sample sizes and volumes, which is an essential advantage for this common method in the SNS region. Therefore, Margalef diversity was chosen as the best performing benthic index for application in the Southern North Sea. By selecting only one index, the benthic assessment method becomes very straightforward, which is recommended by the "Ockham's razor principle": choose the simplest model which gives similar results.

3.4. Assessment results

The final assessment results of the Southern North Sea (SNS) areas, with quality code 3, 2 and 1, are shown in Fig. 10 and Supplement 2.

Table 3
Index and multi-metric index evaluation results. The evaluation criteria used are: the sensitivity and precision to indicate bottom fishing activity, the precision to indicate organic enrichment by pulp mill effluents, the precision of the depth-reference value relationship, and transparency. All criteria have a weight of 1, except organic enrichment which has a weight of 0.5. The two best performing indices and the two best performing MMIs are printed bold in the last column.

Index/MMI	Weights	Fishing sensitivity (slope of exp.fit at 1 sweep/ year)	Fishing sensitivity score	Fishing precision (pseudo-R2 of exp.fit)	Fishing precision score	Organic enrichment precision (pseudo-R2 or R2)	Organic enrichment score	Depth-reference model precision (pseudo-R2 or R2)	Depth-reference model score	Transparancy rank	Transparancy score	Total score
S	1	0.194	1.00	0.23	0.81	0.90	0.96	0.48	0.49	11.0	1.00	**3.79**
D	1	0.157	0.81	0.26	0.92	0.87	0.93	0.86	0.89	10.0	0.91	**3.99**
SNA	1	0.071	0.37	0.13	0.46	0.75	0.80	0.95	0.98	9.0	0.82	3.02
H’	1	0.077	0.40	0.10	0.35	0.61	0.65	0.85	0.88	8.0	0.73	2.68
PIE	1	0.034	0.18	0.05	0.18	0.28	0.30	0.00	0.00	7.0	0.64	1.14
AMBI	1	0.039	0.20	0.08	0.28	0.51	0.54	0.45	0.46	5.0	0.45	1.67
ITI	1	0.102	0.53	0.11	0.39	0.90	0.96	0.02	0.02	6.0	0.55	1.96
S + AMBI	0.38 + 0.62	0.117	0.60	0.24	0.86	0.92	0.98	0.66	0.68	5.0	0.45	3.08
S + ITI	0.83 + 0.17	0.148	0.76	0.25	0.90	0.94	1.00	0.48	0.49	6.0	0.55	3.21
D + AMBI	0.39 + 0.61	0.098	0.51	0.27	0.96	0.89	0.95	0.97	1.00	5.0	0.45	**3.40**
D + ITI	0.83 + 0.17	0.130	0.67	0.28	1.00	0.91	0.97	0.86	0.89	6.0	0.55	**3.58**

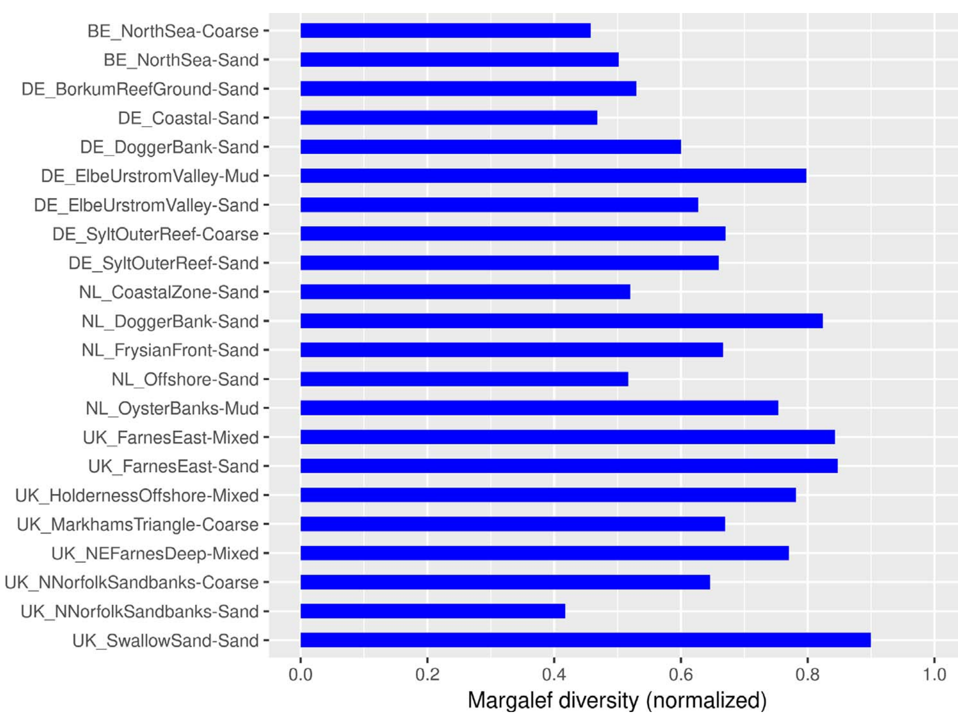

Fig. 10. Assessment results for selected assessment areas in the Southern North Sea using normalized Margalef diversity values. The period averages of normalized Margalef values for 2010–2015 are shown.

The assessments show that the benthic quality in coastal SNS areas is the lowest. In several offshore areas, a relatively higher benthic quality is observed. In several deeper offshore assessment areas relatively high benthic quality can be observed.

The average 90% confidence interval of normalized Margalef diversity values for all areas-habitats-years is +/-0.06 (Supplement 2).

4. Discussion

4.1. Benthic and fishing pressure data

All participating countries in this study essentially use comparable sampling, sieving and analysis methods of macrobenthic fauna: sampling using a grab of 0.1 m^2 or a box core of 0.078 m^2, sieving of samples over 1 mm, and taxonomic analysis of infauna and sessile epifauna at preferably the species level.

The sampling volume and penetration depth of a boxcorer (16–23L; 20–30 cm) is larger than of the Van Veen grab (7–17L; 5–10 cm), whereas the reverse is the case for the sampling area, with respectively 0.078m^2 and 0.1m^2 (Steels et al., 2009). Despite these differences, the sampling efficiency is estimated to be comparable, as 90% of the infauna is found in the top 5 cm of the sediment (Steels et al., 2009). In addition, the remaining differences in sample volume are corrected for by using the Margalef index, which adjusts the species richness using the observed total abundance for each sample. In view of this, it is estimated that no standardization of sampling device within the SNS region is needed, which is an important advantage for this regional assessment.

The current results show that the current amount and quality of ICES fishing pressure data are very useful and sufficient to perform the bottom fishing activity-index response testing at the spatial level of the current assessment areas.

4.2. Reference value estimation

One of the aims of this study was to develop a reference value model using median depth. This depth-reference value model is intended to produce sufficiently reliable reference values in case the index value set is too small (< 3 data years). For Margalef diversity and species richness, sigmoid curve fitting (Richards, 1959) appeared to give the most appropriate curve fit. The sigmoid curve fit is made using a generalized logistic function, which was originally developed for growth modelling using flexible S-shaped curves (Richards, 1959). This type of curve has a lower asymptote, an upper asymptote which is called the carrying capacity, and a growth rate. This type of model is proposed in this study to be applicable to the biodiversity of benthic communities, which is limited in shallow waters due to higher natural pressures, and is allowed to expand to its full diversity in deeper waters with low natural pressures. This sigmoid depth-reference model is supported by the results of Leonardsson et al. (2016) and Armonies et al. (2014), who reported that depth is the dominant habitat factor, and that sediment composition appeared to be a habitat property of secondary importance. It was concluded in this study that this model is plausible, and meets one of the aims set for this method to model reference values for Margalef diversity, and if desired for other benthic indices.

The relationship between median depth and SNA reference values has been linearized due to the log-transformation of both species richness and total abundance in this index (Rygg, 2006; Walvoort and van Loon, 2017c). Note that this depth-SNA relationship suggests that Margalef reference values may further increase at greater depths. However, this has to be tested in the future with additional UK reference values, with at least a quality code 2 (3 data years), for the deep Northern North Sea areas (median depth > 50 m). The increase of the AMBI reference value at greater depth may be caused by larger depositions of sediment, including organic matter, at greater depths due to lower current velocities and bottom shear stress. For example for the Dutch Oyster banks, which are the deepest Dutch marine area (median depth 48 m), it was reported by Van Raaphorst et al. (1998) that this area serves as a mid-shelf temporary depocenter for organic matter, and that mineralization in this and similar areas may play a crucial role in the carbon budget of the North Sea. It is well known that the AMBI has been primarily designed to indicate organic enrichment (Borja et al., 2000).

The Margalef depth-reference value model is probably most representative of the sand and mud habitats which provides most data points in the model, but also appears to be applicable to coarse habitats. The very high Margalef reference values for the deeper UK mixed habitats (Farnes East mixed, 11.1; North East of Farnes Deep, 10.5) appear to fall outside the calibrated depth range of this model (up to 50 m median depth). Since these UK reference values are only based on a single survey (corresponding to quality code 1), at least two more surveys per area are needed to obtain a reference value with quality code 2, which can then be used to extend the depth range of this model.

The new variable percentile method is arguably an improvement over the use of fixed percentile values, as it takes into account the known (in this study) or estimated human pressure level, and its estimated effect on the reference value (Hering et al., 2006). Note that in case of pristine benthic communities (which are probably not present in the SNS), the 50 percentile index value should be used as reference value.

4.3. Evaluation of indices/MMIs

The use of a bottom fishing activity and benthic dataset on the relatively large geographical scale of the SNS offered an important opportunity to obtain highly significant ($p < 0.0000$) fishing activity-benthic index response relationships. Josefson et al. (2009) reported for several human pressures threshold values, up to which benthic communities show resistance in the form of a relatively low degradation rate. Above this threshold value, the benthic community degraded more rapidly. With fishing pressure however, such a threshold value is not observed (Fig. 6 and Supplement 3). Degradation of the benthic community starts immediately at very low fishing pressure.

Resilient benthic communities were predominantly observed in this study in coastal areas where relatively high (> 2 sweeps/year) fishing activities occur (Fig. 6). It is also known that in coastal areas the benthic communities are more robust and less biodiverse by nature, due to higher natural pressures occurring there compared to offshore benthic communities (Armonies et al., 2014; Duineveld et al., 1991; Van Denderen et al., 2015). The fishing pressure is judged to work on top of the natural pressures in an additive/cumulative way, by probably transforming the naturally more robust coastal benthic communities into degraded communities resilient to (rapidly recovering from) further increasing fishing activity. However, in the coastal zone fragile biogenic reefs may occur in specific areas (Collie et al., 2000), which are easily destroyed by beam trawling and recover poorly. In a more historical perspective, fishing in the past may have modified habitat structures in coastal and offshore areas by removing large bivalve shells such as oysters (Handley et al., 2014). These mixed habitats possibly had higher reference values than the currently assessed predominantly sandy and muddy habitats.

For the other human pressures tested (organic enrichment, sedimentation and Titanium dioxide), Margalef was in all cases one of the best performing benthic indices. These result show that Margalef diversity can be regarded as a generally applicable multi-pressure index with relatively high sensitivity and precision.

Margalef diversity was theoretically designed as an index which is less sensitive to the sampling area and volume compared to species richness (Margalef, 1958; Peet, 1974). It was observed in this study that Margalef diversity indeed gives a more precise reference value modelling (Table 3), and a higher assessment precision, compared to species richness. This can be explained by the correction of Margalef diversity for (a) natural variations of total abundance and species richness within an assessment area, and for (b) small differences in sample area and volume. These two explanations were confirmed in this study, in which for each assessment area-year and for each sample the species richness was related with the natural logarithm of the total abundance, as is done within the Margalef index formula. These results show that for all assessment areas-years very significant linear relationships were found, with a median $R^2 = 0.59$ and a median $p < 0.00001$ (see for an example Walvoort and van Loon, 2017b, paragraph 3.14.1). As a result of this correction of species richness for total abundance, the precision of the Margalef assessment values is on average 30% higher than of species richness for all SNS assessment areas. This probably results in a smaller number of samples needed to achieve a comparable precision as when using species richness, which enables more cost-effective monitoring.

4.4. Assessment

The results in general show that the benthic fauna quality in coastal SNS assessment areas is lower compared to deeper offshore areas. This is probably mainly due to the relatively high coastal fishing pressure, but possibly also due to other local coastal pressures such as organic enrichment, sediment dumping, sand extraction and chemical pollution.

The lowest Margalef assessment values were found in coastal areas, higher assessment values were found in several deeper (25–50 m) offshore areas in the Southern North Sea, and the highest values were found in deep (> 50 m) offshore areas in the Northern North Sea. In general, these assessment results are considered plausible, in view of the available fishing pressure information. Several assessment areas in the coastal zone show the asymptotic normalized Margalef value of around 0.45, indicating the induction of a benthic community resilient for (rapidly recovering from) higher fishing activities. These assessment values clearly indicate a moderate benthic fauna condition, which is regarded plausible in coastal areas at the current relatively high levels of bottom surface and subsurface fishing activity (> 2 sweeps/year). The otherwise excellent conditions for oxygen and nutrition in the coastal zone presumably limit the further degradation of these benthic communities to an insufficient or bad state (e.g. normalized Margalef value < 0.4–0.2).

It is proposed to consider the UK Farnes East, North East of Farnes Deep and Swallow Sand areas as having a relatively high benthic condition, also considering their very low bottom fishing activities at these larger median depths (69 to 77m). It is recommended to perform at least two more surveys in these areas to obtain reference values with quality code 2 (> 3 data years), which can then be added to the depth-reference Margalef model. The use of more detailed VMS-data (raw data points) is planned by UK to test the sensitivity and precision of the fishing pressure-benthic index relationships at the benthic sample level. Furthermore, the Margalef benthic assessment method is currently tested in the Danish part of the SNS; and is also proposed for testing in other European marine regions.

The obtained average 90% confidence interval of the normalized Margalef assessment values per area-habitat-year of ± 0.06 is regarded as reasonably precise, and useful for the practice of OSPAR and MSFD assessments.

5. Conclusions

The developed new assessment method using Margalef diversity, reference value estimation and depth modelling, and the BENMMI software framework appears to work effectively for the regional SNS benthic fauna assessment for OSPAR and MSFD. This benthic fauna assessment method is therefore recommended for testing in other European marine regions.

Acknowledgements

Laurant Guerin, Sophie Elliot and Jan Witt are acknowledged for valuable comments and information. Frank van de Ende is acknowledged for expert advice and procedural guidance. Joel Cuperus is acknowledged for expert advice on benthic sampling and sample processing.

Cefas (www.cefas.co.uk) are acknowledged for sharing the 2013 North Norfolk Sandbanks and Saturn Reef Cefas/JNCC survey data (© JNCC/Cefas 2013). Cleveland Potash is acknowledged for sharing their data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolind.2017.09.029.

References

Armonies, W., Buschbaum, C., Hellwig-Armonies, M., 2014. The seaward limit of wave effects on coastal macrobenthos. Helgol. Mar. Res. 68, 1–16.

Bergman, M.J., Ubels, S.M., Duineveld, G.C., Meesters, E.W., 2015. Effects of a 5-year trawling ban on the local benthic community in a wind farm in the Dutch coastal zone. ICES J. Mar. Sci.: J. Conseil 72 (3), 962–972.

Borja, A., Tunberg, B.G., 2011. Assessing benthic health in stressed subtropical estuaries, eastern Florida, USA using AMBI and M-AMBI. Ecol. Indic. 11 (2), 295–303.

Borja, A., Franco, J., Pérez, V., 2000. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Mar. Pollut. Bull. 40 (12), 1100–1114.

Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., et al., 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. Mar. Pollut. Bull. 55 (1), 42–52.

Borja, A., Barbone, E., Basset, A., Borgersen, G., Brkljacic, M., Elliott, M., et al., 2011. Response of single benthic metrics and multi-metric methods to anthropogenic pressure gradients, in five distinct European coastal and transitional ecosystems. Mar. Pollut. Bull. 62 (3), 499–513.

Coates, D.A., Kapasakali, D.A., Vincx, M., Vanaverbeke, J., 2016. Short-term effects of fishery exclusion in offshore wind farms on macrofaunal communities in the Belgian part of the North Sea. Fish. Res. 179, 131–138.

Collie, J.S., Hall, S.J., Kaiser, M.J., Poiner, I.R., 2000. A quantitative analysis of fishing impacts on shelf-sea benthos. J. Anim. Ecol. 69 (5), 785–798.

De Groot, S.J., 1984. The impact of bottom trawling on benthic fauna of the North Sea. Ocean Manage. 9 (3), 177–190.

Duineveld, G.C.A., Künitzer, A., Niermann, U., De Wilde, P.A.W.J., Gray, J.S., 1991. The macrobenthos of the North Sea. Neth. J. Sea Res. 28 (1–2), 53–65.

Eigaard, O.R., Bastardie, F., Breen, M., Dinesen, G.E., Hintzen, N.T., Laffargue, P., et al., 2015. Estimating seabed pressure from demersal trawls, seines, and dredges based on gear design and dimensions. ICES J. Mar. Sci. 173 (Suppl. 1), i27–i43=.

Gittenberger, A., van Loon, W., 2013. Sensitivities of marine macrozoobenthos to environmental pressures in The Netherlands. Nederlandse Faunistische Mededelingen 41, 79–112.

Handley, S.J., Willis, T.J., Cole, R.G., Bradley, A., Cairney, D.J., Brown, S.N., Carter, M.E., 2014. The importance of benchmarking habitat structure and composition for understanding the extent of fishing impacts in soft sediment ecosystems. J. Sea Res. 86, 58–68.

Hering, D., Feld, C.K., Moog, O., Ofenböck, T., 2006. Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. Hydrobiologia 566 (1), 311–324.

Hiscock, K., Langmead, O., Warwick, R., 2005. Identification of Seabed Indicator Species from Time-series and Other Studies to Support Implementation of the EU Habitats and Water Framework Directives. Report to the Joint Nature Conservation Committee and the Environment Agency from the Marine Biological Association. Second edition. Marine Biological Association, Plymouth (JNCC Contract F90-01-705).

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

ICES, 2016. Updated ICES Fisheries GIS Maps Version. (Available from: http://ecosystemdata.ices.dk/download/OSPAR_mapping_bottom_fishing_intensity_data_outputs_2015.zip).

ISO 16665, 2014(en) Water quality – Guidelines for quantitative sampling and sample processing of marine soft-bottom macrofauna.

Josefson, A.B., Blomqvist, M., Hansen, J.L., Rosenberg, R., Rygg, B., 2009. Assessment of marine benthic quality change in gradients of disturbance: comparison of different Scandinavian multi-metric indices. Mar. Pollut. Bull. 58 (9), 1263–1277.

Kaiser, M.J., Clarke, K.R., Hinz, H., Austen, M.C.V., Somerfield, P.J., Karakassis, I., 2006. Global analysis of response and recovery of benthic biota to fishing. Mar. Ecol. Prog. Ser. 311, 1–14.

Kröncke, I., Bergfeld, C., 2003. North Sea benthos: a review. Mar. Biodivers. 33 (1), 205–268.

Kröncke, I., Reiss, H., Eggleton, J.D., Aldridge, J., Bergman, M.J., Cochrane, S., et al., 2011. Changes in North Sea macrofauna communities and species distribution between 1986 and 2000. Estuar. Coast. Mar. Sci. 94 (1), 1–15.

Lambert, G.I., Jennings, S., Kaiser, M.J., Davies, T.W., Hiddink, J.G., 2014. Quantifying recovery rates and resilience of seabed habitats impacted by bottom fishing. J. Appl. Ecol. 51 (5), 1326–1336.

Leonardsson, K., Blomqvist, M., Rosenberg, R., 2016. Reducing spatial variation in environmental assessment of marine benthic fauna. Mar. Pollut. Bull. 104 (1), 129–138.

Long, D., 2006. BGS Detailed Explanation of Seabed Sediment Modified Folk Classification. (Accessible online at: http://ec.europa.eu/maritimeaffairs/emodnet/documents/standards/mesh_ geology. pdf).

Margalef, R., 1958. Information theory in ecology. Int. J. Gen. Syst. 3, 36–71.

Maurer, D., Nguyen, H., Robertson, G., Gerlinger, T., 1999. The Infaunal Trophic Index (ITI): its suitability for marine environmental monitoring. Ecol. Appl. 9 (2), 699–713.

Muxika, I., Borja, A., 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 (1), 16–29.

OSPAR, 2012. MSFD Advice Manual and Background Document on Biodiversity. A Living Document – Version 3.2 of 5 March 2012.

OSPAR, 2017. CEMP Guidelines. Common Indicator: BH3 Extent of Physical Damage to Predominant and Special Habitats. Version 5.1.

Olsgard, F., Gray, J.S., 1995. A comprehensive analysis of the effects of offshore oil and gas exploration and production on the benthic communities of the Norwegian continental shelf. Mar. Ecol. Prog. Ser. 122, 277–306 (heavy metals have most effects, determined with MVA).

Olsgard, F., Hasle, J.R., 1993. Impact of waste from titanium mining on benthic fauna. J. Exp. Mar. Biol. Ecol. 172 (1–2), 185–213.

Peet, R.K., 1974. The measurement of species diversity. Annu. Rev. Ecol. Syst. 285–307.

Richards, F.J., 1959. A flexible growth function for empirical use. J. Exp. Bot. 10 (2), 290–301.

Rosenberg, R., Blomqvist, M., Nilsson, H.C., 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 (9), 728–739.

Rosenberg, R., 1973. Succession in benthic macrofauna in a Swedish fjord subsequent to the closure of a sulphite pulp mill. Oikos 244–258.

Rumohr, H., Kujawski, T., 2000. The impact of trawl fishery on the epifauna of the southern North Sea. ICES J. Mar. Sci.: J. Conseil 57 (5), 1389–1394.

Rygg, B., 2006. Developing Indices for Quality-status Classification of Marine Soft-bottom Fauna in Norway.

Shannon, C.E., Weaver, W., 1963. The Measurement Theory of Communication.

Steele, J.H., Thorpe, S.A., Turekian, K.K. (Eds.), 2009. Elements of Physical Oceanography: A Derivative of the Encyclopedia of Ocean Sciences. Academic Press, pp. 53 (Fig. 9).

Teixeira, H., Neto, J.M., Patrício, J., Veríssimo, H., Pinto, R., Salas, F., Marques, J.C., 2009. Quality assessment of benthic macroinvertebrates under the scope of WFD using BAT, the Benthic Assessment Tool. Mar. Pollut. Bull. 58 (10), 1477–1486.

Tillin, H.M., Hiddink, J.G., Jennings, S., Kaiser, M.J., 2006. Chronic bottom trawling alters the functional composition of benthic invertebrate communities on a sea-basin scale. Mar. Ecol. Prog. Ser. 318, 31–45.
Van Denderen, P.D., Bolam, S.G., Hiddink, J.G., Jennings, S., Kenny, A., Rijnsdorp, A.D., van Kooten, T., 2015. Similar effects of bottom trawling and natural disturbance on composition and function of benthic communities across habitats. Mar. Ecol. Prog. Ser. 541, 31–43.
Van Hoey, G., Permuy, D.C., Vandendriessche, S., Vincx, M., Hostens, K., 2013. An ecological quality status assessment procedure for soft-sediment benthic habitats: weighing alternative approaches. Ecol. Indic. 25, 266–278.
Van Loon, W.M.G.M., Boon, A.R., Gittenberger, A., Walvoort, D.J.J., Lavaleye, M., Duineveld, G.C.A., Verschoor, A.J., 2015. Application of the Benthic Ecosystem Quality Index 2 to benthos in Dutch transitional and coastal waters. J. Sea Res. 103, 1–13.
Van Raaphorst, W., Malschaert, H., Van Haren, H., 1998. Tidal resuspension and deposition of particulate matter in the Oyster Grounds, North Sea. J. Mar. Res. 56 (1), 257–291.
Walvoort, D.J.J., van Loon, W.M.G.M., 2017a. BENMMI Benthos Assessment Software. (Available from: https://cran.r-project.org/web/packages/BENMMI/index.html).
Walvoort, D.W.W., van Loon, W.M.G.M., 2017b. BENMMI User Manual Version 4. (Available from: http://publicaties.minienm.nl).
Walvoort, D.J.J., van Loon, W.M.G.M., 2017c. Short Note on the SNA Biodiversity Index: a Modified SN Index. (Available from: http://publicaties.minienm.nl).
Word, J.Q., 1979. Classification of benthic invertebrates into infaunal trophic index feeding groups. Coast. Water Res. Project Biennial Rep. 1980, 103–121.