WORKSHOP TO EVALUATE PROPOSED ASSESSMENT METHODS AND HOW TO SET THRESHOLDS FOR ASSESSING ADVERSE EFFECTS ON SEABED HABITATS (WKBENTH3)

Please note: List of participants and authors were updated in December 2022

VOLUME 4 | ISSUE 93

ICES SCIENTIFIC REPORTS

RAPPORTS SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA

CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H.C. Andersens Boulevard 44-46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

ISSN number: 2618-1371

ICES Scientific Reports

Volume 4 | Issue 93

WORKSHOP TO EVALUATE PROPOSED ASSESSMENT METHODS AND HOW TO SET THRESHOLDS FOR ASSESSING ADVERSE EFFECTS ON SEABED HABITATS (WKBENTH3)

Recommended format for purpose of citation:

ICES. 2022. Workshop to evaluate proposed assessment methods and how to set thresholds for assessing adverse effects on seabed habitats (WKBENTH3). ICES Scientific Reports. 4:93. 102 pp. https://doi.org/10.17895/ices.pub.21666260

Editors

Ellen Kenchington • Saša Raicevich

Authors

Aurélien Boyé • Paul Coleman • Grete Elisabeth Dinesen • Ulla Fernández • Jan Geert Hiddink • Andrew Kenny • Marie-Louise Krawack • Axel Kreutle • Pascal Laffargue • Liam Matear • Henrik Nygård Antonia Nystrom Sandman • Nadia Papadopoulou • Andrea Pierucci • Maider Plaza • Marina Pulcini Sofia Reizopoulou • Giada Riva • Marie-Julie Roux • Petra Schmitt • Chris Smith • Daniel van Denderen Gert Van Hoey • Sandrine Vaz • Elina Virtanen • Sander Wijnhoven

ICES
CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

Contents

i Executive summary

The Marine Strategy Framework Directive (MSFD) requires EU Member States to achieve and maintain good environmental status (GES) across their marine waters. WKBENTH3 convened as a hybrid meeting to evaluate benthic assessment methods and indicators for their potential to meet the criteria described under the MSFD Descriptor 6 (seabed integrity). They evaluated a suite of indicator methods, proposed by participants. Those included five indicator methods describing the 'Condition of the Benthic Habitat', primarily linked to D6C5, and six indicator methods for 'Physical Disturbance on Benthic Habitats', primarily linked to D6C3. Variants of some of the indicators as well as some other commonly used diversity indices were also assessed.

A common dataset with broad regional representation was used to compare and contrast indicator performance with 17 benthic invertebrate datasets drawn from a range of pressure gradients (14 over gradients of commercial bottom trawling intensity, 2 over gradients of eutrophication and 1 over a pollution gradient). A meta-analysis of the mean response to trawling across all locations showed that most indicators had, on average, declined at the high trawl impact relative to the baseline and a significant effect of trawling was detected for the indicators Community Biomass, Species Richness, Fraction of long-lived species, Median longevity, Fraction of sentinel species - SoS, Relative Margalef diversity index DM', Shannon Index and Inverse Simpson. The complementarity of the different indices was computed using Spearman correlation coefficients between each of the indices for all gradients, ordering indicators with Ward's hierarchical clustering. One of the key findings was the identification of four groups of indices that showed clear patterns of association. Considering the link of indicators to different benthic community properties, WKBENTH3 proposed that the assessment of D6 should be carried out selecting a number of indicators drawn from different cluster groups to ensure that components of diversity, species sensitivity and abundance (density and/or biomass – or other proxy linked to benthic habitat functioning) are addressed.

WKBENTH3 further ranked model-based benthic sensitivity and impact outputs across broad habitat types (BHTs) in eight different subdivisions in order to contrast indicator responses. The ranking showed a broad congruence, however, every subdivision had variation in ranking of BHTs among indicator methods. Further work is needed to determine the cause of those discrepancies and to look more closely at the values and the response curves generated.

WKBENTH3 developed a worked example of how to estimate thresholds for GES based on the approach of 'detectable change'. The approach was applied to each of the different pressure gradients and to muddy sand habitats. It was not able to estimate thresholds for all gradients datasets as the confidence intervals around some relationships were very wide. Experts highlighted that the assessment of seabed integrity needs to ensure that cross-regional, regional, national and local scale assessments can "talk" to each other and that they are complementarity in terms of what aspects of the ecosystem the respective indicators are capturing and what pressure they are tracking (linked to manageable human activity). Cross-regional assessments will inform whether assessments are measuring the same or similar things, allowing for such crosschecking.

ii Expert group information

Expert group name	Workshop to evaluate proposed assessment methods and how to set thresholds for assessing adverse effects on seabed habitats (WKBENTH3)
Expert group cycle	Annual
Year cycle started	2022
Reporting year in cycle	1/1
Chairs	Ellen Kenchington, Canada
	Saša Raicevich, Italy
Meeting venue and dates	3–7 October 2022, ICES HQ, Copenhagen Denmark (30 participants)

1 Background on MSFD Descriptor 6

WKBENTH3 – Workshop to evaluate proposed assessment methods and how to set thresholds for assessing adverse effects on seabed habitats

The Marine Strategy Framework Directive (MSFD; 2008/56/EC) requires Member States to achieve and maintain good environmental status (GES) across their marine waters in relation to the eleven 'descriptors' set out in MSFD Annex I. Descriptor 1 (benthic habitats) and Descriptor 6 (seabed integrity) are the main descriptors for assessing the state of the seabed, while other descriptors address particular pressures and impacts on the seabed (e.g. D2 – non-indigenous species, D5 – eutrophication).

Commission Decision (EU) 2017/848 (the 'GES Decision') sets out criteria and methodological standards for determining GES and assessing the extent to which it has been achieved. It defines that benthic habitats (D1) and seabed integrity (D6) are to be addressed together via the assessment of 22 benthic 'broad habitat types' (BHTs) and at the scale of biogeographically relevant 'subdivisions' of each MSFD region or subregion. If desired, EU Member States can add so-called 'other habitat types' (OHTs) to their assessments. OHTs were not part of the workshop discussions here.

The GES Decision sets out the following criteria for benthic habitats:

- D6C1 Physical loss of the seabed;
- D6C2 Physical disturbance to the seabed;
- D6C3 Adverse effects of physical disturbance on benthic habitats (spatial extent);
- D6C4 Benthic habitat extent (extent of habitat loss from anthropogenic pressures);
- D6C5 Benthic habitat condition (extent of adverse effects from anthropogenic pressures).

As stated in the revised Art.8 guidance document, "The overall status is represented by the assessment of D6C5 per BHT, including the assessment of D6C3 and D6C4. GES of the BHT is achieved when these criteria have met the respective threshold values (extent threshold for D6C4, and quality and extent thresholds for D6C5). The extent of adverse effects from disturbance (D6C3) and the state (impact) assessment, and inputs from other descriptors (either as spatial impact analysis or qualitative description, as deemed appropriate) contribute to D6C5".

The 'quality' and 'extent' threshold values and the method for assessing overall status of a habitat (integration of criteria D6C4 and D6C5) are being established through a European Union-level process, considering regional or subregional specificities. The process is overseen by the MSFD Common Implementation Strategy (CIS), particularly the Technical Group on seabed habitats and seabed integrity (TG Seabed), the Working Group on Good Environmental Status (WGGES) and the Marine Strategy Coordination Group (MSCG).

The work of TG Seabed has been supported by ICES, through a number of ICES Advice documents. In early 2022, the EU (DG ENV) requested ICES to "advise on methods for assessing adverse effects on seabed habitats". This work is being carried out through a number of steps, including:

- The organization of the WKBENTH2 workshop (24–26 May and 8–10 June 2022, ICES 2022b);
- The production of a technical service (June 2022, ICES 2022);
- The scientific peer-review of the technical service and WKBENTH2 workshop report (July–August 2022), and comments on operational applicability from EU TG Seabed (see Annex 3);

- The organization of this WKBENTH3 workshop (3–7 October);
- The scientific peer-review of the WKBENTH3 workshop report (14–26 October), with opportunity to comment on operational applicability from EU TG Seabed; and
- The production of formal advice by ICES Advisory Committee (ACOM), 1–4 November 2022, to be published as ICES Advice and delivered to the EU DGENV by December 2022.

At WKBENTH2, two sets of criteria were developed to evaluate the suitability of indicators and threshold-setting methods, respectively, to assess adverse effects on seabed habitats for MSFD purposes. The criteria were designed for evaluation at a subregional or regional level. The scoring for these criteria was meant as guidance when choosing indicators and thresholds, so failure to meet one criterion would not necessarily prevent the use of the indicator or threshold in an assessment. The framework was evaluated for six indicators and for 11 methods for setting thresholds. The methods for setting thresholds were found to be more suitable for setting 'quality' rather than 'extent' thresholds. Preferred methods identified ecologically-motivated differences between a good and degraded status, rather than another transition. Quality thresholds based on the lower boundary of the range of natural variation were considered the most promising. This approach can be used for most, but not all, indicators with an extensive temporal and/or spatial database.

WKBENTH2 collated a standardized dataset to test the specificity, sensitivity and/or responsiveness of sampling-based benthic indicators to largely fishing intensity pressure gradients for evaluation by WKBENTH3 (Annex 5). Participants provided input into the selection of indicators, consequently, indicator selection was user-driven and not the result of systematic review.

WKBENTH3 convened as a hybrid meeting from 3–7 October 2022 with 30 participants (Annex 1). The workshop suffered from inconsistent attendance but was able to address the five terms of reference (ToR a–e) detailed in Annex 2. In brief, WKBENTH3 met to:

a) Evaluate proposed assessment methods and how to set thresholds for assessing adverse effects on seabed habitats produced in WKBENTH2 [Sections 4, 5, 6];
b) Prepare worked examples using suitable methods on how to set threshold and assess adverse effects on seabed habitat quality for relevant pressures and impacts [Section 7];
c) Prepare an overview of recommended assessment methods for application to MSFD Descriptor 6 [Section 9, Annexes 4 and 6];
d) Provide higher level guidance on future directions for improvements to the recommended methodology and for developing scientifically-based 'extent' indicators [Section 8]; and
e) Provide higher-level guidance as to a set of criteria, and methods to analyse the performance of assessment methods and how to set thresholds for assessing adverse effects on seabed habitats [Sections 7, 8, 9].

1.1 Selection of relevant benthic indicators

The criteria D6C4 and D6C5 assess environmental status, encompassing impacts from all relevant pressures (including e.g. physical pressures, eutrophication, contaminants, non-indigenous species and hydrographical alterations), in terms of habitat loss (D6C4) and as affecting the structure and function of habitats (D6C5). The suite of indicators needed to assess D6 thus needs to be broad and cover responses to all of these pressures. Through the evolution of WKBENTH1, WKBENTH2 and now WKBENTH3, ICES has reviewed approximately 600 indicators for their suitability as methods for reporting on D6. The 19 indicators examined here (Section 2.1) were largely a product of participant (user) interest and in many cases were informed through WKBENTH2 as noted above. This legacy of indicator selection means that the evaluated indicators are not themselves the product of a systematic review and that other national or regional indicators may be useful for the D6 assessment. Nevertheless, they reflect use by a number of EU member states and ensured that experts were present to explain their analytical outcomes. Indicators analysed in this report are applicable for the assessment of D6C3 (an estimate of the extent of adverse effect by physical disturbance per habitat type in each assessment area) and D6C5. There were no extent indicators included for criterion D6C4.

Another important gap is the paucity of indicators for habitat properties other than 'absence of particularly sensitive or fragile species'. The GES Decision highlights, under criteria for D6C3, five properties which capture the biotic structure of the BHT and its functions (Table 1.1.1). The selection of indicators evaluated here are most often based on species composition of the benthic community, accounting for relative abundance or biomass (Table 1.1.1.). Many indicators also include measures of species sensitivities to pressures allowing for more detailed pressure-impact evaluation based on the community composition (Table 1.1.1). Some indicators utilize trait-based information, such as size, which may correlate to functional responses such as habitat provision. However, none of the indicators examined specifically linked species to key ecological functions. Only one indicator considered size structure in its calculation.

Inevitably, the selection of relevant indicators reviewed by WKBENTH3 (Section 2.1) and the properties of the common dataset used to test them (Section 2.2), produced coverage gaps. Not all indicators could be applied to every dataset. Further, time constraints meant that a number of indicators were tested for one part of the analyses but not another (Sections 4 and 6).

Table 1.1.1. Properties of benthic habitats referred to in Commission Decision (EU) 2017/848 Descriptor 6 Criterion C3 (D6C3) cross-referenced to the indicators evaluated in WKBENTH3. Note that 'species composition' and 'relative abundance' refer to 'typical species' in D6C5 which are not defined in the GES Decision.

Habitat Properties Associated with D6C3	Level of Ecological Organization	Indicator Method for Evaluation
'species composition'	Community	D_M', M-AMBI, DKI
'relative abundance' of species	Community	DKI, SoS
'absence of particularly sensitive or fragile species'	Community	M-AMBI, AMBI, DKI, BENTIX, TDIs (TDI, mTDI, pTDI, mT), L1, L2, CumI, SoS
absence of 'species providing a key function'	Community	
'size structure of species'	Community	PD2

2 Methods for the evaluation of the performance of benthic indicators and metrics

2.1 Indicators evaluated

The indicators that were evaluated by WKBENTH3 are summarized in Annex 4. They included five indicator methods describing the 'Condition of the Benthic Habitat' primarily linked to D6C5 (M-AMBI, DK1, D_M', BENTIX and PD2) and six indicator methods for 'Physical Disturbance on Benthic Habitats' (SoS, TDI and variants, L1, L2, CumI and BH3), primarily linked to D6C3. Of those indicators, a number are risk-based (L1, CumI, BH3, PD2), or include a risk-based version as well as a sample-based or empirical version (L2 and SoS (risk- and sample-based)). Risk-based indicators evaluate seabed integrity using an underlying data layer that describes benthic sensitivity to bottom trawling (or any type of seabed abrasion), where sensitivity varies with environmental conditions and/or habitat types, as well as a prediction of benthic impact. WKBENTH2 suggested that an evaluation of risk-based approaches can be realized by visually comparing maps and through a ranked score per BHT and spatial subdivision (See Section 2.3).

In addition to the 11 listed above and detailed in Annex 4, we included variants of some of the indicators (i.e. AMBI, mTDI, pTDI and mT; Annex 4) as well as some other commonly used diversity indices easily calculated from the data (Biomass, Abundance, Species Richness, Shannon Index, Inverse Simpson, Simpson Index). In total we compared 19 benthic indicators using the gradient datasets compiled by WKBENTH2 (Table 2.1.1). Some of those indicators have been used previously by EU member states in relation to the WFD and/or MSFD, while other indicators have been used within OSPAR or ICES. It should be noted that indicators developed by OSPAR (SoS, D_M', BH3) are currently under revision for inclusion in the OSPAR Quality Status Report 2023 (QSR 2023). Results shared here by indicator leads are preliminary and do not represent the final OSPAR position or QSR results.

Table 2.1.1. Indicators used in the analyses reported by WKBENTH3 (Section 2.2.) with details of where they have been used.

Context of Use	Benthic Indicators
OSPAR, Spain	SoS – Sentinels of the Seabed (SoS)
OSPAR, Netherlands	DM' – relative Margalef diversity index
ICES	L1 – Fraction of community longevity exceeding trawling interval
	L2 – Reduction in median community longevity
	PD2 – Population Dynamics Model
Denmark	DKI – Danish Quality Index
France	TDI – Trawling Disturbance Index
France	mTDI – Modified Trawling Disturbance Index
France	pTDI – Partial Trawling Disturbance Index
France	mT – Modified Vulnerability Index

Context of Use	Benthic Indicators
Greece	BENTIX
General	AMBI – AZTI's Marine Biotic Index
	M-AMBI – Multivariate AZTI's Marine Biotic Index
	Biomass, abundance, richness, Shannon Index, Inverse Simpson, Simpson Index

The lack of abundance or biomass data for some of the datasets provided (Section 2.2) meant that indicators could not be calculated for all datasets. Below we describe for each of the indicators how the values were estimated where further information is warranted (see also Annex 4).

M-AMBI and AMBI

AMBI and M-AMBI can be calculated using abundance or biomass. To calculate AMBI, cephalopods and one fish were removed since the method is for macroinvertebrates. Also, some records at a very high taxonomic level (e.g. 'bivalves') were removed.

M-AMBI was calculated for different geographical regions using abundance data only. M-AMBI needs reference conditions, which are associated with the area and characteristics of the habitat (depth, grain size, community, etc.). The geographical areas were only merged in cases where both the sampling size and the units were the same.

In all cases, the 'bad reference conditions' are 6 for AMBI and 0 for diversity and richness. Then, in each area, calculations of M-AMBI were made using as the 'best reference conditions' those within the dataset with the lowest AMBI value and highest Species Richness and diversity. However, some areas may already be degraded. Hence, a second M-AMBI estimate (referred to as M-AMBI-plus) was made using as reference conditions the lowest AMBI value −15%, and the highest richness and diversity +15%, as proposed by Borja *et al.* (2008) and Borja and Tunberg (2011) when monitoring has been done in areas assumed to be impacted by human activities.

DKI - Danish Quality Index

DKI (Josefson, 2009) is a multimetric quality index composed of a diversity component represented by the Shannon-diversity (H′) and a sensitivity component, the AMBI (Borja *et al.* 2000). The two indices are weighted equally in the calculation of DKI which ranges between 0 (poorest ecological/environmental quality) and 1 (highest ecological quality). Calculation of DKI is furthermore adjusted for very low abundances in the sample (equation below). DKI has been intercalibrated against other Scandinavian quality indices where all indices were tested against each other in various pressure gradients (Josefson *et al.*, 2009). The measurements of each of the two indices (AMBI and Shannon) are scaled relatively to the highest expected value expected at the sampling location calculated as $H'_{observed}/H'_{max}$ and AMBI - $AMBI_{min}$ (the AMBI scales from 0 to 7, where 0 represents the highest ecological quality and therefore the observed value is scale to minimum value).

$$DKI\ v.2 = \frac{\left(1 - \left(\frac{AMBI - AMBImin}{7}\right)\right) + \left(H'\frac{1}{H'max}\right)}{2} \times (1 - \left(\frac{1}{N}\right))$$

As DKI was initially developed for Danish waters, where there are very strong salinity gradients which influences both the species diversity (H′) and the AMBI value, the later version of DKI (DKI v.2) used in this assessment has been normalized to the ambient salinity. Species diversity the Baltic Sea, and in general, declines almost linearly with declining salinity. In addition, calculation of AMBI has been shown to have dependencies on salinity or other estuarine environmental gradients that parallels the salinity gradient in estuaries. For AMBI, it has been shown that species tolerant of organic enrichment also are relatively more abundant in brackish environments. $AMBI_{min}$ and H'_{max} are therefore normalized to the salinity using an empirical established relationship with salinity obtained from a large number of datasets in across the Baltic Sea (Carstensen *et al.*, 2014). For the dataset from the southern Baltic (OxyTrawl) and from Gulf of Finland (Finland), salinity was set to 13 and 3 respectively. For the rest of the gradient datasets, salinity was set to 30.

Software tools from ATZI (version 2014) were used to calculated AMBI values (without replication in the input data, i.e. single sample). The values of DKI were estimated for each individual sample and DKI values were afterwards averaged per sampling location.

Relative Margalef's index of diversity (D_M')

Margalef's index of diversity (D_M) is given by:

$$D_M = \frac{S - 1}{\ln(N)}$$

where S is the Species Richness and N is the total abundance for each sample. Margalef's index of diversity (D_M) is an absolute measure of diversity. To improve comparability and consider methodological and 'natural' variability, the relative Margalef index of diversity was proposed:

$$D_M{}' = \frac{D_{ass} - D_{bad}}{D_{ref} - D_{bad}}$$

where D_{ass} is the assessed value for the Margalef diversity index, D_{bad} is the Margalef value for a bad ecological state and D_{ref} is the reference value for a good ecological state. This method is identical with the calculation of the OSPAR indicator BH2b on relative Margalef diversity for the Quality Status Report (QSR 2023), however deviates from Margalef diversity calculated for the Intermediate Assessment (IA 2017). D_{ref} is not a pristine reference, but rather a good quality status within reach, considering the current benthic community compositions and species pools, particularly of use for standardization of the assessment methodology, where sampling and laboratory approaches (e.g. identification of species and recording of specimens) might differ between datasets. The value for D_{ref} is estimated based on low pressure observations. Initially D_M is calculated at the level of samples. According to D_M' methodology, usually results for D_M at the sample level are combined with pressure mapping after which the D_{ref} is achieved as a value dependent on the (low) pressure level and data availability (number of samples and years covered) from a case-specific selection of low-pressure data. Herewith taking 'natural' (and potentially other sources) of variability into account, as there were only a limited number of data available per gradient, and reference values at the level of Broad Habitat Types were only available for the North Sea region, it was decided to use the median of the D_M values with the least pressure for each of the different gradient datasets, as D_{ref}. In all gradient test sets related to fishing pressure, samples with SAR < 0.25 were selected to estimate D_{ref}. Exceptions are FC, NIC2 and TH datasets, where D_{ref} was based on samples with SAR = 0.0, for PH, SAR = 0.06, and for SEL, SAR = 0.5 and SP, where samples < 5% on the relative disturbance frequency gradient were selected. For Finland samples with oxygen levels > 8 mg/l, for Saronikos samples with total N < 0.1%, and Vigo samples with CPI index < 0 were selected to estimate D_{ref}.'

BENTIX

BENTIX was calculated using the abundance datasets only. In calculating BENTIX, all cephalopods were removed, as well as a number of crustaceans and molluscs usually assigned to megafauna, since this index was developed for the benthic macroinvertebrates. In addition, records identified at a higher level of taxonomic resolution (e.g. 'bivalves') were accounted for in the analysis as a part of the benthic community, but they were not scored.

The stations outside the confidence limits of BENTIX (samples with 3 species or less, and/or 6 individuals or less, and/or non-assigned species exceeding 20% of total community abundance) were not included in the analysis. However, an exemption to the latter criterion was applied for Finland's data (Finland), since this dataset was the only one providing an eutrophication pressure gradient other than Saronikos.

The scale of the BENTIX index ranges from 0 (azoic conditions) to 6 (reference conditions). The same boundaries of the BENTIX ecological status classification could be applied to assess several habitat types of coastal waters, since the index is based on the relative proportions of 'sensitive' and 'tolerant' groups of species. However, the boundaries should be further evaluated across habitat and pressure types, especially in cases where the benthic fauna are naturally dominated by tolerant species. Currently, a modification of High/Good and Good/Moderate boundaries is suggested for the purely muddy habitats (silt and clay particles > 90%), but since the data of granulometry were not available for the datasets provided, such adjustment hasn't been applied.

SoS – Sentinels of the Seabed

The indicator Sentinels of the Seabed (SoS) is an ecological indicator developed in the frame of the OSPAR expert groups (BH1 in OSPAR nomenclature) which assess the status of benthic habitats based on the proportion of a set of sentinel species (see Serrano *et al.* (2022) for complete details). The indicator determines sentinel species based on two criteria; 1) species that can be frequently found in the natural habitat and 2) species that are sensitive to the studied pressure. To define "frequent or typical species", two different metrics were applied, i) relative contribution of species to intra-habitat similarity between stations sampled in the target habitat within reference condition areas (no disturbance or very low disturbance) using the Similarity Percentages procedure (SIMPER; Clarke 1993), and ii) relative frequency for each species within the target habitat under reference conditions. This initial set of "frequent or typical species" is filtered by prioritizing species according to a SoS sensitivity index (species responses to the analysed pressure), avoiding, when possible, tolerant species (i.e. those whose abundance does not show a clear response to the pressure) and always avoiding opportunistic species (i.e. those whose abundance increases with the pressure). SoS sensitivity index is calculated from available classifications of sensitivity to a pressure or pressures group. Previous to WKBENTH2, the SoS indicator was tested (Serrano *et al.*, 2022) using two sensitivity indices, the BESITO Index (González-Irusta *et al.*, 2018) for trawling disturbance, and the AMBI groups for chemical pollution (Borja *et al.*, 2000). Here we also used the AMBI groups to test the SoS indicator for eutrophication and we developed a new Sensitivity Index group based on the longevity classes developed by the WGFBIT (Bolam *et al.*, 2014; 2017) to test the case studies based on infauna from the Baltic and North Sea (see Table 2.1.2).

Although ideally BESITO should be used for these case studies, the current cover of the BESITO species list (mainly epifaunal species) of the species present in the test datasets evaluated here (mainly infauna) was low (< 40% total richness) or very low (< 20% total richness) so we decided to test a new proxy to sensitivity based on longevity, in the same way as is being testing in the WGFBIT. For doing this we subjectively defined as sensitive species (Sensitivity Index group of 3) all species with a proportion of biomass assigned to the higher longevity class (> 10 years) equal or higher than 0.7. We defined as tolerant (Sensitivity Index group of 2) all species with a proportion of biomass assigned to the higher longevity class lower than 0.7 but higher than this value when the two highest longevity classes were combined (> 3 years). Finally, we defined as opportunistic species (Sensitivity Index group of 1) all species not included in any of the previous categories. This classification is a temporary solution to the lack of BESITO data for these species and cannot be considered final. Therefore, caution is needed when interpreting the results for the SoS indicator based on longevity data. Future work is needed for extending the BESITO Index to a larger list of species or working on a more robust (and further tested) method to assign sensitivity scores based on longevity classes. In any case, this new proxy for the Sensitivity Index has allowed testing of the SoS indicator in a wider group of case studies and therefore we think that is useful to compare the indicator with other approaches. SoS values could not be estimated where there were relatively high trawling intensities at the least fished sampling stations.

Table 2.1.2 Sentinel species used to estimate SoS for the different gradient datasets.

Dataset Code	Sensitivity Index (S.I.)	S.I. Group	Sentinel Species
AS1	BESITO groups	2	*Anadara demiri*
		3	*Ascidia, Ascidia virgínea, Holothuria tubulosa, Ophiothrix fragilis*
AS2	BESITO groups	2	*Anadara demiri, Anadara inaequivalvis, Anadara kagoshimensis,Corbula gibba*
		3	*Thyone fusus*
FC	BESITO groups	3	*Bathybiaster vexillifer*
		4	*Anthoptilum grandiflorum, Duva florida, Funiculina quadrangularis, Phelliactis* spp., *Heteropolypus sol*
		5	*Actinoscyphia saginata, Stryphnus fortis, Thenea* spp., *Mycale lingua*
NIC1	BESITO groups	3	*Gracilechinus acutus, Lytocarpia myriophyllum , Ophiothrix fragilis, Parastichopus regalis, Actinauge richardi*
NIC2	BESITO groups	3	*Parastichopus tremulus, Araeosoma fenestratum, Hymenodiscus coronata, Nym-phaster arenatus*
		4	*Funiculina quadrangularis, Kophobelemnon stelliferum*
		5	*Acanella arbuscula, Asconema setubalense, Pheronema carpenteri, Geodia* sp.
DB	Longevity groups	2	*Astropecten* spp, *Hydrozoa., Magelona mirabilis, Nephtys* spp., *Phaxas pellucidus*
FG	Longevity groups	2	*Nephtys, Thyasira, Lumbrineris*
		3	*Pennatula phosphorea, Arctica islandica, Virgularia mirabilis*
Gotland	Longevity groups	2	*Macoma balthica*
OxyTrawl	Longevity groups	2	*Macoma balthica, Scoloplos armiger*
PH	Longevity groups	2	*Owenia fusiformis, Parvicardium scabrum*
SP	Longevity groups	2	*Ennucula tenuis, Glyceridae, Kurtiella bidentata, Nereididae, Phaxas pellucidus, Polynoidae, Sigalionidae, Synaptidae*
		3	*Thracia convexa*

Dataset Code	Sensitivity Index (S.I.)	S.I. Group	Sentinel Species
TH	Longevity groups	2	*Glyceridae, Nucula nitidosa, Tellimya ferruginosa*
Vigo	AMBI groups	3	*Ampelisca* sp., *Atylus* sp. *Calyptraea chinensis, Chamelea striatula, Euclymene oerstedi, Eudorella truncatula, Lumbrineris scopa, Maldane glebifex, Metaphoxus fultoni, Musculus costulatus, Nucula* sp.
Finland	AMBI groups	2	*Marenzelleria* sp.
		3	*Bylgides sarsi*
Saronikos	AMBI groups	2	*Aponuphis brementi, Drilonereis filum* *Eunice vittata, Glycera unicornis, Goniada maculata, Harmothoe antílopes, Paralacydonia paradoxa*
		3	*Pilumnus* sp., *Pista cristata, Timoclea ovata*

mTDI, pTDI, TDI and mT

Trawl disturbance indices (TDI) may be computed on either abundance or biomass, each of which may be log-transformed or not. All outputs are provided following Jac *et al.* (2020a) calculations. mTDI, pTDI, TDI and mT indices are based on species biological traits relevant to this pressure (position, size, mobility, fragility, feeding) and were developed focusing on mega-epifauna. As a result, little to no trait scores were available for endofauna.

The biological trait information may be linked to a given taxa at species, genus and/or family level. When it was not possible to find trait information for some taxa (even when trying to degrade the identification to genus or family level), it was decided to use a cut off level. If there was less than 75% of the station summed biological metric (abundance or biomass, log or not) that was informed in terms of species traits, the station was removed. All cephalopods were also removed as they may in some instance largely dominate the community biomass.

As a result, some datasets were only partially informed and other were fully missing. Only the datasets with at least 50% of observations informed and covering 90% of the trawling disturbance gradient were analysed.

Median longevity and fraction long-lived organisms

For each of the locations (Section 2.2), benthic species were linked to a species-by-trait matrix with trait information on longevity (maximum lifespan). Benthic trait information was derived from the ICES Working Group WGFBIT. For most locations, longevity was subdivided into four trait classes (< 1 year, 1–3, 3–10, and > 10 years). For the Flemish Cap, longevity was subdivided into five classes (< 1 year, 1–3, 3–10, 10–50, and > 50 years) based on information from Murillo *et al.* (2020). For each species-longevity combination, a score of one was assigned to a single class when a species longevity matched a longevity class. Otherwise, fractional scores that summed to one were assigned to multiple longevity classes, following Bolam *et al.* (2014). From this species-by-longevity matrix, including in some cases higher taxonomic levels, a table of locations by biomass-weighted trait longevity classes was calculated by multiplying the total biomass per species by the longevity score. These were then summed by longevity class and divided by the total biomass of the location to produce a proportional biomass-weighted longevity table for all locations.

The fraction of long-lived organisms was estimated as the proportional biomass in the most long-lived trait class. For some gradient studies, biomass was zero in this long-lived trait category in most or all sampling stations (e.g. in gradients in the Baltic Sea). In these cases, we also included the proportional biomass in the second most long-lived trait class.

Median longevity was estimated by converting the biomass by longevity to a cumulative biomass by calculating the biomass proportion with longevity that is smaller than or equal to 1, 3, and 10 years (and 50 years for the Flemish Cap) in each location. We estimated the biomass–longevity composition using a statistical model, with the cumulative biomass proportions as the response variable and longevity as the predictor variable. Following Rijnsdorp *et al.* (2018), we used a binomial model where longevity is ln-transformed. We afterwards calculated the median longevity from each statistical model/location. The median longevity describes the longevity in years where the cumulative biomass proportion is 0.5.

Biomass, Abundance, Species Richness, Shannon Index, Inverse Simpson, Simpson Index

Biomass, abundance and Species Richness were obtained from the data. Abundance and richness were further used to calculate the Shannon, Inverse Simpson and Simpson indices. The diversity indices were estimated in R using the "vegan" package (Oksanen *et al.*, 2020).

2.2 Common datasets and analyses for evaluating indicator performance

WKBENTH2 compiled 17 benthic datasets; 14 over gradients of commercial bottom trawling intensity, 2 over gradients of eutrophication and 1 over a pollution gradient (Table 2.2.1, Figure 2.2.1, Annex 5). Eight of these gradients targeted a specific area at relatively small spatial scales and were designed to examine differences in benthic community composition along the gradient. The other datasets were derived from benthic monitoring programs that often covered larger spatial scales. For the latter studies, suitable sampling locations were selected from the larger monitoring program by finding locations with similar BHTs and environmental conditions.

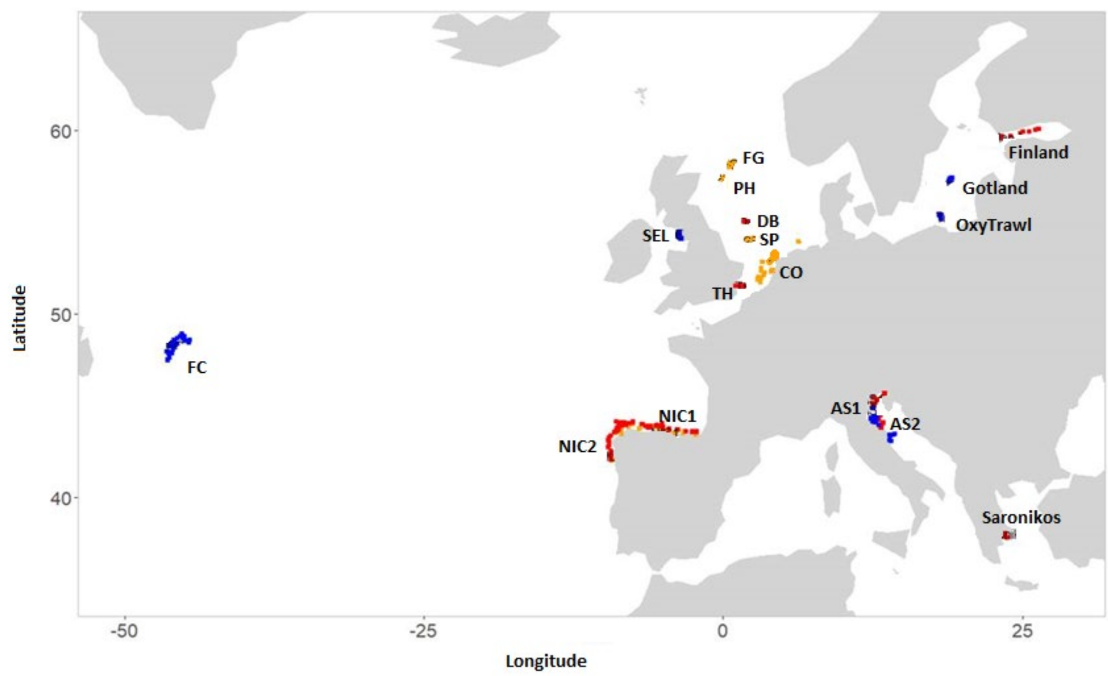

Figure 2.2.1. Location of the different gradient datasets used to test indicator performance by WKBENTH3 (See Annex 5 for more details and Table 2.2.1 for name codes.

For each of the gradients, it was expected that the effects of trawling/eutrophication/pollution on benthic communities will have a larger impact than any differences in the environmental conditions. The pressure gradients are therefore assessed in isolation. However, since there is substantial variation in seabed depth for some datasets, we also evaluated how depth, in combination with the pressure gradient, affected the benthic community in a few relevant locations.

To examine the complementarity of the different indices, we computed the Spearman correlation coefficients between each of the indices for each gradient. We afterwards plotted the average correlation across all gradients, ordering indicators with Ward's hierarchical clustering.

Lastly, and only for the trawling gradients, we calculated the mean response to trawling across all locations. This response was estimated by calculating the change in indicator values from low vs. high trawl disturbed stations at each location. The low and high stations were manually selected. For all areas where trawling intensity was quantified as the swept-area-ratio (SAR), the mean SAR was 5.4 $year^{-1}$ in the high trawl disturbed stations (range: 1.5–13.4) and 0.27 $year^{-1}$ in the control area (range: 0–1.6). Not all indicators provided output for all stations and/or locations. Both the correlations and the responses to trawling were derived from the data that were available.

A number of the gradient datasets took replicate samples at each sampling location (Annex 5). Those data were summated to make the exercise manageable. That aggregation may have caused a bias in some indicators related to diversity due to confounding the alpha and beta-diversity contributions. This is a concern especially for intermediate trawl intensities where there is a chance that some subsamples are affected by trawling and others not.

Table 2.2.1 Brief description of the 17 datasets used to test the indicator's performance by WKBENTH3. Further details are found in Annex 5. For each dataset a list of indicators that were not calculated is given.

Name Code	Location	Sampling Method	Pressure	Sediment Type (BHT)	Depth Range (m)	Indicators Not Evaluated
AS1	Adriatic Sea	Otter trawl	Bottom trawling	Circalittoral sand	9–56	DKI
AS2	Adriatic Sea	Otter trawl	Bottom trawling	Circalittoral mud	8–87	DKI
CO	Dutch EEZ high tidal stress area	Boxcore	Bottom trawling	Sand	22–36	SoS
DB	Dogger Bank	Hamon grab	Bottom trawling	Sand	25–30	
FC	Flemish Cap, NW Atlantic	Lofoten bottom trawl	Bottom trawling	Unknown bathyal sediment	786–1236	D_M', DKI, TDI, mTDI, pTDI, mT, BENTIX, AMBI, M-AMBI, Biomass, Abundance, Shannon Index, Inverse Shannon, Simpson Index
FG	Fladen Ground	Day grab	Bottom trawling	Mud	143−153	
Finland	Gulf of Finland	van Veen grab	Eutrophication	-	56–84	TDI, mTDI, pTDI, mT

Name Code	Location	Sampling Method	Pressure	Sediment Type (BHT)	Depth Range (m)	Indicators Not Evaluated
Gotland	Gotland	van Veen grab	Bottom trawling	Muddy sand	37–59	TDI, mTDI, pTDI, mT
NIC1	Northern Iberian Coast	Otter trawl	Bottom trawling	Offshore circalitoral sand	71–202	DKI
NIC2	Northern Iberian Coast	Otter trawl	Bottom trawling	Several, but mainly mud	186–936	DKI
OxyTrawl	Southern Baltic Sea	Boxcore	Bottom trawling	Sand	70–85	
PH	Long Forties, North Sea	Hamon grab	Bottom trawling	Gravelly sand	74–83	
Saronikos	Saronikos Gulf	Boxcore	Eutrophication	Mixed sand / mud	20–94	
SEL	Sellafield, Irish Sea	Day grab	Bottom trawling	Muddy sand	21–42	SoS
SP	Silver Pit	Boxcore	Bottom trawling	Muddy sand	68–78	
TH	Thames	Boxcore	Bottom trawling	Sand	16–40	
Vigo	Vigo Estuary	BOUMA boxcore	Pollution	Infralittoral Mud	< 30	Median longevity, fraction long-lived, TDI, mTDI, pTDI, mT, Biomass

The mean response was analysed using weighted meta-analysis via linear mixed-effects models (a standard approach for meta-analysis, using the rma.uni function in r package 'metafor'). The response variable for the candidate indicator (I) was the log response-ratio (lnRR), calculated as $\ln(I_{trawled}/I_{control})$ (following Hiddink *et al.* (2020) and references therein). Studies were weighted by the inverse of variance of the original study. This was calculated from the standard deviation (SD) and number (n) of observations in the low and the high impact groups. Two indicator methods were based on proportions, and each had one $I_{trawled}$ value very close to zero. Since this inflates the log response-ratio (and the confidence interval), we set these two $I_{trawled}$ values to 0.01.

We also visualized the response of the indicators in each area. For trawling gradients, the visualization was done using linear regression where trawling intensity was log10(x+1)-transformed. Outputs were only shown when the AIC of the trawl model was lower than the null model (Indicator ~ 1). We verified that similar results were obtained with a generalized additive model (GAM) where there is more freedom in the shape of the fitted curve. For eutrophication and pollution, we estimated the decline with a GAM and, as with trawling, we compared the GAM against the null model.

The data from the Adriatic Sea (AS1, AS2) and Northern Iberian Coast (NIC1, NIC2) were obtained by bottom trawls and included data on cephalopods. Most indicator methods were unable to deal with cephalopods as the methodology was developed for macroinvertebrates (see Section 2.1). Therefore, cephalopods were removed from the datasets when estimating biomass, abundance, and different diversity metrics (richness, Simpson Index, etc.). However, in calculating the SoS indicator (Section 2.1) the cephalopod information was retained for some species considered more linked to the substratum (see González-Irusta *et al.* (2018) for more details).

Some of the datasets resulted from specific sampling campaigns that aimed to sample a gradient of pressure while keeping environmental conditions such as depth and sediment type similar (e.g. DB and FG), while other gradients were the result of ongoing monitoring that covered a wide range of environmental conditions (e.g. NIC1 and NIC2). The second type of gradient may produce confounding effects, where the apparent effect of the human pressure is masked or amplified by underlying environmental gradients.

2.3 Comparative risk-based sensitivity and impact scoring

Accurately assessing ecological condition requires attention to the key factors and processes that drive ecosystem dynamics and their relationship with ecosystem functioning. For seabed ecosystems important relevant functions have been suggested to be bioturbation, nutrient cycling, provision of shelter and food provisioning (secondary production) for higher trophic levels (Rice *et al.,* 2012). However, the application of biological traits analysis (BTA) has been shown to be useful in helping to determine and quantify the relationships between species populations, processes, functions, and the associated ecosystem goods and services they support (Thrush *et al.,* 2014; Bolam *et al.,* 2016; Dimitriadis *et al.,* 2012; van Son *et al.,* 2013). Accordingly, the application of BTA can be used to assess the ecological significance of risk-based indicators by understanding the role that individual species-traits (e.g. longevity, size and feeding mode etc.) play in supporting important ecosystem functions. For example, naturally undisturbed seabed ecosystems are expected to have a range of species present, with each species having a distribution of abundance and biomass over different age and size classes, largely determined by a combination of the prevailing environmental conditions and species interspecific competition for resources (e.g. space and food). Risk-based indicators that are particularly responsive to seabed disturbance are therefore more likely to incorporate some aspects of species size and longevity, as it is typically the largest and oldest individuals in the community that will be disproportionately affected during the initial stages of disturbance, particularly in those environments which are naturally stable. Such impacts may give rise to large drops in the total biomass of the community, and large drops in the rates at which ecosystem processes occur, which may persist for many years following the cessation of a disturbance.

On the other hand, benthic communities predominantly composed of short-lived, fast-growing and small individuals can play a significant role in supporting vital secondary ecosystem production functions (Hiddink *et al.,* 2008). Indeed, while such assemblages may recover from localized disturbances very quickly (owing to their predominantly *r*-selected characteristics), spatially extensive and persistent disturbance of such assemblages can result in a significant ecological risk. In these cases, the persistence and spatial extent of a disturbance event is arguably of greater ecological significance than the event itself.

Evaluation of the different risk-based approaches was done by evaluating ranked sensitivity/impact scores per BHT and spatial subdivision (Table 2.3.1). We obtained five different risk-based approaches that cover eight subdivisions (Figure 2.3.1; note that not all approaches are in all regions). Habitat-specific scores were compared and ranked per subdivision, with the aims of highlighting consistencies and differences between indicator outputs when identifying habitats considered most sensitive, and therefore, at risk to adverse effects from physical pressure. The underlying pressure layers that were used to obtain impact scores differed between assessment methods (Table 2.3.2).

Most risk-based approaches examined risk at the grid cell level, whereas CumI uses polygons. For the purpose of this comparison, grid cell estimates were aggregated to the BHT. The methodology used to aggregate to the BHT is presented in Table 2.3.2. Scatterplots of Relative Benthic Status (RBS) scores of grid cells (0.05°x0.05°) were estimated for SoS, PD2 and L1 indicators by BHT in the Iberian region, and the Spearman Rank correlation coefficients were calculated. For the analyses only the BHTs with more than five grid cells were plotted.

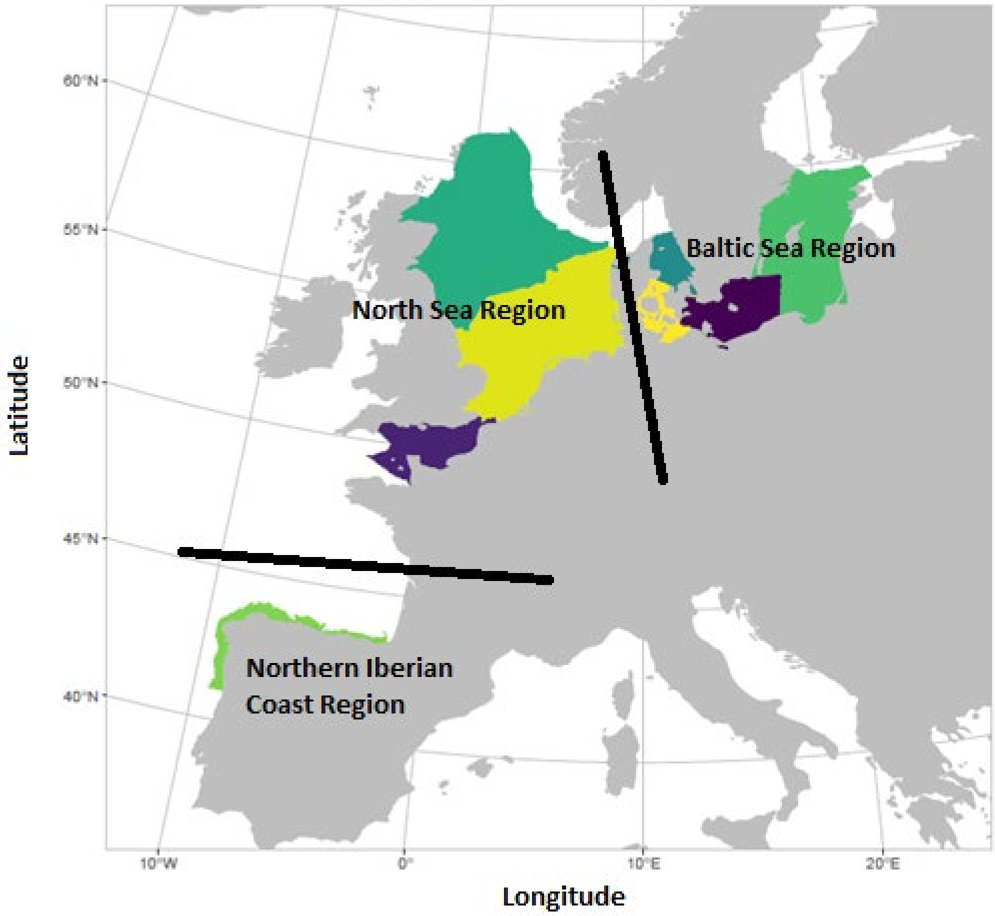

Figure 2.3.1. Spatial subdivisions (dark green, light green, olive, purple) within each of three regions (Table 2.3.1) with multiple risk-based approaches that are compared in WKBENTH3.

Table 2.3.1. Underlying pressure layers used to estimate impact per assessment method.

Region	Assessment Method	Pressure Layer
Northern Iberian Coast (no subdivision)	SoS, L1, BH3	VMS layers analysed were derived from the total aggregated VMS layers from 2016 to 2020, published by ICES (2021a).
	PD2	VMS layers analysed were derived from the total aggregated VMS layers from 2016 to 2020, published by ICES (2021a). A depletion rate for OT_Mix was used, which is the dominant fisheries in the area.
North Sea (3 subdivisions)	BH3	VMS layers analysed were derived from the total aggregated VMS layers from 2016 to 2020, published by ICES (2021a).
	L1, PD2	Métier-specific depletion rates using ICES VMS data averaged for 2013–2018, following ICES (2021b).
Baltic Sea	CumI	VMS layers and “other” physical pressures

Region	Assessment Method	Pressure Layer
(4 subdivisions)	L1, PD2	Métier-specific depletion rates using ICES VMS data averaged for 2013–2018, following ICES (2021b).

Table 2.3.2. Methods used for aggregating the different assessment approaches to a sensitivity and impact score per BHT and subdivision.

Assessment Method	Habitat Sensitivity Estimation	Risk-based Impact Estimation
BH3	Estimated as an average sensitivity score weighted by the habitat area with that score. Final value is averaged for subsurface and surface sensitivity (higher values are more sensitive habitats).	BH3 disturbance is calculated as the potential impacts of the type and intensity of pressures, based on the resilience and resistance of habitat and species. It covers all habitat types, including fine-scale biotopes and BHTs. At present only operational for bottom gear type of fishing activities and aggregate extraction. Other activities will be included. All the results are evaluated according to their levels of confidence based on the type and quality of the underlying datasets. For the purpose of WKBENTH3, data were compared by ranking habitats with the largest areas with high disturbance as most at risk, followed by moderate disturbance and low disturbance.
PD2	Estimated as the average median longevity value across grid cells (higher values are more sensitive habitats).	Estimated as the average decline in B/K (biomass divided by carrying capacity) across grid cells (lower values are habitats more at risk).
L1	Estimated as the average median longevity value across grid cells (higher values are more sensitive habitats).	Estimated as the average proportion of organisms that are disturbed by trawling during their lifespan (lower values are habitats more at risk).
L2	Estimated as the average critical trawling intensity across grid cells (the intensity at which the biomass proportion of long-lived taxa (longevity ≥ 10 yr) is reduced to 50% of the untrawled reference (higher values are less sensitive habitats).	Output is not included in comparison. Estimated as the decrease in median longevity in response to trawling. Median longevity is the longevity where 50% of the community biomass is above/below. The decrease is based on a statistical relationship between trawling intensity and benthic longevity from the North Sea.
SoS	Sensitivity is estimated from the pressure-state curve per habitat type. The pressure-state curve is compared with five theoretical sensitivity curves (i.e. five theoretical curves how state can change with pressure) and the most similar is selected. Highest score is most sensitive. Sensitivity scores are not used in the computation of the RBS using SoS and it was provided only for comparison purposes in the frame of this exercise.	Estimated as the average proportional decline of sentinel species by trawling across grid cells (lower values are habitats more at risk).
CumI	The area of each polygon in the dataset was multiplied by a score derived from the four sensitivity categories. This assumes a linear and equidistant scale for the sensitivity categories (which is not the case in reality but the simplest way to convert the categories into numbers). The sum of these products was divided by the total area in order to get an average value per habitat.	The area of each polygon in the dataset was multiplied by a score derived from the six impact categories. This assumes a linear and equidistant scale for the sensitivity categories (which is not the case in reality but the simplest way to convert the categories into numbers). The sum of these products was divided by the total area in order to get an average value per habitat.

Assessment Method	Habitat Sensitivity Estimation	Risk-based Impact Estimation
D_M'	N/A	Quality status estimation: Median values of D_M' are estimated at the level of national parts of BHTs within AUs, suggesting a representative (benthic community observation) monitoring and further combined to median values at the level of BHTs x AUs. Standardized quality coding of datasets based on number of samples and monitoring years involved indicates relative level of confidence of assessment results (in terms of high, good, sufficient, low or poor).

2.4 Comparing risk-based impact with empirical status assessment

Potentially, all empirical benthic indicators tested in gradient studies (Section 2.2) and considered suitable to assess the quality status with regards to the MSFD (and D6 in particular) can also be applied at larger scales to obtain a quality status assessment. Quality status assessments can be compared with impact assessments coming from risk-based assessments for the same region.

WKBENTH3 compared one empirical assessment, the 'Relative Margalef diversity' indicator (D_M'), with the ranked risk-based impacts for four subdivisions in the Greater North Sea. The D_M' status assessment required certain steps to calculate (e.g. model or inter- and/or extrapolate) assessment results to the level of BHTs and subdivision. The quality assessment using D_M' is aligned with the OSPAR spatial assessment units AU, which slightly deviate from the subdivisions as presented in Figure 2.3.1, but no large effects on the assessment results are expected from that.

To estimate the DM' status assessment in the Greater North Sea, it was observed that regardless of standardization efforts, independent of pressure levels, differences in assessment results (D_M values) were observed between countries (as an effect of slight differences in benthic community observation methodology and identification and recording of specimens). To compensate for (potential) differences between countries, country specific reference values for Margalef diversity (D_{ref}) were estimated at the level of BHTs within each AU for calculation of D_M'. An extensive set of benthic community data (27 890 samples) including grab and core data were available for the period 1998–2021 for the Greater North Sea region as collected via an official OSPAR data call, identified as suitable for D_M' assessments. From these, in the current WKBENTH3 assessments, data from the period 2016–2021 were used to present the current quality status in case at least 30 samples were available, otherwise data from 2009–2015 were used (the case for results for the English Channel and a few BHTs within Southern North Sea and Northern North Sea subdivisions). Only data from grabs and cores of approximately 0.1 m^2 were used except for half of the assessments for the Kattegat where small core data (0.0143 m^2) were used. The current methodology of assessing relative diversity at BHT x subdivision scale suggests a representative monitoring. Although in general data availability was large, it should be noticed that spatial distribution and/or coverage of environmental or pressure gradients or differences in management is possibly not always representative for entire assessment units.

Ranked assessment scores for each subdivision and BHT were estimated at the level of samples or combined to median values of D_M' at the level of BHT x AU x country, after which those results were combined to median values of D_M' at the level of BHT x AU (in case results from more countries are involved). Confidence in terms of sufficient number of data and potentially covering year-to-year variation (independent of quality related to pressure) was included in a standardized way. Confidence results were not taken into account in the ranked comparison with the risk-based approaches.

We note that the presented D_M′ results are intermediate/preliminary and do not necessarily represent the final OSPAR position or QSR result. All results, including assessment results for D_M′ (BH2b) and references to methodology (updated OSPAR CEMP protocol BH2) will be available from https://oap.ospar.org/en/ospar-assessments/quality-status-reports/qsr-2023/indicator-assessments/ in 2023. The methodology to calculate D_M′ at the level samples is available from the ICES WHBENTH2 report (ICES, 2022) and the related Technical Service (https://doi.org/10.17895/ices.advice.21070975) in particular.

3 Indicator performance

3.1 Sources of uncertainty in data inputs

The effects of trawling were examined in all areas over a gradient of commercial bottom trawling intensity. Such a comparative analysis can result in differences in community composition along the trawling gradient that seem to be related to fishing impact, while in fact these patterns may result from the fishery selecting areas with a particular community composition where they catch the most fish. However, not all unfished habitats are necessarily unsuitable for fishing (Dinmore *et al.,* 2003) and it has been suggested that fisheries often return to areas that are free from obstructions that could damage the gear (Holland and Sutinen 2000). Tillin *et al.* (2006) further postulated that trawl effects on benthic communities can have a much larger impact than is expected from small changes in environmental conditions. We therefore expect that most of the observed signal in the analyses of the fishing intensity gradient datasets is from trawling.

About half of the tested indicator methods were developed for a specific region and experts had to adjust their methodology to cover the other geographic areas. This created limitations and uncertainty in the testing of indicator performance. Several indicator methods were not used for some analyses due to time constrains and/or missing information on species sensitivity (Table 2.2.1). In future, the work on the gradient datasets could be extended to other areas and other indicators could also be tested.

Both the trawling and eutrophication gradients varied in pressure unit. This limited the comparability between studies. Most trawl studies used the swept-area ratio (SAR) as the unit to express trawling intensity. Yet, these ratios were estimated for different grid cell sizes and different periods over which the swept-areas were estimated. Any direct comparison of benthic impact at the same pressure value was thus avoided.

The risk-based comparison also had several underlying uncertainties (on top of the uncertainty of each individual method). First, sensitivity and impact scores were summarized per BHT. For each method there are several ways of aggregation. We used the mean score for all methods where an average could be calculated. This was done to standardize the output as much as possible. Median values and/or fractions of area below a certain (threshold) value will likely rearrange the ranks of some of the BHTs. Second, most models were run with different pressure layers (Table 2.3.1) and any variation in underlying pressure information may drive differences in impact estimates. The influence of this effect is unclear and needs to be tested further. For some subdivisions, the ranked results of impact showed a good match between assessment methods, indicating that the (potential) variation in pressure layer had a limited effect. For other areas, there was larger variation in the ranked scores. This may both be driven by assessment method disagreement as well as variation in underlying pressure layer. Furthermore, the comparison based on ranking benthic BHTs according to each risk assessment output provided a general view on the assessment made by each indicator but can mask local differences which can be potentially important for a management perspective in certain areas.

3.2 Geographic overlap

The gradient datasets covered a substantial range of regions, habitats, and depth zones (Annex 5). Nonetheless, most studies were taken from the Greater North Sea, from relatively shallow and sandy habitats. In future, the work on the gradient studies could be extended to other areas to expand both the range of habitats and the depth coverage. WKBENTH3 recommends that the gradient study dataset be developed further within the working group FBIT and used to test performance of newly developed indicators.

Risk-based comparison were done for the Greater North Sea, Baltic Sea, and Iberian Coast regions (Section 2.3). Since BH3 is available for all Atlantic regions, similar comparisons can be developed for other areas where other assessment methods are available. It is recommended to use similar pressure layers for these comparisons. In the Mediterranean Sea, several countries have now developed risk-based assessments using a variety of methodologies. These can also be compared against each other when they overlap in geographic region. WKBENTH3 recommends that the working group FBIT continue comparing sensitivity layers and risk-assessment outputs.

4 Indicator response to pressure gradients

We examined how the different indicators responded to the pressure gradients in the Common Dataset (Section 2.2 , Annex 5). For all trawling gradients, we calculated the mean response to trawling across all locations. Additionally, we visualized the response of each indicator in each area for both trawling, eutrophication, and pollution. The mean response to trawling was estimated by calculating the change in indicator values from low vs. high trawl disturbed stations at each location as described in Section 2.2.

Biomass and Abundance fluctuated the most of all indicators. This was especially observed at intermediate trawl intensities in NIC2, which had 3 orders of magnitude variation in Biomass and Abundance as a consequence of *Munida sarsi* aggregations which seemed to be (at least in part) an opportunistic response to trawling (Ortiz 2021). NIC2 biomass/abundance values were therefore log-transformed in the response visualization. We did not use the log-transformed values in the meta-analysis (a log response-ratio of log biomass/abundance is a difficult variable) but only selected the fished stations in NIC2 with the highest intensity (SAR >10, n= 13), avoiding the stations with very high biomass/abundance at intermediate pressure level. However, those high values were supported by multiple sampling events (Ortiz 2021) and inclusion of those locations in future analyses could provide more insight into the behaviour of the Biomass and Abundance indicators.

For AS1, high biomass stations associated to specific (and localized) biocenoses were present in the gradient study at intermediate fishing effort level. Given the relatively limited number of such sampling locations, this condition possibly generated inconsistent response along the gradient study as compared to the outcomes of more extensive analyses conducted by Riva (2022). We therefore analysed the meta-analysis without AS1 but verified indicator sensitivity to AS1.

4.1 Mean response to trawling

The meta-analysis output showed that most indicators had, on average, declined at the high trawl impact relative to the baseline (Figure 4.1.1). A significant effect of trawling is present when the 95% confidence intervals do not overlap with zero. This was the case for Biomass, Species Richness, Fraction long-lived, Median longevity, SoS, D_M′, Shannon Index and Inverse Simpson. M-AMBI and DKI had a weak overlap with zero.

The significant effect of trawling was most evident for the two indicators that targeted the “sensitive” fraction of the community, i.e. Fraction long-lived and SoS. These showed an average decline > 50%. Biomass also showed a large decline of approximately 40%. Inverse Simpson, D_M′, Median longevity, and Species Richness all showed an average decline of about 15–20%.

Abundance and the abundance-based pTDI were, on average, increased at the high trawl impact sites but those effects were insignificant. Potentially, the increase in abundance was due to the proliferation of small and opportunistic species.

Biomass and Abundance were most sensitive to the inclusion or exclusion of areas and the inclusion of AS1 will cause Biomass confidence intervals to overlap with zero. Most of the other indicators have less fluctuations in values and seem to have a built-in constraint in indicator values. This is clearly the case for the proportional indicators that only fluctuate between zero and one. Biomass (and Abundance) are therefore less useful in areas where data variability is high.

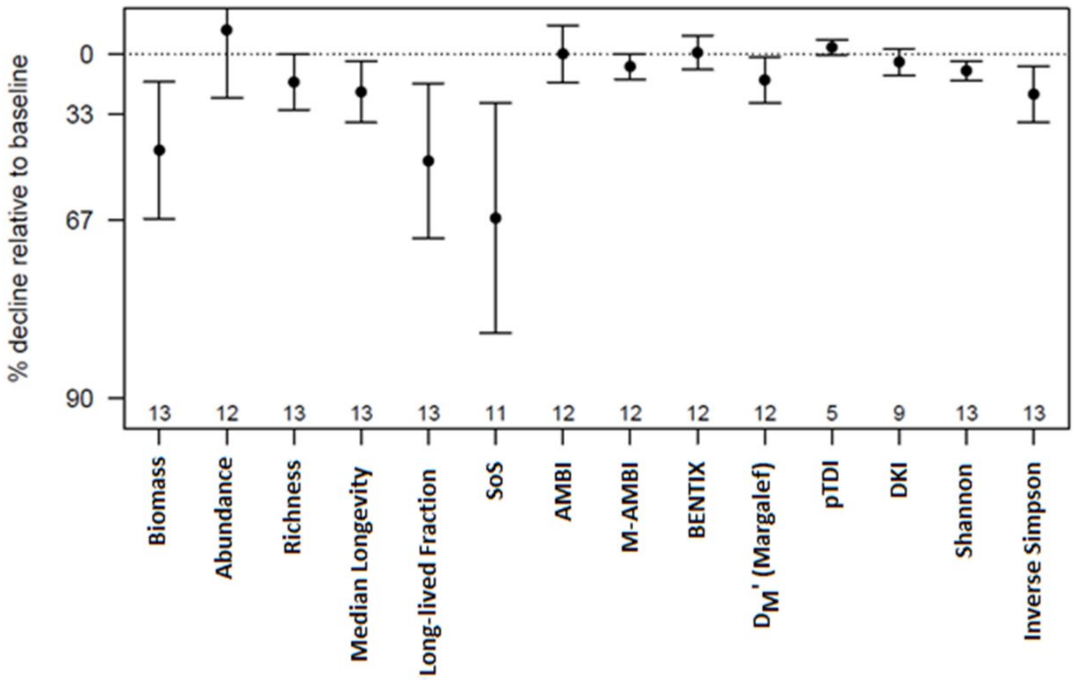

Figure 4.1.1 Mean response to trawling and 95% confidence intervals based on the log response-ratio (lnRR) for the indicators. If the confidence interval overlaps 0 the effect is not significant. *N* (=number of areas with indicator information) is given under each bar. The y-axis gives % changes for ease of interpretation. Model is estimated without AS1; the biomass and abundance confidence intervals become larger with the inclusion of AS1.

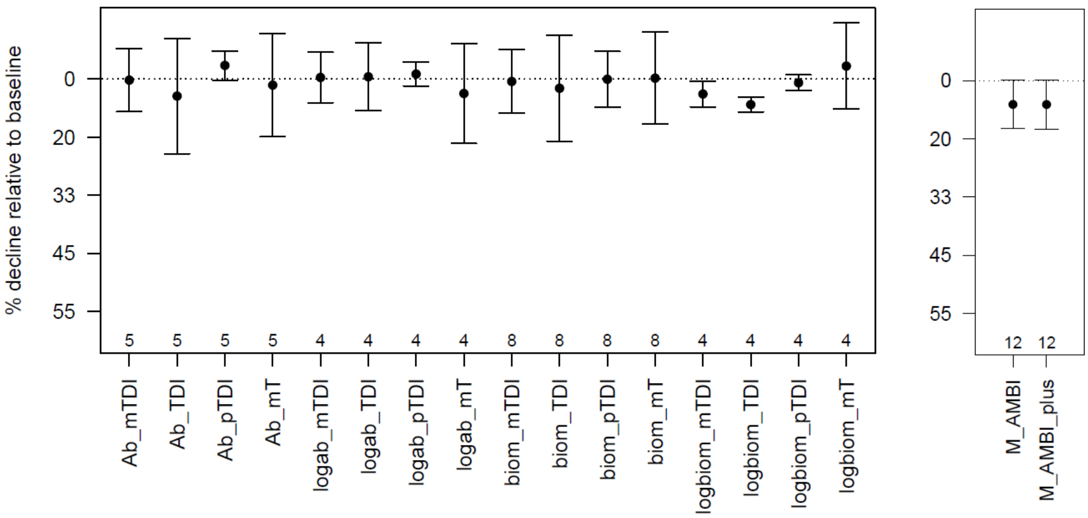

Figure 4.1.2. Mean response to trawling and 95% confidence intervals based on the log response-ratio (lnRR) for different mTDI, pTDI, TDI and mT (left panel) and M-AMBI (right panel) versions. If the confidence interval overlaps 0 the effect is not significant. *N* (=number of areas with indicator information) is given under each bar. The y-axis gives % changes for ease of interpretation. Model is estimated without AS1. Ab = abundance, logab = log(abundance), biom = biomass and logbiom = log(biomass).

The indicators mTDI, pTDI, TDI and mT may be computed on either abundance or biomass, each of which may be log-transformed or not and all outputs were provided. We therefore verified the response of those indicators against each other (Figure 4.1.2). The results showed that most of the indicator versions overlapped with zero. mTDI and TDI versions that are based on the log-transformed biomass values showed a significant negative relationship with trawling but note the small number of locations (n=4). M-AMBI also provided two versions that differed in how the reference conditions were estimated. Both versions showed the same mean response to trawling (Figure 4.1.2).

4.2 Bottom-trawl disturbance gradients

When analysing the trawling pressure gradients individually, we found that each individual indicator showed a decline in at least three of the areas (Figure 4.2.1). As observed in the meta-analysis, indicators that targeted a specific "sensitive" fraction of the community were declining most strongly in response to trawling in many of the datasets. Yet, in some cases, community biomass or BENTIX showed a stronger decline.

None of the indicators showed a decline in the OxyTrawl and Thames datasets. The Thames area is influenced by natural disturbance of tidal waves whereas the OxyTrawl area is influenced by low oxygen concentrations. In both cases, we hypothesize that these other environmental conditions have already selected for a community that is largely insensitive to trawling. This may explain why we find no signal. Other areas with relatively high natural disturbance are CO and DB and, for these areas, only two indicators indicated a decline along the pressure gradient.

As mentioned above, it remains unclear how representative are benthic data at the BHT level in relation to the pressure gradient for AS1, given the presence of some outliers that can be ascribed to localized biocenoses. Nevertheless, both SoS and the long-lived fraction showed a clear and negative decline with trawling (Figure 4.2.1). This suggests that there is a shift in the community composition. To what extent this is driven by trawling needs to be analysed further.

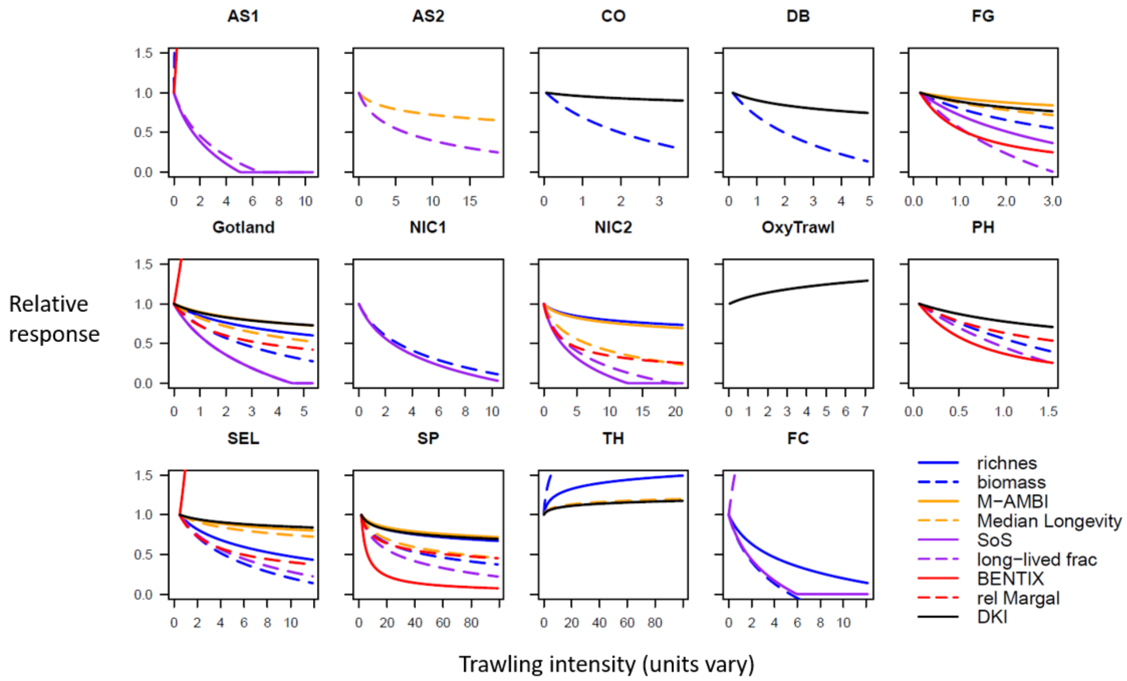

Figure 4.2.1. Trawl gradient response for each area. The visualization is done using linear regression where trawling intensity was log10(x+1)-transformed. Outputs are only shown when the AIC of the trawl model is lower than the null model (Indicator ~ 1). Note that not all indicators are available in all areas (see Table 2.1.1). Biomass is log10-transformed for NIC2.

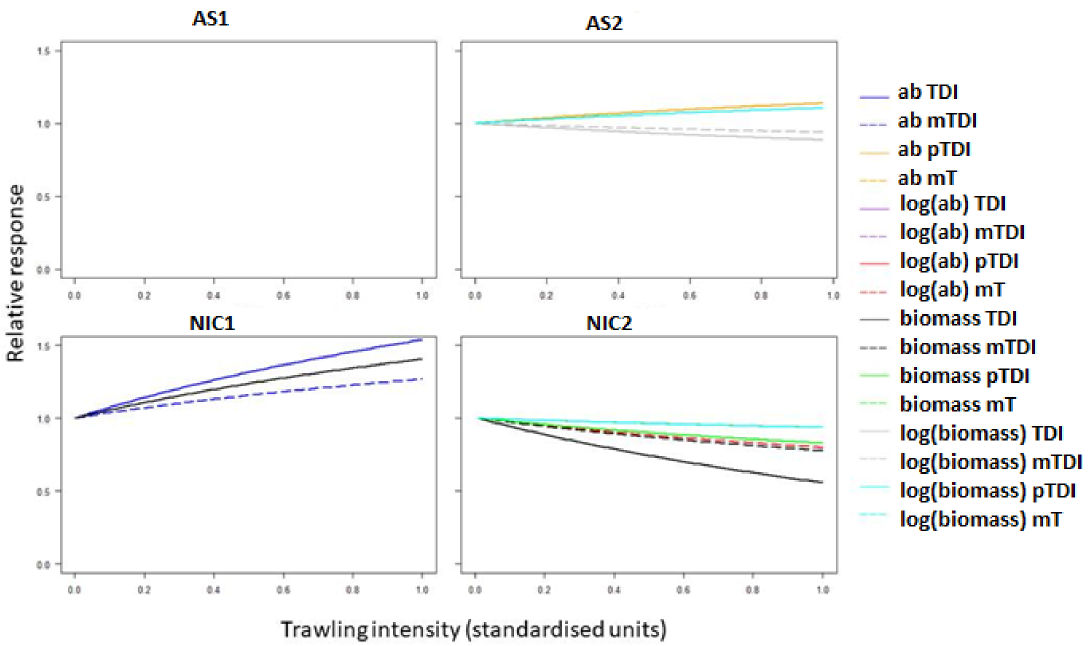

Figure 4.2.2. Trawl gradient response for each area for TDI-related indicators. The visualization was done using linear regression where trawling intensity was log10(x+1)-transformed. Outputs are only shown when the AIC of the trawl model is lower than the null model (Indicator ~ 1). Note that only four areas were sufficiently informed to carry out the gradient analyses. a=abundance; b=biomass.

Several locations showed a clear response to trawling for most of the indicators. If needed, these stations can be used to standardize the indicators against each other.

Several TDIs were fitted against the trawling pressure gradient in four different areas (Figure 4.2.2). As with most other indicator methods, none of the TDIs showed an effect with trawling in AS1. The TDIs that showed a decline in AS2 and NIC2 were indicators that were developed using biomass or log-transformed biomass.

4.3 Eutrophication and pollution gradients

When analysing the pollution and eutrophication gradients, the Relative Margalef index showed a consistent decline across all three gradients. BENTIX showed the strongest decline in 2 out of 3 datasets (Figure 4.3.1). M-AMBI showed a clear decline in the eutrophication datasets. SoS showed the strongest decline in relation to trawling and significantly declined in 2 out of 3 datasets for eutrophication and pollution. Biomass, median longevity, and long-lived fraction only showed a response in the Finland dataset (but note that these indicators were not estimated due to missing biomass data for the Vigo dataset).

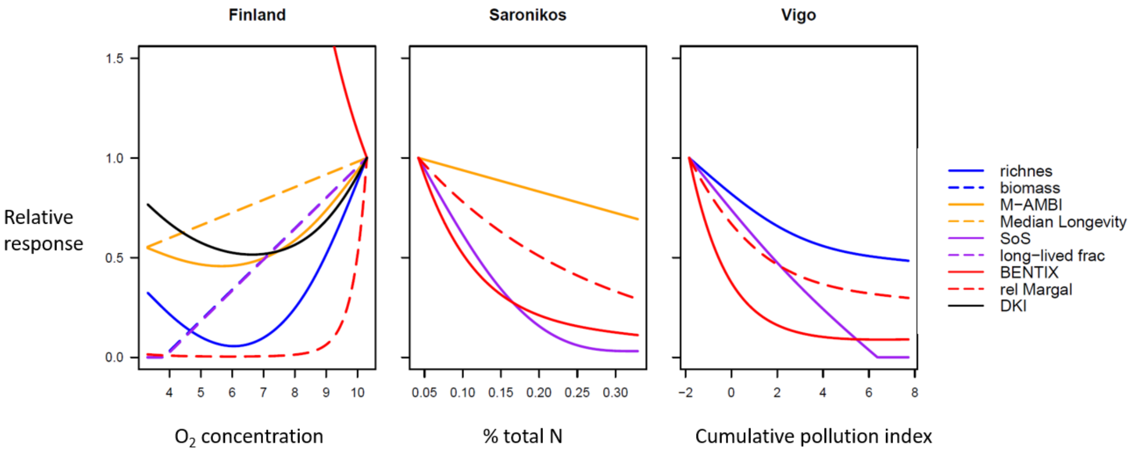

Figure 4.3.1. Eutrophication and pollution gradient response for each area. The visualization is done using generalized additive models. Outputs are only shown when the AIC of the model is lower than the null model (Indicator ~ 1). The increase in DKI and richness at low oxygen concentration in Finland is likely the result of visualizing (and overfitting) the result with a generalized additive model.

5 Complementarity of the selected indices

Different indicators may provide the same response whereas other may respond differently or to different pressures. This was analysed by comparing the correlation coefficients between indicators. The complementarities of the different indices are shown in Figure 5.1.1. We afterwards plotted the average Pearson correlation across all gradients, ordering indicators with Ward's hierarchical clustering. The results are shown in Figure 5.1.2.

All indicators were correlated to a certain extent and three major cluster groups were recognized, wherein the indicators seem to be more correlated than with indicators outside those clusters. The first group (central in the plot) included the indicators with a diversity component in them (e.g. M-AMBI) or based on diversity measurement (e.g. Shannon, Inverse Simpson). Within this group Abundance (log(x+1) and Biomass (log(x+1) were also included. A second group included the TDI-based indicators (which are trait based) and highly correlated. Within this group the pTDI, showed high correlation with SoS and some correlation with median longevity (Med.longevity) and the fraction of long-lived species (Long-lived frac.). This is understandable, as pTDI is focused on the most sensitive species fraction, as are those other indicators. The indicators focused on the most sensitive species fraction, the sentinel species (SoS) or long living species clearly formed a third cluster-group of indicators. The fourth group consisted of AMBI and BENTIX which are highly correlated and also showed a good correlation with the diversity- and TDI-based indicators, especially AMBI. AMBI and BENTHIX were less correlated with the indicators focusing on the most sensitive species.

These results are indicative, as the complementary of certain indicators differs sometimes between the different datasets (see Figure 5.1.1, for example), making the interpretation not straightforward. Also, not all indicators could be calculated on all datasets, so for certain correlations less data were available. Nevertheless, the robustness of those analyses was tested by rerunning the analyses with 1) all datasets and 2) on subsets of data (datasets where all indicators were determined for). The cluster-groups highlighted here were consistent across all analyses. From analytical and ecological points of view, the clustering is logical, as it is straightforward that indicators consisting of similar parameters tend to correlate better than indicators consisting of different parameters.

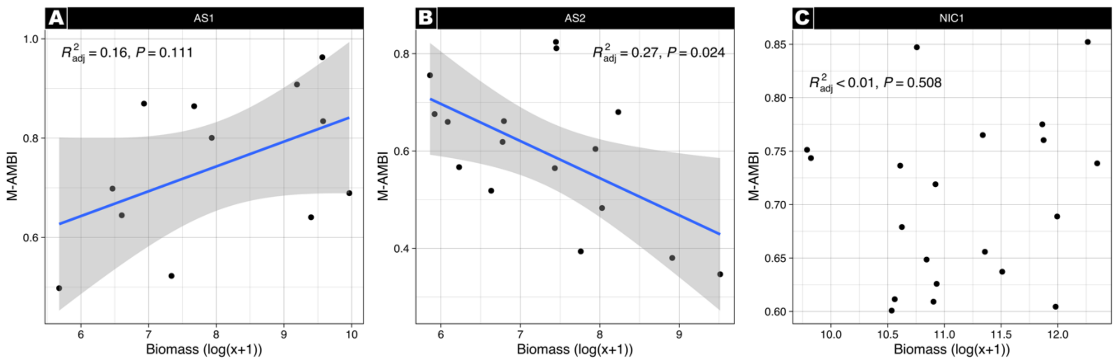

Figure 5.1.1. Relationship between biomass (log(x+1)- transformed) and M-AMBI values for three different gradient studies A) Adriatic Sea – Circalittoral sand, B) Adriatic Sea – Circalittoral mud, and C) Northern Iberian Coast - Offshore circalittoral sand.

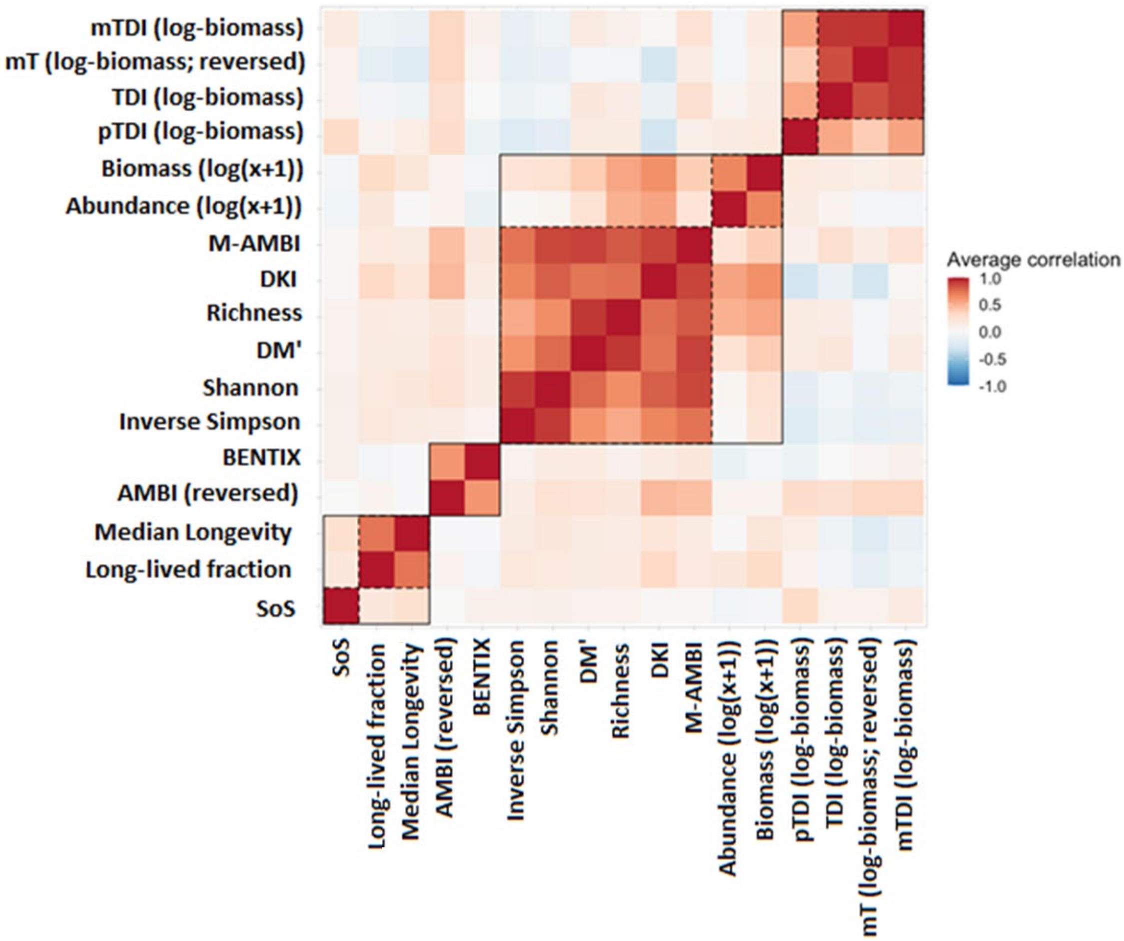

Figure 5.1.2. Hierarchical clustering of the different indicators based on the average Pearson correlation across datasets. For each pair of indicators, all gradients with sufficient data were included in the analysis. Thus, not all pairs of indicators are compared using the same datasets due to missing data.

Indicators have the purpose of detecting changes in the ecosystem in a rapid and transparent way. It is a warning tool for scientists and policy-makers, initiating further investigations or management measures in the cases where the indicator values deviate from the normal (e.g. cross a threshold). As there are consequences coupled to indicator outcomes, it is advisable to have a confident assessment. Therefore, it is commonly stated that an assessment of an ecosystem, in this case the broad habitat type (BHT), should rely on more than one indicator, covering different aspects of the habitat condition (Table 1.1.1). Another reason is that there does not exist an indicator detecting all ecosystem responses in one number, or that shows a similar response across (sub-)regions or is even sensitive in detecting changes to all kind of human pressures. This is also clearly reflected in the outcomes of these comparability analyses, where the responses of the indicators differ across datasets and are not even sensitive to certain human pressures (cf. fishery abrasion and pollution). Nevertheless, **there are clear indicator cluster-groups determined, illustrating and supporting the advice to rely on several indicators to assess benthic habitat condition, ideally picked from different cluster groups (or indicator families)**. Based on these comparability analyses, the WFD experience and MSFD needs, benthic habitat condition indicators should ideally have a component of diversity, species sensitivity and abundance (density and/or biomass). The indicator components can be combined in one algorithm (as most of the WFD indicators) or as different indicators. When combined in one algorithm, there is the probability that the indicator response to a pressure is imbalanced, compared with when they are analysed separately. This can arise through different indicators cancelling out the effects of other indicators. This was observed in the response to pressure gradients analyses (see Section 4).

The final selection of an appropriate set of indicators for the benthic habitat condition assessment in the area of interest depends on the needs of the policy (managing specific human pressures) or nature of the area (specific environmental conditions, e.g. shallow or deep sea areas; Baltic vs. Mediterranean). So, in the indicator set, there can be the need to rely on generic and specific indicators to accomplish the assessment. Whereas generic indicators are in play for detecting overall responses to raise an overall alarm bell, they can have difficulties detecting the cause of the response in relation to a specific human activity. Therefore, specific indicators are also required to follow up the responsiveness to a specific human activity or measurement measure (e.g. has the measure the desired effect on improved status?). Within the set of general and specific indicators, both type of approaches outlined in this report (status and risk-based indicators) can be catalogued under this and potentially applied hierarchically.

6 Risk-based comparisons

To facilitate understanding of relevant risk-based benthic indicator methods (Table 2.3.1) and their capacity for assessing seabed physical abrasion pressure, outputs were collated and compared across the Northeast Atlantic and Baltic Sea. Comparisons were made in subregions where indicator outputs were available at the time of assessment and of sufficient resolution to facilitate comparisons: North Iberian Atlantic, Northern/Central North Sea, Southern North Sea, Kattegat, Channel, Baltic Sea - Arkona Bornholm, Baltic Sea – Western Baltic and Baltic Sea – Baltic Proper (Section 2.3). The ranked scores show if the different risk-based approaches currently available in EU waters find the same type of seabed habitats most sensitive to bottom fishing and/or most at risk of adverse effects.

It should be noted that the risk-based indicators compared have been developed for unique applications, with varying designs, and therefore, measure distinct factors when assessing environmental change. For example, SoS (uses key traits to inform BESITO index), PD2/L1 (longevity) and BH3 (key traits from habitat characterizing species to inform resistance and resilience) derive sensitivity from different sources of information, each requiring unique data inputs. Therefore, observed outputs likely reflect variations in method design (e.g. use of longevity alone, or a range of traits, including longevity), in addition to data availability in each region; hence why certain habitats appear more, or less sensitive in the same regions between methods.

North Iberian Atlantic

In the Northern Iberian Atlantic, Sentinels of the Seabed (SoS), Population Dynamic (PD2), L1 and BH3 indicators were assessed. All indicators identified 'Upper bathyal sediment' as being the most sensitive habitat to physical abrasion pressure (Table 6.1). 'Offshore circalittoral mud and sand' were also identified as having consistent rankings of pressure assessments between indicators. However, differences were observed when assessing sensitivity for 'Offshore circalittoral coarse sediment', where SoS and BH3 each ranked sensitivities of 5, whereas PD2 and L1 indicated sensitivities of 4. Conversely, PD2 and L1 both indicated 'Circalittoral sand' as being the least sensitive habitat, whereas SoS and BH3 ranked sensitivity as 3 and 4, respectively (Table 6.1).

Table 6.1. Indicator sensitivity information from North Iberian Atlantic; ranked where 1 is the most and 5 being the least sensitive. The three habitats in SoS with rank 3 have the same sensitivity score.

BHT	Fraction of total area	SoS	PD2	L1	BH3
Upper bathyal sediment	0.06	1	1	1	1
Offshore circalittoral mud	0.17	3*	2	2	2
Offshore circalittoral sand	0.3	3*	3	3	3
Offshore circalittoral coarse sediment	0.05	5	4	4	5
Circalittoral sand	0.07	3*	5	5	4

All indicators identified 'Offshore circalittoral mud' as being the habitat most impacted by physical abrasion pressure and 'Circalittoral sand' as being the least impacted (Table 6.2), showing consistency between indicator outputs. When comparing outputs from 'Offshore circalittoral sand', 'Upper bathyal sediment' and 'Offshore circalittoral coarse sediment', SoS and BH3 indicated the same rank of impact, with 'Upper bathyal sediment' being the second-most impacted habitat. However, PD2 and L1 differed in their outputs, with 'Offshore circalittoral sand' being the second-most impacted habitat (Table 6.2).

Table 6.2. Indicator impact information from North Iberian Atlantic; ranked where 1 is the most impacted and 5 the least sensitive.

BHT	Fraction of total area	SoS	PD2	L1	BH3
Offshore circalittoral mud	0.17	1	1	1	1
Offshore circalittoral sand	0.3	3	2	2	3
Upper bathyal sediment	0.06	2	3	4	2
Offshore circalittoral coarse sediment	0.05	4	4	3	4
Circalittoral sand	0.07	5	5	5	5

Correlation plots between SoS, PD2 and L1 for the Iberian Coast

Scatterplots of Relative Benthic Status (RBS) scores of grid cells (0.05˚x0.05˚) were estimated for SoS, PD2 and L1 indicators by BHTs, and showed that the RBS estimated strongly correlated across the three indicators analysed for the North Iberian Atlantic region (Figures 6.3, 6.4). SoS and PD2 had the highest Spearman Rank correlation coefficient (r_{sp}= 0.98), followed by PD2 and L1 (r_{sp} =0.95) and SoS and L1 (r_{sp}=0.92). When the analysis was performed per BHT, a significant correlation was also found for most of the BHTs analysed. SoSvs.PD2 showed again the strongest correlation where all the BHT's coefficient values, except 'Circalittoral mud' (r_{sp} =0.8), were close to r_{sp}=1.

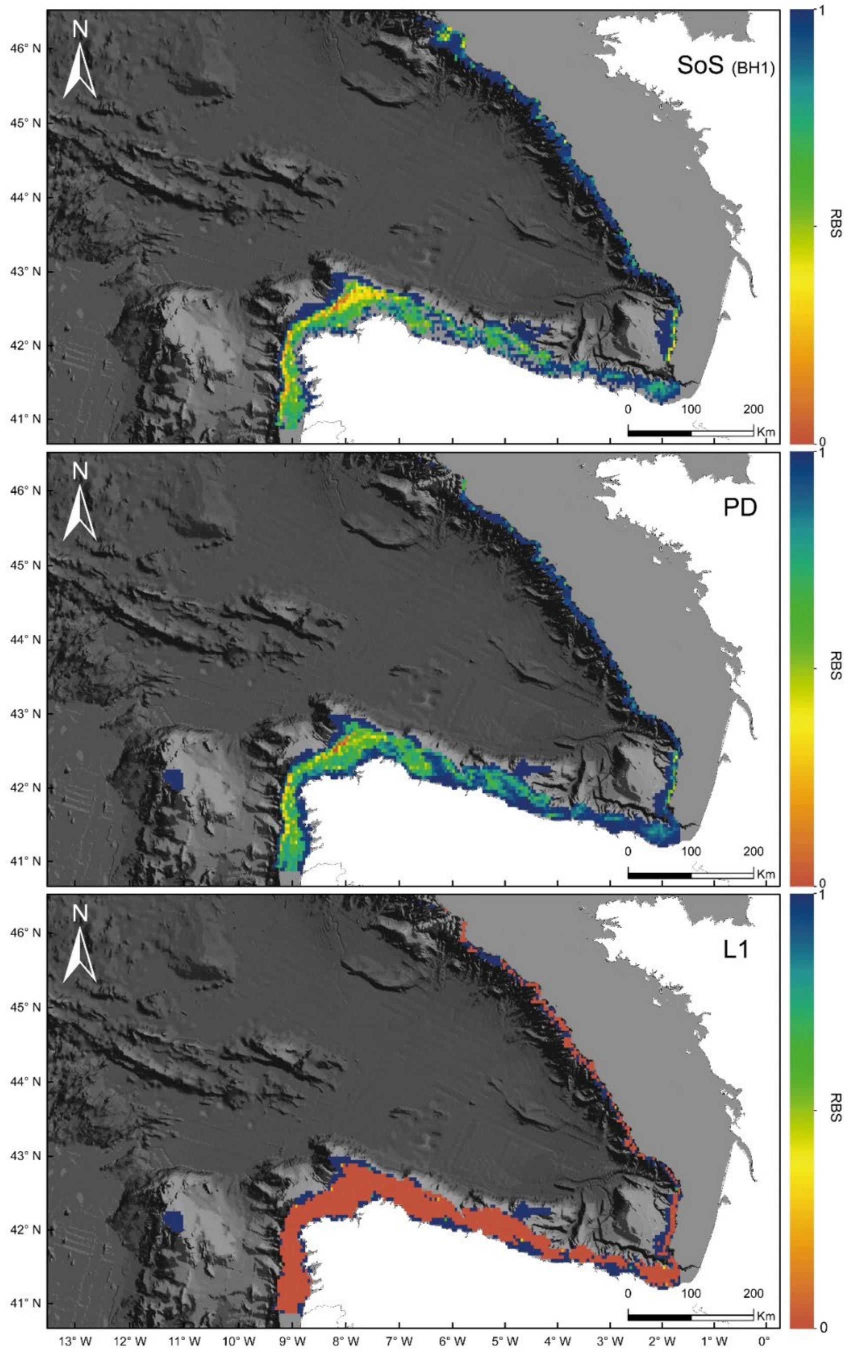

Figure 6.3. Relative Benthic Status (RBS) determined by SoS, PD2 and L1 indicators for the North Iberian Atlantic assessment unit.

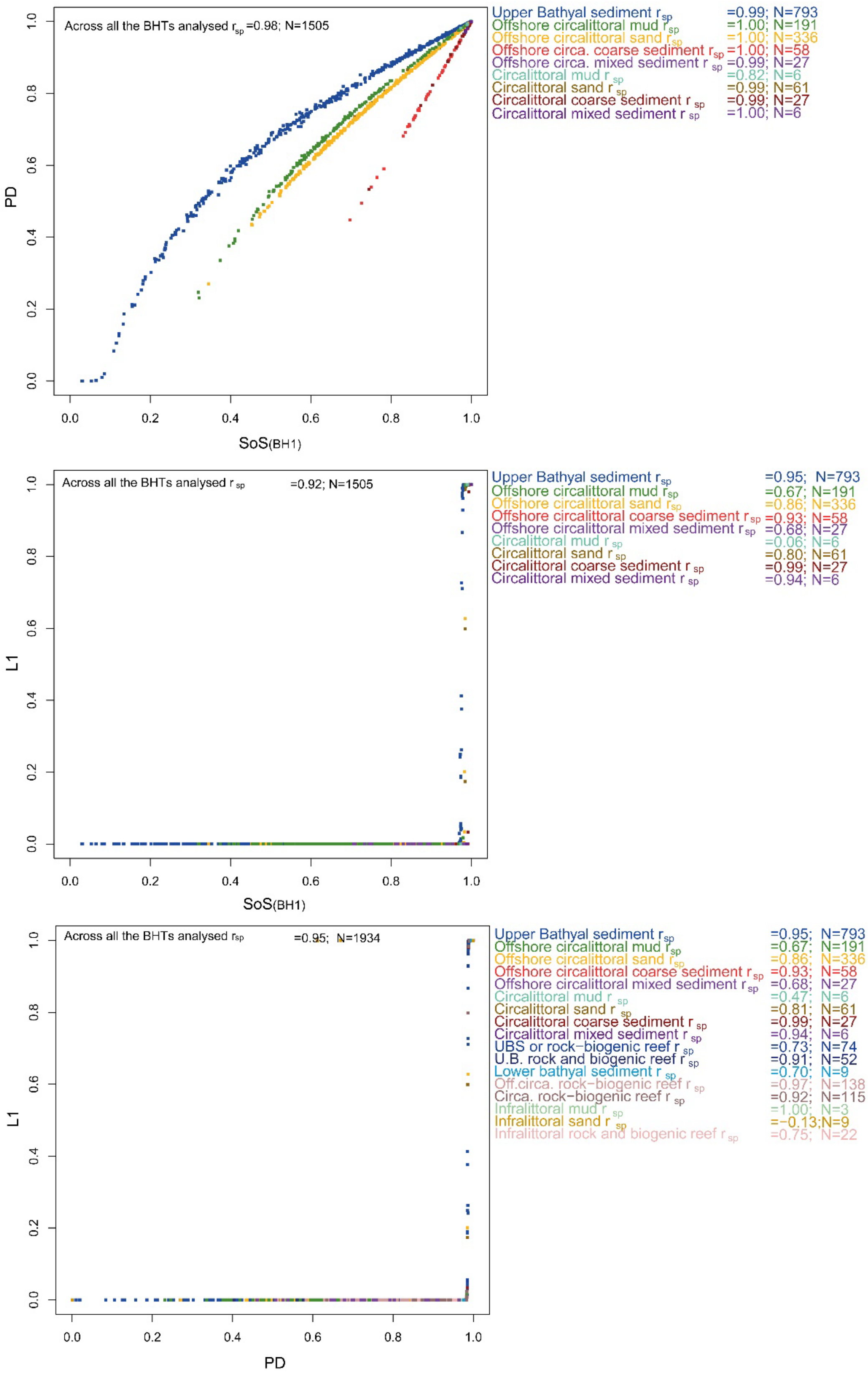

Figure 6.4. Scatterplots of Relative Benthic Status (RBS) scores of grid cells (0.05˚x0.05˚) estimated by methods SoS, PD2 and L1 by BHT, and the Spearman Rank correlation coefficients. Only the BHTs with more than five grid cells were plotted.

The high Spearman rank correlations between L1 and the two others indicators mask, however, highly non-linear relationships (Figure 6.4). L1 only varies within a narrow range of SoS and PD2 values; most of the L1 gradient is observed near RBS of 1 for SoS and PD2 and then L1 remains zero, even when the two other indicators suggest good RBS. This leads to a large portion of the Iberian Coast with a RBS of zero for L1 (Figures 6.3). This could suggest that L1 is particularly sensitive to early degradations, but is less sensitive to more profound impacts as well as to the recovery phases, compared with SoS and PD2.

Note that SoS scores were not calculated for areas where commercial bottom trawling is banned (EU Regulation 2016/2336; Real Decreto 502/2022) or for the areas where the fleet does not operate (rocky and biogenic reef habitats).

Northern/Central North Sea

In the Northern/Central North Sea, PD, L1, L2, BH3 and D_M′ indicators were assessed. Sensitivity outputs from PD and L1 were identical, with 'Offshore circalittoral sand' being the most sensitive and 'Offshore circalittoral mixed sediment' being the least sensitive habitats for these indicators, specifically. PD, L1 and L2 all identified 'Offshore circalittoral coarse sediment' and 'Circalittoral sand' as being the second and third-most sensitive habitats. However, L2 differed from PD and L1, with 'Offshore circalittoral mixed sediment' being the most sensitive and 'Offshore circalittoral sand' the fourth-most sensitive. BH3 sensitivity differed from all other indicators, with 'Offshore circalittoral mud' identified as the most and 'Offshore circalittoral coarse sediment' being the least sensitive habitats (Table 6.3).

Table 6.3. Indicator sensitivity information from the Northern/Central North Sea; ranked where 1 is the most and 5 being the least sensitive.

BHT	Fraction of total area	PD2	L1	L2	BH3
Offshore circalittoral sand	0.63	1	1	4	3
Offshore circalittoral coarse sediment	0.1	2	2	2	5
Circalittoral sand	0.01	3	3	3	2
Offshore circalittoral mud	0.24	4	4	5	1
Offshore circalittoral mixed sediment	0.01	5	5	1	4

PD, L1 and BH3 identified 'Offshore circalittoral mud' as being the habitat most impacted by physical abrasion pressure; D_M′ also indicated a similar impact, with a slightly lower ranking score of 2 (Table 6.4). BH3 and PD were broadly similar, identifying 'Offshore circalittoral sand' and 'Offshore circalittoral coarse sediment' as being the least impacted habitats. Impact rankings from L1 were similar to those of BH3 and PD. However, L1 identified 'Circalittoral sand' as the least impacted habitat, which was ranked for PD and BH3 as 3 and 2, respectively. D_M′ differed from other indicators, identifying 'Offshore circalittoral coarse sediment' as being the most impacted and 'Offshore circalittoral mixed sediment' the least impacted habitats. Impact rankings were not available for L2 at the time of assessment.

Table 6.4. Indicator impact information from the Northern/Central North Sea; ranked where 1 is the most impacted and 5 the least sensitive.

BHT	Fraction of total area	PD2	L1	BH3	D_M'
Offshore circalittoral mud	0.24	1	1	1	2
Offshore circalittoral mixed sediment	0.01	2	2	3	5
Circalittoral sand	0.01	3	5	2	4
Offshore circalittoral sand	0.63	4	3	4	3
Offshore circalittoral coarse sediment	0.1	5	4	5	1

Southern North Sea

In the Southern North Sea, PD, L1, L2, BH3 and D_M' indicators were assessed. Sensitivity outputs from PD, L1 and BH3 were broadly similar for some habitats, such as 'Circalittoral mixed sediment' and 'Offshore circalittoral mud', which were ranked as having the highest levels of sensitivity by these indicators. However, L2 indicated these habitats as being the least sensitive, contrasting to sensitivity outputs from PD, L1 and BH3 (Table 6.5).

Table 6.5. Indicator sensitivity information from the Southern North Sea; ranked where 1 is the most and 5 being the least sensitive.

BHT	Fraction of total area	PD2	L1	L2	BH3
Circalittoral mixed sediment	0.02	1	1	7	2
Offshore circalittoral mud	0.14	2	2	8	1
Circalittoral coarse sediment	0.08	3	3	3	6
Offshore circalittoral sand	0.33	5	5	4	5
Circalittoral sand	0.28	4	4	5	7
Infralittoral sand	0.03	6	6	6	4
Circalittoral mud	0.02	8	8	2	3
Offshore circalittoral coarse sediment	0.07	7	7	1	8

Outputs from PD, L1 and BH3 were broadly similar, with all three indicators identifying 'Offshore circalittoral mud' as being the most impacted habitat. In addition, 'Circalittoral Coarse Sediment' was found to be the least impacted habitat in PD and L1 outputs, and the second-least impacted for BH3. In contrast, the most and least impacted habitats identified by D_M' differed from those identified by PD, L1 and BH3, with 'Offshore circalittoral mud' being one of the least impacted habitats (ranking 7 of 8). D_M' identified 'Infralittoral sand' as being the most impacted habitat, which was ranked as the third-most impacted habitat by PD and L1, and the fourth-most impacted by BH3 (Table 6.6).

Table 6.6. Indicator impact information from the Southern North Sea; ranked where 1 is the most impacted and 5 the least sensitive.

BHT	Fraction of total area	PD2	L1	BH3	D_M'
Offshore circalittoral mud	0.14	1	1	1	7
Circalittoral mud	0.02	2	4	2	8
Infralittoral sand	0.03	3	3	4	1
Offshore circalittoral sand	0.33	4	2	5	5
Circalittoral mixed sediment	0.02	7	7	3	6
Circalittoral sand	0.28	6	5	6	2
Offshore circalittoral coarse sediment	0.07	5	6	8	4
Circalittoral coarse sediment	0.08	8	8	7	3

Kattegat

In the Kattegat, PD2, L1, L2, BH3 and D_M' indicators were assessed. PD2 and L1 showed the same sensitivity ranking with 'Offshore circalittoral mud' the most and 'Infralittoral mud' the least sensitive. L2 had broadly similar sensitivities with the four most sensitive habitats corresponding to PD and L1. The BH3 outputs strongly differ, with only 'Offshore circalittoral mud' having a higher sensitivity as in PD2, L1 and L2. Other habitats considered as sensitive in BH3 (Offshore circalittoral sand, Circalittoral mixed and coarse sediment) are ranked as the least sensitive in PD2 and L1.

Table 6.7. Indicator sensitivity information from the Kattegat; ranked where 1 is the most and 5 being the least sensitive.

BHT	Fraction of total area	PD2	L1	L2	BH3
Offshore circalittoral mud	0.35	1	1	4	2
Offshore circalittoral mixed sediment	0.06	2	2	3	7
Offshore circalittoral coarse sediment	0.03	3	3	2	8
Offshore circalittoral sand	0.13	4	4	1	6
Infralittoral mixed sediment	0.05	5	5	6	1
Infralittoral sand	0.21	6	6	5	4
Circalittoral mud	0.05	7	7	7	3
Infralittoral mud	0.05	8	8	8	5

PD2, L1 and BH3 all indicated the highest impact for 'Offshore circalittoral mud', and 'Circalittoral mud' was ranked second (PD2, BH3) or third (L1) highest impact (Table 6.8). Least impacted in all assessments were the 'Infralittoral' sediments, with the exception of 'Infralittoral mixed sediment'. This habitat was ranked with the highest sensitivity in BH3 and also showed a higher impact in BH3 compared to PD2 and L1. The results of the sample-based indicator D_M' are quite contrary to the risk-based assessments: 'Infralittoral sediments' have the highest impact rankings, whereas the offshore sediments and 'Circalittoral mud' have a better status than indicated with PD, L1 and BH3.

Table 6.8. Indicator impact information from the Kattegat; ranked where 1 is the most impacted and 5 the least sensitive.

BHT	Fraction of total area	PD2	L1	BH3	D_M'
Offshore circalittoral mud	0.35	1	1	1	7
Circalittoral mud	0.05	2	4	2	5
Offshore circalittoral coarse sediment	0.03	3	2	5	
Offshore circalittoral mixed sediment	0.06	4	3	6	6
Offshore circalittoral sand	0.13	5	5	4	8
Infralittoral mixed sediment	0.05	8	7	3	3
Infralittoral sand	0.21	7	6	8	4
Infralittoral mud	0.05	6	8	7	2

Channel

In the Channel, PD2, L1, L2 and BH3 indicators were assessed. 'Offshore circalittoral sand' and 'Circalittoral sand' were among the third most sensitive habitats in all indicators (Table 6.9). 'Infralittoral coarse sediment' was considered less sensitive. There were larger differences for 'Offshore circalittoral coarse sediment', that was ranked as the second most sensitive habitat in PD, L1 and L2, but the least sensitive in BH3. On the other hand, 'Infralittoral sand' was considered the most sensitive habitat in BH3, but the least sensitive in the other indicators.

Table 6.9. Indicator sensitivity information from the Channel; ranked where 1 is the most and 5 being the least sensitive.

BHT	Fraction of total area	PD2	L1	L2	BH3
Offshore circalittoral sand	0.05	1	1	3	2
Offshore circalittoral coarse sediment	0.51	2	2	2	7
Circalittoral sand	0.09	3	3	6	3
Offshore circalittoral mixed sediment	0.03	5	5	1	5
Circalittoral coarse sediment	0.19	4	4	4	4
Infralittoral sand	0.03	7	7	7	1
Infralittoral coarse sediment	0.03	6	6	5	6

Least impacted in all risk-based indicator assessments were 'Circalittoral and infralittoral coarse sediment'. 'Offshore circalittoral sand' was among the two most impacted habitats. Although the ranking was a little different between indicators, the outputs of risk-based methods were considered broadly similar. The sample-based indicator D_M' also ranked 'Circalittoral sand' as the habitat most impacted and 'Infralittoral coarse sediment' as least impacted. There are larger differences for other habitats such as 'Circalittoral sand' that is considered to have the second highest impact in D_M', but is ranked at position 4 in all risk-based indicators (Table 6.10).

Table 6.10. Indicator impact information from the Channel; ranked where 1 is the most impacted and 5 the least sensitive.

BHT	Fraction of total area	PD2	L1	BH3	D_M'
Offshore circalittoral sand	0.05	1	1	3	1
Offshore circalittoral mixed sediment	0.03	5	2	1	7
Offshore circalittoral coarse sediment	0.51	2	3	5	4
Infralittoral sand	0.03	3	5	2	3
Circalittoral sand	0.09	4	4	4	2
Circalittoral coarse sediment	0.19	7	6	6	5
Infralittoral coarse sediment	0.03	6	7	7	6

Baltic Sea

In the Baltic Sea three assessment areas (Arkona Bornholm, Western Baltic and Baltic Proper) were compared. Indicators calculated for this region were PD2, L1 and CumI.

Baltic Sea - Arkona Bornholm

Sensitivity rankings of PD2, L1 and CumI are broadly similar to 'Infralittoral sediments' regarded as more sensitive than 'Offshore circalittoral sediments' (Table 6.11). However, the CumI ranked 'Circalittoral mud and sand' as the most sensitive, while in PD2 and L1 these habitats had medium sensitivity ranks. 'Infralittoral mixed sediment' was the most sensitive in the PD2 and L1, but the least sensitive in the CumI.

Table 6.11. Indicator sensitivity information from the Baltic Sea - Arkona Bornholm; ranked where 1 is the most and 5 being the least sensitive.

BHT	Fraction of total area	PD2	L1	CumI
Infralittoral sand	0.22	2	2	4
Infralittoral mixed sediment	0.05	1	1	7
Infralittoral coarse sediment	0.05	3	3	3
Circalittoral sand	0.15	4	4	2
Circalittoral mud	0.03	5	5	1
Circalittoral mixed sediment	0.17	6	6	5

BHT	Fraction of total area	PD2	L1	CumI
Offshore circalittoral mud	0.16	7	7	6
Offshore circalittoral mixed sediment	0.06	8	8	7

‘Offshore circalittoral mixed sediment’ and ‘Offshore circalittoral mud’ were among the habitats that are most impacted (Table 6.12), while ‘Infralittoral mixed sediment’ was less impacted in all indicators. For other habitat types, there were larger differences. In some cases, PD2 and L1 showed similar results, e.g. for ‘Circalittoral mixed sediment’, but also L1 and CumI correspond well, e.g. for ‘Infralittoral coarse sediment’.

Table 6.12. Indicator impact information from the Baltic Sea - Arkona Bornholm; ranked where 1 is the most impacted and 5 the least sensitive.

BHT	Fraction of total area	PD2	L1	CumI
Offshore circalittoral mixed sediment	0.06	1	3	2
Offshore circalittoral mud	0.16	3	1	4
Circalittoral mud	0.03	6	2	3
Circalittoral mixed sediment	0.17	8	6	1
Infralittoral coarse sediment	0.05	2	7	7
Circalittoral sand	0.15	7	4	5
Infralittoral sand	0.22	4	5	8
Infralittoral mixed sediment	0.05	5	8	6

Baltic Sea – Western Baltic

In the Western Baltic, ‘Offshore circalittoral mud’ was one of the least sensitive habitats over all indicators and ‘Circalittoral mud’ was among the most sensitive habitats. Otherwise, the ranking was quite different, with the more sensitive habitats in PD2 and L1 considered less sensitive with CumI and vice versa (Table 6.13).

Table 6.13. Indicator sensitivity information from the Baltic Sea – Western Baltic; ranked where 1 is the most and 5 being the least sensitive.

BHT	Fraction of total area	PD2	L1	CumI
Circalittoral mud	0.1	3	3	1
Circalittoral sand	0.11	2	2	4
Circalittoral mixed sediment	0.03	1	1	7
Infralittoral sand	0.45	5	5	3
Infralittoral mixed sediment	0.18	4	4	6

BHT	Fraction of total area	PD2	L1	CumI
Infralittoral mud	0.06	6	6	5
Infralittoral coarse sediment	0.04	8	8	2
Offshore circalittoral mud	0.01	7	7	8

With regard to the impact, there is a good agreement between L1 and CumI, except for one habitat type ('Infralittoral mixed sediment'). PD2 showed different results, with 'Infralittoral' habitats assessed as more impacted compared to 'Circalittoral sediments'. In L1 and PD2, 'Circalittoral sediments' as well as 'Offshore circalittoral mud' were more impacted (Table 6.14).

Table 6.14. Indicator impact information from the Baltic Sea – Western Baltic; ranked where 1 is the most impacted and 5 the least sensitive.

BHT	Fraction of total area	PD2	L1	CumI
Offshore circalittoral mud	0.01	4	1	2
Circalittoral mud	0.1	7	2	1
Circalittoral sand	0.11	6	3	3
Infralittoral coarse sediment	0.04	2	5	6
Infralittoral sand	0.45	1	6	8
Infralittoral mixed sediment	0.18	3	8	4
Circalittoral mixed sediment	0.03	8	4	4
Infralittoral mud	0.06	5	7	7

Baltic Sea – Baltic Proper

In the Baltic Proper, the offshore habitats were assessed as less sensitive than infra- and circalittoral habitats (Table 6.15). The only infralittoral habitat type was considered the most sensitive in PD2 and L1, but had a medium rank in CumI. In contrast, 'Circalittoral mud' had the highest sensitivity in the CumI, but a much lesser rank in PD2 and L1.

Table 6.15. Indicator sensitivity information from the Baltic Sea – Baltic Proper; ranked where 1 is the most and 5 being the least sensitive.

BHT	Fraction of total area	PD2	L1	CumI
Infralittoral mixed sediment	0.04	1	1	4
Circalittoral sand	0.08	2	2	2
Circalittoral mud	0.05	5	5	1
Circalittoral mud or Circalittoral sand	0.11	4	4	3
Circalittoral mixed sediment	0.2	3	3	8

BHT	Fraction of total area	PD2	L1	CumI
Offshore circalittoral mixed sediment	0.11	6	6	7
Offshore circalittoral mud or Offshore circalittoral sand	0.21	7	7	5
Offshore circalittoral mud	0.06	8	8	5

Impact outputs varied greatly among indicators (Table 6.16). 'Offshore circalittoral mud' was assessed as more impacted by all indicators, but otherwise there were not many similarities between all three indicators. For some habitat types the ranking of PD2 and L1 matched better, but for others there was a larger agreement between L1 and CumI. There was a particularly large discrepancy between the ranking of PD2 and CumI for the most and the least sensitive habitat.

Table 6.16. Indicator impact information from the Baltic Sea – Baltic Proper; ranked where 1 is the most impacted and 5 the least sensitive.

BHT	Fraction of total area	PD2	L1	CumI
Offshore circalittoral mud	0.06	3	1	2
Circalittoral mud	0.05	2	2	6
Offshore circalittoral mixed sediment	0.11	4	3	4
Infralittoral mixed sediment	0.04	1	6	8
Circalittoral mixed sediment	0.2	5	7	3
Offshore circalittoral mud or Offshore circalittoral sand	0.21	8	8	1
Circalittoral mud or Circalittoral sand	0.11	7	5	5
Circalittoral sand	0.08	6	4	7

Data Constraints

The risk indicators aim to provide solutions to meet the political requirements where the empirical indicators are not available. However, their caveats, limitations, and knowledge gaps should be considered. These indicators tend to be area-, habitat-, pressure-, and method-specific and, in most cases, must be fed with empirical data. In some regions, time-series of International Bottom Trawl Survey (IBTS) monitoring data provide information on the abundances of invertebrates, however, in many regions similar data are not available and knowledge gaps creating uneven geographical coverage affect the vast majority of risk-based indicators. The lack of spatial monitoring coverage is both an inter-regional and an intra-regional problem, making it necessary to improve sampling efforts in specific habitats where IBTS or other monitoring coverage is low. Therefore, the selection of indicators should consider various levels of data availability (e.g. some indicators will perform better in cases of data scarcity); data availability is critical to understanding how indicators can be operationalized.

All risk-based indicators have an ecological basis, despite having different approaches concerning source biological data, sensitivity, habitat data, and activity/pressure/impact inputs. These differences result in the strengths and suitability of each indicator for national and regional specificities. The outputs from different indicators regarding the ranking of the most affected habitats do not always agree but, in most cases, are similar (being different, data-driven and data-dependent). However, it would be essential to know the intensity and degree of affiliation for each BHT since this determination could make a significant difference in the outputs. Therefore, it is necessary to complement the ranking analyses with correlations between the indicators' results for each BHT in each region. Indicators should be seen as complementary and, in some cases, nested at different scales or by types of methods (risk assessments calibrated or ground-truthed with sample-based indicators), assessing different aspects of benthic ecosystems. Therefore, a combination of individual indicators is needed to assess benthic habitats under D6.

Comparison of Methods and Outputs

Northeast Atlantic (North Iberian Atlantic, Northern/Central North Sea, Southern North Sea, Kattegat, Channel)

Comparisons between risk-based indicators highlighted a variety of similarities and variations between methods; please note, SoS was not trialled in the Kattegat, Channel and North Sea regions, and D_M' was not applied in the North Iberian Atlantic. In areas such as the North Iberian Atlantic, all assessed approaches (SoS, PD2, L1 and BH3) identified 'Upper bathyal sediment' as being the most sensitive habitat to physical abrasion pressure. Conversely, in areas such as the Kattegat, and Northern and Central North Sea, sensitivity was most similar between PD2, L1 and L2, which derive sensitivity from the longevity trait alone at a BHT-level, differing from approaches that consider a range of biological traits (including longevity) and define sensitivity from the broad- to biotope/community-level (e.g. EUNIS Level 3 to Level 6), such as BH3. PD, L1 and BH3 all produced broadly similar sensitivity results in the Southern North Sea, contrasting to outputs from L2, indicating that outputs were likely data driven, such as the resolution of underlying habitat information to inform the sensitivity and extent of assessed habitats.

Although sensitivities varied between indicators, impact outputs broadly aligned; it should be noted that L2 was not analysed for impacts in this study. In the North Iberian Atlantic, SoS, PD2, L1 and BH3 all identified 'Offshore circalittoral mud' as being the habitat most impacted by physical abrasion pressure and 'Circalittoral sand' as being the least impacted. In addition, in the Northern and Central North Sea, PD2, L1 and BH3 identified 'Offshore circalittoral mud' as being the habitat most impacted habitat. Results in the Southern North Sea and Kattegat were also similar, where PD2, L1 and BH3 all identified the same habitats as being the most impacted; D_M' also indicated a similar impact, with a slightly lower ranking score of 2. In the Channel the ranking of most impacted varied between indicators, although patterns were broadly similar between methods; all indicators identified 'Circalittoral and infralittoral coarse sediment' as being the least impacted habitats. Key differences were observed in the Channel, where habitats such as circalittoral sand that was considered to have the second highest impact in D_M' was ranked at position 4 in PD, L1 and BH3.

Observed variances in comparisons were likely the result of key methodological differences between indicators, such as the scale on which sensitivity was calculated and the traits considered (single-trait vs. multi-trait assessments). In addition, how pressure was assessed also varied between methods; for example, PD2 considered gear-specific depletion from select types of fishing, whereas SoS and BH3 analysed all bottom-contacting gear as the total aggregated fishing layers from ICES. Furthermore, the availability and resolution of the data required to apply each indicator likely varied between assessment areas. Therefore, it should be recognized that these risk-based indicators assessed and measured different components with differing approaches, each having application-specific advantages with varied capacities for use in specific contexts. No 'one-size-fits-all' approach could be determined from the comparisons made between outputs; how each method is applied should be informed by the nature of the application, the biogeographic context and the level and resolution of available data required. Multiple indicators should be considered for use in a complementary fashion, where feasible, to maximize scientific integrity and accuracy when assessing environmental change in response to seabed physical pressure.

The outputs from different indicators and their associated ranking, for the purpose of comparison in WKBENTH3 did not always agree but, in most cases, were similar (being different by indicator design, data-driven and data-dependent). Examining the correlation between indicator values for each BHT may provide insights as to whether the ranks reflect the actual indicator values.

Baltic Sea (Baltic Proper, Arkona Bornholm and Western Baltic)

In the Baltic Sea region, we observed less consistency in risk-based outcomes compared to most subdivisions of the Northeast Atlantic. In the Northeast Atlantic subdivisions, sensitivities typically varied between indicators, but impact outputs broadly aligned indicating clear fishing pressure differences between BHTs. The BHT 'Offshore circalittoral mud' was heavily fished in all subdivisions and this drove its high impact score across assessment methods. Most of the Baltic Sea is fished at much lower intensity and with smaller differences between BHTs (ICES 2021b).

PD2 and L1 share the same sensitivity layer and were estimated with the same fishing pressure layer. Nonetheless, impact scores were not much better aligned between PD2 and L1 as compared with CumI. The CumI impact scores were not only based on bottom trawling but on all physical pressures considered and this may drive some additional variation in outcome.

7 Establishing thresholds

WKBENTH2 identified 11 approaches that have been proposed to set thresholds for good status (or to avoid an adverse effect or degraded state) in environmental management. Most approaches were seen as suitable for setting quality thresholds. Preferred methods identified an ecologically-motivated difference between a good and degraded state, rather than another transition. Methods that defined habitat- (and region-) specific thresholds, estimated with their uncertainty, were also preferred. This means that a single approach for choosing thresholds that yields different thresholds for different habitats/regions is desirable.

7.1 Indicators as ecological quality ratio

The Ecological Quality Ratio (EQR) is used within the Water Framework Directive (WFD) to obtain standardized outputs through an intercalibration process (WKBENTH2 Section 2). The intercalibration process in the WFD was aimed at ensuring comparability of the classification results of the WFD assessment methods developed by the Member States for the biological quality elements. The essence of intercalibration is to ensure that the high-good and the good-moderate boundaries in all Member States' assessment methods correspond to comparable levels of ecosystem alteration.

All indicators evaluated were expressed as an EQR, with values ranging between 0 and 1. The results showed that some indicators show a much stronger decline along the pressure gradient as compared to other indicators. One option, as done in the WFD, is to intercalibrate all indicator methods against each other and obtain an overarching EQR. The WFD intercalibration was rather complex and technical (see further WKBENTH2 Section 2). An additional downside of intercalibration is that the "overarching" EQR has lost biological meaning, making it difficult to set a scientifically justified quality threshold. Alternatively, and as discussed in the next section, the setting of quality thresholds can be indicator specific.

7.2 Way forward with setting quality thresholds

Indicators represent different aspects of a benthic community and without intercalibration each indicator will need a different quality threshold. Each quality threshold can be estimated using many different approaches, each with their own strengths and weaknesses as identified in WKBENTH2.

Establishing scientifically-justified quality thresholds between good and degraded for the indicators discussed in this report requires the application of the most-promising approaches for setting thresholds to each indicator. This will require the collation (and possibly even collection) of datasets and the estimation of the relationships for estimating the thresholds. One suggested way forward is to prioritize the most promising approaches for setting thresholds from WKBENTH2, and the most promising indicators from WKBENTH3, and attempt to estimate thresholds for each indicator using those approaches. This is a large amount of work for all threshold-indicator combinations, and a task beyond what could be achieved in WKBENTH3. WKBENTH2 explored operationalizing one threshold-indicator combination: the natural variability threshold in relation to community biomass and the PD2 assessment method. In WKBENTH3, we explored one threshold approach for two indicators (Section 7.3). Beyond these worked examples, WKBENTH3 was unable to estimate quality thresholds.

Once several threshold-indicator combinations have been established, it is likely that quality threshold values for each indicator do not correspond to comparable levels of pressure. Rather than intercalibrating the threshold values, WKBENTH3 suggests estimating the probability that the state is above the threshold for a combined set of indicators, which are all deemed suitable to assess aspects of D6. A worked example to show how GES could be approached as a probability is presented in Section 7.4.

7.3 Worked example: the 'Detectable Change' approach for setting quality thresholds

A statistically-detectable approach in a gradient dataset is the point where the upper 95% confidence interval crosses the level of the lower 95% confidence interval at the lowest level of pressure. Graphically this is represented by the point where a horizontal line extending the lower confidence interval at the lowest pressure crosses the upper confidence interval. We applied this approach for whole community biomass using the bottom-trawling gradient datasets (Annex 5) and we used Relative Margelef diversity index, D_M' for the pollution and eutrophication gradients.

Whole Community Biomass and Bottom Trawling

We applied the approach to three datasets (SEL, Gotland and SP), which were all on muddy sand (the most prevalent sediment type among the trawl gradient datasets). The GES thresholds defined using this approach varied between 0.76 and 0.72 times the biomass at the lowest level of trawling (Figure 7.3.1; any proper implementations of the approach should estimate the threshold at zero trawling rather than at the lowest level of trawling in the dataset). This approach to setting a GES threshold has some parallels to the 'natural variation' approach, as the confidence intervals at no trawling would represent the natural variation in biomass in situations where many stations with no trawling would have been sampled (which is not necessarily the case in the datasets presented here).

The approach sets a threshold as a pressure threshold rather than as a state threshold. A level of trawling pressure that corresponded to a detectable change could only be estimated for two of the datasets (Gotland and SEL), and this level of intensity was 2.76 and 3.17 y^{-1} respectively (both quantified as the swept-area-ratio although using different cell sizes). For the SP dataset, the confidence intervals are very wide, and the upper confidence interval never crosses the threshold, and therefore there is no level of fishing pressure at which a detectable change exists (note the different trawling pressure units).

We used these pressure thresholds to evaluate what fraction of C-squares in the North and Baltic Sea are above the pressure threshold and would therefore be considered to be in a 'good' state. Because we estimated the pressure threshold for muddy sand, we limited this analysis to the BHTs that match this sediment (Infralittoral sand, Circalittoral sand, Offshore circalittoral sand, Circalittoral mud or Circalittoral sand, Offshore circalittoral mud or Offshore circalittoral sand, Infralittoral mud or Infralittoral sand).

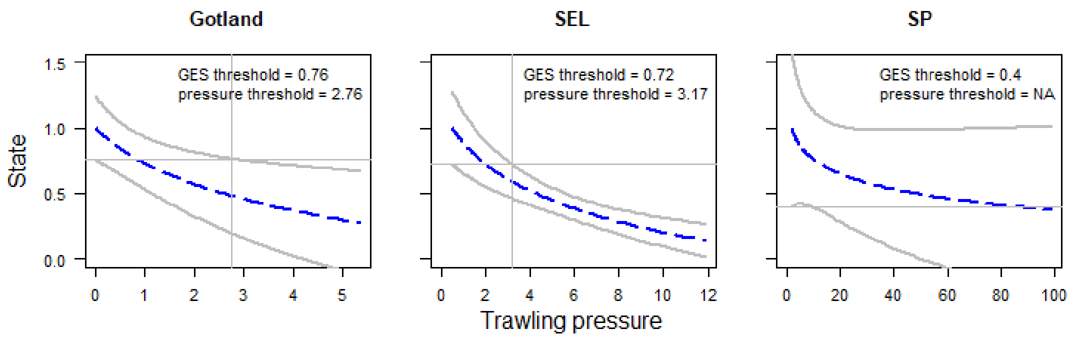

Figure 7.3.1. Illustration of setting threshold based on detectable change for biomass from gradient studies on muddy sand. The tick blue line is the mean response, thick grey line 95% confidence interval. The thin grey lines indicate the GES threshold and the matching pressure thresholds.

For the North Sea, the fraction of the area that has a fishing intensity above the threshold for good status is around 0.11 to 0.17 depending on the threshold used and the BHT (Figure 7.3.2 and Table 7.3.1). For the Baltic, most BHT had only a very minor fraction of the area with a fishing intensity above the threshold, except for Offshore circalittoral sand, where 14% of the area had a fishing intensity above the threshold (Figure 7.3.2 and Table 7.3.1).

Table 7.3.1. Fraction of the area below the pressure thresholds for the North Sea and the Baltic Sea by BHT.

North Sea		
	TV1 = 2.76 y-1	TV2 = 3.17 y-1
Circalittoral sand	0.17	0.14
Infralittoral sand	0.13	0.12
Offshore circalittoral sand	0.17	0.15
Baltic Sea		
	TV1 = 2.76 y^{-1}	TV2 = 3.17 y^{-1}
Circalittoral sand	0.03	0.03
Infralittoral sand	0.02	0.02
Offshore circalittoral sand	0.14	0.14
Circalittoral mud or Circalittoral sand	0	0
Infralittoral mud or Infralittoral sand	0	0
Offshore circalittoral mud or Offshore circalittoral sand	0.001	0.001

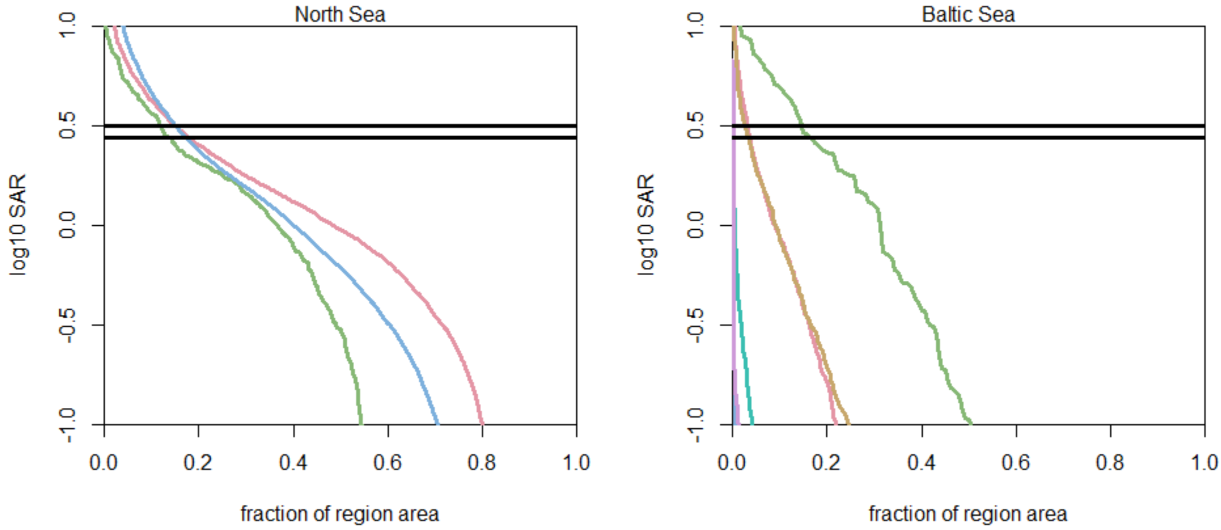

Figure 7.3.2. The fraction of the area that has a fishing intensity (log SAR) above the threshold for good status (heavy black lines) for the North Sea and Baltic Sea. *Y*-axis limited to display the data but does not display the full range of SAR values and therefore not the full fraction of the area.

Relative Margalef Diversity Index and Pollution/Eutrophication

We also applied this approach for the relative Margalef diversity index, D_M', using pollution and eutrophication gradient datasets from Finland, Saronikos and Vigo. The GES thresholds defined using this approach varied between 0.57 and 0.09 times the index at the lowest level of pressure (Figure 7.3.3, noting that low oxygen is high pressure).

A level of pressure that corresponded to a detectable change could be estimated for two out of three datasets. For the Finland dataset, GES can be assumed to have been achieved at an oxygen concentration > 8.25. For Saronikos, no pressure level matching GES could be defined using this approach. For Vigo, GES would be achieved at a cumulative pollution index of <0.57.

No regional maps of these pressures were available to the WKBENTH3, so no regional assessment using these thresholds could be presented here.

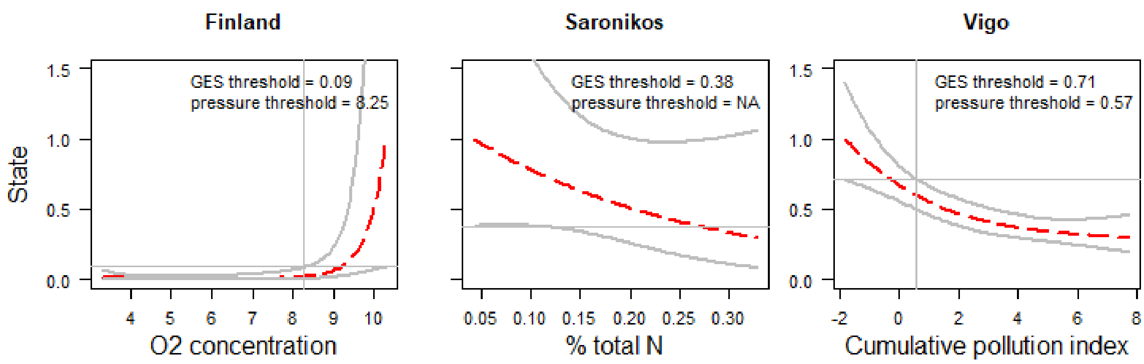

Figure 7.3.3. Illustration of setting threshold based on detectable change for the Relative Margalef diversity (D_M') from pollution and eutrophication gradient studies. The red dashed line is the mean response, and the thick grey lines the 95% confidence interval. The thin grey lines indicate the GES threshold and the matching pressure thresholds.

7.4 Estimating the probability that state is above the quality threshold

Both the state/pressure relationship and the threshold between good and degraded state have an associated uncertainty. This implies that GES is not an on/off function, but instead a probability. Here we illustrate how it is possible to evaluate the probability that the state is above the threshold based on the confidence intervals around the state/pressure relationship.

Figure 7.4.1 shows the state-pressure relationship with its confidence intervals for the M-AMBI indicator from the SP dataset in relation to bottom-trawling pressure. M-AMBI has defined a good status threshold that was developed for detecting the effect of pollution in the WFD. We use this value here to illustrate the process of estimating the probability of achieving GES, but we do not imply that this threshold is a relevant threshold between good and degraded state for ecosystems that are experiencing physical abrasion. The probability that the state was below the threshold was evaluated based on what % confidence interval matched the threshold. For example, the point where the orange curve, which is the mean response, crosses the threshold, represents a 50% chance of the state being above the threshold. The point where the bottom grey curve, which is the lower 95% confidence interval, crosses the threshold, represents a 95% chance of the state being above the threshold. In this example, a trawling pressure above 20 would result in a 0 probability of achieving GES.

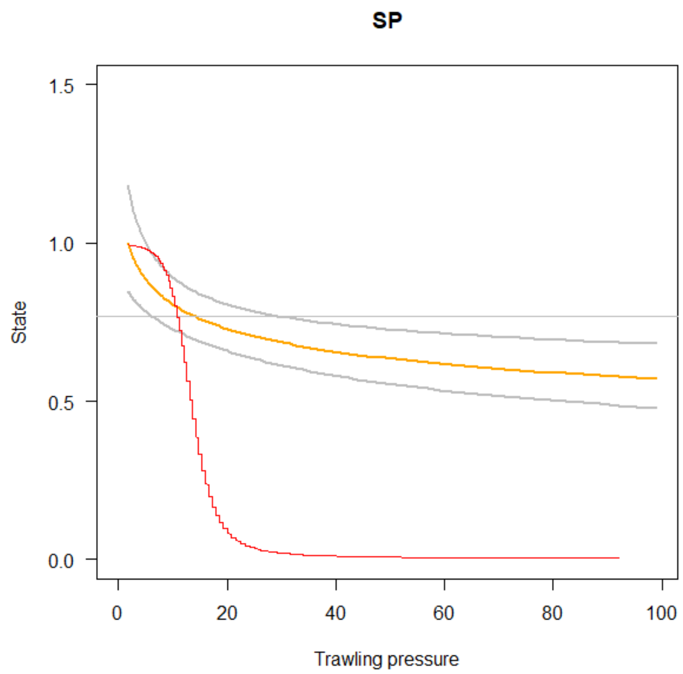

Figure 7.4.1 The relationship between the M-AMBI index and trawling pressure for the Silver Pit dataset (SP). The orange line indicates the mean response fitted using a linear model of log(M-AMBI/pressure) and the thick grey lines indicate the 95% confidence intervals. The thin grey line at state = 0.77 indicates a threshold between good and degraded state for M-AMBI that was developed for detecting the effect of pollution in the WFD. We use this value here to illustrate the process of estimating the probability of achieving GES, but we do not imply that this threshold is a relevant threshold between good and degraded state for ecosystems that are experiencing physical abrasion. The red line indicates the probability the state is above the threshold and declines from 1 to 0 as pressure increases.

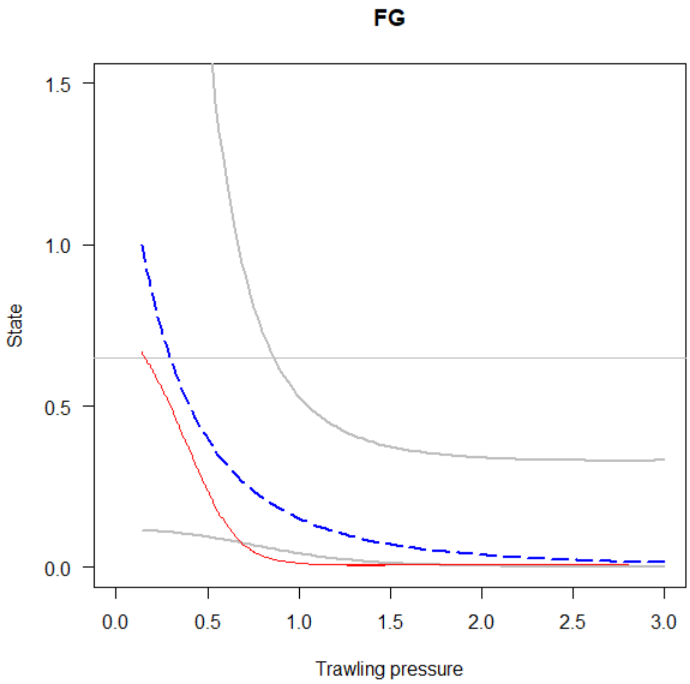

Figure 7.4.2. The relationship between the biomass and trawling pressure for the Fladen Ground dataset (FG). The dashed blue line indicates the mean response fitted using a linear model of log(biomass/pressure) and the thick grey lines indicate the 95% confidence intervals. The thin grey line at state = 0.65 indicates a threshold between good and degraded state for biomass that is estimated based on the natural temporal variation of benthic biomass in undisturbed ecosystems that was presented in WKBENTH2 and obtained from (Nichols 2022). We use this preliminary value here to illustrate the process of estimating the probability of achieving GES but emphasize that further work is required to estimate this threshold. The red line indicates the probability the state is above the threshold and declines from 1 to 0 as pressure increases.

Figure 7.4.2 represents the probability of biomass being above the threshold for the FG dataset. In this example, a threshold between good and degraded state for biomass was estimated based on the natural temporal variation of benthic biomass in undisturbed ecosystems that was presented in WKBENTH2 and obtained from (Nichols 2022). We use this preliminary value here to illustrate the process of estimating the probability of achieving GES but emphasize that further work is required to estimate this threshold. The probability of the state being above the threshold is about 0.65 and declines to 0 at trawling pressure = 1.

Using a probabilistic approach like this, decision-makers can decide what level of risk or precaution they are willing to take by choosing a different probability of achieving GES. Further developments of this approach should take account of the uncertainty in both the pressure-state relationship and the threshold itself.

7.5 Quality thresholds in relation to the definition of habitat loss

The review group of WKBENTH2 pointed out that "*there is not much policy steer to define thresholds of good environmental status, but the little that has been agreed and published (primarily in Commission Decision 2017/848) should be directly addressed as a priority, especially the specific statement that physical loss shall be understood as a permanent change to the seabed which has lasted, or is expected to last, for a period of two reporting cycles (12 years) or more.*"

WKBENTH2 suggested one approach that could be used to estimate a threshold in relation to the recoverability of a benthic community. However, and although it is a complex interaction when dealing with chronic disturbance, WKBENTH3 aimed to keep the discussion on habitat quality and thresholds away from the 12-year definition of habitat loss. Primarily, since the 12-year definition is not a scientifically justified threshold. Nonetheless, some risk-assessment methods are also able to estimate recoverability and may be able to define the level of disturbance that a community can withstand before it is physically lost (according to the 12-year definition in the GES Decision). Science could thus consider the consequences of this definition and advice on what the specific thresholds or the level of precaution should be. However, neither the assessment or thresholds for criterion D6C4 nor indicators assessing D6C4 were part of the workshop analyses.

7.6 Quality thresholds in relation to risk-based indicator methods

In several of the risk-based indicator methods a threshold for adverse effects has been defined or is currently under development (Table 7.6.1). These established threshold values do not necessarily comply with the recommendations from WKBENTH2, and most of these thresholds do not provide an ecologically meaningful distinction between a good and a degraded status that is required in the MSFD, and are instead based on subjective choices or expert judgment. As the methods of the indicators vary greatly, also the methods used for threshold setting are different.

Table 7.6.1. List of risk-based indicators and their associated quality thresholds.

Indicator	Quality Threshold Method	Indicator Output	Habitat-specific Threshold?	Threshold Status
SoS (BH1)	Statistical threshold based on pressure-state curves, arbitrary setting of tipping point and distances to tipping point	continuous (0–1)	yes, based on different habitat sensitivity	under development
BH3	Not yet defined. Threshold could be set between disturbance categories, ranging from 0 (no disturbance) and 9 (highest disturbance) or between groups of disturbance categories: none (0), low (1–4), moderate (5–7), high (8–9).	Ordinal (categories 1–9)	no	-
L1	None	continuous	-	-
L2	None	continuous	-	-
PD2	GES is defined by quantifying the annual range in natural variation (RNV) and its lower threshold of benthic invertebrate abundance in undisturbed seabed ecosystems; and to assess whether these measures co-vary with environmental variables.	Continuous (0–1)	no	under development
CumI	Threshold is set between categories of low and moderate impact, indicating the degree of impact that leads to adverse effects. The pressure-response relationship is incorporated in the impact matrix that combines pressure and sensitivity of habitats.	Ordinal (categories 1–6)	no	approved by HELCOM

7.6.1 SOS

Once the predicted values of sentinel species proportion across the habitat were generated, they were converted into high disturbance, moderate disturbance and low disturbance areas by using a quality threshold specific for each BHT (minimum proportion of sentinel species acceptable to keep ecosystem processes). To establish the quality threshold for each habitat, it was necessary to make three determinations based on the pressure-state curves of each habitat:

Habitat sensitivity determination. The threshold must be defined based on the specific sensitivity of the habitats to guarantee the habitat quality. The habitat sensitivity was calculated by comparing the response curve for each habitat with five theoretical models using an R function developed for this purpose (see https://github.com/Gonzalez-Irusta/SoS). The theoretical models represent five possible responses to pressure, from a sensitivity of 1 (not sensitive) to 5 (very sensitive). The function assigns a value from 1 to 5 to each habitat based on the best fit between the theoretical model and the observed response to the pressure for that specific habitat (lowest sum of squares of the differences between them). This calculation is repeated 1000 times using bootstrapping, obtaining the mean sensitivity of each habitat and its standard deviation based on the type of response observed in the BH1 indicator.

Degradation point calculation. The method consists of identifying the point at which the habitat has lost most of its quality (degradation point) and establishing the quality thresholds at different distances to this point depending on its sensitivity, giving the most sensitive habitats the highest distance to degradation (Figure 7.6.1). The degradation point is the point at which the pressure-state curves change their trend, decreasing the rhythm at which the reduction in the habitat state is observed. Although several statistical tools are being explored to obtain this point, currently, the method relies on the 45 degrees slope of the tangent to the curve, previously used in different works to determine the tipping point in aggregation curves (Colloca *et al.*, 2009; González-Irusta and Wright 2017).

Quality thresholds definition. Once this point has been computed, the condition threshold is established as a percentile of the distance between the origin of the curve and the degradation point. The thresholds generated must respond to the range of sensitivities of the different habitats, so a more conservative one will be used for sensitive responses, while a more permissive one will be applied for habitats with more tolerance to the pressure. For that, three thresholds were defined: (i) the standard which corresponds to the middle point between the beginning of the curve and the tipping point (p.50), (ii) the precautionary located in the first third of that range (p.33) and (iii) the tolerant threshold (p.66).

This assessment used the precautionary threshold for habitats with a sensitivity value of 4, the standard for habitats with a sensitivity of 3 and the tolerant for habitats with a sensitivity of 2. The criteria that support the BH1 methodology for setting quality thresholds are the most appropriate to date, but it is temporary and may be modified in future by expert agreements related to criteria to define the suitability of thresholds values.

The disturbance classes were distinguished as follows: (i) no pressure, the value of the pressure on the area is zero, (ii) low disturbance when the proportion of sentinel species was higher than the threshold, even after removing the standard error; (iii) high disturbance when the proportion of sentinel species was minor than the threshold, even after adding the standard error and (iv) moderate disturbance areas when the position (higher or lower) of the proportion of sentinel species related to the threshold changes after adding/removing the standard error.

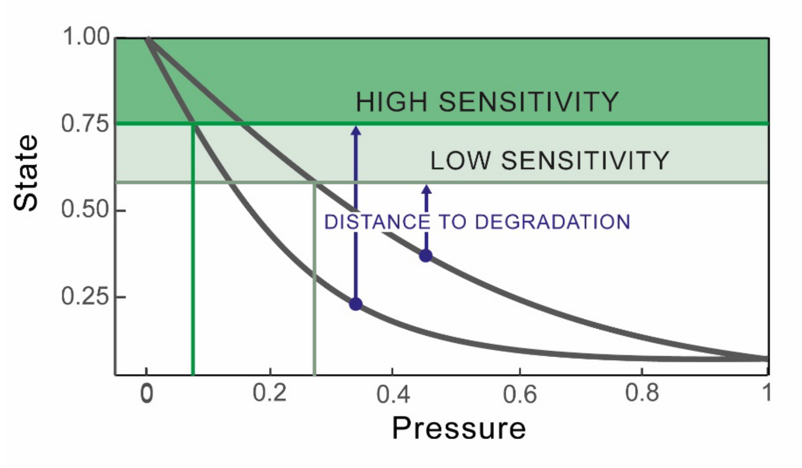

Figure 7.6.1. Distance to degradation approach methodology for setting thresholds to evaluate disturbance on seabed habitats. The green lines show the quality threshold at different distances from the degradation point depending on the BHT's sensitivity.

7.6.2 BH3

The BH3 indicator assesses the spatial extent and magnitude of potential physical disturbance to benthic habitats caused by human activities, where a known pressure-activity link is established. The indicator combines pressure data with information on receptor sensitivity, derived from traits-based assessments of biological communities that characterize assessed biotopes. Both habitat and species sensitivity are considered within BH3, in terms of resistance (the ability to withstand a given pressure) and resilience (the ability to return to an unimpacted state), in response to assessed physical pressures.

A disturbance matrix combines pressure and sensitivity of species and habitats and represents the pressure-response relationship. The matrix was created from previous studies that analysed the impacts of pressures on sensitive species and habitats when applied at different intensities (Table 7.6.2). The resulting disturbance categories are summarized into four groups ('Zero' = disturbance category 0, 'Low' = disturbance categories 1–4, 'Moderate' = disturbance categories 5–7, and 'High' = disturbance categories 8 and 9). These groupings are currently used for comparative interpretations of disturbance outputs across the OSPAR Maritime Area only.

An OSPAR-scale quality threshold for BH3 is still in discussion and will be explored further through future work. Nationally agreed thresholds for the indicator in the UK and DE used the boundary between disturbance categories 4 and 5 to distinguish low and highly disturbed habitat areas.

Table 7.6.2. BH3 disturbance matrix with summary groups; 'Low' (1–4), 'Moderate' (5–7), and 'High' (8–9). Note 'Zero' = zero SAR values in VMS data.

Disturbance matrix		Sensitivity				
		1	2	3	4	5
Pressure	1	1	2	3	4	6
	2	1	2	4	6	7
	3	1	3	5	7	9
	4	1	4	6	8	9
	5	2	4	7	9	9

7.6.3 PD2

WKBENTH2 provided a worked example how to define GES by quantifying the annual range in natural variation (RNV) and its lower threshold of benthic invertebrate abundance in undisturbed seabed ecosystems; and to assess whether these measures co-vary with environmental variables. This is just one approach that could be explored to estimate a threshold for the PD method, but other approaches exist and could be applied. Further work is required to estimate this threshold.

A literature search was conducted to find benthic invertebrate abundance time-series. The annual variation in abundance was used to calculate the RNV (defined as the 95% confidence interval) for each time-series. It was hypothesized that the more stable the ecosystem, the smaller the RNV and therefore the higher the lower threshold (0.025 quantile) of benthic invertebrate abundance. Multiple linear regression analyses were performed with 52 studies; 402 studies were screened and examined against a set of eligibility criteria for inclusion in statistical analysis. Depth of the study site and benthic response (individual species or whole community abundance) were included as explanatory variables after backward model selection. RNV significantly decreases and lower threshold significantly increases as depth increases, which is expected as ecosystems are typically more stable with increasing depth (Figure 7.6.2). This trend is seen across benthic responses, though the individual species studies have significantly higher RNVs and smaller lower thresholds than the whole community studies. Neither the RNV or lower threshold varied significantly with latitude, average species lifespan or substratum type.

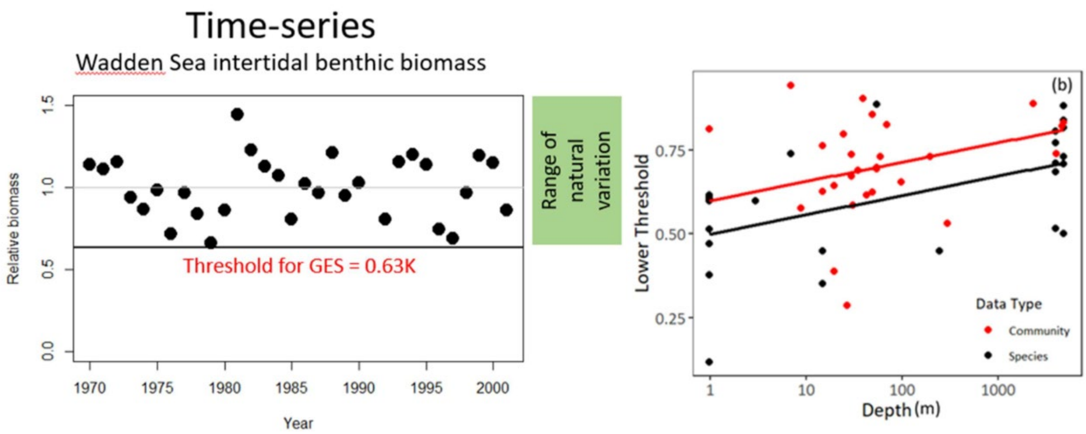

Figure 7.6.2. Example of the treatment of a time-series of intertidal community biomass. The raw data are scaled around 1 and a regression line fitted. The residuals around the regression line are then used to estimate the range of natural variation, which in this example would be at EQR = 0.63. Data from (Beukema and Cadee 1997).

On average the RNV of benthic invertebrate abundance was 1.06, translating to a lower threshold of conservatively estimated as 0.8. This means that the abundance can be reduced to around 80% of the mean abundance and still be considered as having a GES. There is potential to explore other environmental variables that may explain more of this variation.

7.6.4 CumI

The threshold value setting logic is based on the ordinal approach applied within this indicator. The ordinal approach results in 6 final disturbance ‘impact’ categories into which the indicator results fall. The division between these categories is the basis for the threshold value application.

Table 7.6.3. CumI intersection matrix to combine magnitude of pressure and biotope sensitivity to potential impact on the benthic biotopes with subclasses for resulting moderate impacts into three classes.

Impact intersection matrix		Magnitude of pressure			
		High	Moderate	Low	Very low
Sensitivity	High	High	High	Moderate/m2	Moderate/m1
	Moderate	High	Moderate/m3	Moderate/m1	Low
	Low	Moderate/m2	Moderate/m1	Low	Very low
	Very low	Moderate/m1	Low	Very low	Very low

In practical terms the threshold value applied is placed at the division between low and moderate categories within the indicator result categories (Impact intersection matrix, Table 7.6.3). Thus, all moderate and high category results are considered to fail the threshold values, while those achieving low or very low overall outcomes achieve the threshold value. This threshold value placement is based on the defined scale that in categories below this threshold value no adverse effects on benthic biotopes are expected, through a lower level of both sensitivity (biologically defined as a combination of resilience and resistance) and the magnitude of the pressure (abiotically defined as a combination of pressure intensity and frequency; Table 7.6.4).

Table 7.6.4. Classification of disturbance according to the different impact categories in the assessment procedure. The boundary between low and moderate is the boundary of adverse/significant impacts.

Impact (simple)	Impact (extended)
high	high
	moderate 3
moderate	moderate 2
	moderate 1
low	low
very low	very low

7.6.5 Conclusions

The methods for quality threshold setting differ between indicators, as the indicator methods themselves vary greatly. Not all threshold methods are suitable for all indicators considered here. Statistical methods are not applicable for indicators based on an ordinal scale like BH3 and CumI. The 'range of natural variation' method used for the PD2 cannot be applied for L1, as this indicator cannot be estimated without reference to a particular fishing intensity level. As highlighted in the WKBENTH2 report, some threshold-setting methods are to be preferred with regard to e.g. their ecological meaning. A data-based threshold is also desirable, however, availability of spatial and temporal data of undisturbed areas may not always be sufficient to determine a reliable threshold.

It is recommended that in the indicator method the rationale for the ecological meaning of the quality threshold is included. It should be clear what the difference between good status and adverse effects is, with regard to e.g. Species Richness, abundance, biomass, presence of sensitive/long-lived species, recoverability or other species/community parameters. Mostly, there is no tipping point that leads to a degraded benthic status, but a gradual transition from good status to degraded state. A description of the ecological meaning would also improve the comparability of thresholds from different indicators.

8 Next steps needed to operationalize benthic status assessment

8.1 Policy drivers

Seabed ecosystems account for > 7 944 000 km² in Europe, which is over 1.8 times larger than the EU's land area. The seabed ecosystem is home to some 2500 species of benthic organism that represent virtually all known phyla. These species and their populations form a wide variety of communities across distinct habitat types. The MSFD defines Good Environmental Status for seabed habitats according to a number of criteria (COM Decision 2017/848), and in particular criterion D6C5 which requires that the extent of adverse effects from anthropogenic pressures on the condition of the habitat type does not exceed a specified proportion of the natural extent of the habitat type in the assessment area. The condition of the habitat type is further specified as encompassing its biotic and abiotic structure and its functions, including (but not exclusively) typical species composition, typical species relative abundance, absence of particularly sensitive or fragile species or species providing a key function and size structure of species.

Achieving GES for seabed integrity therefore aims to safeguard both benthic community structure and function. Structure and function, commonly used in environmental sciences, are not mutually exclusive of each other, they are both vital. They ensure that viable populations of native or typical species exist across the seabed, representative habitats are distributed across their natural range of variation, including the presence and the size structure of species of particularly sensitive or fragile species or species providing a key function, ecological processes (e.g. nutrient cycles) are maintained and, ecoregions and benthic species are able to respond to short- and long-term environmental change.

The overarching aim of safeguarding benthic community structure and function can be linked with two broadly cited objectives. The first objective is the *protection, and where practicable restoration, of* a number of seabed species and associated habitat that are valued due to their intrinsic value to global biodiversity. The second objective is the *sustainable use of* seabed habitats for ecosystem functions and ecosystem services that are essential to our current way of living (such as food and nutrition, transport, raw material extraction, genetic material extraction, pollution control, leisure, etc.). The MSFD clearly states that such use of marine ecosystems should be kept within levels compatible with the achievement of good environmental status while ensuring that the capacity of marine ecosystems to respond to human-induced changes is not compromised.

Other relevant policies for defining GES for the condition of seabed habitats

The protection of certain seabed habitat types which have intrinsic value to marine biodiversity has been ensured in EU waters since 1992 through the Habitats Directive (Directive 92/43/EEC). The objective of the Habitats Directive with regard to these habitat types is to achieve or maintain favourable conservation status (FCS). Criteria for the assessment of FCS of a habitat type include:

- Its natural range and areas it covers within that range are stable or increasing.
- The specific structure and functions which are necessary for its long-term maintenance exist and are likely to continue to exist for the foreseeable future.
- The conservation status of its typical species is favourable.

Since 2000, the Water Framework Directive, which like the MSFD includes a dimension of sustainable use, has an objective of reaching 'good ecological status' of coastal habitats. The WFD defines 'ecological status' as an expression of the quality of the structure and functioning of aquatic ecosystems associated with surface waters. Elements to take into account include specified biological quality elements, specified physico-chemical quality elements and hydro-morphological quality elements.

At regional level, the Regional Seas Conventions, such as OSPAR (Agreement 2008-6) and HELCOM (BSEP 138), have also conferred specific protection status to a number of habitat types, largely following the Red List criteria and assessment principles of the International Union for Conservation of Nature (IUCN).

In a global context, most particularly valued and sensitive habitats and communities are also defined as Vulnerable Marine Ecosystems. The United Nations General Assembly (UNGA) calls for adoption of conservation and management measures to prevent significant adverse impacts on VMEs. Selecting VME habitats and indicator species is based on the sensitivity of areas holding these communities or habitats, which is such that human activities that cause disturbance may severely or permanently damage and degrade them.[1]

In 2021, TG Seabed produced a comprehensive review of relevant methods for assessing habitat status under other policies, which looked in particular at the Habitats Directive and the Water Framework Directive.[2]

8.2 Ecological relevance of functioning seabed ecosystems

An ecosystem may be considered as a unit within which an assemblage of living organisms interact with each other and with the chemical and physical environment, resulting in natural processes and establishment of a series of complex ecological balances. The Convention on Biological Diversity (CBD 2001) defines an ecosystem as “a dynamic complex of plant, animal and micro-organism communities and their non-living environment interacting as a functional unit” (Article 2 of the Convention). An ecosystem may be considered as a unit within which an assemblage of living organisms interact with each other and with the chemical and physical environment, resulting in natural processes and establishment of a series of complex ecological balances. Ecosystems may operate at a wide range of spatial and temporal scales, from long-term global systems, to very small, localized or ephemeral systems. It is the interactions and processes within ecosystems that afford the delivery of a wide range of environmental services and benefits.

Given the extent to which ecosystems are connected across different spatial and temporal scales it is often difficult to define precise boundaries between ecosystems, especially when applied to the development of management measures. To overcome some of the fuzzy nature of ecosystem boundaries, spatial management units have tended to be defined on the basis of their physiographic and habitat features first, followed then by a definition of their associated biology.

Table 8.2.1. The main functions, ecosystem processes and goods and services of marine ecosystems that are particularly pertinent to trawling impacts on sedimentary habitats.

Functions	Ecosystem Processes	Ecosystem Services and Benefits
Regulation functions		
Nutrient regulation	Role of fauna in storage and re-cycling of nutrients (e.g. N, P, S)	Enhanced benthic-pelagic coupling and maintenance of healthy systems
Gas/climate regulation	Role of fauna in carbon fluxes, CO_2 sequestration	Maintenance of favourable climate for humans

[1] FAO (2009): International Guidelines for the Management of Deep-sea Fisheries in the High Seas

[2] See: Circabc (europa.eu)

Functions	Ecosystem Processes	Ecosystem Services and Benefits
Habitat functions		
Refugium function	Suitable living space for some species	Maintenance of biodiversity and some commercially harvested species
Nursery function	Suitable habitat for some species to reproduce	
Sediment stability	Stabilization or destabilization through direct (e.g. tubes) or indirect (e.g. diatom predation) processes	Reduced temporal shift in sediment balance within coastal areas.
Production functions		
Food provision	Conversion of energy to prey for animals	Provision of seafood; maintaining food production

In this respect, we may refer to services and benefits provided by the marine ecosystem; an increasingly common method of classifying exactly what we may gain and/or lose when we exploit the environment (Holmlund and Hammer 1999). There are a number of methods used for classifying goods and services, and although the typology devised by de Groot *et al.* (2002) was primarily based on terrestrial functions, it aids us to classify those pertinent to the marine environment in a clear manner. The main services and benefits we would consider as the most important for marine policy to safeguard for marine shelf seas are listed in Table 8.2.1. While the relative importance of each of these will vary between different habitats, additional functions may be regarded as essential in some situations.

In terms of societal goals, what we want from seabed ecosystems are for the services and benefits to be maintained at a level that ensures human and overall ecosystem welfare in the long term (i.e. suitability). Significant steps have been made towards developing assessment methods that can track change in response to increased pressure caused by a manageable human activity from pristine to degraded condition in terms of the structure and functioning of the seabed ecosystems and its associated assemblages' state. In terms of societal goals, we are thus able to set limits in terms of how much we are willing or not willing to accept in terms of degradation for a certain amount of the habitat area in in the key functions, ecosystem processes and services and benefits that marine ecosystems (i.e. BHTs) provide.

8.3 From pristine to degraded

All indicators should describe the same type of pressure/state shape of relationship if using the same type of biological observations and metrics. An undisturbed ecosystem is expected to have many species present, with each species having a natural distribution of abundance and biomass over the different age and size classes, with ecosystem processes at high rates (stage 1). Initially, when pressure from human activity is introduced, the ecosystem is indistinguishable from undisturbed in biodiversity, structure (age, size, species) and function because any changes fall within the range of natural variation (2). When the pressure increases further, it is expected that the largest and oldest individuals in the community will be lost, but all species will be present and ecosystem processes are likely to continue at rates that are near natural (3). Sustainable human use of the ecosystem can involve intense activities and is likely to result in widespread changes in size, age and species composition, with values generally outside the range of natural variation (4). Progressing pressure may result in the loss of the largest and most-long-lived species, resulting in large drops in the total biomass of the community, and large drops in the rates at which ecosystem processes occur (5). With further pressure, more species will be lost, and therefore overall Species Richness continues to drop, and all parameters are likely to be much lower than in undisturbed systems (6). At some level of pressure, the ecosystem would not be able to recover to its undisturbed state on human time-scales, even if the pressure was totally removed (7), and at the highest levels of pressure, the ecosystem can be considered lost and transformed into another ecosystem altogether (8).

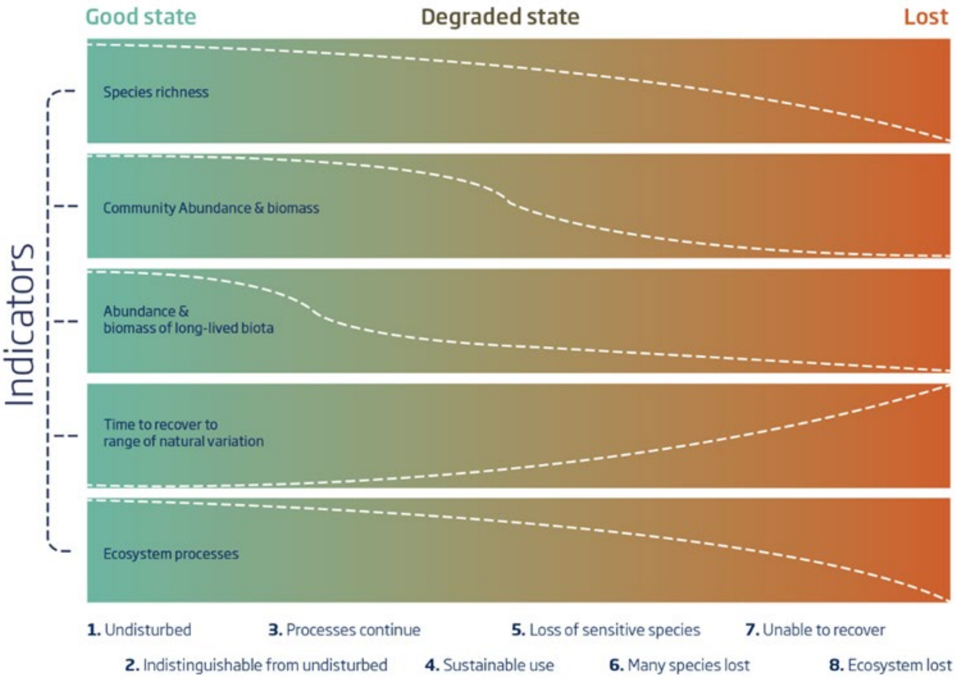

Figure 8.3.1. The indicator trends presented here assume a stochastic environment with no directional (and human-induced) environmental change. From: Hiddink JG, Valanko S, Delargy AJ, and van Denderen D. 2022. How to set thresholds for good status and significant adverse impacts in marine ecosystems? ICES Journal of Marine Science. Submitted.

The challenge is to manage the ecosystem so that ecosystems/communities/habitats are at a sufficiently 'good' state to ensure we sustain overall ecological integrity. The degradation from an undisturbed to a degraded and then lost ecosystem is described in Figure 8.3.1. Stage 1 and 2 both ensure biodiversity, structure, and function and can be considered 'good'. Most people would probably agree that stage 7 and 8 are degraded or even lost. Any changes from stage 3 to 6 may be considered as 'good enough' when part of a socio-economic trade-off and where a prioritization of the management actions is needed.

8.4 General directions as to where GES should in principle be

Natural processes can result in fluctuations of the ecosystem state across space and time, and it is generally agreed that while some change in the state can be compatible with a system being in a good status, as well as some human use, larger change would lead to a degraded state (Folke *et al.*, 2003).

Table 8.4.1. Elements in policy for the assessment of the condition of a habitat type.

Policies	Habitats Directive	Water Framework Directive	Marine Strategy Framework Directive
Overarching objectives	Favourable Conservation Status:	High status: The values of the biological quality elements for the surface water	Good Environmental Status: D1: Biological diversity is maintained. The quality and

Policies	Habitats Directive	Water Framework Directive	Marine Strategy Framework Directive
	natural range and area of habitat type are stable or increasing specific structure and functions which are necessary for its long-term maintenance exist and are likely to continue to exist for the foreseeable future the conservation status of its typical species is favourable	body reflect those normally associated with that type under undisturbed conditions, and show no, or only very minor, evidence of distortion. Good status: The values of the biological quality elements for the surface water body type show low levels of distortion resulting from human activity, but deviate only slightly from those normally associated with the surface water body type under undisturbed conditions.	occurrence of habitats and the distribution and abundance of species are in line with prevailing physiographic, geographic and climate conditions. D6: Seabed integrity is at a level that ensures that the structure and functions of the ecosystems are safeguarded and benthic ecosystems, in particular, are not adversely affected.
Specific elements to assess the condition of habitats	Article 17 HD Guidance document (2017) Range of habitat Area of habitat covered within range Structure and function Structures are considered to be the physical components of a habitat type. These will often be formed by assemblages of species (both living and dead) but can also include abiotic features. Functions are the ecological processes occurring at a number of temporal and spatial scales and they vary greatly between habitat types. Future prospects.	Annex V - Quality elements for the classification of ecological status: Biological elements (composition, abundance and biomass of phytoplankton, other aquatic flora and benthic invertebrate fauna). Hydromorphological elements supporting the biological elements (morphological conditions depth variation structure and substratum of the coastal bed structure of the intertidal zone and tidal regime direction of dominant currents wave exposure). Chemical and physico-chemical elements supporting the biological elements (including general ones – transparency, thermal conditions, etc. – and specific pollutants).	COM Decision –Condition of habitat (D1C6/D6C5) The condition of the habitat type, including: Biotic and abiotic structure. Functions, including: Typical species composition Typical species relative abundance Absence of particularly sensitive or fragile species or species providing a key function Size structure of species.

The threshold for good environmental status (GES) should identify the indicator value at which an ecosystem transitions from a good to a degraded state. Three different types of thresholds exist (quality, extent and connectivity), of which we will discuss quality threshold here in detail. Quality is defined by the indicator value on a local, point or cell, scale. A quality threshold defines at what value of the indicators the local quality can be considered to be 'good'. Desirable characteristics of thresholds for good status are that they are habitat specific and estimated with their uncertainty. This means that a single approach for choosing thresholds that yields different thresholds for different habitats is desirable.

Effective thresholds need to be ecologically meaningful, and therefore separate good and degraded state based on the characteristics of the ecosystem that management aims to conserve. The 'quality' threshold value under D6C5 represents an acceptable degree of deviation in habitat quality from reference state (at a specific location, not on the whole of the habitat type). When the quality of the habitat falls below the threshold value it is considered to be adversely affected. The deviation in habitat quality includes changes in its biotic and abiotic structure and its functions (e.g. through changes in species composition and their relative abundance, absence of particularly sensitive or fragile species or species providing a key function, size structure of species), and may be due to individual or multiple anthropogenic pressures; in some cases the cause of deterioration may not be known. The acceptable degree of deviation in habitat quality from reference state at a specific location will depend largely on the habitat's resilience and needs to consider the habitat's ability to recover to avoid permanent effects.

An EU-wide boundary between the 'good condition/quality' and 'bad condition/quality' of a habitat type represents an acceptable deviation from the reference state at a specific location and ensures that there is an equivalence in the degree of deterioration in habitat quality across habitat types, pressures and regions. Such a boundary can be expressed in qualitative terms, by reference to Commission Decision 2017/848, or in quantitative terms, by making use of a 'normalized' EQR scale, like in the Water Framework Directive (see also Tables 8.4.1, 8.4.2).

Table 8.4.2. Elements in the scientific literature for the assessment of the condition of a habitat type.

Scientific literature	Hiddink 2022	Levin 2016	Grumbine 1994
Overarching objectives	An undisturbed ecosystem is expected to have many species present, with each species having a natural distribution of abundance and biomass over the different age and size classes, with ecosystem processes at high rates	Key metrics that may serve as threshold indicators are measures of biodiversity, abundance, habitat quality, population connectivity, heterogeneity levels, and community productivity. If information is not available to set particular ecological thresholds, a suite of other indicators can be used to determine the likelihood of significant adverse change and impacts, including those that address species-, community- or ecosystem-level impacts.	Within the overall goal of sustaining ecological integrity, five specific goals were frequently endorsed: Maintain viable populations of all native species in situ. Represent, within protected areas, all native ecosystem types across their natural range of variation. Maintain evolutionary and ecological processes. Manage over periods of time long enough to maintain the evolutionary potential of species and ecosystems. Accommodate human use and occupancy within these constraints.
Specific elements to assess the condition of habitats	Elements: Species Richness Community abundance and biomass Abundance and biomass of long-lived biota Time to recover to range of natural variation Ecosystem processes	Significant species-level changes or impacts include: (i) extinction; (ii) significant decline in abundance; (iii) decline in foundation species; (iv) reduction below critical reproductive density; (v) loss of source populations; and/or (vi) loss of critical stepping-stone populations. Community-level impacts include (i) alteration of key trophic linkages among species in a community; (ii) reduction in species diversity beyond natural levels of variability; and/or (iii)	Elements: Populations of all native species Native ecosystem types across their natural range of variation Evolutionary and ecological processes (i.e. disturbance regimes, hydrological processes, nutrient cycles, etc.) Evolutionary potential of species and ecosystems. Human use and occupancy

Scientific literature	Hiddink 2022	Levin 2016	Grumbine 1994
		regional declines in habitat heterogeneity, such as loss of entire habitats or community types. At the ecosystem-level, impairment of important ecosystem functions such as biomass production, nutrient recycling or carbon burial can lead to loss of major ecosystem services upon which society depends. They may include loss of carbon sequestration capacity, genetic resources, or fisheries production.	

8.5 Extent thresholds

Guidance in the GES Decision for Descriptor 6 includes estimation of the spatial extent of pressures and affected habitats. For the risk-based methods described in Sections 2.3 and 6, area calculations are relatively straightforward as indicated in the tables associated with Section 6. However, determining thresholds for spatial extent to conserve ecosystem functions is more challenging, while it must be recalled that under the EU Habitat Directive, extent thresholds values associated to the assessment of Favourable Conservation Status for habitats (and species) are set. There are clear relationships between species diversity and area and plotting that relationship can help to determine optimal spatial extent thresholds. Presumably similar approaches could be taken for other ecosystem functions. Under the common implementation strategy (CIS) for the MSFD a policy driven spatial extent threshold is under discussion for Descriptor 6 (seabed integrity) with the provision that this spatial extent threshold could be revised once science progresses to be able to inform this threshold.

While WKBENTH3 struggled to make significant progress on demonstrating operationally how reviewed assessment methods could be used in the context of setting thresholds for spatial extent, discussions based on existing scientific work outlined a way forward and the aspects that need to be considered when setting thresholds for spatial extent. One potential way forward was exemplified at the workshop with a presentation.

Potential way forward: incorporating spatial extent

The extent of biodiversity features can be quantified with species distribution models (SDMs), which tell the likely distribution of species, habitats or communities (depending on the type of modelled biodiversity feature). SDMs use information on the environmental conditions expected to influence the occurrence patterns of biodiversity feature in question and are dependent on suitable observational data on species or communities, and on environmental factors. Marine SDMs have been developed for both the deep sea and coastal seas (e.g. Virtanen *et al.*, 2018; Kenchington *et al.*, 2019). SDMs rarely consider potential harmful effects caused by activities leading to loss or degradation of habitats, such as bottom trawling or dredging (etc). Tuning such models in a proper way would require monitoring data before and after the development has taken place. The models tell the most likely distribution of biodiversity features, should activities never have occurred. In this sense, SDMs in most of the cases already tell of the areas where biodiversity features are in good environmental status, as where they are not likely located, the environmental conditions are poor (e.g. eutrophication, hypoxia) for their potential occurrence.

In addition to biodiversity features, information on activities which lead to loss or degradation of biodiversity features, is needed. Typically, such information is based on expert judgement, where sensitivity of biodiversity features is defined to different type of pressures, caused by human activities. In such assessments, spatial footprint and disturbed area of activities is estimated, if exact information is not available. Spatial footprint and the impact zone of activities can be defined based on e.g. satellite images and underwater devices.

When information on both biodiversity features and spatial footprints of activities (and impact zones) are available, thresholds for seabed integrity can be defined based on the fraction of biodiversity feature potentially disappeared based on the spatial footprints and impact zones, relative to the known occurrence area and the time frame considered. The approach can even be modified with the information on recovery and resilience of biodiversity features.

In the simplest form, the disappeared and disturbed fraction of biodiversity features could be calculated based on the weighted range size rarity (Williams *et al.*, 1996;) for each area without considering the detrimental effects of human activities (the so-called "pristine condition"), and with considering the effects. The acceptable percent change between the two could be used to define the spatial extent needed to ensure the integrity of the seabed, based on the biodiversity features. The weighted range size rarity can be calculated automatically e.g. based on Zonation analyses where weighted range size rarity for each cell, i, is defined as:

$$wrsr_i = \sum_j w_j q_{ij},$$

where w_j is the weight assigned to species j and qij is the fraction of species j's range falling within cell i. This measure highlights the areas that have a relatively large proportion of narrow-ranged species and lowers the contribution of species with wider distribution ranges (Veach *et al.*, 2017). Information required can be species, habitats, or communities, Or even the proportion of seabed representing potential substrata known to influence the occurrence of certain biotopes.

With the information on (1) biodiversity feature ranges without the detrimental impacts and (2) biodiversity feature ranges modified with the detrimental impacts, extinction thresholds for biodiversity features could be set based on an approach introduced by Kuipers *et al.* (2019). They developed a global extinction probability (GEP) index, to evaluate how local impacts on species contribute to regional and global extinction probability of species. GEP uses species range sizes, global conservation status and Species Richness, to indicate thresholds needed to avoid species losses. The approach could be modified to setting thresholds for seabed integrity. GEP is calculated as follows:

$$GEP_{g,j} = \frac{\sum_s \frac{\sum_i A_{s,j,i} \bullet O_{s,j,i} \bullet TL_s}{\sum_{j,i} A_{s,j,i} \bullet O_{s,j,i}}}{\sum_s TL_s}$$

where As,j,i is the part of the range area of species s (belonging to species group g) in the area j and grid cell i (i.e. As,j,I corresponds to the area of grid cell i in e.g. ecoregion j occupied by species s). If j equals a grid cell, j contains a single grid cell, i, only. Oi,j is the occurrence-weight value, which can be set individually to highlight the occurrence probability class O of species s in pixel i and region j. TLs here is the IUCN threat level weight value of species s (belonging to group g), but which could be modified to reflect national/regional threat assessments. If the sum of the regional GEPs of a certain species group (∑j(GEPg,j)) equals one, all species of the group are lost in all regions, the species group will be extinct globally (Verones *et al.*, 2022).

Other Considerations

Setting thresholds for spatial extent for the seabed integrity would first require information on the location of species/communities/habitats (from hereafter: biodiversity features), and of the activities causing loss or degradation of biodiversity features within the sea area in question. This is required to get an overview of both habitat quality, amount of habitat in good quality and spatial configuration within an area in question. Linked to this, one would need information derived from species distribution models (SDMs). Hydrodynamic models can then be used to estimate connectivity among the areas in the region. As seen in the above example such work is being applied in the marine environment using tools such as Zonation 5 (Moilannen *et al.,* 2021). Similarly, for the deep sea, work on connectivity is helping to understand the role of habitat configuration in maintaining viable populations of deep-sea invertebrates. The role of habitat fragmentation (the number and size of "healthy undisturbed" areas in a mosaic of disturbed patches of varying bottom fishing intensity) is being examined using indicators of landscape ecology. WKBENTH2 (ICES 2022) also provides a short update on the status of some of the current marine connectivity research activities, and suggestions on how connectivity may be incorporated into MSFD in future.

The assessment of Descriptor 6 of the EU's Marine Strategy Framework Directive (MSFD) ensures the sustainable management of human activities affecting our seas, it should be "at a level that ensures the structure and functions of ecosystems are safeguarded and benthic ecosystems, in particular, are not adversely affected". Thus, in terms of spatial management decisions, the threshold should likely be driven also by societal or management choices and cannot be chosen purely based on ecological/biological science. Spatial management tools for achieving a balance between state and human activities should be developed in order to it may be possible to identify a situation where a balance between the seabed state and socio-economic benefits is found that is optimal from a societal point of view. In 2021, ICES developed a series of methodologies to give examples of spatial thresholds based on trade-off evaluations, where different management scenarios were analysed based on impact, and percentage of the unfished area against fisheries value. This type of approach could be conducted at a sub/regional scale if the methods and data required are available and should help with the management approach. It is planned that further work on trade-offs will be carried out in response to a request from DGENV in late 2022 and in the first half of 2023.

8.6 Approaches to integrate assessments from different spatial scales and components of the ecosystem

Based on WKBENTH3 findings there does not seem to be one indicator that can alone address all the requirements to assess seabed integrity. Experts found that the assessment as a whole need to ensure that cross-regional, regional, national and local scales assessments can "talk" to each other and that they are complementarity in terms of what aspects of the ecosystem the respective indicators are capturing and what pressure they are tracking (linked to manageable human activity). Cross-regional assessments will inform whether assessments are measuring the same or similar thing, allowing crosschecking. While cross-regional methods can be linked to regional and local scale methods that can be used to provide specificity and links to *in situ* monitoring and assessment. Below Figure 8.6.1. illustrates the discussions had at WKBENTH3 and how the various components to be considered are linked to each other when operationalizing an assessment procedure for MSFD D6.

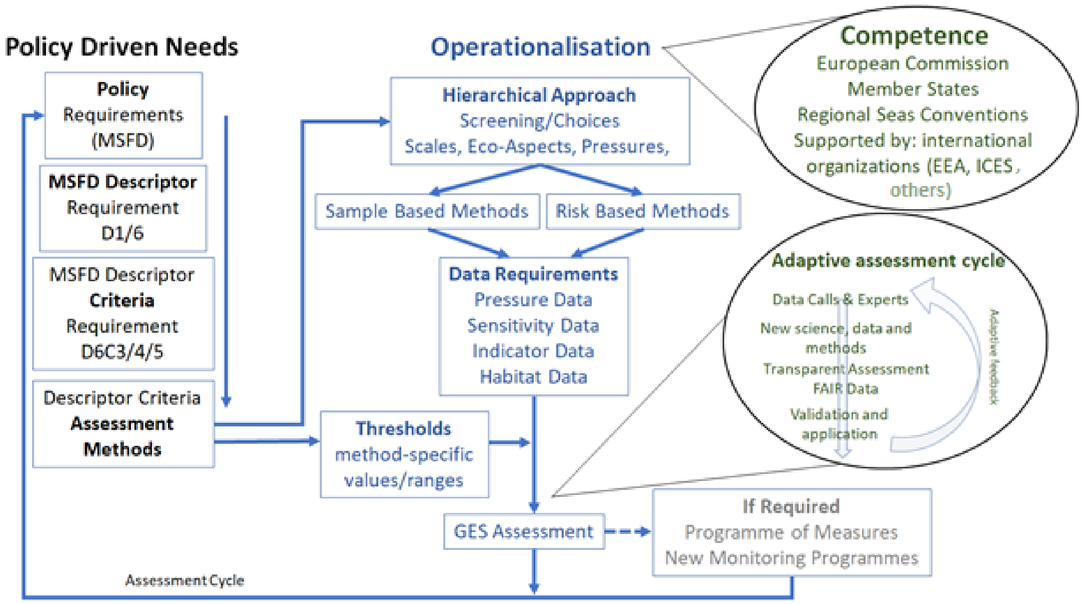

Figure 8.6.1. Conceptual diagram summarizing the requirements in terms of higher level "policy driven needs" and mapping of the process that WKBENTH3 has presented in this report as a potential way to operationalize an assessment procedure that is linked to management measures. Competence as well as properties of the assessment procedure are highlighted.

It was further discussed by the workshop that one of the potential vehicles that could adapted to take forward an assessment framework for seabed integrity (Descriptor 6) could come from either integrated ecosystem assessment or cumulative impact assessments. Similarly EU's TG Seabed have also explored in their longer version of the Article 8 guidance (how to run a D6 assessment) a structural framework for aggregation and integration of indicators (European Commission 2022).

TG Seabed's proposed framework

This draft framework provides conceptual guidance on how to integrate different types of indicators and assessments into the extent of adverse effects according to D6C5. The broad concept behind the proposed structure is that where suitable and appropriately designed monitoring data exist at the appropriate scale, this provides the strongest input to the overall assessment of D6C5, and where such data are missing or considered insufficient, the assessment is supported by a combination of state indicators under other descriptors or policies, and/or risk-based methods.

Integrated Ecosystem Assessment (IEA)

IEA is a tool for implementing ecosystem-based management. Levin *et al.* (2009) define IEA as an incremental iterative process for 'formal synthesis and quantitative analysis of information on relevant natural and socio-economic factors, in relation to specified ecosystem management objectives'. IEAs are proposed as a framework 'for organizing science in order to inform decisions in marine EBM at multiple scales and across sectors', enhancing the ability of managers to evaluate cumulative impacts and carry out trade-off analyses (Levin *et al.*, 2009). IEAs in their purest form intend to take a comprehensive multi-sectoral, multi-pressure ecosystem view of the entire social-ecological system, involving stakeholders to identify management objectives. Conceptually, IEA is both simple and sensible; however, the data, monitoring and modelling requirements of full ecosystem-based management are many and daunting (Hilborn 2011; Hobday *et al.*, 2011; McQuatters-Gollop 2012; Dickey-Collas 2014; Borja *et al.*, 2016; Harvey *et al.*, 2017).

Both NOAA and ICES have adopted the 'Levin cycle' as their framework for IEA (ICES 2012). The cycle outlines 5 stages of IEA: scoping, indicator development, risk analysis, management strategy evaluation, and ecosystem assessment (Levin *et al.,* 2009, 2014; Samhouri *et al.,* 2014). The idea of the loop is useful as it highlights IEA as an iterative process; however, the framework is not prescriptive, instead adapting to regional requirements and various data situations (Levin *et al.*, 2014; Holsman *et al.*, 2017). Despite this, the imagery of the cycle and description of 'steps' can present an obstacle to progress as a lack of progress in one step can hamper development in another. IEA has been described as a process in which a management objective is assessed in an ecosystem context; therefore, an entire IEA process may not be required to inform management measures (Harvey *et al.*, 2017). Instead, it has been proposed that we think of IEA as a toolbox or 'cloud' (Dickey-Collas 2014) moving towards improved ecosystem understanding by progressing each of the critical elements (while maintaining effective communication).

There are several approaches to mapping cumulative impacts, e.g. the Baltic Sea Impact Index (HELCOM 2010), HARMONY (Andersen *et al.,* 2013), SYMPHONY (Hammar *et al.,* 2018) and Cumulative impact from physical pressures on benthic biotopes (CumI, HELCOM 2022). The CumI has been reviewed by WKBENTH3 as it specifically targets D6.

9 Discussions and conclusions

To evaluate the indicators for their suitability to assess D6, the WGECO/WGBIODIV indicator evaluation criteria revised by WKBENTH2 were used. Instead of scoring the indicators with only numbers, the suitability and shortcomings of the indicators were presented in writing to give a more accurate description of their properties (Annex 6). The criteria consolidated in WKBENTH2 were then consulted to ensure extracting the relevant information for the description of indicators and their qualitative comparative assessment.

The criteria and methods developed under WKBENTH2 proved to be useful in elucidating indicator methods performance and threshold setting approach. These phases encompass both a methodological development cycle and an operational cycle.

Five of the indicators (SoS, M-AMBI, TDI, PD2 and DKI) had been scored already during WKBENTH2 and this information was taken into account also here. It is important to note that some of the shortcomings listed, e.g. applicability of the indicator, are more related to the availability of data than to the properties of the indicators themselves. Common for all the empirically based indicators is that to be effective the data collection and monitoring programs need to be designed accordingly. Whereas lacking pressure-specific responses can be seen as a shortcoming, indicators integrating overall condition of the benthic habitats are useful in pointing out cumulative effects.

A synopsis of the WKBENTH3 evaluation is provided in Table 9.1. One of the key findings of WKBENTH3 is the complementarity of groups of indices, which showed clear patterns of association. One group of indicators showing positive correlation included those based on a diversity component (e.g. m-AMBI) or based on diversity measurement (e.g. Shannon, inverse Simpson) and also included abundance and biomass (log x+1) indicators. Another group included the TDI family of indicators (mTDI, mT, TDI and pTDI) which were highly correlated with each other and address the 'absence of particularly sensitive or fragile species' (Table 1.1.1). They have low correlation with other indicators addressing that property, that is with M-AMBI, DK1, BENTIX, long-lived fraction, median longevity, CumI and SoS. Of those, BENTIX and AMBI (reversed) are highly correlated with each other, while the long-lived fraction, median longevity and SoS form a third distinct clade, with the remainder, which also capture properties of species composition, forming a fourth clade (Figure 5.1.2).

Indicator performance was variable when applied to gradients within each dataset (Section 4.2), although most indicators followed the same trends. For instance, while most indicators had declined at the high trawl impact relative to the baseline (Figure 4.1.1), only a few indicators showed a significant decline. These included biomass, richness, fraction long-lived, median longevity, SoS, DM', Shannon Index and Inverse Simpson). Indicators that were developed for, and used traits selected for their sensitivity to trawling disturbance, SoS and long-lived fraction, showed a stronger response to the trawling gradients in the common datasets than others (Figure 4.1.1.). Biomass also showed a strong response to the trawling gradients assessed. On the other hand some indicators showed no response in this exercise, as in the case of most of the TDI family indicators.

When analysing the pollution and eutrophication gradients, the Relative Margalef Index showed a consistent decline across all three gradients. BENTIX showed the strongest decline in 2 out of 3 studies. The indicators that showed the strongest decline in relation to trawling showed less consistent responses (but not all could be estimated for all studies due to missing biomass data).

Several indicators are specific for certain BHTs, (sub)regions (see Table 9.1 and Annex 5). In general, there is a need for testing and validation for broad application at MSFD-scale (e.g. extending indicators to offshore areas and/or (sub)regional indicators to other (sub)regions). Previously-defined threshold values are often pressure-specific and cannot directly be adapted for use in other pressure gradients (e.g. TVs estimated for eutrophication cannot directly be used for physical disturbance by fisheries).

A worked example of how to estimate thresholds for GES using data from gradient studies based on the approach of 'detectable change' was developed, and applied for muddy sand habitats for trawling pressure, and for pollution and eutrophication gradients. The approach was not able to estimate thresholds for all gradients studies when the confidence intervals around the relationships were very wide.

Comparison of the output of risk-based assessment methods also showed that, although sensitivities varied between indicators, impact outputs broadly aligned in many tested areas. This was the case of the North Iberian Atlantic. In this area SoS, PD2, L1 and BH3, all identified 'Offshore circalittoral mud' as being the habitat most impacted by physical abrasion pressure and 'Circalittoral sand' as being the least impacted. However, some variances in the comparisons were observed and may be the result of key methodological differences between indicators, including scale and resolution of the sensitivity assessment and pressure data. Our results should be considered as an initial scoping exercise.

Examples of spatial extent thresholds currently established under environmental policies can be found in relation to the thresholds associated to the assessment of Favourable Conservation Status for habitats (and species) under the HD. While WKBENTH3 struggled to make significant progress on demonstrating how assessment methods could be used to in the context of setting thresholds for spatial extent, discussions based on existing work outlined a way forwards and the aspects that need to be considered when setting thresholds for spatial extent. Consideration of threshold values often need to be informed from data drawn from reference or low-pressure conditions. Such areas are not always available and may reflect historical pressures affecting the region, creating different baseline conditions within regions.

Considering the link of indicators to different benthic community properties, the assessment of D6 should be carried out selecting a number of indicators from different cluster groups to ensure that components of diversity, species sensitivity and abundance (density and/or biomass – or other proxy linked to benthic habitats functioning) are addressed.

Experts found that the assessment of seabed integrity as a whole needs to ensure that cross-regional, regional, national and local scales assessments can "talk" to each other and that they are complementarity in terms of what aspects of the ecosystem the respective indicators are capturing and what pressure they are tracking (linked to manageable human activity). Cross-regional assessments will inform whether assessments are measuring the same or similar thing, allowing for this crosschecking.

Table 9.1. Synthesis table summarizing the main elements emerging from the comparative indicator methods assessment.

Indicator method	Synopsis of WKBENTH3 Evaluation
Multivariate AZTI Marine Biotic Index (M-AMBI)	The multimetric index M-AMBI is based on benthic macroinvertebrates and combines AMBI, diversity and Species Richness. M-AMBI integrates the response of the three metrics: Species Richness, Shannon diversity and the biotic index AMBI based on the relative proportion of sensitive/tolerant species (five ecological groups). The index is compliant with the WFD and successfully applied in coastal, and marine waters over a wide variety of geographical areas and habitat types against multiple pressures.

Indicator method	Synopsis of WKBENTH3 Evaluation
	The index is suitable for the MSFD criteria D6C5 and D5C8. The indicator performance analyses support the correlation between AMBI/M-AMBI and the common taxonomic diversity indices (Margalef, Shannon Index, Simpson Index) and DKI and BENTIX. These indicators seem to respond better to eutrophication and pollution gradient rather than trawling, supporting the fact that they are not specifically defined for response to trawling. The metric renders a five-step numerical scheme for the classification of benthic communities: Bad ecological status: <0.2; Poor: 0.2–0.39, Moderate: 0.39–0.53, Good: 0.53–0.77, High: >0.77. Depth-specific and/or habitat-specific reference conditions and thresholds should be further tested and evaluated.
Danish Quality Index (DKIver2)	The Danish quality index (DKI) has been developed to assess the condition of a water area in accordance with the EU Water Framework Directive, expressing the ecological state of an area, using routine soft bottom benthic monitoring data. Responds directly to D6C5 (but can also address e.g. D5C8), restricted to soft bottom habitats. Describes overall status of an area, but not proportion of a specific habitat type. Indicator and threshold values adjusted for salinity (as this affects species diversity and abundance regardless of status). In case of use for MSFD assessments, adjustment of indicator and monitoring programmes to capture differences between habitat types. The interaction relationship between salinity and diversity is not fully known (now assumed to be linear). Also, some combinations of salinity and exposure levels are not available in Danish waters, and thus not included in testing procedures. The monitoring is not designed to assess specific BHTs. The meta-analysis output shows no significant effect of trawling on the mean response.
Relative Margalef Diversity (D_M') (OSPAR BH2b)	D_M' estimates community diversity relative to a case (monitoring technique x habitat x assessment unit) specific reference diversity. Potentially applicable on basis of all standardized benthic community monitoring data. Community diversity reflecting general quality status as a resultant of all pressures impacts at stake. Potentially applicable in all regions in case of monitoring in no-/low-pressure areas as well; operational in Greater North Sea. Indicator for D6C5 and D1C6. The outcomes from WKBENTH3 analyses show a significant mean reduction of D_M' values in response to trawling. The Relative Margalef index shows also a consistent decline across pollution and eutrophication gradients, suggesting a no-specific response to a particular type of pressure. No evidence-based suggestions for quality threshold values yet, although DM' values > 0.8 indicate relative high diversity and < 0.6 relative low diversity. Confidence of diversity assessment is good in case of sufficient representative data; indicator will assess total quality status (not only including manageable human activity). Options to define more detailed habitat classes or include gradients in reference values to compensate for natural variability more accurately. Reliability of scaling and aggregation of results highly dependent of representativity of monitoring. Need for setting thresholds comparable to other indicator assessments and in need of representative monitoring and reference areas.
BENTIX	BENTIX is a biotic index based on the concept of indicator groups. The index uses the relative contribution of tolerant and sensitive taxa of benthic macroinvertebrates (two ecological groups) weighting them accordingly to the ratio of their occurrence in the benthic fauna. It is compliant with the WFD and successfully applied for the classification of ecological quality of coastal waters in the Eastern Mediterranean over a variety of habitat types and against various pressures. The index is suitable for the MSFD criteria D6C5 and D5C8. The metric renders a five-step numerical scheme for the classification of benthic communities: Bad ecological status: <0.42; Poor: 0.42–0.58, Moderate: 0.58–0.75, Good: 0.75–1, High: =1. Depth-specific and/or habitat-specific reference conditions and thresholds should be further tested and evaluated. BENTIX doesn't show a significant mean reduction in response to trawling, but analysing the trawling pressure gradients individually, it seems to decline in some of the datasets, showing in some cases a strong decline, in particular in relation to the pollution and eutrophication gradients.
Population Dynamic 2 (PD2)	PD2 is a mechanistic model; it estimates the decrease in biomass in response to trawling and the recovery time. The model needs depletion and recovery parameters, of trawl impact relating to infauna and epifauna. Tested at the North and Baltic Sea region level; can also be tested at other scales. Responds directly to D6C3 and D6C5. Strong rooting in general concepts of population dynamics and the fact that it is a single indicator summarizing impact across the entire benthic community. The current implementation assumes an equilibrium between benthic state and trawling but when large changes in trawling pressures

Indicator method	Synopsis of WKBENTH3 Evaluation
	occur this may not be accurate. Mainly relevant to trawling impacts, however, it has been used to map impacts of hypoxia in the Baltic Sea. Concept is peer reviewed; however it is still in development under WGFBIT. The PD analysis, expressed in terms of median longevity, shows a significant decline in relationship with trawling, both for the mean reduction response and the pressure gradient analysis. The indicator could be considered trawling pressure specific as it shows a stronger response in relation to trawling than to the other pressures gradients. Impact estimates are affected by the uncertainty in the habitat-specific longevity composition of the benthic community. Further testing on whether the model can be extrapolated to other regional seas is needed.
Sentinels of the Seabed (SoS) (OSPAR BH1) Fraction of sensitive species	SoS is based on selection of most sensitive and distinguishing (sentinel) species comparing known pressure (gradients) to no-/low-pressure situations. Potentially applicable on basis of all standardized benthic community monitoring data in case of presence of no-/low-pressure situations or reference has been defined before. Very sensitive in case of clear pressure gradients and single/dominant pressure situations (operational in Bay of Biscay and Iberian Coast, examples from Mediterranean and North-Atlantic); not operational in multi-pressure situations and lack of reference areas yet. Indicator for D6C3 with valuable input to D6C5. SoS shows a significant mean response (showing an average decline >65%). Analysing the pressure gradients SoS shows a strong decline in relation to trawling in several locations, providing consistent results with regard to a trawl-specific disturbance, as it targets the most 'sensitive' fraction of the community. No threshold values yet, although providing valuable information on responses in gradients of use to set TVs. The confidence of the model is good and reliability of assessment result high in case results are based on clear (single) pressure gradients including presence of reference areas. Uncertainty is taken into account as the standard error when classifying the areas into low, moderate and high disturbance. Areas are represented with a moderate disturbance, when uncertainty does not allow discerning between low or high disturbance. Scaling opportunities are good in case of accurate pressure (and habitat) mapping available, and only the pressure as taken into consideration is at stake. Priority work for future includes evidence-based proposals for TVs and testing/optimizing application in multi-pressure situations (including need for data from reference areas).
Trawling Disturbance Index (TDI)	TDIs are indicators based on a combination of species response traits to trawling pressure. The 5 traits (position in the sediment, feeding mode, mobility, adult size, fragility) should be defined for each taxa and indicators computation needs quantitative variable for weighting (biomass or abundance) from observation dataset. Traits are scored to reflect species vulnerability to trawling but they do not inform about ecosystem functions. TDIs are particularly applicable to benthic epi-megafauna of the soft-bottoms and have been successfully applied for the Mediterranean Sea and the Northeast Atlantic. TDIs are suitable for the MSFD criteria D6C3, D6C4 and D6C5. Different types of TDIs indicators could be computed on either abundance or biomass. The analysis shows no significant mean reduction in response to trawling except for the mTDI and TDI biomass-based values that shows a significant negative relationship with trawling. No qualitative threshold of the indicator values themselves have been developped. Habitat/regional specific pressure thresholds in the English Channel, the North Sea and northwestern Mediterranean (Jac *et al.,* 2020b) were proposed but not evaluated within this report.
Fraction of Long-lived Species and median community longevity	Fraction of long-lived species and median community longevity are sampling-based indicators. The indicators are associated with the risk-based methods L1, L2 and PD2 and were therefore evaluated in the gradient studies. The long-lived fraction and median longevity show a significant mean reduction in response to trawling and also across the pressure gradient for several regions. The long-lived fraction shows a strong decline in relation to trawling as it targets the "sensitive" fraction of the community.
Risk-based longevity indicators L1 and L2	L1 and L2 are risk-based indicators. They utilize benthic data from boxcore and grab samples to estimate effects of trawling on longevity biomass composition. L1 uses the reference (undisturbed) longevity biomass composition as a sensitivity layer and estimates impact as the proportion of the benthic community with a lifespan exceeding the time interval between trawling events. L2 estimates sensitivity as a critical trawling intensity (the intensity at which the biomass proportion of long-lived taxa, longevity ≥ 10 yr, is reduced to 50% of the untrawled reference) and estimates impact as a decline in median longevity. Tested

Indicator method	Synopsis of WKBENTH3 Evaluation
	at the North Sea level (Rijnsdorp *et al.,* 2020); approach can be used to monitor and assess impact and seabed status at the regional scale and broad scale habitat types. Responds directly to D6C3 and can also be used to inform D1. Indicators are still in development stage; however, approach can be used to monitor and assess impact and seabed status at the regional scale and broad scale habitat types. L1 and L2 are highly correlated with PD, all based on longevity distribution. No thresholds have been evaluated. Current methodology concerned with the uncertainty in the habitat specific biomass–longevity relationships, random variation within the sampling stations not assessed.
Cumulative Impact from physical pressures on benthic biotopes (CumI)	CumI evaluates the cumulative potential/expected impact of several physical pressures on the benthic biotopes of the Baltic Sea, (partly) based on pressure-specific sensitivities. The sensitivities used in the indicator apply to the whole community, while actual sensitivities (from literature reviewed for the CumI) are species based. CumI addresses D6C3, and to some extent D6C4. The method is empirically based and applicable to all pressure gradients. Thresholds value for adverse effects is based on the categorial approach applied within this indicator, and set to the boundary between 'moderate' and 'low' impact in the cumulation process, whereas the category 'very high' is considered as functional loss. Future developments include a more rigid approach to assess the uncertainty and more accurate resolution in pressure mapping. The comparison with other risk-based assessments in three Baltic Sea regions showed some similarities with the L1 assessment, but low consistency with PD2. Confidence is assessed on the basis of a combination of a data quality score plus a score for the temporal data coverage and the spatial data coverage.
Extent of physical disturbance to benthic habitats (BH3)	The BH3 indicator assesses the spatial extent and magnitude of potential physical disturbance to benthic habitats caused by human activities, where a known pressure-activity link is established. The indicator combines pressure data with habitat maps and sensitivity information, derived from peer-reviewed traits-based sensitivity assessments of biological communities that characterize assessed biotopes. BH3 assesses responses in sensitive benthic communities following pressure events through assessments of *in situ* species and biotope data and predictive habitat polygons developed using EUSeaMap, EMODnet Bathymetry, EMODnet Geology Copernicus marine and data on light attenuation, light at the seabed and kinetic, current and wave energy datasets. BH3 is operational at an OSPAR Region-scale, is applied for the MSFD by the UK, Germany, Ireland, and is referenced in assessments by other Member States in MSFD Regions: Greater North Sea, including the Kattegat and the English Channel, Celtic Seas and the Bay of Biscay and Iberian Coast. BH3 mainly applies to D6C3 and informs D6C5; BH3 also has relevance to components of D1. Outcomes of the risk-based comparisons undertaken in WKBENTH3 highlighted that outputs were broadly similar to those of the other indicators used in the comparisons. Nationally agreed thresholds are used for BH3 in the UK and DE, using the boundary between disturbance categories 4 and 5 to distinguish low and highly disturbed habitat areas. Outputs are developed with accompanying confidence maps to indicate uncertainty in component data layers used in assessments. Numeric confidence scores are assigned to each of the attributes: confidence based on underlying data; confidence within data source; and confidence in the sensitivity of the habitat to a pressure. Future developments include adaptation to analyse new human activities, further exploration of threshold values and integration with wider OSPAR-scale indicator assessment methods and ground-truthing of findings.
Biomass, Abundance, Shannon Index, Inverse Simpson, Simpson Index	Traditional total occurrence and diversity indicators potentially applicable on basis of all standardized benthic community monitoring data. Respond to the total of pressures and natural variability at stake, but not necessarily unidirectional and neglect natural differences in community structure. Potentially applicable in all regions. The analysis output shows that Species Richness, Shannon, and Simpson's index clearly decline along all the different types of pressure gradients (trawling, pollution, and eutrophication) mostly of the datasets, supporting their more generic use. The meta-analysis of the mean response shows a significant effect of trawling on these indicators. The analysis highlight also that biomass and abundance have higher fluctuations in values and therefore they are less useful in areas where data variability is high. No evidence-based suggestions for quality threshold values yet. Many examples of application already. Confidence of diversity and/or occurrence assessment (the initial function of these indicators) is good in case of sufficient representative data; the relatedness to quality status is however less clear. Reliability of scaling and aggregation of results highly dependent of representativity of monitoring. Need for

Indicator method	Synopsis of WKBENTH3 Evaluation
	setting thresholds comparable to other indicator assessments in case considered sufficiently suitable as quality indicators.

10 References

Beiras, R., Durán, I., Bellas, J., Sánchez-Marín, P. (2012) Biological effects of contaminants: Paracentrotus lividus sea urchin embryo test with marine sediment elutriates. ICES Techniques in Marine Environmental Sciences, No. 51, 13 pp. DOI: http://dx.doi.org/10.25607/OBP-262

Bellas, J., Nieto, Ó., Beiras, R. (2011). Integrative assessment of coastal pollution: development and evaluation of sediment quality criteria from chemical contamination and ecotoxicological data. Continental Shelf Research, 31(5), 448–456.

Beukema, J.J., Cadée, G.C. (1997). Local differences in macrozoobenthic response to enhanced food supply caused by mild eutrophication in a Wadden Sea area. Food is only locally a limiting factor. Limnology and Oceanography, 42(6), 1424–1435.

Bolam, S.G., Eggleton, J. D. (2014). Macrofaunal production and biological traits: spatial relationships along the UK continental shelf. Journal of Sea Research, 88, 47–58.

Bolam, S.G., McIlwaine, P.S.O., Garcia, C. (2016). Application of biological traits to further our understanding of the impacts of dredged material disposal on benthic assemblages. Marine Pollution Bulletin, 105(1), 180–192.

Bolam, S.G., Garcia, C., Eggleton, J., Kenny, A.J., Buhl-Mortensen, L., Gonzalez-Mirelis, G., ... Rijnsdorp, A.D. (2017). Differences in biological traits composition of benthic assemblages between unimpacted habitats. Marine Environmental Research, 126, 1–13.

Borja, A., Dauer, D., Diaz, R., Llanso´, R.J., Muxika, I., Rodriguez, J.G., Schaffner, L., (2008). Assessing estuarine benthic quality conditions in Chesapeake Bay: a comparison of three indices. Ecological Indicators 8, 395–403.

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

Borja, A., Elliott, M., Andersen, J.H., Berg, T., Carstensen, J., Halpern, B. S., Heiskanen, A.-S., *et al.* (2016). Overview of Integrative Assessment of Marine Systems: The Ecosystem Approach in Practice. Frontiers in Marine Science, 3. http://journal.frontiersin.org/article/10.3389/fmars.2016.00020 (Accessed 26 January 2018)

Borja, A., Franco, J., Perez, V. (2000). A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Marine Pollution Bulletin 40, 1100–1114.

Carstensen, J., Andersen, J.H., Gustafsson, B.G., Conley, D. J. (2014). Deoxygenation of the Baltic Sea during the last century. Proceedings of the National Academy of Sciences, 111(15), 5628–5633.

Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure. Australian journal of ecology, 18(1), 117–143.

Colloca, F., Bartolino, V., Lasinio, G. J., Maiorano, L., Sartor, P., Ardizzone, G. (2009). Identifying fish nurseries using density and persistence measures. Marine Ecology Progress Series, 381, 287–296.

De Groot, R.S., Wilson, M.A., Boumans, R.M.J. (2002). A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecological Economics, 41, 393–408.

Dickey-Collas, M. 2014. Why the complex nature of integrated ecosystem assessments requires a flexible and adaptive approach. ICES Journal of Marine Science, 71: 1174–1182.

Dimitriadis, C., Evagelopoulos, A., Koutsoubas, D. (2012). Functional diversity and redundancy of soft bottom communities in brackish waters areas: localvs.regional effects. Journal of Experimental Marine Biology and Ecology, 426, 53–59.

Dinmore, T. A., Duplisea, D. E., Rackham, B. D., Maxwell, D. L., Jennings, S. (2003). Impact of a large-scale area closure on patterns of fishing disturbance and the consequences for benthic communities. ICES Journal of Marine Science, 60(2), 371–380.

European Commission, 2022. MSFD CIS Guidance Document No. 19, Article 8 MSFD, May 2022. Pages 152 and 156 and on talk about spatial aggregation and integration for D6. https://circabc.europa.eu/ui/group/326ae5ac-0419-4167-83ca-e3c210534a69/library/36bab4b1-c8d3-4f47-8a62-4977309bb7dd/details

Folke, C., Carpenter, S., Walker, B., Scheffer, M., Elmqvist, T., Gunderson, L., Holling C.S. (2003). Regime shifts, resilienc and biodiversity in ecosystem management. Annual Review of Ecology, Evolution, and Systematics, 35, 557–551.

González-Irusta, J.M., Wright, P.J. (2017). Spawning grounds of whiting (*Merlangius merlangus*). Fisheries Research, 195, 141–151.

González-Irusta, J.M., de la Torriente, A., Punzón, A., Blanco, M., Serrano, A. (2018). Determining and mapping species sensitivity to trawling impacts: the BEnthos Sensitivity Index to Trawling Operations (BESITO). ICES Journal of Marine Science, 75 (5),: 1710–1721.

Hammar, L., Schmidtbauer Crona, J., Kågesten, G., Hume, D., Pålsson, J., Aarsrud, M., Mattson, D., Åberg, F., Hallberg, M., Johansson, T. (2018). Symphony: Integrerat planeringsstöd för statlig havsplanering utifrån en ekosystemansats (Havs-och vattenmyndighetens rapport No. 2018:1). Havs- och vattenmyndigheten, Göteborg.

Harvey, C.J., Kelble, C.R., Schwing, F.B. (2017). Implementing "the IEA": using integrated ecosystem assessment frameworks, programs, and applications in support of operationalizing ecosystem-based management. ICES Journal of Marine Science. https://academic.oup.com/icesjms/articlelookup/doi/10.1093/icesjms/fsw201 (Accessed 26 January 2018).

HELCOM (2010). Ecosystem Health of the Baltic Sea 2003–2007: HELCOM Initial Holistic Assessment. Baltic Sea Environmental Proceedings No. 122.

HELCOM (2022). Cumulative impact from physical pressures on benthic biotopes. HELCOM indicator report. Online. https://portal.helcom.fi/workspaces/EN-BENTHIC-191/Shared%20Documents/CumI/CumI%20indicator%20report/Cumulative-impact-indicator-report-2022-04-13.docx

Hiddink, J.G., Kaiser, M.J., Sciberras, M., McConnaughey, R.A., Mazor, T., Hilborn, R., ... Jennings, S. (2020). Selection of indicators for assessing and managing the impacts of bottom trawling on seabed habitats. Journal of Applied Ecology, 57(7), 1199–1209.

Hiddink, J.G., Rijnsdorp, A.D., Piet, G. (2008) Can bottom trawling disturbance increase food production for a com-mercial fish species? Canadian Journal of Fisheries and Aquatic Sciences, 65, 1393–1401.

Hinz, H., Prieto, V., Kaiser, M.J. (2009). Trawl disturbance on benthic communities: chronic effects and experimental predictions. Ecological Applications, 19(3), 761–773.

Hilborn, R. 2011. Future directions in ecosystem based fisheries management: A personal perspective. Fisheries Research, 108: 235–239.

Hobday, A. J., Smith, A. D. M., Stobutzki, I. C., Bulman, C., Daley, R., Dambacher, J. M., Deng, R. A., *et al.* 2011. Ecological risk assessment for the effects of fishing. Fisheries Research, 108: 372–384.

Holland, D.S., Sutinen, J.G. (2000). Location choice in New England trawl fisheries: old habits die hard. Land Economics, 133–149.

Holmlund, C.M., Hammer, M. (1999). Ecosystem services generated by fish populations. Ecological Economics, 29, 253–268.

IA (2017) (OSPAR). https://oap.ospar.org/en/ospar-assessments/intermediate-assessment-2017/introduction/ospar-and-intermediate-assessment-2017/

ICES (2012). Report of the Workshop on Benchmarking Integrated Ecosystem Assessments (WKBEMIA). http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/SSGRSP/2012/WKBEMIA12.pdf (Accessed 16 November 2020).

ICES (2017) – Bottom trawl survey

ICES (2021). OSPAR request on the production of spatial data layers of fishing intensity/pressure. In Report of the ICES Advisory Committee, 2021. ICES Advice 2021, ospar.2021.11. https://doi.org/10.17895/ices.advice.8297

ICES (2021). ICES advice to the EU on how management scenarios to reduce mobile bottom fishing disturbance on seafloor habitats affect fisheries landing and value. In: Report of the ICES Advisory Committee, 2021. ICES Advice 2021. sr.2021.08. https://doi.org/10.17895/ices.advice.8191

ICES (2022). Workshop to scope assessment methods to set thresholds (WKBENTH2). ICES Scientific Reports. 4:70. 99 pp. http://doi.org/10.17895/ices.pub.20731537

Jac, C., Desroy, N., Certain, G., Foveau, A., Labrune, C., Vaz, S. (2020a). Part 1: Generic sensitivity indices to measure the effect of trawling on benthic mega-epifauna. Ecological Indicators, 117, 106631.

Jac, C., Desroy, N., Certain, G., Foveau, A., Labrune, C., Vaz, S. (2020b). Detecting adverse effect on seabed integrity. Part 2: How much of seabed habitats are left in good environmental status by fisheries? Ecological Indicators, 117, 106617.

Jennings, S., Dinmore, T. A., Duplisea, D.E., Warr, K. J., Lancaster, J.E. (2001). Trawling disturbance can modify benthic production processes. Journal of Animal ecology, 70(3), 459–475.

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. Marine Pollution Bulletin, 58(9), 1263–1277.

Kenchington, E., Callery, O., Davidson, F., Grehan, A., Morato, T., Appiott, J., Davis, A., Dunstan, P., Du Preez, C., Finney, J. González-Irusta, J.M., *et al*. (2019). Use of Species Distribution Modeling in the Deep Sea. Canadian Technical Report of Fisheries and Aquatic Sciences 3296, ix + 76 p.

Kuipers, K. J., Hellweg, S., Verones, F. (2019). Potential Consequences of regional species loss for global Species Richness: A quantitative approach for estimating global extinction probabilities. Environmental Science and Ttechnology, 53(9), 4728–4738.

Levin, P.S., Fogarty, M.J., Murawski, S.A., Fluharty, D. (2009). Integrated Ecosystem Assessments: Developing the Scientific Basis for Ecosystem-Based Management of the Ocean. PLoS Biology, 7: e1000014.

Levin, P.S., Kelble, C.R., Shuford, R.L., Ainsworth, C., deReynier, Y., Dunsmore, R., Fogarty, M. J., *et al.* (2014). Guidance for implementation of integrated ecosystem assessments: a US perspective. ICES Journal of Marine Science, 71: 1198–1204.

McQuatters-Gollop, A. (2012). Challenges for implementing the Marine Strategy Framework Directive in a climate of macroecological change. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370: 5636–5655.

Moilanen, A., Lehtinen, P., Kohonen, I., Virtanen, E., Jalkanen, J., Kujala, H. (2022). Novel methods for spatial prioritization with applications in conservation, land use planning and ecological impact avoidance. Methods in Ecology and Evolution, 13(5), 1062–1072.

Muñoz, P.D., Sacau, M., García-Alegre, A., Román, E. (2020). Cold-water corals and deep-sea sponges by-catch mitigation: Dealing with groundfish survey data in the management of the northwest Atlantic Ocean high seas fisheries. Marine Policy, 116, 103712.

Murillo, F.J., Serrano, A., Kenchington, E., Mora, J. (2016). Epibenthic assemblages of the Tail of the Grand Bank and Flemish Cap (northwest Atlantic) in relation to environmental parameters and trawling intensity. Deep Sea Research Part I: Oceanographic Research Papers, 109, 99–122.

Murillo, F.J., Weigel, B., Bouchard Marmen, M., Kenchington, E. (2020). Marine epibenthic functional diversity on Flemish Cap (north-west Atlantic)—Identifying trait responses to the environment and mapping ecosystem functions. Diversity and Distributions, 26(4), 460–478.

Nichols, A. (2022) How large are natural variations in the abundance of biota in seabed ecosystems? Bangor University.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt R., Legendre, P., ... Wagner, H. (2020). Vegan: Community Ecology Package. R package version 2.5–7. https://CRAN.R-project.org/package=vegan

Ortiz, J.J. (2021). Especies oportunistas, ¿mito o realidad? Estudio del impacto de la pesca de arrastre sobre especies de sensibilidad baja.Master's thesis, Universidade de Santiago de Compostela.

Pavlidou, A., Simboura, N., Pagou, K., Assimakopoulou, G., Gerakaris, V., Hatzianestis, I., ... Borja, A. (2019). Using a holistic ecosystem-integrated approach to assess the environmental status of Saronikos Gulf, Eastern Mediterranean. Ecological Indicators, 96, 336–350.

QSR (2023) (OSPAR). https://www.ospar.org/work-areas/cross-cutting-issues/qsr2023

Queirós, A.M., Hiddink, J.G., Kaiser, M.J., Hinz, H. (2006). Effects of chronic bottom trawling disturbance on benthic biomass, production and size spectra in different habitats. Journal of Experimental Marine Biology and Ecology, 335(1), 91–103.

Rice, J., Moksness, E., Attwood, C., Brown, S. K., Dahle, G., Gjerde, K. M., ... Westlund, L. (2012). The role of MPAs in reconciling fisheries management with conservation of biological diversity. Ocean and Coastal Management, 69, 217–230.

Rijnsdorp, A.D., Bolam, S.G., Garcia, C., Hiddink, J.G., Hintzen, N.T., van Denderen, P.D., van Kooten, T. (2018). Estimating sensitivity of seabed habitats to disturbance by bottom trawling based on the longevity of benthic fauna. Ecological Applications, 28(5), 1302–1312.

Rijnsdorp, A.D., Hiddink, J.G., Van Denderen, P.D., Hintzen, N.T., Eigaard, O.R., Valanko, S., … Garcia, C., (2020). Different bottom trawl fisheries have a differential impact on the status of the North Sea seafloor habitats. ICES Journal of Marine Science, 77(5), 1772–1786.

Riva, G. 2022. Applicazione e confronto di indici per la valutazione degli impatti della pesca a strascico demersale sulle comunità epibentoniche dell'Adriatico. MSc Thesis in Marine Biology. Università di Padova. Supervisors: C. Mazzoldi, S. Raicevich. Pp. 106.

Samhouri, J.F., Haupt, A.J., Levin, P.S., Link, J.S., Shuford, R. (2014). Lessons learned from developing integrated ecosystem assessments to inform marine ecosystem-based management in the USA. ICES Journal of Marine Science, 71, 1205–1215.

Serrano, A., de la Torriente, A., Punzón, A., Blanco, M., Bellas, J., Durán-Muñoz, P., ... González-Irusta, J. M. (2022). Sentinels of Seabed (SoS) indicator: Assessing benthic habitats condition using typical and sensitive species. Ecological Indicators, 140, 108979.

Thrush, S.F., Hewitt, J.E., Parkes, S., Lohrer, A.M., Pilditch, C., Woodin, S.A., ... Van Colen, C. (2014). Experimenting with ecosystem interaction networks in search of threshold potentials in real-world marine ecosystems. Ecology, 95(6), 1451–1457.

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. Marine Ecology Progress Series, 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. Marine Ecology Progress Series, 541, 31–43.

Van Denderen, P.D., Bolam, S.G., Friedland, R., Hiddink, J.G., Noren, K., Rijnsdorp, A.D., ... Valanko, S. (2020). Evaluating impacts of bottom trawling and hypoxia on benthic communities at the local, habitat, and regional scale using a modelling approach. ICES Journal of Marine Science, 77(1), 278–289.

Van Denderen, P.D., Törnroos, A., Sciberras, M., Hinz, H., Friedland, R., Lasota, R., … Hiddink, J.G. (2022). Effects of bottom trawling and hypoxia on benthic invertebrate communities. Marine Ecology Progress Series, 694, 13–27.

van Son, T.C., Oug, E., Halvorsen, R., Melsom, F. (2013). Gradients in traits composition and their relation to environmental complex-gradients and structuring processes: a study of marine sediment species communities. The Open Marine Biology Journal, 7(1).

Veach, V., Di Minin,E., Pouzols, F.M., Moilanen, A. (2017). Species Richness as criterion for global conservation area placement leads to large losses in coverage of biodiversity. Diversity and Distributions, 23(7), 715–726.

Verones, Fk, Kuipers, K., Núñez, M., Rosa, F., Scherer, L., Marques, A., Michelsen, O. *et al.* (2022). Global extinction probabilities of terrestrial, freshwater, and marine species groups for use in Life Cycle Assessment. Ecological Indicators 142, 109204.

Virtanen, E.A., Viitasalo, M., Lappalainen, J., Moilanen, A. (2018). Evaluation, gap analysis, and poten-tial expansion of the Finnish marine protected area network. Frontiers in Marine Science, 5, 402.

Williams, P., Gibbons, D., Margules, C., Rebelo, A., Humphries, C., Pressey, R. (1996). A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British birds. Conservation Biology 10, 155–174.

Annex 1: List of participants

Participant	Institute	E-mail	Country (of Institute)
Alice Belin (*observer*)	European Commission DGENV	alice.belin@ec.europa.eu	Belgium
Andrea Pierucci	COISPA Tecnologia and Ricerca	pierucci@coispa.eu	Italy
Andrew Kenny	Cefas	andrew.kenny@cefas.co.uk	UK
Antonia Nystrom Sandman	AquaBiota Water Research	Antonia.Sandman@aquabiota.se	Sweden
Aurélien Boyé	Ifremer, Plouzané	Aurelien.Boye@ifremer.fr	France
Axel Kreutle	Federal Agency for Nature Conservation	axel.kreutle@bfn.de	Germany
Chris Smith	Hellenic Centre for Marine Research	csmith@hcmr.gr	Greece
Daniel van Denderen	DTU	pdvd@aqua.dtu.dk	Denmark
Elina Virtanen	SYKE	Elina.A.Virtanen@syke.fi	Finland
Ellen Kenchington (*Chair*)	Fisheries and Oceans Canada	ellen.kenchington@dfo-mpo.gc.ca	Canada
Gert Van Hoey	The Flanders Research Institute for Agriculture, Fisheries and Food	Gert.vanhoey@ilvo.vlaanderen.be	Belgium
Giada Riva	University of Padua	giada.riva@studenti.unipd.it	Italy
Grete Elisabeth Dinesen	DTU Aqua, National Institute of Aquatic Resources	gdi@aqua.dtu.dk	Denmark
Henrik Nygård	Finnish Environment Institute	henrik.nygard@syke.fi	Finland
Jan Geert Hiddink	Bangor University	j.hiddink@bangor.ac.uk	UK
Kenneth Patterson (*observer*)	European Commission DGMARE	Kenneth.Patterson@ec.europa.eu	Belgium
Liam Matear	Joint Nature Conservation Committee	Liam.Matear@jncc.gov.uk	UK
Maider Plaza	Spanish Institute of Oceanography	maider.plaza@ieo.csic.es	Spain
Marie-Julie Roux	Fisheries and Oceans Canada	Marie-Julie.Roux@dfo-mpo.gc.ca	Canada
Marie-Louise Krawack	Ministry of Environment and Food of Denmark	makra@mfvm.dk	Denmark
Marina Pulcini	Institute for Environmental Protection and Research	marina.pulcini@isprambiente.it	Italy
Nadia Papadopoulou	Hellenic Centre for Marine Research	nadiapap@hcmr.gr	Greece
Pascal Laffargue	Ifremer	Pascal.Laffargue@ifremer.fr	France

Participant	Institute	E-mail	Country (of Institute)
Paul Coleman	Marine Institute	paul.coleman@marine.ie	Ireland
Petra Schmitt	BIOCONSULT Schuchardt and Scholle GbR	schmitt@bioconsult.de	Germany
Sander Wijnhoven	Ecoauthor	sander.wijnhoven@ecoauthor.net	Netherlands
Sandrine Vaz	Ifremer	Sandrine.Vaz@ifremer.fr	France
Saša Raicevich (*Chair*)	Institute for Environmental Protection and Research	sasa.raicevich@isprambiente.it	Italy
Sebastian Valanko	International Council for the Exploration of the Sea	sebastian.valanko@ices.dk	Denmark
Sofia Reizopoulou	Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC)	sreiz@hcmr.gr	Greece
Ulla Fernández	Centro Oceanográfico de Santander	ulla.fernandez@ieo.csic.es	Spain

Annex 2: Resolutions

Workshop to evaluate proposed assessment methods and how to set thresholds for assessing adverse effects on seabed habitats (WKBENTH3)

2022/WK/HAPISG The Workshop to evaluate proposed assessment methods and how to set thresholds for assessing adverse effects on seabed habitats (WKBENTH3), chaired by Ellen Kenchington (Canada) and Saša Raicevich (Italy), will be established and will meet in Copenhagen, Denmark, 3–7 October 2022 to:

a. Evaluate proposed assessment methods and how to set thresholds for assessing adverse effects on seabed habitats produced in WKBENTH2 and peer-reviewed, using the agreed upon criteria, methods and analysis of their performance therein established, with an emphasis on relevance to the MSFD Descriptor 6 (Seabed integrity is at a level that ensures that the structure and functions of the ecosystems are safeguarded and benthic ecosystems, in particular, are not adversely affected) and associated Annexes III and IV;
b. Prepare worked examples using suitable methods on how to set threshold and assess adverse effects on seabed habitat quality for relevant pressures and impacts listed in Annex III and environmental targets listed in Annex IV of the MSFD;
c. Based on ToRs a and b, prepare an overview of recommended assessment methods for application to MSFD Descriptor 6;
d. Provide higher level guidance on future directions for improvements to the recommended methodology presented in ToR c and for developing scientifically-based ‘extent’ indicators for assessing adverse effects of human-induced pressures on seabed habitats;
e. Provide higher-level guidance as to a set of criteria, and methods to analyse the performance of assessment methods and how to set thresholds for assessing adverse effects on seabed habitats. Documentation should ensure that the exercise to evaluate methods can be redone in future.

WKBENTH3 will report by 14 October 2022 for the attention of the Advisory Committee.

Supporting Information

Priority	High, in response to the stepwise process of delivering guidance on seabed integrity for the Marine Strategy Framework Directive (MSFD). The workshop outputs will feed into ICES WGFBIT and the ongoing efforts to provide guidance on assessment methods to set threshold and assess adverse effects on seabed habitats in the operational implementation of the MSFD.
Scientific justification	Term of Reference a-d) Based on WKBENTH2 and technical service, as well as a peer review of the reports TOR a will evaluate proposed assessment methods and how to set threshold for assessing adverse effects on seabed habitats using agreed upon criteria, methods and analysis of their performance, with an emphasis on the MSFD. Agreed and peer review of criteria on what makes a good indicator, in general (e.g. WGECO, Rice and Rochet 2005) and specifically for assessing the seabed habitats (e.g. WKBENTH 2017). The evaluation should facilitate production of formal ICES advice on the suitability and shortcomings of any proposed indicators for

	MSFD assessment purposes, reflecting their performance to assess the parameters specified in Commission Decision (EU) 2017/848 on condition of seabed habitats and the adverse effects of key pressures. Evaluation should also take into account the indicators applicability across MSFD broad habitat types (or subtypes), their suitability for large sea areas (i.e. all marine waters of MS, marine regions or subregions). Quantitative and qualitative analytical approaches, as well as worked examples, should be used to illustrate suitability of methods to set threshold and assess adverse effects on seabed habitats. Options for setting thresholds should take into accouns as far as possible recent work by EU's TG SeaBed on threshold values for adverse effects on habitat condition (D6C5) and for the maximum allowable extent of habitat loss (D6C4) and of adverse effects (D6C5) Ref. document GES_26-2022-13. Options on setting thresholds should identify higher level criteria that can be used to identify values (or ranges of values) for the indicators which would distinguish a habitat in good condition from the one which is adversely affected or lost (in general or by specific pressures) to set thresholds. This should, for example, reflect on whether there is a linear or non-linear response of the habitat to particular pressures. Consolidate a review of proposed assessment methods based on peer review of WKBENTH2 and technical service. The aim of this TOR is to agree for advice production purposes a detailed review of indicators used, or under development, by Regional Sea Conventions, Member States and ICES, for assessing the state/condition of seabed habitats suitable for MSFD assessments. The indicators considered can also include peer-reviewed indicators which have large-scale application. Provide a detailed review of indicators used, or under development, by Regional Sea Conventions (RSCs), Member States and ICES, for assessing the state/condition of seabed habitats and relevant existing literature. This should include indicators based on both direct observational data and on models. Relevant indicators to be reviewed include those of RSCs for quality status assessments, of Member States for MSFD purposes such as under the Water Framework Directive (WFD) and the Habitats Directive (HD), and those used by ICES. The review should specify the input data, how it is processed, the parameters of habitat quality used, how quality is quantified, any threshold values used, the applicable seabed (habitat) and pressure types, how the output is expressed, and how confidence and uncertainty are handled.
Resource requirements	ICES secretariat and advice process.
Participants	Workshop with researchers and RSCs investigators If requests to attend exceed the meeting space available ICES reserves the right to refuse participants. Choices will be based on the experts' relevant qualifications for the Workshop. Participants join the workshop at national expense.
Secretariat facilities	Data Centre, Secretariat support and meeting room.
Financial	Covered by DGENV special request.
Linkages to advisory committees	Direct link to ACOM.
Linkages to other committees or group	Links to HAPISG and SCICOM.
Linkages to other organizations	Links to RSCs and EC.

Annex 3: Feedback from TG SEABED on WKBENTH2 report and technical service

Below an overview of the main comments arising from the feedback from TG SEABED on WKBENTH2 report and technical service is provided. More detailed comments are given in the document 'feedback from TGSEABED submitted to ICES 05092022' that is annexed to the WKBENTH2 report. In general, the ICES WKBENTH2 report and technical service presents an interesting and useful compilation of benthic indicators currently in use and various methods of setting threshold values. The workshop was attended by ecologists, academics and scientists involved in implementing the MSFD and/or RSC assessments but the representation of the regions was very unbalanced. The technical service offers no analysis or summary statistics e.g. how many of these address D6C3, D6C5 or both, have thresholds already, work with macrobenthos or larger epibenthos or both (this is not a trivial issue as monitoring time and resources are hugely different). The report needs some revision and tiding up of the terminology to be aligned with the MSFD D6 terms (e.g. good environmental status – not state; broad habitat type – not broad scale habitat) and increasing the specific link to D6 criteria and thresholds and as much as possible.

The benefits from the scoring indicators exercise for TG Seabed work are not clear. There are many national and regional indicators developed for MSFD assessments. Each of them is designed to assess a specific MSFD criterion, specific pressures and specific components of benthic habitats and a set of indicators is needed to assess benthic habitat status and impacts. ICES advice should therefore include a compilation of pros and cons of each indicator, and not a simple score for each indicator, which is rather meaningless. The advantage of the comparison of risk-based indicators is also not clear. Habitat types shall be ranked according to their sensitivity and impact. Again, it would be preferable to describe pros and cons of the approaches.

Numerous methods for threshold setting are described and this was a valuable part of the workshop/work. However, the methods seem primarily suitable for individual indicators with single metrics (abundance/biomass), and not for establishing quality and extent thresholds for D6C4 and D6C5. Most threshold setting approaches are statistical methods that need extensive datasets (time-series), also from undisturbed areas. Several of the presented methods were not considered useful to determine the overall GES of habitats by workshop participants (chapter 4.3). The report seems to be implying that devoting more time to get more data to look at more trends and to perform more analyses to be able to further develop methods and to arrive to a 'perfect science' defined threshold is somewhat desirable/favoured. However, over promoting data hungry approaches will lead to delays, imbalances and reduced coherence between data rich and data poor regions. The advice therefore should also include suggestions on setting threshold values for quality and extent of adverse effects, when data availability is poor, as is the case in most regions and habitats. Concerning the extent threshold, there is a lot of scientific literature available on the requirement for undisturbed habitat area and the extent of these areas with regard to conservation objectives. This has not been considered in the workshop report. Proposing suggestions and solutions that can be practical, realistic and both doable across Europe and doable now, should be the top priority and the main concern/guiding principle following also the precautionary approach – this is important as indicators and thresholds are needed now.

Any final proposals should include data requirements (e.g. length of time-series, data from undisturbed areas, BHTs and depth coverage, type of monitoring and benthic data, needs for validation of risk-based approaches for habitats lacking actual empirical data etc.) and areas/BHTs where proposed methods need further development before they can be applied with confidence. In this way operationality is explicitly considered in a transparent way.

Annex 4: Characteristics of evaluated benthic habitat quality indicators

Table 1. Description of 10 benthic habitat quality indicator methods (and their variants) applied to a common dataset (Annex 5) representing primarily trawling intensity, but also eutrophication and pollution gradients. Indicator methods are grouped by link to the EU Decision 2017/848 Descriptor 6 (Seabed Integrity), by the indicator, and whether the indicator method was scored against the evaluation criteria proposed and applied by WKBENTH2. *Indicates indicator method scored against criteria established in and by WKBENTH2.

Predominantly linked to 2017/848 Descriptor 6: D6C5[3]		
Habitat Quality Indicator: Condition of Benthic Habitats		
Indicator Method	**Description**	**Method(s) to Calculate**
Multivariate AZTI Marine Biotic Index (M-AMBI)* M-AMBI ≈ M-AMBI*(n) = (S (n) + H’(n) + AMBI-BC(n))/3 Where M-AMBI*(n) is the normalized value where the Species Richness (number of species, S), diversity (Shannon Index, H’) and AMBI-BC are calculated or each sample (n). Diversity is calculated from numerical abundances, with logarithm set to base 2. BC is calculated on the basis of the list of taxa with the assigned ecological groups supplied by AZTI-Tecnalia (http://ambi.azti.es/), with a null weighting given to the species that are not listed. *See also M-AMBI (n) and bivariate versions. Reference: Sigovini *et al.* (2013). Hydrobiologia. https://link.springer.com/article/10.1007/s10750-013-1565-y	M-AMBI is a multimetric index for assessing the ecological quality status of marine waters. It is based on benthic macroinvertebrates and integrates AMBI, a biotic index based on species sensitivity/tolerance, with Shannon diversity and richness.	- The origins of the M-AMBI algorithm integrates the three metrics by means of factor analysis (FA). When M-AMBI is calculated, no factor is discarded after PCA is performed, and the Varimax rotation is applied to the original space. It is now suggested that this step by omitted - M-AMBI is also closely approximated by the simple mean of the normalized metrics with no need for multivariate techniques. -A bivariate version highly correlated with M-AMBI, can be calculated whereby the constitutive metrics are reduced to a diversity measure and a species sensitivity index.

[3] D6C5: The extent of adverse effects from anthropogenic pressures on the condition of the habitat type, including alteration to its biotic and abiotic structure and its functions (e.g. its typical species composition and their relative abundance, absence of particularly sensitive or fragile species or species providing a key function, size structure of species), does not exceed a specified proportion of the natural extent of the habitat type in the assessment area.

Danish Quality Index (DKIver2)* $$DKI = ((1- ((AMBI-AMBI_{min})/7))+ (H/H_{max}))/2 * (1-(1/N))$$ where H is the Shannon–Wiener index with log base 2, and H_{max} is the highest value that H reaches in undisturbed condition, N is the number of individuals in the sample. DKI can attain values between 0 and 1. If N=1 then DKI=0. Where H_{max} = f (salinity), $AMBI_{min}$ = f (salinity), N = Number of individuals. Also: H/H_{max} must never be > 1, if so it should be set to 1. 2) $AMBI_{min}$ must never be negative, if so it should be set to 0. DKI values can vary between 0 and 1 and may be regarded as EQR values where the 'reference' is the best value we can get at a given salinity. Reference: Carstensen *et al.* (2014). Danish Centre for Environment and Energy. http://dce2.au.dk/pub/SR93.pdf	DKIver2 a multimetric index consisting of a sensitivity component (AMBI), a diversity component (Shannon–Wiener diversity) and a factor including number of individuals to compensate for low densities. The diversity and sensitivity components are adjusted for salinity, as changes in this parameter influences both species number and composition regardless of status.	- The diversity component (H) and the sensitivity component (AMBI) are both normalized to attain a value between 0 and 1, and the diversity is normalized against the highest diversity observed in the area. The components are also adjusted to fit low salinity and low diversity environments.
Relative Margalef Diversity (D_M') (OSPAR BH2b) Assessed Margalef value: D_{ass} = (S-1)/ln(N) where, S is the number of species and N is the total number of individuals in the sample. Relative Margalef value: $D_M' = D_{ass} / D_{ref}$ where, D_{ref} is the case-specific reference Margalef value estimated based on no- or low-pressure areas, estimated by selecting a standardized specific percentile value for Margalef diversity from the reference set dependent on relative pressure level. Updated methodology applied for OSPAR QSR2023: assessment results for the Greater North Sea Region for the period 1998–2021 including updated CEMP protocol coming available from OSPAR website soon (https://oap.ospar.org/en/ospar-assessments/quality-status-reports/qsr-2023/).	This is a biodiversity index which gives the Species Richness of a sample divided by the logarithm of the total abundance within the same sample.	- Margalef's index of diversity calculated as Species Richness (S-1) divided by abundance (ln(N)); a relative value (Relative Margalef diversity) is calculated by dividing the Assessed Margalef by a Reference Margalef (defined at the level of Assessment Units (AU), Broad Habitat Types (BHT), Monitoring techniques, etc, or combinations of those) to make results comparable. ('Relative Margalef diversity' used to be called 'Normalized Margalef diversity' in former versions of the indicator). Benthic community diversity is expected to be representative for the benthic habitat quality status.

BENTIX $$BENTIX = \{ 6 \ X \ \%GI) + 2 X \ (\% GII + \% GIII)\}/100.$$ Where GI includes species sensitive to disturbance in general, GII includes species tolerant of disturbance or stress, and GIII includes opportunistic species. Reference: Simboura and Zenetos (2002). Mediterranean Marine Science. https://ejournals.epublishing.ekt.gr/index.php/hcmr-med-mar-sc/article/view/12221	BENTIX is a biotic index based on the concept of indicator species and uses the relative percentages of two general ecological groups of species, the 'tolerant' and the 'sensitive' grouped according to their sensitivity or tolerance to disturbance factors.	-The BENTIX index uses the relative contribution of generally tolerant and sensitive taxa, recombining the five ecological groups (GI-GV) used in the AMBI index and weighting them according to the ratio of their occurrence in the benthic fauna. - The selection of the weight coefficients in the BENTIX formula is not random and it is based on the realization that the probability of a benthic species picked up randomly, to be tolerant of stress is 3:1. This ratio is multiplied by 2 to create a scale ranging from 2 to 6. The 'sensitive' taxa group GS, including all sensitive (GI) and indifferent (GII) species is weighted by 6 and the 'tolerant' taxa group GT, including all tolerant (GIII), first (GV) and second order opportunistic species (GIV) are equally weighted by 2. - In order to include structural components of benthic communities for the purposes of the MSFD, a formula has been developed combining the BENTIX index with diversity indices using specific reference values for different ecotypes (which still has to be tested and validated with pressure data bathymetric zones and habitat types outside the coastal zone).
Preliminary link to 2017/848 Descriptor 6: D6C3[4]		
Habitat Quality Indicator: Impact of Physical Disturbance on Benthic Habitats		
Indicator Method	**Description**	**Method(s) to Calculate**
Population Dynamic 2 (PD2)* RBS = B/K = 1- F d/r Relative benthic status (RBS, defined as the biomass B relative to the carrying capacity K), derived by solving the logistic population growth equation for the equilibrium state. Trawling effort F= SAR, defined as the total area swept by trawl gear	Assesses the sensitivity of the seabed to bottom-contacting fishing gear, in order to determine the impact/status of the seabed. Uses habitat- and gear-specific mortality and recovery dynamics to	- The PD method is a risk-based mechanistic model that estimates the total reduction in community biomass (B) relative to carrying capacity (K), corresponding to the estimated fishing intensity. Total community biomass relative to carrying capacity (B/K) describes the equilibrium state, i.e. the interaction between the depletion caused by fishing and the recovery of the benthic community. The impact is given by 1−B/K. The depletion rates are estimated from a meta-analysis providing gear-specific depletion rates. Recovery rates are derived from a longevity-specific meta-analysis.

[4] D6 criteria element: Benthic broad habitat types or other habitat types, as used under Descriptors 1 and 6. D6C3: Spatial extent of each habitat type which is adversely affected, through change in its biotic and abiotic structure and its functions (e.g. through changes in species composition and their relative abundance, absence of particularly sensitive or fragile species or species providing a key function, size structure of species), by physical disturbance.

within a given area of seabed in one year divided by that area of seabed (units y-1=). Depletion *d* is the fraction mortality per trawl pass estimated from experimental trawling studies. *r* is the intrinsic rate of population increase. Reference: ICES. 2019. Interim Report of the Working Group on Fisheries Benthic Impact and Trade-offs (WGFBIT), 12–16 November 2018, ICES Headquarters, Copenhagen, Denmark. ICES CM 2018/HAPISG:21. 74 pp. https://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/HAPISG/2018/01%20WGFBIT%20-%20Report%20of%20the%20Working%20Group%20on%20Fisheries%20Benthic%20Impact%20and%20Trade-offs.pdf	derive local impact scores that can be used to assess this indicator or separately to describe aspects of the Condition of Benthic Habitats.	- This method assumes that the sensitivity to trawling depends on the longevity of species and communities.
Sentinels of the Seabed (SoS) (OSPAR BH1)* Fraction (proportional occurrence) of sensitive species. “Frequent or typical species” (species frequently found in the natural habitat and species sensitive to studied pressure) selected on basis of *i*) relative contribution of species to intra-habitat similarity in target habitat under reference condition (no disturbance or very low disturbance) using Similarity Percentages procedure (SIMPER; Clarke, 1993) and *ii*) relative frequency for each species within the target habitat under reference conditions. ‘Sentinel species“ filtered from “frequent or typical species” by prioritizing species according to a SoS sensitivity index (species responses to the analysed pressure). Sensitivity to trawling disturbance based on BEnthic Sensitivity Index to Trawling Operations (BESITO, González-Irusta *et al.*, 2018) classification of species, Sensitivity to pollution defined by AMBI classification of species (Borja *et al.*, 2000), Potentially other sensitivity classifications related to other pressures can be used. Reference: Serrano *et al.* (2022). Ecological Indicators. https://www.sciencedirect.com/science/article/pii/S1470160X22004502?dgcid=rss_sd_all	Assesses the environmental status of a habitat using the proportion (in biomass or number) of sentinel species in the habitat across a pressure gradient.	- The SoS indicator requires three types of data: 1) biological data (quantified species inventory generally of biomass or abundance), 2) environmental data, 3) pressure data. First, a ‘typical species set’ is computed using intra-habitat similarity and frequency under reference conditions. Second, the ‘sentinel species set’ is generated by selecting the most sensitive species from the typical species set. This selection is made using specific indices able to assess species sensitivity to a particular pressure. Changes in the proportion of sentinel species across a pressure gradient can be computed, allowing to generate pressure-state curves. - Also relevant to D6C5.
Trawling Disturbance Index (TDIs)* TDI: Groups (Gx) of sensitivity (SI: based on the sum of scores of traits) G1: 0≤SI≤4, G2: 5≤SI≤7, G3: 8≤SI≤10, G4: 11≤SI≤13, G5: 14≤SI≤15. $TDI_x = \frac{Log1 \times Log(G1_x+1) + Log(G2_x+1) + Log4 \times Log(G3_x+1) + Log8 \times Log(G4_x+1) + Log16 \times Log(G5_x+1)}{Log(N_x+1)}$ where G1x-G5x were the total abundances/biomasses of each group in the xth observation and Nx the total abundance/biomass of the xth observation	TDI is based on the biological traits composition of epifaunal species that determine vulnerability to trawling: mobility, fragility, position on substrata, average size and feeding mode. Less disturbed communities	- Each biological trait is assigned a score depending on its vulnerability to trawling: score 0 – traits advantageous to support trawling; 1– traits that determine low vulnerability to trawling; 2 – moderate vulnerability; 3 – high vulnerability. A total score for each species is calculated. The abundances of all organisms with the same total score within a replicate are summed (and recorded as a % of total) and the species abundance data are transformed into a scores’ abundance dataset. The scores were classed into 5 groups (G1–5) with G5 having the highest scores and hence vulnerability to trawling. Principal Coordinates Analysis based on the Bray–Curtis similarity (PCO) ordinated the samples in two dimensions and each axis is correlated with the environmental variables. The

mTDI is a weighted average of the Sis of the community with Nx, the number of taxons in the xth observation; Bix, abundance of the ith taxon in the xth observation; Bnx, summed abundances of the xth observation and SIi, the sensitivity index (SI) of the ith taxon: $mTDI_x = \sum_1^{N_x} \frac{Bi_x}{Bn_x} \times SI_i$ pTDI is a modification of mTDI to focused only on sensitive species (SI >7). $pTDI_x = \sum_1^{N_x} \frac{Bij_x}{Bn_x} \times SI_{ij}$ with Bijx, abundance of the ith taxon of the list j of sensitive taxon (SI>7) in the xth observation; and SIij, SI of the ith taxon of the list j of vulnerable taxon, ; Bnx, summed abundance of the xth observation (including all observed taxa) mT = the modified vulnerability index results from a generic framework that can be adapted to any kind of pressure and related trait. An addidtional trait describing the protection status of the species was also added. Traits were classified as direct primary (F_{i1-3}: position, mobility and size) or agravating (F_{i4}: fragility) or indirect primary (F_{i5-6}: feeding and status) factors that were combined as follow: The direct component of the index, t_i, of each individual taxIn i, is obtained by applying: $t_i = a_i^{1-g_i/(g_i+\gamma)}$ with $a_i = F_{i1} \times F_{i2} \times F_{i3}$, $g_i = F_{i4}$ and $\gamma = 0.5$. The indirect component of the index, si, of the ith taxon is obtained by applying the same equation with ai = $(F_{i5} + F_{i6})/2$ and $g_i = 0$. The modified vulnerability Index (mTx) in then calculated as: $mT_x = -\sum_{i=1}^{N_x} \frac{Bri_x}{t_i \times s_i}$	have the highest abundance of groups considered vulnerable to trawling. mT utilized rescaled traits scores and introduces an additional trait based on the protection status of each species.	PCO was used to test whether fishing effort is the variable having the highest correlation with the functional group composition. - Also relevant to D6C5.

with Bri_x, relative abundance of the ith taxon of the station x and N_x the total number of taxon of the station x. Reference: Jac *et al.* (2020a). Ecological Indicators. https://doi.org/10.1016/j.ecolind.2020.106631		
Fraction of community longevity exceeding trawling interval (L1) The impact of trawling is estimated as the proportion of biomass of those taxa with a longevity exceeding the reciprocal trawling intensity (L = 1/T), such that: $I_{L1} = 1 - \frac{\exp(\alpha + \beta_L \ln(\frac{1}{T}) + \beta_H H + \beta_T \ln(T_0) + \beta_{HL} H \ln(\frac{1}{T}) + \beta_{HT} H \ln(T_0))}{(1 + \exp(\alpha + \beta_L \ln(\frac{1}{T}) + \beta_H H + \beta_T \ln(T_0) + \beta_{HL} H \ln(\frac{1}{T}) + \beta_{HT} H \ln(T_0)))}$ described by the cumulative biomass (B) as a function of longevity (L), habitat (H) and trawling intensity (T). Reference: Rijnsdorp *et al.* (2020). PLoS One. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7703930/pdf/pone.0228528.pdf	L1 estimates the proportion of the biomass of the benthic community that is potentially impacted by trawling. It assumes that benthic taxa with a longevity of more than the average interval between two successive trawling events will be potentially affected by bottom trawling.	- A more precautionary indicator than L2, related to the longevity composition of the benthic community - Because the impact is estimated relative to the untrawled community, a value of T0 = 0.01 was included to avoid taking the log of zero. - This indicator may be correlated with the ability of the community to recover after disturbance.
Longevity-based indicator (L2) $T_c = \exp\left[(\ln(1 - pB_{ref}/(pB_{ref})) - (\beta_0 + \beta_1 \ln(10) + \beta_2 H + \beta_5 \ln(10) H)/(\beta_3 + \beta_4 H + \beta_6 \ln(10))\right]$ Seabed sensitivity was estimated as the critical trawling intensity (T_c) at which the biomass of long-lived taxa (>10 yr) is reduced to a proportion (p) of the untrawled biomass B_{ref}. Reference: Rijnsdorp *et al.* (2018). Ecological Applications. https://esajournals.onlinelibrary.wiley.com/doi/10.1002/eap.1731	The indicator estimates the decrease in median longevity in response to trawling. Median longevity is the longevity where 50% of the community biomass is above/below. The decrease is based on a statistical relationship between trawling intensity and benthic longevity from the North Sea.	- Mixed effect modelling with cumulative biomass as a response variable and longevity, trawling and environmental conditions as predictor variables. - This indicator may be correlated with the ability of the community to recover after disturbance.
Risk-based approaches linking predominantly to 2017/848 Descriptor 6: D6C3		

Habitat Quality Indicator: Impact of Physical Disturbance on Benthic Habitats		
Indicator Method	**Description**	**Method(s) to Calculate**
Cumulative Impact from physical pressures on benthic biotopes (CumI) Reference: Hoppe *et al.* (2015) HELCOM portal. https://portal.helcom.fi/meetings/IN%20BENTHIC%202-2018-573/MeetingDocuments/3-2%20ATT.2%20Cumulative%20impact%20on%20benthic%20biotopes-IndicatorReport-FINAL-CORESET%20II%20version.pdf	This indicator evaluates the combined effect from several physical pressures on the benthic biotopes, as only studying single pressures in isolation does not provide an adequate evaluation of seabed integrity. This HELCOM indicator reflects the pressures from physical damage in the form of the altered environmental status (in short: state change) resulting from the respective pressures.	- A biotope map with sensitivity information will be evaluated against the individual pressure maps (using the magnitude of pressure) separately. This results in a set of maps with potential impacts on the benthic biotopes. The different impact maps are then cumulated using a hierarchical approach. Currently, the underlying data do not allow to represent the indicator using exact numerical values. Until this is possible a more descriptive approach can be applied. The process of determination of the cumulative impact on benthic biotopes is done spatially using vector data and GIS software. The spatial resolution of the pressure data should correspond to the resolution of the given biotope map.
Extent of physical disturbance to benthic habitats (BH3) Reference: OSPAR portal. https://oap.ospar.org/en/ospar-assessments/intermediate-assessment-2017/biodiversity-status/habitats/extent-physical-damage-predominant-and-special-habitats/	This Indicator is designed to assess impacts on all subtidal habitat types at a subregional level. It uses a combination of spatial analysis, subdividing the sea arlas into a grid, to extrapolate data and knowledge from existing monitoring and local studies to larger areas.	- The indicator uses two types of information: the distribution and sensitivity of habitats (resilience and resistance); and the distribution and intensity of human activities and pressures that cause physical damage (e.g. mobile bottom gear fisheries, sediment extraction and offshore constructions). Sensitivity and pressure (literature-based) are combined to calculate the potential damage to a given seabed habitat, and the trends across a six-year period (2010–2015).

Annex 5: Metadata for common dataset

Name Code: AS1

Area: Adriatic Sea
Broad Habitat Type: Circalittoral sand
Depth Range: 9–56 m
Sampling Gear: Otter trawl
Sampling Program: Italian bottom trawl survey (SoleMon)
Number of stations (samples per station): 12 (1)
Maximum distance between stations (km):
Sampling year (month): 2016 (11)
Data Provider(s): Saša Raicevich
Reference: Riva (2022)
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
TBB/OT	SAR per year average from 2012 to 2016	Sum of OTB and TBB (VMS and AIS data) (Russo *et al.*, 2020; 2021)	Kg wet weight per km^2	Numbers per km^2	Details on depth and NPP available

Name Code: AS2

Area: Adriatic Sea
Broad Habitat Type: Circalittoral mud
Depth Range: 8–87 m
Sampling Gear: Otter trawl
Sampling Program: Italian bottom trawl survey (SoleMon)
Number of stations (samples per station): 16 (1)
Maximum distance between stations (km):
Sampling year (month): 2016 (11)
Data Provider(s): Saša Raicevich
Reference: Riva (2022)
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
TBB/OT	SAR per year average from 2012 to 2016	Sum of OTB and TBB (VMS and AIS data) (Russo *et al.*, 2020; 2021)	Kg wet weight per km^2	Numbers per km^2	Details on depth and NPP available

Name Code: CO

Area: Dutch EEZ high tidal stress area
Broad Habitat Type: Sand
Depth Range: 22–36 m
Sampling Gear: Boxcore (0.078 m²)
Sampling Program: Dutch infauna sampling program (MWTL)
Number of stations (samples per station): 15 (1)
Maximum distance between stations (km): 329
Sampling year (month): 2007 (3–6)
Data Provider(s): Daniel van Denderen
Reference: van Denderen *et al.* 2015
Open Access: Data can be obtained by contacting RWS, Netherlands (https://www.informatiehuismarien.nl/)

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
Beam and Otter Trawls	SAR per year - last 365 days	Estimated from interpolated VMS data from Dutch fisheries on 0.001 x 0.001 grid	Gram ash-free dry weight per 0.078 m²	Numbers per 0.078 m²	Details on sediment type, depth, tidal bed stress are available

Name Code: DB

Area: Dogger Bank
Broad Habitat Type: Sand
Depth Range: 25–30 m
Sampling Gear: Hamon grab (0.1 m²)
Sampling Program: Scientific cruise
Number of stations (samples per station): 7 (5)
Maximum distance between stations (km): 20
Sampling year (month): 2003 (9)
Data Provider(s): Jan Hiddink
Reference: Queirós *et al.* 2006
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
Beam and Otter Trawls	SAR per year		Gram wet weight per 0.5 m²	Numbers per 0.5 m²	Details on sediment type, depth, tidal bed stress are available

Name Code: FC

Area: Flemish Cap, Northwest Atlantic
Broad Habitat Type: Unknown (middle bathyal sediment)
Depth Range: 786–1236 m
Sampling Gear: Lofoten bottom trawl
Sampling Program: Scientific survey
Number of stations (samples per station): 26 (1)
Maximum distance between stations (km):
Sampling year (month): 2007 (5–7)
Data Provider(s): P. Durán Muñoz, M. Sacau
Reference: Murillo *et al.* (2016)
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
Bottom Trawls	Number of pings by square km	Estimated from VMS data from international fisheries and calculated as the sum of pings by cell and year (can be translate into hous/km^2)	Gram wet weight per km^2		Details on depth are available

Name Code: FG

Area: Fladen Ground
Broad Habitat Type: Mud
Depth Range: 143−153m
Sampling Gear: Day grab (0.1 m^2)
Sampling Program: Scientific survey
Number of stations (samples per station): 14 (5)
Maximum distance between stations (km): 41
Sampling year (month): 2004 (6)
Data Provider(s): Jan Hiddink
Reference: Tillin *et al.* 2006
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
	SAR per year		Gram wet weight per 0.5 m^2	Numbers per 0.5 m^2	Details on sediment type, depth, tidal bed stress are available

Name Code: Finland

Area: Gulf of Finland
Broad Habitat Type:
Depth Range: 56–84 m
Sampling Gear: van Veen grab (0.112 m²)
Sampling Program: Finnish monitoring program
Number of stations (samples per station): 8 (3)
Maximum distance between stations (km):
Sampling year (month): 2015 (5)
Data Provider(s): Henrik Nygard
Reference: SYKE database and ICES database
Open Access: yes - https://www.syke.fi

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
	mg O_2/l	Measured dissolved oxygen concentration (1 meter above seabed)	Gram wet weight per m²	Numbers per m²	Detalis on depth are available

Name Code: Gotland

Area: Gotland
Broad Habitat Type: Muddy sand
Depth Range: 37–59 m
Sampling Gear: van Veen grab (0.1 m²)
Sampling Program: Swedish benthic sampling program
Number of stations (samples per station): 8 (3)
Maximum distance between stations (km):
Sampling year (month): 2012
Data Provider(s): Katja Noren and Mattias Skold
Reference: van Denderen *et al.* 2020
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
Otter Trawl	SAR per year - average over the last 2.5 years	Estimated from interpolated VMS data from Swedish and Danish fisheries and calculated as the sum of the area swept by trawls within a 250 m radius at each sampling location	Gram wet weight per 0.1 m²	Numbers per 0.1 m²	Detalis on depth are available

Name Code: NIC1

Area: Southern Bay of Biscay/Northern Iberian Coast
Broad Habitat Type: Sand (Offshore circalitoral sand)
Depth Range: 71–202 m
Sampling Gear: Otter trawl (see ICES, 2017 for gear specifications)
Sampling Program: Spanish IBTS (DEMERSALES)
Number of stations (samples per station): 20 (1)
Maximum distance between stations (km):
Sampling year (month): 2016
Data Provider(s): José Manuel González, Maider Plaza Morlote, Ulla Fernández
Reference: ICES 2017, Serrano *et al.* (2022)
Open Access: No, under request.

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
Otter Trawl	SAR per year - average over the last 5 years	Estimated from interpolated VMS data from Spanish fisheries and calculated as the sum of the area swept by cell and year	Gram wet weight per km^2	Numbers per km^2	Detalis on depth are available

Name Code: NIC2

Area: Southern Bay of Biscay/Northern Iberian Coast
Broad Habitat Type: Several, but mainly mud (Upper bathyal sediment)
Depth Range: 186–936 m
Sampling Gear: Otter trawl (see ICES, 2017 for gear specifications)
Sampling Program: Spanish IBTS (DEMERSALES)
Number of stations (samples per station): 52 (1)
Maximum distance between stations (km):
Sampling year (month): 2016
Data Provider(s): José Manuel González, Maider Plaza Morlote, Ulla Fernández
Reference: ICES 2017, Serrano *et al.* (2022)
Open Access: No, under request

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
Otter Trawl	SAR per year - average over the last 5 years	Estimated from interpolated VMS data from Spanish fisheries and calculated as the sum of the area swept by cell and year	Gram wet weight per km^2	Numbers per km^2	Detalis on depth are available

Name Code: OxyTrawl

Area: Southern Baltic Sea
Broad Habitat Type: Sand
Depth Range: 70–85 m
Sampling Gear: Boxcore (0.06 m²)
Sampling Program: Scientific cruise
Number of stations (samples per station): 11 (5)
Maximum distance between stations (km):
Sampling year (month): 2018 (9)
Data Provider(s): Daniel van Denderen
Reference: van Denderen *et al.* 2022
Open Access: yes - https://doi.org/10.14284/567

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
Otter Trawl	SAR per year - average of 2012–2018	ICES VMS at 0.05 x 0.05 resolution (data call year is 2019)	Gram wet weight per 0.3 m²	Numbers per 0.3 m²	Details on sediment type, depth and trawling intensity are available

Name Code: PH

Area: Long Forties, northern North Sea
Broad Habitat Type: Gravelly sand
Depth Range: 74–83 m
Sampling Gear: Hamon grab (0.1 m²)
Sampling Program: Scientific cruise
Number of stations (samples per station): 5 (5)
Maximum distance between stations (km): 19
Sampling year (month): 2003 (9)
Data Provider(s): Jan Hiddink
Reference: Tillin *et al.* 2006
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
	SAR per year		Gram wet weight per 0.5 m²	Numbers per 0.5 m²	Details on sediment type, depth, tidal bed stress are available

Name Code: Saronikos

Area: Saronikos Gulf
Broad Habitat Type: Mixed sand / mud
Depth Range: 20–94 m
Sampling Gear: Boxcorer (0.1 m²)
Sampling Program: Saronikos sampling programme 2012
Number of stations (samples per station): 8 (2)
Maximum distance between stations (km):
Sampling year (month): 2012 (2)
Data Provider(s): Sofia Reizopoulou
Reference: Pavlidou *et al.* (2019)
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
	% total N	Data/pressure indicators: % of orgC and total N in the sediment at each sampling location. Pressure is based on total N (correlates with orgC)	Mg dry weight per 0.2 m²	Numbers per 0.2 m²	Details on sediment type, %org C and depth are available

Name Code: SEL

Area: Sellafield
Broad Habitat Type: Muddy sand
Depth Range: 21−42 m
Sampling Gear: Day grab (0.1 m²)
Sampling Program: Scientific cruise
Number of stations (samples per station): 15 (5)
Maximum distance between stations (km): 42
Sampling year (month): 2009 (6)
Data Provider(s): Jan Hiddink
Reference: Hinz *et al.* 2009
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
Otter Trawl	SAR per year		Gram weight per 0.5 m²	Numbers per 0.5 m²	Details on sediment type, depth, tidal bed stress are available

Name Code: SP

Area: Silver Pit
Broad Habitat Type: Muddy sand
Depth Range: 68–78 m
Sampling Gear: Boxcorer (0.078 m^2)
Sampling Program: Scientific cruise
Number of stations (samples per station): 6 (4)
Maximum distance between stations (km): 40
Sampling year (month): 2002 (7)
Data Provider(s): Stefan Bolam
Reference: Jennings *et al.* 2001
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
	Relative frequencies of disturbance on a linear scale	Estimated from aerial survey data collected by fisheries inspection services	Mg dry weight per 0.31 m^2	Numbers per 0.31 m^2	Details on sediment type, depth, tidal bed stress are available

Name Code: TH

Area: Thames
Broad Habitat Type: Sand
Depth Range: 16–40 m
Sampling Gear: Boxcorer (0.078 m^2)
Sampling Program: Scientific cruise
Number of stations (samples per station): 6 (4)
Maximum distance between stations (km): 49
Sampling year (month): 2002 (7)
Data Provider(s): Stefan Bolam
Reference: van Denderen *et al.* 2015
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
	Relative frequencies of disturbance on a linear scale	Estimated from aerial survey data collected by fisheries inspection services	Mg dry weight per 0.31 m^2	Numbers per 0.31 m^2	Details on sediment type, depth, tidal bed stress are available

Name Code: Vigo

Area: Vigo Estuary
Broad Habitat Type: Infralitoral Mud
Depth Range: < 30 m
Sampling Gear: BOUMA boxcorer (0.0175 m²)
Sampling Program: Pollution monitoring program
Number of stations (samples per station): 20
Maximum distance between stations (km):
Sampling year (month): 2004–2006
Data Provider(s): José Manuel González'
Reference: Bellas *et al*. 2011, Beiras *et al*. 2012
Open Access: no

Dominant bottom fisheries in area, if any	Pressure intensity unit	Method to estimate pressure gradient	Biomass unit	Abundance unit	Other environmental information
Small fisheries with no VMS. No trawling	Cumulative pollution index	The CPI index combine several pollutants (e.g. Cd, Hg) in one index. See Bellas *et al.*, 2011; 2012 for a complete description of CPI method		Numbers per km²	Details on depth and specific pollutants are available

Annex 6: Suitability and shortcoming of evaluated benthic indicators

Table 1. Suitability and shortcomings for each of the 10 benthic habitat quality indicator methods (Annex 4). Indicator methods are grouped by link to the EU Commission Decision 2017/848 Descriptor 6 (Seabed Integrity) and by the indicator.

Predominantly link to 2017/848 Descriptor 6: D6C5		
Habitat Quality Indicator: Condition of Benthic Habitats		
Indicator Method	**Suitability**	**Shortcomings**
Multivariate AZTI Marine Biotic Index (M-AMBI)	- Tested along a gradient of organic enrichment and oxygen depletion with some fishing activity differentials - Can be applied across biogeographical boundaries or in individual locales without a pre-existing benthic index - Able to disentangle community changes due to variation along environmental gradients from those due to anthropogenic disturbances in most cases. - Successfully applied for the classification of ecological quality of coastal waters over a wide variety of geographical areas and habitat types against multiple pressures. - M-AMBI*(n), which involves substituting standardization with minimum–maximum normalization, the results become independent of other samples, so no minimum number of samples has to be enforced, and samples can freely be added to (or removed from) the dataset.	- Indicator not specifically defined for response to trawling, difficult to attribute response to specific pressure. - High taxonomic expertise is required; restricted to infauna and thus to grab samples. - One limitation of M-AMBI using FA is that the results depend on the whole set of samples considered, and the addition of new data always leads to different results. - Successfully applied for coastal waters however still to be evaluated for other bathymetric zones and habitat types. - For the application of M-AMBI the results of AMBI are combined with Species Richness and Shannon diversity, also both requiring reference conditions to be set depending on habitat types and depth zones.
Danish Quality Index (DKI)	-- Utilizes routine soft bottom benthic monitoring data (grab samples). Haps cores from the national monitoring can be used if pooled. - Based on quantitative data of the community (abundance). - Based on number and abundance of sensitive species in relation to total abundance, showing responses to pressures at community level. - Originally developed for use in poly- to euhaline benthic environments characterized by a relatively high species diversity, further developed to fit low salinity and low diversity environments. - Developed to measure benthic macrofauna 'quality', in response to any disturbance gradient. - Operationally applied in DK.	- Complicated computation which requires normalization. - Restricted to grab samples (soft bottom habitats) and assessment of whole waterbodies. - Response to pressure and/or management actions is slow

Relative Margalef Diversity (D_M') (OSPAR BH2b)	- Traditional benthic community diversity indicator (using common benthic monitoring data), attempts to compensate for the effects of sample size by dividing the number of species in a sample by the natural log of the number of organisms collected. Achieving comparability among various assessments (e.g. for specific sampling methodologies/procedures, BHTs, AUs) by dividing assessed diversity (at sample level) by a case specific reference diversity, resulting in the relative Margalef diversity (D_M'). - Well established. Used for a variety of human activities (bottom trawling, organic enrichment, inorganic pollutants). - Sufficiently sensitive to distinguish quality differences for a range of pressure levels. - Assessment results easy to communicate in terms of diversity.	- Reference values may be improved by explicitly incorporating depth or salinity gradients as key factors influencing assessment results. - Confidence of assessment results depend on reliability of distinguishing no- or low-pressure areas to estimate reference levels. - Indicator responds to any type of pressure and does not specifically indicate the type of pressure at stake. - Diversity is only one aspect of benthic habitat quality. It is expected that relative high diversity is indicative for presence of all ecosystem functions and minor deterioration.
BENTIX	- Tested along a gradient of organic enrichment and the long-term trends of decline or recovery of the community health in response to dumping. - Successfully applied for the classification of ecological quality of coastal waters in the Eastern Mediterranean over a wide variety of habitat types against various pressures.	- High taxonomic expertise is required. - Successfully applied in coastal waters, however, still to be tested in other bathymetric zones. Depth-specific or habitat-specific reference conditions and thresholds should be tested and evaluated.
Predominantly link to 2017/848 Descriptor 6: D6C3		
Habitat Quality Indicator: Impact of physical disturbance on Benthic Habitats		
Indicator Method	**Suitability**	**Shortcomings**
Population Dynamic 2 (PD2)	- PD2 is a mechanistic model; it estimates the decrease in biomass in response to trawling and the recovery time. - The model needs depletion and recovery parameters of trawl impact relating to infauna and epifauna. - Tested at the North and Baltic Seas at a region level; can also be tested at other scales as parameters required for the PD method were derived from the globally available trawl impact studies. - Strong rooting in general concepts of population dynamics and the fact that it is a single indicator summarizing impact across the entire benthic community. - The biomass component of the PD2 method is a proxy for ecosystem (functioning) processes, for example, nutrient cycling or energy flow through foodwebs. - Estimating relative benthic status requires only maps of fishing intensity and habitat type and parameters for impact and recovery rates,	- Approach requires estimates of the longevity of all species in a community which is not known for many deep-living species. - The current implementation assumes an equilibrium between benthic state and trawling (benthos is adapted to certain level to a certain continuous trawling frequency), but when large changes in trawling pressures occur this may not be accurate. A dynamic implementation of the logistic model may be more appropriate in that case. - Mainly relevant to trawling impacts (has been used to map impacts of hypoxia in the Baltic Sea). - It requires macrobenthic biomass data/longevity from undisturbed areas (which can be hard to find for all BHT). No threshold available yet and the proposed natural variation method for setting a threshold is data heavy (trends from undisturbed areas) and probably area specific.

	which have been taken from meta-analyses of all available studies of towed-gear impacts.	
Sentinels of the Seabed (SoS) (OSPAR BH1) Fraction of sensitive species	- Applicable to any kind of standardized benthic community observation data (applied to trawl, boxcorer and video observation data). - Once Sentinel species have been defined (monitoring technique and Assessment Unit specific) for certain pressure and habitat combinations, these can be used for assessments in other areas, provided that the species are also distributed in those areas. - Potentially applicable with regards to any type of disturbance (physical or chemical), and then highly specific for the type of disturbance. - Highly sensitive to distinguish reduced quality from good (reference) quality situations in the relative low-pressure range (i.e. by use of most sensitive species). - Based on quantitative data of the benthic communities (densities or biomass). - Assessment results easy to communicate with regards to specific pressures (to which methodology is applied). - Operational in Bay of Biscay and Iberian Coast, examples of application from Mediterranean and North-Atlantic, providing ecological input for risk-based approaches like BH3 application in Greater North Sea region. - Could be used as early warning detecting decrease in most sensitive species.	- Definition of sensitive species may leave room for different interpretations. - In case no sentinel species have been defined yet for the specific situation; benthic community data should come with detailed pressure information at the level of individual samples (potentially make use of D6C2 input). - Need for reference (no- or low-pressure) areas and known (single) pressure gradients to identify sentinel species. - Does not provide information about the general quality status (D6C5); provides important results with regards to specific disturbances investigated, rather independent of potential other disturbances at stake (i.e. other pressures in multi-pressure situations might determine the quality status as well). - Sentinel species not necessarily indicative to distinguish quality changes under poor quality conditions (sensitive species might be gone anyway); although there might be options to select less sensitive (distinguishing) species indicative for other part of pressure range (in alternative approach). - Thresholds linked to recovery of habitats not yet available (however assessment results in gradients provide valuable information to justify setting of TVs).
Trawling Disturbance Index (TDIs: TDI, mTDI and pTDI)	- Empirically based state indicator, strong link between species traits and trawling/abrasion pressure. - Epi-megafauna dataset already available for large parts of the continental shelves at European scale (*e.g.* based on the IBTS fisheries surveys network). - Potential utilization of data coming from non-destructive method (e.g. video). - Tested and published results for various soft-bottom ecosystems types (Atlantic/Channel and Mediterranean) - Published species response traits dataset for already analysed set of ecosystems. - mTDI and pTDI: weighting based on abundance or biomass - Variants include mTDI (Modified TDI) and pTDI (Partial TDI) giving flexibility and able to identify early effects on the most sensitive species (pTDI>mTDI): - pTDI : highest theorical sensitivity (indicator restricted to the most sensitive species groups.	- Definition of some traits (e.g. 'fragility') may leave room for different interpretations. - 'Feeding mode' trait is not directly selected for by the gear and is a correlated trait. -Variability of disturbance at intermediate levels of trawling was less evident. - For TDI: Log- relationships between sensitivity groups are scientifically unfounded. - mT indicator introduces criteria (protection status) independent of trawling pressure which makes it less specific to the pressure and difficult to interpret.

Longevity-based indicators L1 and L2	- Risk-based methods that utilize benthic data from boxcore and grab samples and species longevity information to estimate sensitivity (L1 and L2) and impact (L2) using statistical model outputs. - Derived benthic sensitivity metrics to estimate indicators on a continuous scale for total community and specific functional groups. - Tested at the North Sea Area; approach can be used to monitor and assess impact and seabed status at the regional scale and broad scale habitat types. - Relationship between trawling intensity and longevity distribution and trawling intensity, mortality and recovery rates allow for changes in the indicator values over time, regional and broad scale habitat scales. - Uncertainty in habitat specific biomass longevity relationship confidence intervals are estimated on model prediction of fixed effects. - Concept is peer reviewed and further utilized in conjunction with PD2 as a quantitative and mechanistic framework to assess trawling impact.	- Determination of longevity may leave room for different interpretations. - Data demanding, if benthos sampling is not within the range of habitat variables included within model variables. - Trawling Intensity is the only pressure assessed. - Statistical method; with reduced statistical power due to mismatch of trawling and benthos temporal sampling range. - L2 indicator shows a wide variation across grid cells trawled that reflect the variation in bed shear stress. - Indicator still under development. - No threshold information present.
Risk-based approaches linking predominantly to 2017/848 Descriptor 6: D6C3		
Habitat Quality Indicator: Impact of physical disturbance on Benthic Habitats		
Indicator Method	**Suitability**	**Shortcomings**
Cumulative Impact from physical pressures on benthic biotopes (CumI)	- Considers all major and relevant physical pressures, not only bottom trawling. - Method uses (partly) pressure-specific sensitivities. - Method uses exact extent of pressures when known and not only approximations with rasters/grids. - Applicable to all pressure gradients. - Honors both intensity and frequency of pressures when data are available. - Works in both data-rich and data-limited areas. - Method also identifies (functional) loss as a consequence of multiple cumulative pressures.	- High computational demands due to the use of vector data (polygons). - Sensitivity values are summarizing a community while the actual sensitivity is species-based. - Definition of some senstivity may leave room for different interpretations. - Difficult to aggregate the ordinal CumI scale with other indicators which are using a cardinal scale.
Extent of physical disturbance to benthic habitats (BH3)	- BH3 can calculate disturbance at a range of spatial resolutions. - BH3 can be particularly useful for assessing large sea areas where currently only limited data are available. - BH3 currently assesses physical disturbance from bottom-contact fishing and aggregate extraction and can be adapted to new human activities where data are available'.	- Habitat damage and modification, which took place before the period 2010–2015 is not explicitly considered. - Assessments influenced by the availability and resolution of input data: pressure data (e.g. VMS C-squares); sensitivity information and habitat map resolution. - Assessments can be process and data heavy, requiring high levels of computing power when running analyses at the scale of the Northeast Atlantic (e.g. due to complexity of detailed habitat maps).

	- Sensitivity data are derived from peer-reviewed literature assessments, which consider biological traits of habitat characterizing species. - A numeric method of calculating confidence is used. Numeric confidence scores are assigned to each of the attributes: confidence based on underlying data; confidence within data source (such as MESH confidence for habitats); and confidence in the sensitivity of the habitat to a pressure. All the results are evaluated according to their levels of confidence based on the type and quality of the underlining datasets. - BH3 is operational at a Northeast Atlantic scale, used in OSPAR and MSFD assessments, and equivalent mechanisms for national-scale reporting in OSPAR.	