OSPAR CEMP Guideline

Common indicator: Condition of benthic habitat communities (BH2) – common approach

OSPAR Agreement 2018-06[1]

This OSPAR biodiversity indicator is still at a relatively recent stages of implementation and as a result of iteration and learning, it is anticipated that there will be evolution of the methods and approaches documented in the CEMP guidelines. Version updates will be clearly indicated and be managed in a phased approach via ICG-COBAM through its expert groups and with the oversight and steer of BDC.

Contents

[1] This document exists in English only. Update 2022/2023

Main author: Laurent Guérin (FR)

1 Introduction

These guidelines outline the monitoring and assessment requirements for the OSPAR common approach: “condition of benthic habitat communities (BH2)”.

More detailed technical specifications for this indicator are in **Annex 1**. As a conceptual “umbrella” (see below), it is recommended that, in the future, a set of guidelines (according to each habitat and pressure types to be assessed) should be developed for this common approach to operationalise it at its full potential (all MSFD/OSPAR issues for benthic habitats) and update OSPAR CEMP to ensure coherence and comparability at (sub)-regional scale.

Specific applications of the indicator to date are described in **Annex 2** and **Annex 3**. These applications contributed to the OSPAR Intermediate Assessment (BH2-A and BH2-B).

1.1 Context and summary description

Depending on habitat and pressure-type to be assessed (according to specific assessment methods and scales), this BH2 common approach can be adapted to enable the analyses of impacts from each human pressure type on the condition of each benthic habitat type (see Table 1). The indicator should thus inform management measures for human activities that generate pressures, notably the maximum acceptable level of pressure for sustainable use.

Regional assessments on benthic habitats are at a relatively recent stage of development and this common approach will be further elaborated on in subsequent assessment cycle(s).

Preliminary testing demonstrated that BH2 is conceptually applicable to several habitat types and is sensitive to various pressure types. Several case studies (per habitat and pressure type) have been realised through Southern North Sea Project, BenthoVAL, EcApRHA and NEA PANACEA research programmes. Nevertheless, more data (geographical coverage and habitat and pressure type) and further testing is required to conclude on relevance and performances of each Multi metric index (MMI) types according to each habitat and pressure types, at (sub)-regional scale, to operationalise BH2 common concept for each specific application (table 1).

Table 1: Relationships between habitat and pressure types, and how / if the relationships are currently assessed or considered. Relationships based on the revised Marine Strategy Framework Directive (Commission Decision 2017/848/UE and MSFD Annex III 2017/845/UE). Broad pressure and habitat types (EUNIS[2] L3) are described at high levels, however assessment of state-pressure relationship should be undertaken on a finer scale at habitat community (EUNIS L5) and pressure subtypes (Annex 1, table 2). Monitoring, assessment and reporting should be completed per pressure and habitat type. Effects of cumulative (different types) pressures are currently not assessed (gap of knowledge). In the future, this could potentially be completed and/or combined to give an overall understanding of the condition of benthic habitats in the North-East Atlantic.

[2] The European Nature Information System habitat classification is a system to classify types of habitats on several nested levels. The higher the level, the more details and sub-types of habitats are included. Level numbering used here is from EUNIS version 10/2016. http://eunis.eea.europa.eu/

Broad Habitat Type	Physical damage	Removal of species	Hydrological changes	Eutrophication (nutrients or organic matter)	Non indigenous species	Contaminants	Litter	Energy, including under water noise
Littoral rock and biogenic reef				A				
Littoral sediment				A				
Infralittoral rock and biogenic reef				A				
Infralittoral sediment	P			A				
circalittoral rock and biogenic reef				A				
circalittoral sediment	P			A				
Offshore circalittoral rock and biogenic reef								
Offshore circalittoral sediment	P							
Upper bathyal rock and biogenic reef								
Upper bathyal sediment								
Lower bathyal rock and biogenic reef								
Lower bathyal sediment								
Abyssal								

(Column group heading: Broad Pressure Type)

Key:

A	Assessed and reported under the European Union Water Framework Directive (WFD)
	Considered under the European Union Marine Strategy Framework Directive (MSFD)
P	Partially assessed in the OSPAR Quality Status Report 2023
	Main risk (potentially widespread across the OSPAR Maritime Area)
	Relationship identified but not currently assessed

The principle of the common approach "Condition of benthic habitat communities, BH2" is to detect and quantify impact of a pressure by a deviation (of values of the index) compared to (simultaneously assessed) baseline condition (defined as undisturbed or least disturbed), per habitat type.

BH2 has been endorsed as a common regional indicator for the North East Atlantic OSPAR (2013) and is part of a common set of indicators. This is expected to quantify pressure-state relationships (i.e. impacts) and to give, in combination with other benthic indicators, a useful integrated quality score of the condition of the benthic community, for each assessed habitat, calibrated by occurring pressure(s).

This common approach (BH2) will build upon two types of information

- Biological (species composition and relative abundances) and environmental support data (substrate and water metrics) of the sampling station;
- Disturbance (pressure) characteristics: type, intensity and frequency of occurring pressure(s) that potentially causes damage in the sampled area.

The way to combine these basic parameters and metrics in indices (e.g. diversity index, pressure index, etc.) and/or multivariate analysis to compute BH2, depends of the various combinations of habitat and pressure types, and sub-regional specificities. Alternative combinations and proposed options to build MMIs are generally based on the same core datasets (ecological parameters and metrics) and could thus be tested alternatively or simultaneously in the case studies. Various approaches to calculate and test MMI, depending on habitat, pressure types or national approaches are presented in Annexes 1, 2 and 3. The "BENMMI" tool was developed in the Southern North Sea project and could be used to facilitate these calculations and testing (Walvoort and Van Loon, 2016).

Environmental support data (abiotic data, e.g. salinity, depth classes, sand grain size, etc.) are used to define different benthic habitat types. However, the condition of benthic habitats, as defined in MSFD and implied by BH2, needs to be assessed at community level (EUNIS L5). The ICG-COBAM benthic expert group stated during the Hamburg workshop (July 2012, See Annex 1), that the quality assessment of the habitat should be based on its community level (EUNIS level 5) to consider and reduces a factor of the natural variability of communities, when assessing effects of pressure. However, in specific local context (depending of the area to be assessed), where natural variability between communities is lower that effects of pressures, a slightly higher EUNIS level (e.g. L4 or 3) could be tested and considered (See BH2-B, Annex 3). Both biological and physico-chemical data (e.g. oxygen, flow and organic matter rate) may be influenced by human and/or natural pressures, and could be simultaneously considered as state variables and a proxy for the level of disturbance (validated by pressure data). The collection of (semi-)quantitative pressure data, at relevant and compatible temporal and spatial scales, is a key step to integrate this metric and to quantify the pressure-state relationship (i.e. impacts).

Lack of adequate data at (sub)-regional scale compromises the establishment of clear and accurate pressure-state relationships, and thus possibility to define accurate assessment values for each habitat versus pressure pair. This is due to both lack of adapted monitoring and data flows barriers. Even if promising, this indicator requires more development and testing to be fully operational for all OSPAR and MSFD issues. Issues and action plans to further progress were developed through EcApRHA[3] research programme and OSPAR process. For an OSPAR indicator to be operational, the methodological specifications need to be completed and agreed by the expert group, as well as having in place the monitoring and communication arrangements (CEMP).

The development of this common approach is closely linked to those of BH1 (typical species composition – candidate indicator), BH3 (physical damage of predominant and special habitats – common indicator) and BH4 (area of habitat loss – candidate) development, to ensure complementarity and avoid redundancies. There has been no further development for BH5 (size-frequency distribution of bivalve or other sensitive/indicator species), however this indicator, with a specific parameter (size), should also be considered in future development and gap filling. It is anticipated that the overall assessment of benthic habitat requires combining information provided by these indicators (at least BH3 and BH2 in a first step, to combine complementary and respectively wide scale spatial approach to fine scale ground-truthing. See EcApRHA deliverables 2.3 and 4.1).

For a sub-regional scale integrated assessment (See EcApRHA), the common approach has to be used in combination with the indicator "Extent of Physical damage to predominant and special habitats" (BH3). As a positive feedback, the results from BH3 could help optimise and design monitoring assessment areas to assess condition of a habitat type against a pressure gradient. To report GES at the (sub)-regional scale, the method to scale up assessment/monitoring areas/stations results should be discussed in a second step. This

[3] Applying an Ecosystem Approach to (sub) Regional Habitat Assessments (EcApRHA): Addressing gaps in biodiversity indicator development for the OSPAR Region from data to ecosystem assessment. This project is co-financed by the European Union's DG ENV/MSFD/Action Plans 2014, agreement no. 11.0661/2015/712630/SUB/ENVC.2 OSPAR. www.ospar.org/work-areas/bdc/ecaprha

scaling up of assessment at sub-regions requires the development of a method to extrapolate data from assessment areas at a community level (EUNIS 5) to sub-regions, via broad habitat type (EUNIS 3).

2 Monitoring

Regional assessments on benthic habitats are at an early stage of development and this common approach will be further elaborated in subsequent assessment cycle(s).

More details on monitoring and assessment principles of the common approach are described in BH2 technical specification in **Annex 1**, which aims to contribute to the further updating of the CEMP guideline when agreed both at expert and policy levels. Details on already agreed and implemented monitoring and assessment in the context of the European Union Water Framework Directive are described in **Annex 2**. Details on the OSPAR assessment for region II are described in **Annex 3**.

It is anticipated that current monitoring, and pressure data, have gaps to fully assess all habitats, against all pressure type (See Table 1) in all OSPAR maritime area, notably for offshore habitats.

2.1 Purpose and quantitative objectives

Depending on habitat and pressure type to be assessed (according to specific assessment methods and scales), this common approach can be adapted to enable the analyses of impacts from each human pressure type on the condition of each benthic habitat type (see Table 1). It should thus inform of management measures for human activities generating pressures, notably the maximum acceptable level of pressure, for sustainable use (See also paragraph 3.3 on assessment criteria).

This common approach was explicit in the MSFD Commission Decision on GES (2010/477/EU), for the indicator 6.2.2, partly explicit in indicators 1.6.1, 1.6.2 and 6.2.1, and implicit in Criteria (for benthic habitats):

- 2.2. Environmental impact of invasive non-indigenous species
- 4.3. Abundance/distribution of key trophic groups/species
- 5.2. Direct effects of nutrient enrichment
- 5.3. Indirect effects of nutrient enrichment
- 6.1. Physical damage, having regard to substrate characteristics
- 7.2. Impact of permanent hydrographical changes
- 8.2. Effects of contaminants
- 10.2. Impacts of litter on marine life

According to the recently revised Marine Strategy Framework Directive (MSFD) Guidance - COMMISSION DECISION (EU) 2017/848, this common approach could be applied, notably through integration with other indicators (to be further discussed, work in progress), to assess Criteria (applied to benthic habitats):

- D6C1: Physical loss (permanent change) of the natural seabed [*at community level*]
- D6C3: each habitat type which is adversely affected, through change in its biotic and abiotic structure and its functions […], by physical disturbance
- 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

- D4C1: The diversity (species composition and their relative abundance) of the trophic guild is not adversely affected due to anthropogenic pressures
- D2C3: Proportion of the species group or spatial extent of the broad habitat type which is adversely altered due to non-indigenous species, particularly invasive non-indigenous species
- D5C6: The abundance of opportunistic macroalgae is not at levels that indicate adverse effects of nutrient enrichment
- D5C7: The species composition and relative abundance or depth distribution of macrophyte communities achieve values that indicate there is no adverse effect due to nutrient enrichment including via a decrease in water transparency
- D5C8: The species composition and relative abundance of macrofaunal communities, achieve values that indicate that there is no adverse effect due to nutrient and organic enrichment
- D7C2: Spatial extent of each benthic habitat type adversely affected (physical and hydrographical characteristics and associated biological communities) due to permanent alteration of hydrographical conditions
- D8C2: The health of species and the condition of habitats (such as their species composition and relative abundance at locations of chronic pollution) are not adversely affected due to contaminants including cumulative and synergetic effects
- D8C4: The adverse effects of significant acute pollution events on the health of species and on the condition of habitats (such as their species composition and relative abundance)

Further framework development (monitoring and assessment methods, detailed protocols and data flows) is however needed to address all potential MSFD/OSPAR purposes. It should also include the necessary flexibility (both for relevant metrics and indices to be used) to consider the wide range of applications to different habitat types, to target offshore habitats and to assess sensitivity/tolerance to various pressure types. These further works should be made according to the detailed technical specification (Annex 1), which is a living archiving document, to be updated by BH2 lead, describing all already agreed principles made by contributing experts.

2.2 Monitoring Strategy

Relevant assessment scales for BH2 should be at the benthic habitat community scale (EUNIS level 4 to 6) to take into consideration and reduce a factor of the natural variability of communities, which could be higher than effects of pressure. Corresponding geographical scale depends both on habitat type and pressure type to be assessed. To be comparable at (sub)regional level, a proposal of 3 nested scales for monitoring was discussed in the expert group. The three scales include:

- Network of stations at the (sub)-regional scale (common monitoring and inter-calibration for EU reporting)
- Network of stations at the national scale (national funding, specific pressure or habitats, and inter-calibration for national and EU reporting)
- Network of assessment areas, at a site scale, adapted to specific local pressure and habitat types (and according to a risked based approach; fig.1). It is to note that these assessment areas are conditioned by a prioritised habitat-pressure pair and aim to quantify specific state-pressure relationships, and should be adaptive and adapted, depending on the specific habitat and pressure type (both for specific relevant spatial and temporal resolution). Coordinated national and sub-regional networks of assessment area would be more cost-efficient, to share habitat/pressure pairs to be quantified. If no or low impacts are measured at a location during the monitoring, or if another

habitat/pressure is prioritised (e.g. appearance of a new pressure), the network should be updated ((re)moving or adding specific assessment areas). This update requires a close link and high reactivity with local management (policy-makers) and should enable, in the case of newly planned activities, to start monitoring pressures before, during and after potential impact by the occurring activity.

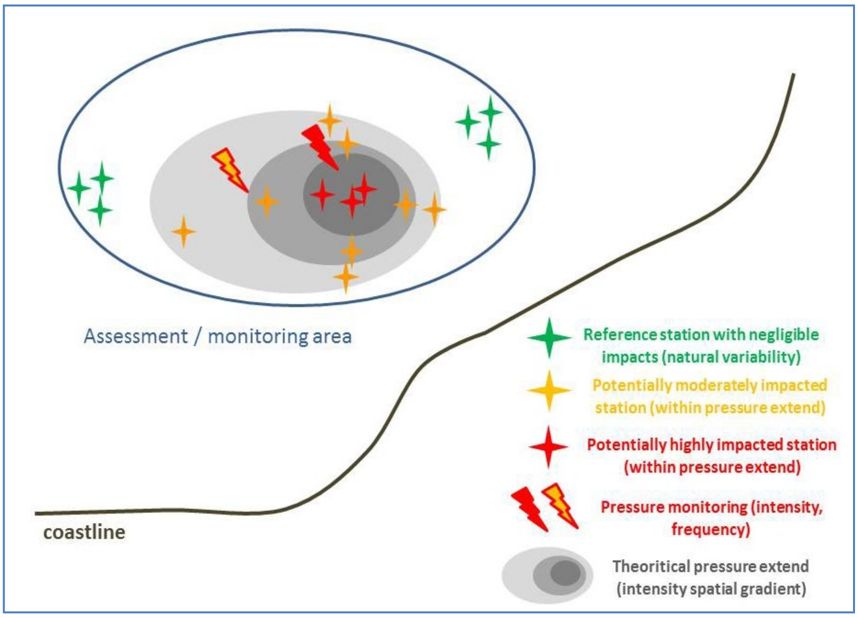

Figure 1: Theoretical principle of an assessment/monitoring area, for specific pressure and habitat types. Scale and spatial pattern may vary from meters to kilometres, depending of the type of pressure and habitat distribution and pattern. Monitoring design (scale, frequencies, etc) must be adapted to the ecological and disturbance (pressure) characteristics (Elliott et al.,2018).

2.3 Sampling Strategy and Quality assurance/ Quality Control

The required methods and effort for ecological characteristics strongly depend on the habitat type. Epibenthic species on hard substrates are generally monitored using quadrats counts or optical, non-destructive methods. Endobenthic communities are generally sampled using standardized grabs or corers. To avoid spurious evaluations, minimum sample estimation (replications of samples) is strongly recommended. It will be a task of the working group to advice on the most relevant method for each habitat, according to the pressure type. Lots of standards already exist for many types of habitats, but still need harmonization at the EU level. For example, it is recommended to use the ISO norms for benthic communities sampling:

- EN ISO 16665. Edition: 2014-08-01. Water quality — Guidelines for quantitative sampling and sample processing of marine soft-bottom macrofauna.
- EN ISO 19493. Edition: 2007-09-01. Water quality — Guidance on marine biological surveys of hard-substrate communities.

Ultimately, a guidance of detailed recommended methods for each pair of habitats per pressure types, including sub-regional specificities, should be produce to ensure comparability at the European or (sub)-regional scale.

The spatial planning of the monitoring (for each nested assessment scale, spatial resolution of sampling locations, etc.) depends of BH2 metrics, habitat type, pressure type, exposure to pressure (scale and intensity) and (sub)-regional potential specificities. These specificities should be further discussed and detailed by expert groups. Furthermore, monitoring budget constraints often play a role. Optimisation with other indicators monitoring requirements (notably BH1 and BH3) and with existing or future monitoring programmes (both on ecological and disturbance parameters) should be investigated.

Monitoring frequencies should also be discussed, depending on habitat and pressure types (See table 1). It will be determined, case by case, by the natural variability of biological and environmental parameters, at the scale of habitat, and by the variability of anthropic disturbances (intensity and frequency). A cost-benefit analysis should also be conducted, including optimised monitoring (both between indicators, and with other biodiversity and pressure/activities monitoring programmes).

Monitoring requirement for BH2 needs further consideration within OSPAR (specifically ICG-COBAM and the relevant expert groups, in close interactions with ICG-C and ICG-MSFD). Ideally, a guidance document, or more probably a set of guidance documents (according to each habitat and/or pressure types), should be produced as a supplement to this guideline to advise on detailed and specific BH2 monitoring requirements and facilitate its implementation. These documents should include potential (sub)-regional specificities. It will be a step by step work, which habitat/pressure pairs will need to be prioritised.

2.4 Data reporting, handling and management

Currently, all datasets are produced and managed by each Contracting Parties, including for Water Framework Directive (200/60/CE[4]) elements of the common approach. Thus, monitoring and data flow at (sub)-regional scale still need to be defined and coordinated for most of habitat and pressure types.

- *Reporting format (Available via a link in the CEMP Appendices)*
  [To be completed, according to each habitat and pressure types to be assessed…]

- *Data metadata schema (Link to ODIMS[5], INSPIRE[6] compliant)*
  [To be completed, according to each habitat and pressure types to be assessed…]

- *Confidence levels in data*
  Biological data: The quality of the biological data depends largely on the sample collection and taxonomic expertise of the analysts and of the quality control for each of the monitoring networks. Hence, caution should be taken when compiling data from different sources. Spatial and temporal resolutions have to be compatible with other dataset (environmental, pressure; See Annex 1).
  Environmental data: spatial and temporal resolutions have to be compatible with other datasets (biological, pressure).
  Pressure data: spatial and temporal resolutions have to be compatible with other datasets (biological, environmental).
  [To be completed, according to each habitat and pressure types to be assessed…]

- *Data flows described (Additional to information in CEMP Appendix)*
  Once the common assessment method is defined, and specific ones (per habitat and pressure type), compatible monitoring programmes need to be in place to produce relevant data (figure 2). There

[4] DIRECTIVE 2000/60/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2000 establishing a framework for Community action in the field of water policy (OJ L327, 22/12/2000).

[5] ODIMS: OSPAR Data and Information Management System. http://odims.ospar.org/

[6] The INSPIRE Directive aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. http://inspire.ec.europa.eu/

are currently many gaps, both in the methods, ongoing relevant monitoring and data availability, to assess or even test BH2 for all habitats and pressure types (table 1). BH2 thus requires further work to be fully operational at (sub)-regional scale to address all MSFD/OSPAR issues. Nevertheless, previous and ongoing work summarised here has enabled the potential of this common approach, and contribution to a partial assessment. Figure 2 highlights data flows will have to be established according to the policy on data sharing of each network and institution.
[To be completed, according to each habitat and pressure types to be assessed…]

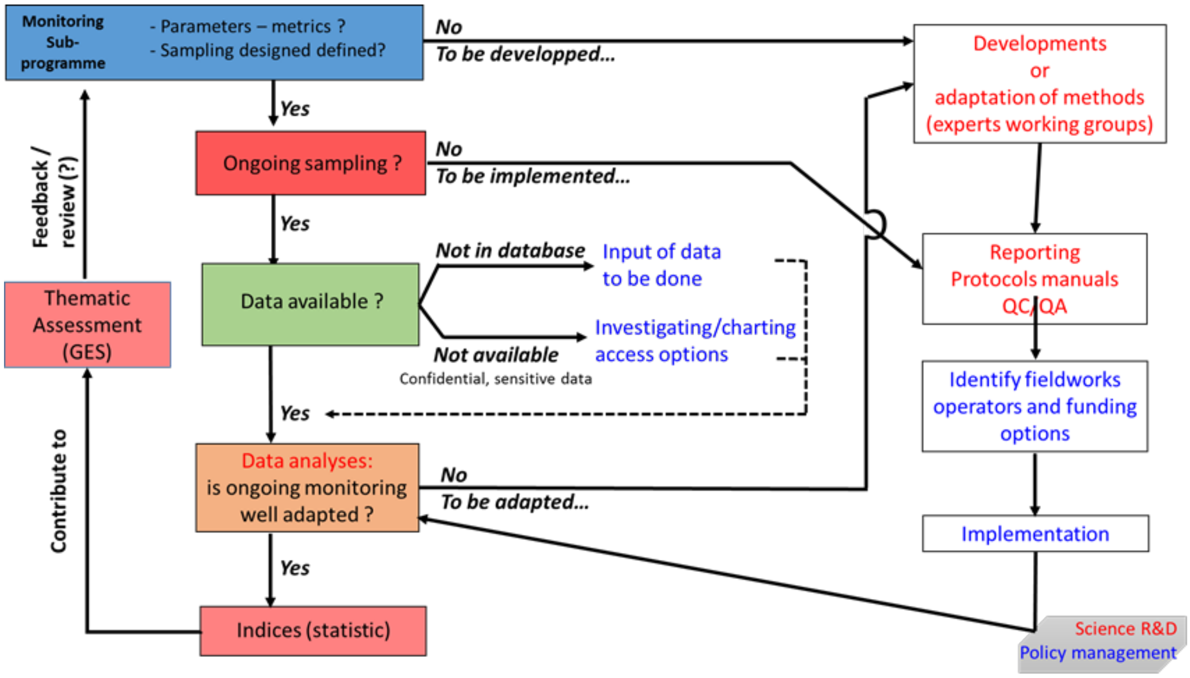

Figure 2: Technical process to implement an operational thematic assessment, at indicator or more integrated levels.

3 Assessment

Regional assessments on benthic habitats are at an early stage of development and this common approach will be further elaborated in subsequent assessment cycle(s).

More details on monitoring and assessment principles of the common approach are described in BH2 technical specification (**Annex 1**). Details on already agreed and implemented monitoring and assessment in the context of the European Union Water Framework Directive are described in **Annex 2**. Details on the OSPAR assessment for region II are described in **Annex 3**.

Results for the OSPAR intermediate assessment (2017) are available at https://oap.ospar.org/en/ospar-assessments/intermediate-assessment-2017/biodiversity-status/habitats/condition-of-benthic-habitat-defining-communities/

3.1 Data acquisition

- *How you extract the data specifically for your assessment question*
  [To be completed and detailed, according to each habitat and pressure types to be assessed…]

3.2 Preparation of data

- *Normalisation of data (If it has come from different monitoring methods)*
  [To be completed and detailed, according to each habitat and pressure types to be assessed…]
- *Aggregation and integration of data acquired*
  [To be completed and detailed, according to each habitat and pressure types to be assessed…]

3.3 Assessment criteria

- *Defining assessment unit/scale (temporal and spatial)*
  The assessment unit is the benthic habitat community, defined at corresponding EUNIS level (4 to 6). Assessment scale should be adapted, depending on pressure type to be assessed. For example, assessment units in Water Framework Directive are "waterbodies", as defined by each Member State to report on them ecological quality status. Benthic assessments contribute then as a "quality element", in addition to other quality elements such as pelagic (plankton) assessment and fish assessment (estuaries), to assess the biological status of the water body. Within the WFD, which by definition aims to assess ecological status of water bodies through quality elements, integration rules (often "one out all out") are applied between biological, chemical and hydrological status to assess each waterbody quality status, against eutrophication (effects of nutrients and organic enrichments). So, even if the benthic habitat is assessed at community level, waterbodies are the relevant assessment scale for management of eutrophication, as it enables to identify source of pressure, mainly from terrestrial or coastal inputs. Thus, results of benthic habitat assessment contribute as one of the qualities elements, to conclude on the ecological quality status of the wider waterbody.

For the specific assessment of the Greater North Sea against abrasion by bottom trawling fisheries (Annex 3), and according to the generally lower alpha (local species richness) and beta (habitats) diversities in this sub-region, it was tested and agreed that this assessment can be conducted on EUNIS level 3, indicated as Broad Habitat Types (BHTs) within OSPAR benthic Assessment Units (AUs) (See Annex 3). It was recognized that in most of other OSPAR area, where diversity of habitats at community level are greater, the common approach recommendation to undertake sampling and assessment at community level (EUNIS L4 to 6) would be required. For other pressure types, no integration rules have been yet developed, however this needs to be further discussed. The combination with other benthic indicators (e.g. for physical disturbance by bottom-trawling fisheries, according to the proposed method developed through EcApRHA research programme, for integrated assessment), would enable a wider scale (spatially) and more complete (improved confidence) assessment, adapted to the management of each pressure type (See EcApRHA deliverable 5.6 – Action plan and OBHEG work plan).

[To be completed and detailed, according to each habitat and pressure types to be assessed…]

- *Baseline/ reference level*
  According to ICG-COBAM (2012) and previous experts' meetings, the following baseline and assessment value methods are strongly recommended for BH2:
  * Baseline setting: "reference" state, with negligible impacts. An existing "reference" state, simultaneous monitored with assessed observations units, for each habitat type, is the most scientifically robust, transparent and comprehensible method. Furthermore, in the context of climate and global changes, past (historical data) or modelling of "reference" states are less robust and relevant, and should be avoided for condition assessment, except for some extreme cases (no actual "references" existing at all assessment scales). As BH2 is conceptually calibrated along

a pressure gradient, the reference may accommodate some impacts as long as the assessed gradient of resulting state cover the GES/no GES range (expected shift of the community at this boundary) (Elliott et al., 2018). As Marine Protected Areas (MPAs) begin to recover to a less impacted state (if adequately managed, enforced and appropriately designated), the number and quality of "reference" areas will increase. Experts' judgement is recommended as an integral part of the baseline setting approach for benthic habitats. However, in specific local context (depending of the area to be assessed) pristine (or least damaged) benthic habitats communities may not be present any more due to widespread and long-term chronic fishing pressure. In this case, an alternative reference setting principle could include the use of a high percentile value of a set of index values from a least disturbed area and period (of 6 years for reporting). The precise percentile value used can be adjusted to the degree of chronic fishing pressure, or any other chronic human pressure, using a well-defined expert judgement rule, in the specific assessed area. If sufficient reference values of good quality are available, the construction of a depth-reference value model is recommended, because it enables the estimation of reference values for assessment areas with insufficient benthos data. See Van Loon et al. (2018) for more details and specific application.

* Assessment value setting: assessment value set as a deviation from the baseline. The specific state assessment values which are set should account for the natural variability in space and time of the habitat type and its potential for recovery (resilience). The complementary use of experts' judgements is recommended.

In addition to these well-defined principles, the setting of baselines and assessment values requires testing/validating by case studies (See Annex 1, II.1) and calibrated by the implementation of BH2 at (sub-)regional scale (See Annex 1, III.2).

[To be detailed, according to each habitat and pressure types to be assessed…]

- *Assessment value*

  According to BH2 definition and stage of development, a qualitative assessment value has actually been discussed in experts' group (See Annex 1) and defined as: "Anthropic pressures are at a level (both intensities and frequencies) that does not adversely compromise recovery (resilience) of benthic habitat communities, both in its diversity, structure and functions".

  The progressive operationalisation of BH2 would enable to quantify and estimate this assessment value and current quality status compared to available baselines, for each habitat and pressure type prioritised. Development of future integrated assessment, by combining other indicators (specifically area-related ones for wider scales) would enable to produce spatial distribution of assessment values, for each assessed habitat and pressure type prioritised (See EcApRHA deliverables 2.3, 4.1 and 5.6, and Elliott et al, 2018).

  [To be further discussed and completed, both by scientist and policy makers, according to each habitat and pressure types to be assessed. Risk-based approach and precautionary principle[7] are strongly recommended in case of lack of knowledge and notably for multiple co-occurring pressures…]

[7] Refer to http://www.ospar.org/about/principles/precautionary-principle

3.4 Spatial Analysis and / or trend analysis

- *Statistical analysis (e.g. Method for trend analysis, Establishment of confidence limits)*
  [To be completed and detailed, according to each habitat and pressure types to be assessed…]

3.5 Presentation of assessment results

- *Consideration of target audience and appropriate communication style*
- *Assessment metadata schema (link to ODIMS)*
- *Confidence in assessment (data and method, including spatial and temporal coverage)*
  [Work and document (assessment sheets) in progress, to be finalised for OSPAR intermediate assessment 2017]

4. Change Management

- *Responsibility for follow up of the assessment is the Biodiversity Committee via the ICG-COBAM expert group for Benthic Habitats;*
- *Links and coherence with ICG-MSFD, EIHA and ICG-C for (cumulative) pressures should be further strengthened. Links and coherence with other relevant European expert groups to be maintained (HELCOM, BARCON, ICES/BEWG, JRC);*
- *This indicator, and links with other benthic, and potentially pelagic and trophic indicators (integrated assessments) was developed with the assistance of the EU-Cofinanced EcApRHA project 2015-2017.*

5. List of Annexes

Annex 1 – Detailed technical specifications of BH2

Annex 2 - Condition of benthic habitat communities (BH2-A): Some coastal habitats in relation to nutrients and/or organic enrichment. CEMP specific guidelines

Annex 3 - Condition of benthic habitat communities (BH2-B): Margalef diversity in region II (Greater North Sea) CEMP specific guidelines

6. References:

Elliott, S.A.M., Guérin, L., Pesch, R., Schmitt, P., Meakins, B., Vina-Herbon, C., González-Irusta, J.M., de la Torriente, A., Serrano, A., 2018. Integrating benthic habitat indicators: Working towards an ecosystem approach. Marine Policy 90, 88-94. https://doi.org/10.1016/j.marpol.2018.01.003

ICG-COBAM, 2012. "OSPAR's MSFD advice manual on biodiversity. Approaches to determining good environmental status, setting of environmental targets and selecting indicators for MSFD descriptors 1, 2, 4 and 6". Version 05/03/2012. 114p. + annexes. http://www.ospar.org/work-areas/cross-cutting-issues/msfd/msfd-advice-manuals

Van Loon W.M.G.M., Walvoort D.J.J., Van Hoey G., Vina-Herbon C., Blandon A., Pesch R., Schmitt P., Scholle J., Heyer K., Lavaleye M., Phillips G., Duineveld G.C.A., Blomqvist M. A regional benthic fauna assessment method for the Southern North Sea using Margalef diversity and reference value modelling. Ecological Indicators, 2018.

Walvoort D.W.W. and Van Loon W.M.G.M., 2016. BENMMI user manual. http://publicaties.minienm.nl/documenten/user-manual-benmmi-3-0

More references included in Annexes 1, 2 and 3 (living documents of the OSPAR Benthic Habitat Expert Group)

ANNEX 1

OSPAR – Common Biodiversity Indicators

Condition of benthic habitat communities (BH2)

Technical specifications (method development) Version 2018/01/12

This document is an updated "technical specifications" of the "condition of benthic habitat communities" OSPAR indicator (BH2), as a "living" document (first drafted in ICG-COBAM, 2013b), presented and adopted at BDC 2013 and which has been regularly updated, according to work progressed. These technical specifications summarise main work conducted since 2011 and conclusions agreed by experts, including contributions of other relevant working groups (e.g. ICES, JRC, EIHA and ICG-C) and associated research projects. This work in progress is part of the Term of Reference of ICG-COBAM and is made in the continuity of its previous works on indicators, of which terminology, principles and results are described in reference documents[8]. Drafted by Laurent Guérin (BH2 lead, FR), integrating experts feedback, from dedicated workshops: Hamburg (10-11/07/2012), Madrid (29-30/10/2013), Dinard (8-10/07/2014), Aberdeen (16-17/06/2015), Warnemünde (29-30/10/2015); Dinard (16-17/03/2016), London (20-22/06/2017), ICES (28/04-02/05/2014, 16-20/02/2015, 04-08/05/2015, 9-13/05/2016); JRC (7-9/09/2015, 20-21/04/2016) and associated research programmes: Southern North Sea Project (2014-2016, coord. Willem Van Loon, NL), ANR BenthoVAL (2014-2017, coord. Olivier Gauthier, FR) and EcApRHA[9] (2015-2017, coord. OSPAR Secretariat).

I.1. Summary description of the indicator (principles and development process)

Name: **Condition of benthic habitat communities**
OSPAR Code: **BH2**
Status: Adopted at BDC and OSPAR 2013 as **OSPAR-wide (sub-regions II, III and IV) common indicator**
State of methodological development:

Development step	**State**
Ecosystem components attributed (species/habitat types)	Defined. Intertidal, shallow and shelf benthic habitats. Applicability/adaptability to bathyal and abyssal benthic habitats (deep sea) is less obvious and would need further discussion
Applicability to sub-regions	Yes, for all, with possible specificities (e.g. for baseline / GES)
Biological and environmental parameters and metrics	Defined for parameters. Metrics may be adapted, according to sampling protocols to be applied

[8] ICG-COBAM, 2011. "OSPAR's MSFD advice manual on biodiversity. Approaches to determining good environmental status, setting of environmental targets and selecting indicators for MSFD descriptors 1, 2, 4 and 6". Version 05/03/2012. 114p. + annexes.

ICG-COBAM, 2013a. "Report by ICG-COBAM on the development of an OSPAR common set of biodiversity indicators. Part A: process for selecting indicators ; Part B: abstracts of proposed common biodiversity indicators". Version 18/02/2013. 40p.

ICG-COBAM, 2013b. "Report by ICG-COBAM on the development of an OSPAR common set of biodiversity indicators. Part C : technical specifications". Version 18/02/2013. 156 p.

[9] Applying an Ecosystem Approach to (sub) Regional Habitat Assessments (EcApRHA): Addressing gaps in biodiversity indicator development for the OSPAR Region from data to ecosystem assessment. This project is co-financed by the European Union's DG ENV/MSFD/Action Plans 2014, agreement no. 11.0661/2015/712630/SUB/ENVC.2 OSPAR. www.ospar.org/work-areas/bdc/ecaprha

Disturbance/pressure parameters and metrics	Defined for parameters. Metrics depends of pressure types, and data availabilities
Monitoring requirements	Partly defined (to be further discussed and agreed, depending of habitat and pressure types)
Assessment scales	Defined. Habitat at community level
Baseline / GES	Principles: defined Case studies: several realised in 2014-2017 Calibration: no (operationalisation needed)

State of operationalisation:

Operationalisation phase	**State**
Work plan	Initial (2013-2015) + COBAM/EcApRHA (2015-2017) + OBHEG multi-year Work plan (submitted to COBAM (2) 2017)
Methodological development	Partly defined, still on-going for full potential. Living document regularly updated from experts meeting (e.g. workshops), for submission to COBAM and BDC
Case studies / data flows	Partly. Several made on existing data. Case studies realised in 2014-2017. Sub-regions coverage and representativeness depends on data availabilities and CP's involvement. Data flows and policy to be further discussed and agreed
Status assessments (baseline and GES setting, reporting)	Requires an updated Joint Assessment and Monitoring Programme (JAMP), according to technical specifications and monitoring requirements. Discussion and agreement with policy-makers needed

Context and description of BH2

Given the complexity of biodiversity, both in its range of character and the number of aspects that contribute to an assessment of state, it is common practice to use a set of indices to assist in monitoring and assessment programmes and to help simplify this complexity. There are a variety of different types of indices: state (including impact), pressure and response. This help limits the number of parameters that need to be monitored to those which can most effectively represent wider functional and structural aspects of the ecosystem. Where possible, state indices should closely respond (in space and time) to a particular anthropogenic pressure (by responding to the impact of the pressure) and hence be linked to associated management requirements.

The assessment of environmental state provided by one or more indices should allow inferences to be made on the wider state of biodiversity components in that ecosystem. State means the actual (measured or otherwise assessed) benthic condition (e.g. of a species, species functional group, community or habitat) in a given geographical area. The assessment of state can be derived by taking direct measurements of the particular biodiversity component ('state indices') or indirectly by measuring the prevailing anthropogenic pressures ('pressure indices'). The best assessment and way to link state changes to pressure would be to measure simultaneously the particular biodiversity component and the prevailing pressure (both human and natural). Additionally, it is usually necessary to consider the cost of implementing the indicator.

Species composition and relative abundances are used to study or assess the biodiversity and state of benthic habitat communities (EUNIS[10] level 4 to 6). A plethora of Benthic Indices (BI) exists, but their use depends on (1) objective of the assessment/study (e.g. pressure or habitat types), (2) geographical/spatial resolution scale and (3) ecological and statistical relevancies. The development of BI, combining various metrics, indices and sensitivity/tolerance classifications, was made mandatory by the EU Water Framework Directive (WFD) for its implementation and assessments. Multi-metric BI (See appendix) classically contains diversity indices (e.g. Shannon or Simpson), species richness indices (e.g. the number of species, Margalef D) and/or classification of sensitive/tolerant species (e.g. AMBI or ITI). This sensitivity of species is a proxy used as disturbance quantification for some pressures on benthic community. This requires a good knowledge, both for community and species considered, of its natural spatial and temporal scales of variation and its sensitivity to a specific pressure gradient.

Multidimensional scaling in community structure from a reference condition, aims to link pressure (for management issues) and pressure-impact calibration of state indices (level of disturbance/resilience), with various pressures types at the community level. The latter eliminates the need for species sensitivity classification given community composition changes can be quantified. In the wider regional context for OSPAR or MSFD requirements, there are actually many knowledge gaps regarding the effect of pressures on different habitats types (Kaiser et al. 2006).

The collection of (semi-)quantitative pressure data and the construction of a suitable pressure index (HELCOM 2010) is a key step to integrate this metric and to validate the pressure-impact evaluation. This common approach is explicit in the indicator 6.2.2 of the Commission Decision on GES (2010/477/UE), partly explicit in indicators 1.6.1, 1.6.2 and 6.2.1, and implicit in Criteria (applied to benthic habitats):

- 2.2. Environmental impact of invasive non-indigenous species
- 4.3. Abundance/distribution of key trophic groups/species
- 5.2. Direct effects of nutrient enrichment
- 5.3. Indirect effects of nutrient enrichment
- 6.1. Physical damage, having regard to substrate characteristics
- 7.2. Impact of permanent hydrographical changes
- 8.2. Effects of contaminants
- 10.2. Impacts of litter on marine life.

According to the recently revised Marine Strategy Framework Directive (MSFD) Guidance - COMMISSION DECISION (EU) 2017/848, this common approach could be applied, notably through integration with other indicators (to be further discussed, work in progress), to assess Criteria (applied to benthic habitats):

- D6C1: Physical loss (permanent change) of the natural seabed [*at community level*]
- D6C3: each habitat type which is adversely affected, through change in its biotic and abiotic structure and its functions […], by physical disturbance
- 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
- D4C1: The diversity (species composition and their relative abundance) of the trophic guild is not adversely affected due to anthropogenic pressures

[10] The European Nature Information System habitat classification is a system to classify types of habitats on several nested levels. The higher the level, the more details and sub-types of habitats are included. Level numbering used here is from EUNIS version 10/2016. http://eunis.eea.europa.eu/

- D2C3: Proportion of the species group or spatial extent of the broad habitat type which is adversely altered due to non-indigenous species, particularly invasive non-indigenous species
- D5C6: The abundance of opportunistic macroalgae is not at levels that indicate adverse effects of nutrient enrichment
- D5C7: The species composition and relative abundance or depth distribution of macrophyte communities achieve values that indicate there is no adverse effect due to nutrient enrichment including via a decrease in water transparency
- D5C8: The species composition and relative abundance of macrofaunal communities, achieve values that indicate that there is no adverse effect due to nutrient and organic enrichment
- D7C2: Spatial extent of each benthic habitat type adversely affected (physical and hydrographical characteristics and associated biological communities) due to permanent alteration of hydrographical conditions
- D8C2: The health of species and the condition of habitats (such as their species composition and relative abundance at locations of chronic pollution) are not adversely affected due to contaminants including cumulative and synergetic effects
- D8C4: The adverse effects of significant acute pollution events on the health of species and on the condition of habitats (such as their species composition and relative abundance)

Further framework development (monitoring and assessment methods, detailed protocols and data flows) is however needed to address all potential MSFD/OSPAR purposes. It should also include the necessary flexibility (both for relevant metrics and indices to be used) to consider the wide range of applications to different habitat types, to target offshore habitats and to assess sensitivity/tolerance to various pressure types. These further works should be made according to these detailed technical specifications, which are a living archiving document, to be updated by BH2 lead, describing already agreed principles made by contributing experts.

Depending on habitat and pressure type to be assessed (according to specific assessment methods and scales), this BH2 common approach can be adapted to enable the analyses of impacts from each human pressure type on the condition of each benthic habitat type (see tables 1 and 2). It should thus inform of management measures for human activities generating pressures, notably the maximum acceptable level of pressure, for sustainable use. Regional assessments on benthic habitats are at an early stage of development and this common approach will be further elaborated in subsequent assessment cycle(s).

Lack of adequate data compromises the establishment of clear and accurate pressure-state relationships, and thus possibility to define accurate assessment values for each habitat versus pressure pair. This is due to both lack of adapted monitoring and data flows barriers. Even if promising, this indicator requires more development and testing to be fully operational for all OSPAR and MSFD issues. Issues and action plans to further progress are defined and available through EcApRHA action plan and OBHEG multi-year work plan. For an OSPAR indicator to be operational, the methodological specifications need to be completed, detailed and agreed by the expert group, as well as having in place the monitoring and communication arrangements (CEMP).

The development of this common approach is closely made according to those of BH1 (typical species composition), BH3 (physical damage of predominant and special habitats) and BH4 (area of habitat loss) development, to ensure complementarity and avoid redundancies. There is actually no further development for BH5 (size-frequency distribution of bivalve or other sensitive/indicator species), however this indicator, with a specific parameter (size), should also be considered. It is anticipated that the overall assessment of benthic habitat requires combining information provided by these indicators (at least BH3 and BH2 in a first step, to combine complementary and respectively wide scale spatial approach to fine scale ground-truthing).

For a sub-regional scale integrated assessment (See EcApRHA), this common approach has to be used in combination with the indicator "Extent of Physical damage to predominant and special habitats" (BH3). As a positive feedback, the results from BH3 could help optimise and design monitoring assessment areas to assess condition of a habitat type against a pressure gradient. To report GES at the (sub)-regional scale, the method to scale up assessment/monitoring areas/stations results should be discussed in a second step. This scaling up to assessment at sub-regions require to develop method to extrapolate from assessment areas of a community level defined habitat (EUNIS 5) to sub-region, via predominant habitat level (EUNIS 3) and ecological subdivision of the sub-region (biogeographic area).

Table 1: Relationships between habitat and pressure types, and how / if the relationships are currently assessed or considered. Relationships based on the revised Marine Strategy Framework Directive (Commission Decision 2017/848/UE and MSFD Annex III 2017/845/UE). Broad pressure and habitat types (EUNIS L3) are described at high levels, however assessment of state-pressure relationship should be undertaken on a finer scale at habitat community (EUNIS L5) and pressure subtypes (See table 2, below). Monitoring, assessment and reporting should be completed per pressure and habitat type. Effects of cumulative (different types) pressures are currently not assessed (gap of knowledge). In the future, this could potentially be completed and/or combined to give an overall understanding of the condition of benthic habitats in the North-East Atlantic.

	Broad Pressure Type							
Broad Habitat Type	Physical damage	Removal of species	Hydrological changes	Eutrophication (nutrients or organic matter)	Non indigenous species	Contaminants	Litter	Energy, including under water noise
Littoral rock and biogenic reef				A				
Littoral sediment				A				
Infralittoral rock and biogenic reef				A				
Infralittoral sediment	P			A				
circalittoral rock and biogenic reef				A				
circalittoral sediment	P			A				
Offshore circalittoral rock and biogenic reef								
Offshore circalittoral sediment	P							
Upper bathyal rock and biogenic reef								
Upper bathyal sediment								
Lower bathyal rock and biogenic reef								
Lower bathyal sediment								
Abyssal								

Key:

A	Assessed and reported under the European Union Water Framework Directive (WFD)
	Considered under the European Union Marine Strategy Framework Directive (MSFD)
P	Partially assessed in the OSPAR Quality Status Report 2023
	Main risk (potentially widespread across the OSPAR Maritime Area)
	Relationship identified but not currently assessed

Table 2: theoretical example of assessment result of condition of a habitat subtype (EUNIS level 5) against sub-types of 'physical damage' pressure. This should be used in combination with 'physical damage to predominant and special habitat' indicator for an integrated assessment of benthic habitats (where and how much impacted?). [1]Maximum acceptable level of pressure: from state-pressure relationship, level of pressure corresponding to the assessment value of the statistical index of condition, considered as inacceptable impact (no resilience, not recoverable, not sustainable). Depending on pressure sub-type, pressure unit could be expressed as intensity, frequency or both

Broad habitat type (EUNIS L3 e.g. offshore circalittoral) sand versus physical damages

* Habitat sub-type (EUNIS L5): e.g. *Abra alba* and *Nucula nitidosa* in circalittoral muddy sand or slightly mixed sediment

Physical pressures sub-types	Condition of benthic habitat community
Permanent change of habitat characteristics (biotic or abiotic) e.g. infrastructure	Habitat loss

Removal of substratum (extraction) with possible recovery of same benthic community	Condition index value / maximum acceptable[1] level of pressure
Surface abrasion (e.g. by bottom trawling fisheries)	Condition index value / maximum acceptable[1] level of pressure
Sub-surface abrasion (e.g. by dredging)	Condition index value / maximum acceptable[1] level of pressure
Changes in suspended solids	Condition index value / maximum acceptable[1] level of pressure
Siltation rate changes	Condition index value / maximum acceptable[1] level of pressure

I.2. Ecosystem components attributed (species/habitat types)

Ecosystem components (fig. 1) relevant for BH2 are intertidal (mediolittoral), shallow and shelf benthic habitats, both special (including OSPAR list of threatened and/or declining habitats, except deep sea habitats; table 3) and MSFD broad habitats types (See tables 1 and 2), but at biological community level (e.g. EUNIS 4 or 5). For the purpose of MSFD, the term habitat addresses both the abiotic characteristics and the associated biological community. Traditionally applied to macro-endofauna (from 1 mm mesh size) of soft bottom sediments, the BH2 could conceptually apply to any community defining a habitat (e.g. including macro-epifauna and macro-flora like macroalgae for rocky habitats, or seagrass beds for angiosperms).

This indicator is thus conceptually applicable to rocky, biogenic reefs and coarse sediments habitats, but some adaptations could be required due to specific monitoring requirements (e.g. % recovery or densities for biological metrics).

Applicability/adaptability to bathyal and abyssal habitats is less obvious and would need further discussion.

Applicability/adaptability to microbenthos (up to 1 mm mesh size) and foraminifera (Schönfeld *et al.*, 2012) would also need further discussion.

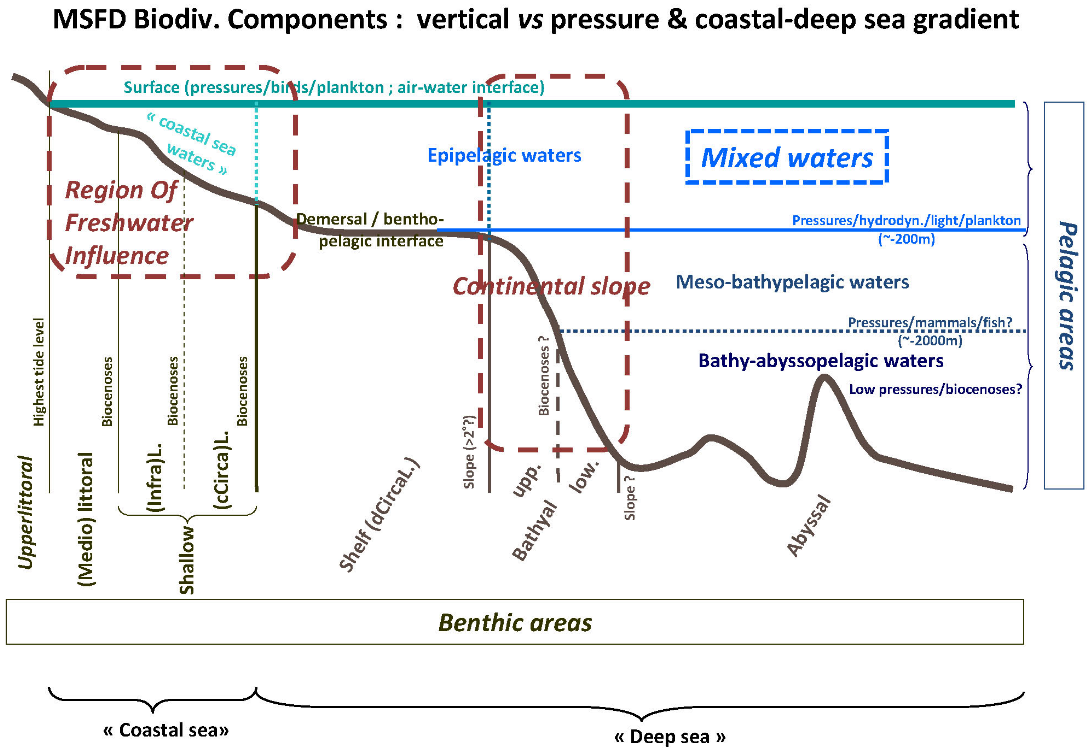

Figure 1: Biodiversity components (benthic and pelagic habitats). This scheme has been developed and adapted from ICG-COBAM works, by the interpretation of common ecological characteristics and MSFD requirements at Regional Sea scale.

Table 3: OSPAR List of threatened and/or declining habitats (Ref. 2008-6) and their theoretical applicability of BH2.

DESCRIPTION	OSPAR Regions where the habitat occurs	BH2 applicability
HABITATS		
Carbonate mounds	**I, V**	**?**
Coral Gardens	**I, II, III, IV, V**	**?**
Cymodocea meadows	**IV**	**Yes**
Deep-sea sponge aggregations	**I, III, IV, V**	**?**
Intertidal *Mytilus edulis* beds on mixed and sandy sediments	**II, III**	**Yes**
Intertidal mudflats	**I, II, III, IV**	**Yes, but at biological community level (e.g. EUNIS 4 or 5) for all subtypes**
Littoral chalk communities	**II**	**?**
Lophelia pertusa reefs	**All**	**?**
Maerl beds	**All**	**Yes**

Modiolus modiolus beds	**All**	**Yes**
Oceanic ridges with hydrothermal vents/fields	**I, V**	**?**
Ostrea edulis beds	**II, III, IV**	**Yes**
Sabellaria spinulosa reefs	**All**	**Yes**
Seamounts	**I, IV, V**	**?**
Sea-pen and burrowing megafauna communities	**I, II, III, IV**	**Yes**
Zostera beds	**I, II, III, IV**	**Yes**

Habitats defined at biological community level, including invertebrate's species listed by OSPAR (*Arctica islandica, Megabalanus azoricus, Nucella lapillus, Ostrea edulis, Patella ulyssiponensis aspera*) and habitat subtypes of Habitats Directive[11] list, should be considered. BH2 would potentially be applicable to all of these habitats, if defined at biological community level.

"Species habitats"[12] for OSPAR list and Habitats Directive lists for mammals, seabirds, turtles and fish should also be considered, where links with benthic habitats exist (e.g. to rest, feed, reproduce, etc). Depending on its characteristics, BH2 could be applied to assess these specific habitats status, and consequently, links with these species' conservation status.

I.3. Indicator parameters and metrics

The common approach BH2 implies two types of information (parameters):

(i) **ecological (state) characteristics**: **biological** (species composition and relative abundances) and **environmental** (substrate and water abiotic characteristics) variables of the sampling station;

(ii) **disturbance characteristics**: type (e.g. nutrients or organic enrichments, mobile bottom gear fisheries, sediment extraction, contaminants, non-indigenous species, etc.), intensity (e.g. concentration of contaminants, abundances of NIS) and frequency (e.g. continuous, dates and periods of year, only one initial, etc.) of occurring pressure(s)that potentially cause damage.

The typical ecological metrics and dataset are composed of several data matrices related to the same observation units (sampling station):

- the community matrix: a biological matrix specifying the abundance (density, biomass), %coverage or presence/absence of all species[13](recommended, but may be accommodated at lower level taxonomic groups) in different samples, pertaining to different positions in space and/or time;

[11]Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora.OJ L 206, 22.7.1992.Directive as last amended by Directive 2006/105/EC (OJ L 363, 20.12.2006, p. 368).

[12]habitats related to mobile species by their functions, e.g. spawning, breeding and feeding areas and migration routes of fish, birds and mammals

[13]It has to be taken into account that species communities may differ locally even if the habitat is similar (e.g. sandbanks). Attention has to be paid to the fact that species lists depend on the expertness of taxonomists in the monitoring teams. Different results could be caused by uneven taxonomic expertise in the teams that could mask the real differences in environmental status. The set-up of the relevant metric also has to be habitat specific and might be (further) developed by each Member State with respect to their (sub)-

- the spatial and temporal matrices: specifying the position of observations in space and time (e.g. Global Positioning System coordinate(s) and date(s) of sampling);
- the environmental matrices: any number of matrices describing environmental conditions, as support parameters (e.g. grain size and/or profile imagery of sediment, organic matter rate, rock structure, water temperature, salinity, depth, hydrodynamics, etc.).

The disturbance characteristics may be indirectly estimate by the ecological ones (sensitivity classification of species or habitats, sediment profile imagery, etc.). But recent approaches recommend estimating the disturbance directly by pressure data. For policy and management issue (e.g. OSPAR, MSFD), this imply that BH2 has to be calculated and validated on ecological **and** pressure data (See above for parameters of disturbance characteristics) to estimate and calibrate impacts. The disturbance metrics and dataset should be acquired as much as possible at coherent and compatible spatial and temporal scales, with the ecological data. Ideally, the sampling strategy (observations units) should be collected along a disturbance gradient (both spatial and temporal) of a specific main pressure, in an area selected by a risk-based approach (pressure gradient within its extent, and close reference locations with negligible impacts; See I.5. fig.2 and Elliott et al., 2018). Anthropogenic pressure data should be related to variation in benthic community structure. Ideally, disturbance data should be of a quantitative nature, but semi-quantitative can be accommodated in statistical models. In whole, the joint analysis of disturbance and community data should allow to adequately describe the EcoQ (Ecological Quality status) of habitats submitted to different anthropogenic pressures.

Besides, some biotic indices (BI) have been developed, including **alternative component** (e.g. foraminifera, Schönfeld *et al*., 2012) **or parameters**. For example, Sediment Profile Images (SPI) provides exclusive in situ views of the sediment-water interface and subsurface sediments (Diaz and Treffy, 2006; Rhoads and Cande 1971). Acquired parameters are both relevant for ecological and disturbance characteristics. This technique can constitute a useful tool to assess sediment features and the activity of benthic organisms (Nilsson and Rosenberg, 1997), optimized for offshore cohesive sediments. SPI has been successfully used to detect benthic habitat changes (Rumohr and Schomann, 1992), including: organic enrichment (Karakassis et al., 2002; Labrune et al., 2012), benthic hypoxia (Nilsson and Rosenberg, 2000), and physical disturbance (Rosenberg et al., 2003). Other analysis like testing sensitivity of BI to fishing impacts have been recently published (Juan and Demestre, 2012). The SPI method could be usefully tested for the pressure monitoring of fisheries and other human pressures on the sea floor. Pressure data are necessary to validate the results.

Multi-metric approach and multivariate analysis

Numerical ecology generally aims to statistically model the variance of the community matrix by way of the environmental matrices, while taking the spatial and/or temporal components into account. In the context of assessing EcoQ (Ecological Quality status), the community matrix is traditionally first summarised to a few or a single metrics that are used in further computations. This step can involve some form of multivariate analysis.

Notwithstanding their inherent differences in terms of elaboration, assumptions, and practical application, all EcoQ indices rely on a fundamental three-step process in establishing the status in a given location:

1. Computation of a Biotic Index (BI) using the community matrix;

regional reference values. Incoherence with metrics used in the indicator BH1 "Typical species composition" should be avoided, and an optimization should be made for monitoring data (common parameters).

2. Computation of an Ecological Quality Ratio (EQR), by comparing the observed BI to that of an undisturbed or least disturbed location (“reference”);
3. Transformation of observed EQR (standardized quantitative variation of the BI) to an EcoQ (target value = deviation from the baseline, limit between GES & non-GES) by relying on any number of conversion scales.

A number of scientific problems still plague each of these steps. This limits our ability to routinely interpret such results in the context of European Framework Directives. Generally, each used index and EcoQ have been developed for a specific issue, and specific assessment scale. Development, testing and implementation has thus to be made carefully, according to each assessment and scale requirements (specific to each Directives).

Experts recognised that many alternative algorithms exist to compute BI, but than most of BI actually used corresponds to one or a combination of these four following alternatives:

- Diversity indices (e.g. Shannon, Margalef d);
- Species sensitivity indices (based on species sensitivity/rarity classification, e.g. AMBI);
- Species biological traits indices (based on species biological traits classification, e.g. ITI);
- Multivariate analyses (on metrics and/or intermediate indices, e.g. Bray-Curtis similarity, BRI).

Thus, the way to combine these basic parameters and metrics in intermediate indices (e.g. diversity index, pressure index, etc.) or global multivariate analysis to build multi-metric indices (MMI), depend on the issue (pressure type), habitat types or sub-regional specificities, has still to be further discussed and agreed, depending on habitat and pressure types assessed. Nevertheless, alternative combinations and proposed options to build alternative BI or MMI are generally based on the same core datasets (ecological parameters and metrics) and could thus be tested alternatively or simultaneously in the case studies.

The “BENMMI” tool was developed in the Southern North Sea project and could be used to facilitate these calculations and testing (Walvoort and Van Loon, 2016).

I.4. Spatial scope: applicability to sub-regions

BH2 is conceptually applicable in all sub-regions, for all relevant type of habitats (See I.2), with possible specificities (e.g. for baseline or GES). It is potentially more sensitive and more specifically linked, by definition, to anthropogenic pressure than the “typical species composition” indicator (BH1). Nevertheless, BH1 is complementary and a necessary preliminary step to characterize ecological state of habitats (natural variation, both at spatial and temporal scales) and high impacts (changes of habitats characteristics). BH3 and BH4, as wider scale area-related indicators, are also complementary and a necessary preliminary step to select prioritized assessment areas for BH2, according to a risk-based approach.

Since same habitat may be dwelled by different species in different sub-regions, a specific set of habitats by sub-regions (OSPAR & MSFD) is expected. This work has to be linked (and avoid redundancies) with EUNIS, which update is in progress. Coordinated works with BH1 development and operationalisation should provide **a cross-table of main habitats ecological characteristics** (using a common typology, e.g. EUNIS), with ‘potential biological communities’ specificities within main biogeographic zones of the OSPAR region and sub-regions. The ultimate complementarity would be to complete BH1 monitoring (which spatial resolution can be higher, as frequency and costs may be lower) with additional monitoring requirements for BH2 (parameters, frequencies), by focussing on prioritized habitat X pressure pairs (with a spatial analyse (risk-based approach) provided by BH3) to calibrate the pressure/impacts relationship (See EcApRHA and OBHEG work plan).

The effect of pressures and relevant associated assessment scales depends of type of pressure (e.g. eutrophication impacts occur generally at a wider scale than sediment extraction). Thus, some technical aspects of BH2 should vary and be specified for each category of pressure. A **classification of pressure types** has been discussed by the expert's groups to test BH2 technical aspects according to each pressure type. As a first step, some pressure types have been prioritized, as mainly occurring at OSPAR (sub-)regional scale, notably physical pressures as noticed for BH3. Coordinated works with BH3 development and operationalisation should provide **a cross-table of main disturbance characteristics** (at least for physical pressures) and spatial analyses to conduct a risk-based approach, with potential vulnerabilities and specific impacts for each habitat type. The ultimate complementarity would be to calibrate the pressure/impacts relationship, by BH2 ground truthing, to fit BH3 vulnerability classification and impacts models (See EcApRHA and OBHEG work plan).

Further works and discussions in the OSPAR benthic habitat expert group and through associated research projects are on-going to progress on the testing of ecologically relevant habitats and pressure types for BH2 case studies. However, the (often limited) data availability often restrict the number of habitats/pressure pairs which can actually be assessed with sufficient statistical confidence and/or geographical coverage.

I.5. Assessment scales

Defining Assessment Scale (from ICG-COBAM, 2011)

Following the ecosystem-based approach required for implementation of the MSFD, the assessment of biodiversity components should be undertaken at ecologically relevant scales, considering the cumulative pressures and their impacts from human activities (Art 8.1b, Annex III Table 2) and based on the criteria provided for assessment in the COM Decision (e.g. habitat/species distribution, species population size, habitat extent and habitat/population condition). […]

Defining scale can be confusing because this term is relevant in different ways depending on several different aspects of the MSFD. Thus, scale should be considered in relation to:

i. **Assessment of state (in relation to the definition of GES and associated state targets)** of one or several biodiversity components, as GES is determined at the level of the region/subregion (Art. 3(5)). These may be linked, by trophic relation for example in Descriptor 4 or functional relation such as between species and habitats (refer Habitats Directive). This aspect could be expressed as an 'ecological assessment area' (or aggregated sub-areas) for reporting purposes

ii **Management measures**, which can be considered at either a local scale, to avoid missing or masking cumulative local impacts that could affect the overall quality status at larger scales, or at a broad scale to manage efficiently biodiversity components or pressures that operate over large areas of a region/subregion (as required by MSFD);

iii **Monitoring** to assess state, expressed as the spatial and temporal resolution of data. These resolutions (number of sampling stations, accuracy of remote detection, sampling frequencies, etc.) are likely to be a compromise between "high resolution", which enable a very accurate, but expensive assessment and a more pragmatic approach, identifying a resolution in accordance with available resources which can then be used to define assessment scale and data needs. […]

There are many different aspects of assessment scale (e.g. habitats occur at a different scale to many of the pressures acting upon them), with respect to the scale required to determine GES, this would depend on the scale of the habitat within a particular assessment area. In practice this would require **consideration on a case by case basis**. […] Assessment scales for habitats **are smaller than, and nested within**, sub-regions.

Relevant assessment scales for BH2 should be at the benthic habitat community scale (EUNIS level 4 to 6) to take into consideration and reduce a factor of the natural variability of communities, which could be higher than effects of pressure. Corresponding geographical scale depends both on habitat type and pressure type to be assessed. To be comparable at (sub-)regional level, a proposal of 3 nested scales for monitoring was discussed in the expert group. The three scales include:

- Network of stations at the (sub-)regional scale (common monitoring and inter-calibration for EU reporting)
- Network of stations at the national scale (national funding, specific pressure or habitats, and inter-calibration for national and EU reporting)
- Network of assessment areas, at a site scale, adapted to specific local pressure and habitat types (and according to a risked based approach; fig.1). It is to note that these assessment areas are conditioned by a prioritised habitat/pressure pair and aim to quantify specific state/pressure relationships, and should be adaptive and adapted, depending on the specific habitat and pressure type (both for specific relevant spatial and temporal resolution). Coordinated national and sub-regional networks of assessment area would be more cost-efficient, to share habitat/pressure pairs to be quantified. If no or low impacts are measured at a location during the monitoring, or if another habitat/pressure is prioritised (e.g. appearance of a new pressure), the network should be updated ((re)moving or adding specific assessment areas). This update requires a close link and high reactivity with local management (policy-makers) and should enable as much as possible, for the case of new planned pressure, to start monitoring before, during and after the pressure duration to have a complete understanding of the impacts.

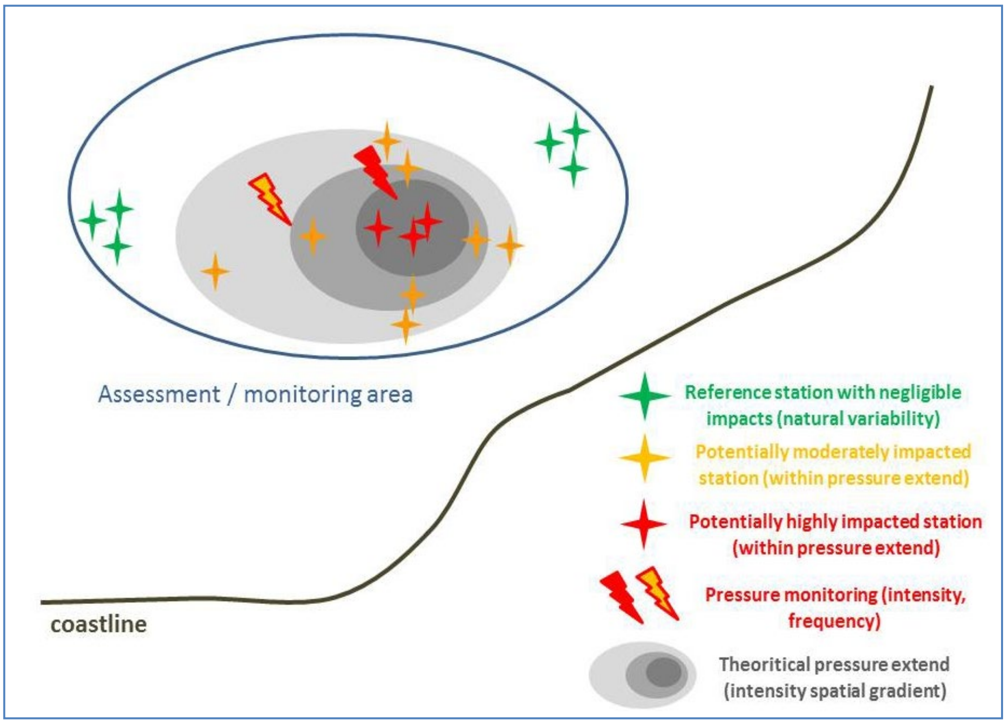

Figure 2: Theoretical principle of an assessment/monitoring area, for specific pressure and habitat types. Scale and spatial pattern may vary from meters to kilometres, depending of the type of pressure and habitat distribution and pattern. Monitoring design (scale, frequencies, etc) must be adapted to the ecological and disturbance (pressure) characteristics (Elliott et al., 2018).

I.6. Monitoring requirements (parameters, frequency, material & methods): guidelines/specificities

The required methods and effort for ecological characteristics strongly depend on the habitat type to be addressed. Epibenthic species on hard substrates are generally monitored using quadrats counts or optical, non-destructive methods. Endobenthic communities are generally sampled using standardized grabs or corers. To avoid spurious evaluations, minimum sample estimation (replications of samples) is strongly recommended. It will be a task of the working group to advice for the more relevant method per habitat, according to the pressure type. Lots of standards already exist for many types of habitats, but still need harmonization at the EU level. For example, it is recommended to use the ISO method (ISO, 2011) for marine soft-bottom macrofauna monitoring. Ultimately, a guidance of recommended methods for each pair of habitat X pressure types, including sub-regional specificities, should be produce to ensure comparability at the European scale.

The spatial planning of the monitoring (for each nested assessment scale, spatial resolution of sampling locations, etc.) depends of BH2 metrics, habitat type, pressure type, exposure to pressure (scale and intensity) and (sub-)regional potential specificities. This issue should be further discussed by expert groups. Furthermore, monitoring budget constraints often play a role. Optimization with other indicators monitoring requirements (notably BH1 and BH3) and with existing or future monitoring programmes (both on ecological and disturbance parameters) should be investigated.

Monitoring frequencies should also be discussed. It will be determined by the natural variability of biological and environmental parameters, at the scale of habitat, and by the variability of anthropic disturbances (intensity and frequency). A cost-benefit analysis should also be conducted, including optimized monitoring (both between indicators, and with other biodiversity and pressure/activities monitoring programmes).

The inventory of existing monitoring programmes, through ICG-COBAM works (monitoring template) and other working groups (e.g. FP7 STAGES, FP7 JMP North Sea and Celtic Sea, JRC MSFD art.12 report on art. 11 reporting, etc.) would be useful, compared to technical requirements (monitoring parameters) of benthic indicators, to identify gaps and potential adaptation needed. These monitoring requirements should be clearly defined, to be submitted to BDC and OSPAR meetings, in order to be integrated in the OSPAR CEMP (Coordinated Environmental Monitoring Programme).

Monitoring requirement for BH2 need further discussion and works in COBAM and the expert group. Ideally, a guidance document, or more probably a set of guidance documents (according to each habitat and/or pressure types), should be produced to advice on detailed and specific BH2 monitoring requirements and facilitate its implementation. These documents should include potential (sub)-regional specificities. It will be a step by step work, which will need to prioritize habitat/pressure pairs.

I.7. Baseline and GES principles

According to ICG-COBAM (2012) and previous experts' meetings, the following baseline and assessment value setting methods are strongly recommended for BH2:

- **Baseline setting: Method A ("reference" state, with negligible impacts)**. An existing "reference" state, simultaneous monitored with assessed observations units, for each habitat type, is the most scientifically robust, transparent and comprehensible method. Furthermore, in the context of climate and global changes, past (historical data) or modelling of "reference" states are less robust and relevant, and should be avoided for condition assessment, except for some extreme cases (no actual "references" existing at all assessment scales). As BH2 is conceptually calibrated along a pressure gradient, the reference may accommodate some impacts as long as the assessed gradient of resulting state cover the GES/no GES range (expected shift of the community at this boundary) (Elliott et al., 2018). As Marine Protected Areas (MPAs) begin to recover to a less impacted state (if adequately managed), the number and quality of "reference" areas will increase. Experts' judgement is recommended as an integral part of the baseline setting approach for benthic habitats. However, in specific local context (depending of the area to be assessed) pristine (or least damaged) benthic habitats communities may not be present any more due to widespread and long-term chronic fishing pressure. In this case, an alternative reference setting principle could include the use of a high percentile value of a set of index values from a least disturbed area and period (of 6 years for reporting). The precise percentile value used can be adjusted to the degree of chronic fishing pressure, or any other chronic human pressure, using a well-defined expert judgement rule, in the specific assessed area. If sufficient reference values of good quality are available, the construction of a depth-reference value model is recommended, because it enables the estimation of reference values for assessment areas with insufficient benthos data. See Van Loon et al. (in press) and Annex 3 (BH2-B) for more details and specific application.

- **Assessment value setting: Method 3 (assessment value set as a deviation from the baseline).** The specific state which are set should account for the natural variability in space and time of the habitat type and its potential for recovery (resilience). The complementary use of experts' judgements is recommended.

According to BH2 definition and stage of development, a qualitative assessment value has actually been discussed in experts' group and defined as: "Anthropic pressures are at a level (both intensities and frequencies) that does not adversely compromise recovery (resilience) of benthic habitat communities, both in its diversity, structure and functions".

The progressive operationalisation of BH2 would enable to quantify and estimate this assessment value and current quality status compared to available baselines, for each habitat and pressure type prioritised. Development of future integrated assessment, by combining other indicators (specifically area-related ones for wider scales) would enable to produce spatial distribution of assessment values for this value, for each assessed habitat and pressure type prioritised (See EcApRHA deliverables 2.3, 4.1 and 5.6, and Elliott et al, 2018).

In addition to these well-defined principles, the setting of baselines and GES requires testing/validating by case studies (See II.1) and calibrated by the implementation of BH2 at (sub-)regional scale (See III.2).

II.1. Case studies

Dataset requirements for BH2 are defined here at parameters level, but need further discussion and works in COBAM and the expert groups for agreement on each specific metrics depending on all specific habitats and pressures types. But methodological development is yet sufficiently advanced so than several case studies have already been conducted (See IV.1) and submitted to BDC 2015 in a "testing report" (BDC

15/3/Info.2 Add.1 - E) and to BDC 2016 as partial contribution to the Intermediate assessment[14]. Sub-regions coverage and representativeness (both on habitat and pressure types) depend on CP's involvement and data availability. Data flows and author copyrights needs to be further discussed and agreed. Ideally, a guidance document, or more probably a set of guidance documents (according to each habitat and/or pressure types), should be produced to advice on BH2 data requirement and facilitate its testing and assessment, for each habitat and pressure type. These documents should include potential (sub)-regional specificities.

II.2. Data flows

Further considerations are needed, depending on habitat and pressure types. Once the common assessment method is defined, comparable monitoring has to be in place to produce data, in order to compute any indicator (fig.3). There are actually many gaps, both on methods, ongoing relevant monitoring and data availability, to assess or even test BH2 for all habitats and pressure types. BH2 requires thus further work and relevant monitoring to be fully operational at (sub-) regional scale to address all MSFD/OSPAR issues. Nevertheless, previous and ongoing works already enabled to demonstrate soundness and potential of this indicator, and contribution to a partial assessment. The aim of this operationalisation phase is to make the best use of the OSPAR Contracting Parties' expertise and propose a global expertise by sharing, helping or conducting analyses on the most as possible dataset, as much as possible representative of sub-regions, to further test and develop the BH2 indicator. Ongoing relevant monitoring (See chapter I) is anyway a preliminary step required to provide relevant datasets (fig. 3). Before sharing the datasets, a charter should be signed by authors to protect those copyrights and guarantee them involvement in the process and further potential publication (scientific articles and/or OSPAR reports). Even for The Water Framework Directive assessment - the actually only implemented and reported (to European Commission) assessment of BH2, applied to some coastal habitats exposed to eutrophication (nutrients inputs) – the completion, accessibility or availability of data was heterogeneous between Member States.

[14] Results for the OSPAR intermediate assessment (2017) are available at https://oap.ospar.org/en/ospar-assessments/intermediate-assessment-2017/biodiversity-status/habitats/condition-of-benthic-habitat-defining-communities/

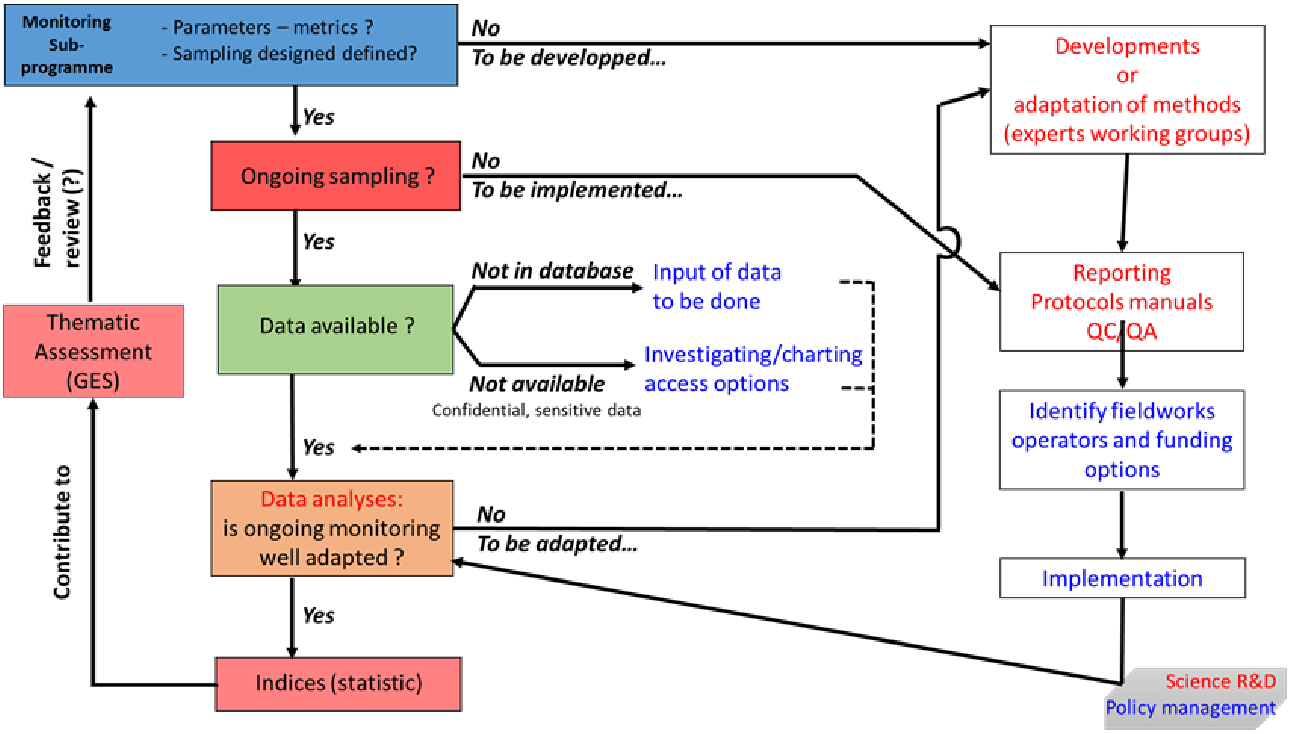

Figure 3: Technical process to implement an operational thematic assessment, at indicator or more integrated levels.

III.1. Protocol for status assessments (principles)

This includes sampling procedure, assessment scales, (sub)regional specificities, rule of combination of parameters, data analysing procedure, computation of the indicator to an Ecological Quality Ratio (EcoQR) and then setting up of an EcoQ (assessment of Ecological Quality) relevant for management (setting up of Ecological Quality Objectives, EcoQO).

For BH2, several protocols will be necessary for each prioritised habitat and pressure types to be assessed.

Annexes 2 and 3 details principles and protocols for these 2 specific assessments.

III.2. Protocol for status assessments (baselines and GES setting, reporting)

This is the last phase of operationalisation, to be further considered when all previous phases will be sufficiently advanced. For threshold between GES and no GES, a preliminary dialogue and agreement with policy-makers and decision committees is needed (setting up of EcoQ and EcoQO, definition of thresholds). Furthermore, case studies at European or Regional scales can be informative, but final calibration need defined and operational monitoring programme: updated OSPAR CEMP (Coordinated Environmental Monitoring Programme), according to technical specificities and monitoring requirements.

Reporting process depends deeply on the temporal planning of the monitoring, and the temporal planning of the overall assessment itself. At least one reporting every six years should be warranted, at the European scale, for all the habitats to which this indicator is applied. The possibility of reporting or assessing with a higher frequency should be further discussed.

IV.1. Work plan for the development of the indicator

France (Muséum National d'Histoire Naturelle) volunteered to lead the development of this indicator for OSPAR, according to previously submitted and updated work plan. Many Contracting Parties (CPs) regularly contributed, but all CPs are more than welcome to participate and involve in this development and operationalisation process. National research projects are also actively contributing with national funds, notably "Southern North Sea Project" (2014-2016, coordinated by Willem Van Loon, NL) and "BenthoVAL" (2014-2016, coordinated by Olivier Gauthier, FR) and are expected to contribute to many steps of the development process, including case studies. Since 2015, the EcApRHA research project (2015-2017, agreement n°. 11.0661/2015/712630/SUB/ENVC.2 OSPAR), co-financed by the European Union, led by OSPAR secretariat with 9 partners, including OSPAR indicators leads, include to forward this indicator project plan. This notably by investigating links with other indicators (benthic, pelagic and trophic) in order to develop more integrated assessment in close links with OSPAR/COBAM related experts' groups. A key element of this work is the involvement of experts from each CP to ensure a scientifically robust, transparent and representative process and facilitate appropriation for management issues. The OSPAR benthic expert group solicited through ICG-COBAM and EcApRHA works, coordinated and animated by all benthic indicators leads, is the core working group to ensure this coordinated development and link with national (similar projects) and international (e.g. ICES, JRC, FP7, etc.) expertise. Further development priorities are described in the OBHEG Terms of Reference and work plan.

Lead expert and contact for France is Laurent Guérin (laurent.guerin@mnhn.fr), for BH2 development coordination.

The role of the indicator lead includes:

- Draft, update and report on the indicator specific documentation, notably CEMP guidelines and Technical specification, and where relevant specific indicator work plan;
- Guide the work on further development of the indicator, keeping in mind and contributing as much as possible to integration methods development (through Basecamp, e-mail, teleconferences and physical meetings), notably ensure coherence and complementarity between indicators and integration methods development;
- Contribute to the organisation (notably agenda items) and participate to dedicated benthic habitat workshops as needed (at least once per year) and possible;
- Ensure information flow and coordinated work between involved experts and with other OSPAR indicators leads and BHEG co-chairs, relevant OSPAR committees (notably ICG-COBAM and BDC) and other relevant benthic expert groups (e.g. ICES, JRC, HELCOM, BARCON);
- Submit specific indicator documentation for reviews to all involved COBAM-BHEG experts, other indicator leads, COBAM-BHEG co-chairs and COBAM. Update and validate documentation, as much as required for reporting, according to progress made (e.g. through workshops or any related working group or project) and comments received, acknowledging all significant contributions.

The role of each involved expert includes:

- Anticipate as much as possible and estimate approximately dedicated individual and/or team resources (dedicated expert time and capacity to attend workshops), available into the timescale and for milestones. Let's keep informed BHEG co-chairs, and any relevant indicator lead (for specific technical task), to enable adapt the work plan to current total OSPAR expertise resources and highlight potential needs of extra resources for specific tasks, to be forwarded to COBAM and BDC;
- Regularly participate to discussions through Basecamp, e-mail, teleconferences and, as much as possible, meetings;

- Identify relevant national similar project, experts, relevant case studies, and facilitating communication and data flows;
- Actively contribute all along the development process of benthic habitat assessment and monitoring methods, as described in terms of reference, work plan and workshops agendas.
- Review, comment and amend, if needed, relevant (according to expert's specific fields of expertise) draft deliverables submitted by BHEG co-chairs or indicator leads.

The detailed work plan and the list of currently identified expertise resources for COBAM-BHEG, as of 10 July 2017, is presented in the OBHEG multi-year work plan.

The need for further method development and the lack of adequate data or access to existing data (biological, environmental and pressure) compromise the establishment of clear and accurate state-pressure relationships, and thus the possibility to define accurate assessment values at OSPAR scale. However, such assessment could be undertaken in certain areas where there is sufficient data. The lack of data is due to both lack of appropriate monitoring (e.g. for offshore habitats) and barriers to data flows.

a. Barriers and gaps:

Method development:

Within an assessment concept, you have to consider the data (see data use and availability), the baseline data ("reference" with low or least disturbance)) and the type of indicator (e.g. variety of WFD and MSFD benthic indicators). Currently, the assessment method/framework is defined, but can benefit from further harmonization regarding practical implementation, regarding indicator types, baseline setting and data gathering. This method harmonization should lead to a more confident, large scale evaluation of the benthic status, which is in line between the involved countries.

One of the critical points is the determination of baseline data, which are lacking for many habitats at the OSPAR scale for different pressures, and especially for abrasion by bottom trawling fisheries. This compromises our knowledge of a sustainable condition for certain habitats, which are continuously under high pressure levels. "Pristine" baselines (no disturbance) are lacking for most of habitats, which compromise our knowledge of the potential best condition of natural habitat communities (Elliott et al, 2018). It is not practical or feasible to use this pristine state as an environmental target everywhere, but is useful to understand its natural dynamic and recovery potential. Next to it, the approach to determine the baseline and to benchmark the "reference" values between countries is a point of attention for future assessments. Increasing establishment and management of Marine Protected Areas (including Marine Natura 2000 sites), notably including "no take or low-pressure areas" could help to provide data and test approaches in the future, for the relevant habitat types.

Cumulative Effects

The cumulative effects of co-occuring pressures (different pressure at the same place and the same time range) are not yet assessed. According to the actual gaps of knowledge on benthic habitat communities, it will be necessary to first understand and quantify effects of each pressure type, where data are available, before investigating and understanding the effects of cumulative pressures (e.g. whether they are synergetic or antagonistic?). This is an important requirement as many habitats are exposed to multiple pressures. Monitoring of benthic habitats exposed to cumulative pressures, where management measure aims progressively for a more sustainable use, even for one pressure type, would help to better understand this in the future.

Data use and availability:

- Many potentially relevant data exist but are not all available (e.g. fishing pressure data at fine spatial resolution or biological data from marine research and marine industry)
- Many biological datasets exist (more or less available), but few have associated proxy or data on pressure at a compatible spatial and temporal scale.
- Most OSPAR Contracting Parties have wide scale marine benthic monitoring (at least for Water Framework Directive). Finer scale networks of state-pressure relationship assessment areas are more heterogeneous and should be further investigated. Developing a coordinate monitoring - or better, a joint monitoring - would be much more cost-efficient and would warranty coherence and robustness for an assessment at (sub)regional scale.
- Each country holds its own data, so a common method (and tools?) still needs to be developed. Development of data flow arrangements to access and analyse data, compute indicators, and compare/inter-calibrate results, is still in progress for WFD. The need for these steps should be anticipated and relevant work should be coordinated at (sub)-regional scale to ensure coherence and facilitate the computation of data for indicator assessment.

b. Cost-effectiveness:

As this indicator is already partly applied and implemented at EU scale for Water Framework Directive (for some coastal habitat types) for eutrophication, it can be considered already partly operational for nutrient inputs and coastal habitats. This common approach is an “umbrella” for several assessments, is based on existing and anticipated monitoring requirements (for OSPAR, MSFD, others European Directives and the Common Fisheries Policy) and is, therefore, cost efficient if coordinated.

c. Further development and roadmap:

The indicators for” Condition of benthic habitat communities” and “physical damage to predominant and special habitats”, are the only OSPAR-wide common agreed benthic habitat indicators. A regional (wide scale) monitoring strategy and technical guidance are partly in place and implemented for some coastal habitats and for nutrient input pressure (Water Framework Directive). The Regional Monitoring Strategy should be further coordinated, extended offshore and adapted for (finer scale) assessment areas, to progressively consider and assess effects of other types of pressures on each benthic habitat type. As geographical (nested) scales are identified, assessment methods would also be progressively adapted and described.

An Action plan identifying gaps, barriers and solutions to further develop OSPAR benthic indicators, and an integrated assessment, for next assessment cycles, is available in EcApRHA research programme deliverables (http://www.ospar.org/work-areas/bdc/ecaprha) and was adapted for the multi-annual OBHEG work plan.

In future development and assessments, consideration could be given to using this common approach to other habitat types, e.g. those listed in the Habitats Directive, OSPAR threatened and declining habitats.

IV.2. Resources needed

It is important to distinguish the different successive phases of the operationalisation of an indicator, notably at (sub)-regional scale. All the operationalisation process and some reference documents are detailed in annexes.

- *Methodological development*

Main resources needed at this stage are working time of lead and involved experts, and financial costs for organization and venue to meetings and workshops. Travel expenses to participate in the workshops and meetings will be needed and should be supported by each CP. It is essential that resources (expert time and travelling costs), from all involved CPs are available to support the development of this indicator.

- *Testing (case studies)*

Main resources needed at this stage are working time of lead, involved experts and authors, availability of relevant datasets and financial costs for choosing, analysing and computing the dataset to test the BH2. It is essential that resources (relevant datasets, working time and analysing costs), from all involved CPs are available to support the development of this indicator.

- *Implementing (monitoring, computation and reporting)*

Depending of habitat and pressure types to be assessed, it is actually impossible to estimate the real specific cost of BH2 implementation, until optimised monitoring programme will be clearly defined and detailed. For example, the cost of a (vessel) survey may be high, but the cost for each sample depends of the number of samples realised during the survey. If other indicators need are covered during the survey (e.g. mammals' observation, pelagic habitat sampling, etc.), the global cost is also reduced. Rather than costs, resources needed should be further discussed, and further detailed as indicator development and associated guidance are refined.

At this stage, a coarse estimation of some basic resources needed is the following:

- Vessels, suitable to work from sublittoral to shelf

- Scuba diving sampling

- Adequate equipment (box core samplers, grabs, dredges, etc.) for sampling collection from intertidal to shelf

- Laboratory infrastructure to analyse samples under Quality Controls

- Qualified personnel for sampling, data input and processing, indicator computing and interpreting. Taxonomical, ecological and statistical skills are key requirements.

The actually implemented coastal WFD monitoring should be used to contribute to MSFD and OSPAR requirements, in order to optimize the resources allocation. Nevertheless, it must be acknowledged that the sampling surface covered by WFD is a very low proportion of the total marine waters covered by MSFD (this proportion is different for each country), even if the sources of this pressure are mainly coastal or terrestrial inputs. For all other habitat and pressure types, additional resources are needed to operationalize this indicator at its full potential (all MSFD/OSPAR issues for benthic habitats). Combination with other indicators (integrated assessment) and complementary technologies under development (such as sediment-profile-imagery or biomolecular tools), could improve cost-effectiveness to assess both environmental and disturbance-impact characteristics in large areas.

<u>References</u>

Boon A.R., A. Gittenberger A. and Van Loon W.M.G.M., 2011, Review of Marine Benthic Indicators and Metrics for the WFD and design of an optimized BEQI, Deltares, report nr 1203801-000.

Bermejo R., Vergara J.J., Hernández I., 2012. Application and reassessment of the reduced species list index for macroalgae to assess the ecological status under the Water Framework Directive in the Atlantic coast of Southern Spain. Ecol. Ind., 12: 46-57.

Borja A., Josefson A.B., Miles A., Muxika I., Olsgard F., Phillips G., Rodríguez J.G., Rygg B., 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive.c: 42-52.

Dauvin J.C., Ruellet T., 2007. Polychaete/amphipod ratio revisited: Implementation of the Water Framework Directive in European marine waters. Marine Pollution Bulletin, 55(1-6), 215-224.

Diaz R.J., Trefry J.H., 2006. Comparison of sediment profile image data with profiles of oxygen and Eh from sediment cores.Journal of Marine Systems 62(3-4) 164-172.Helcom 2010.Towards a tool for quantifying anthropogenic pressures and potential impacts on the Baltic Sea marine environment.BSEP 125.

Díez I., Bustamante M., Santolaria A., Tajadura J., Muguerza N., Borja A., Muxika I., Saiz-Salinas J.I., Gorostiaga J.M., 2012. Development of a tool for assessing the ecological quality status of intertidal coastal rocky assemblages, within Atlantic Iberian coasts. Ecol. Ind. 12: 58-71.

Elliott, S.A.M., Guérin, L., Pesch, R., Schmitt, P., Meakins, B., Vina-Herbon, C., González-Irusta, J.M., de la Torriente, A., Serrano, A., 2018. Integrating benthic habitat indicators: Working towards an ecosystem approach. Marine Policy 90, 88-94. https://doi.org/10.1016/j.marpol.2018.01.003

HELCOM, 2010. Towards a tool for quantifying anthropogenic pressures and potential impacts on the Baltic Sea marine environment: A background document on the method, data and testing of the Baltic Sea Pressure and Impact Indices. Balt. Sea Environ. Proc. No. 125.

ICG-COBAM, 2012. OSPAR's MSFD advice manual on biodiversity. Approaches to determining good environmental status, setting of environmental targets and selecting indicators for MSFD descriptors 1, 2, 4 and 6. Version 05/03/2012. 114p. + annexes. http://www.ospar.org/work-areas/cross-cutting-issues/msfd/msfd-advice-manuals

ISO 16665:2011, Water quality - Guidelines for quantitative sampling and sample processing of marine soft-bottom macrofauna

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

Juan S., Demestre M., 2012. A Trawl Disturbance Indicator to quantify large scale fishing impact on benthic ecosystems. Ecological Indicators, 18, 183-190

Juanes J.A., Guinda X., Puente A., Revilla J.A., 2008. Macroalgae, a suitable indicator of the ecological status of coastal rocky communities in the NE Atlantic. Ecol. Ind, 8: 351-359

Kaiser M., Clarke K., Hinz H., Austen M., Somerfield P., Karakassis I, 2006. Global analysis of response and recovery of benthic biota to fishing. Marine Ecology Progress Series 311:1–14.

Karakassis I., Tsapakis M., Smith C.J., Rumohr H., 2002. Fish farming impacts in the Mediterranean studied through sediment profiling imagery. Marine Ecology Progress Series 227:125-133.

Krause-Jensen D., Sagert S., Schubert H., Boström C., 2008. Empirical relationships linking distribution and abundance of marine vegetation to eutrophication Ecol. Ind. 8: 515-529

Labrune C., Romero-Ramirez A., Amouroux J.M., Duchêne J.C., Desmalades M., Escoubeyrou K., Buscail R., Gremare A., 2012. Comparison of ecological quality indices based on benthic macrofauna and sediment profile images: a case study along an organic enrichment gradient of the Rhône River. Ecological indicators, 12: 133-142.

Muxika I., Borja A., Bald J., 2007. Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive. Mar. Poll. Bull., 55: 16-29

Neto J.M., Gaspar R., Pereira L., Marques J.C., 2012. Marine Macroalgae Assessment Tool (MarMAT) for intertidal rocky shores. Quality assessment under the scope of the European Water Framework Directive Ecol. Ind 19: 39-47

Nilsson H.C., Rosenberg R., 1997. Benthic habitat quality assessment of an oxygen stressed fjord by surface and sediment profile images. Journal of Marine Systems 11:249-264.

Nilsson H.C., Rosenberg R., 2000. Succession in marine benthic habitats and fauna in response to oxygen deficiency: analysed by sediment profile-imaging and by grab samples. Marine Ecology Progress Series 197:139-149.

Rhoads D.C., Cande S., 1971. Sediment profile camera for in situ study of organism-sediment relations.LimnolOceanogr 16:110-114

Rumohr H., Schomann H., 1992. REMOTS sediment profiles around an exploratory drilling rig in the southern North Sea. Marine Ecology Progress Series 91:303-311.

Rosenberg R., Nilsson H.C., Grémare A., Amouroux J.M., 2003. Effects of demersal trawling on marine sedimentary habitats analysed by sediment profile imagery. Journal of Experimental Marine Biology and Ecology 285-286: 465-477.

Scanlan C.M., Foden J., Wells E., Best M.A., 2007: The monitoring of opportunistic macroalgal blooms for the water framework directive Marine Poll. Bull., 55: 162-171

Schönfeld J., Alve E., Geslin E., Jorissen F., Korsun S., Spezzaferri S. and FOBIMO, 2012. The FOBIMO (FOraminiferalBIo-MOnitoring) initiative—Towards a standardised protocol for soft-bottom benthic foraminiferal monitoring studies Marine Micropaleontology, 94/95. pp. 1-13.

Teixeira H., Neto J.M., Patrício J., Veríssimo H., Pinto R., Salas F., Marques J.C., 2009.Quality assessment of benthic macroinvertebrates under the scope of WFD using BAT, the Benthic Assessment Tool Marine Poll. Bull., 58: 1477-1486

Tett P., Carreira C., Mills D.K., Van Leeuwen S., Foden J., Bresnan E., Gowen R.J., 2008. Use of the Phytoplankton Community Index to assess the health of coastal waters. ICES Journal of Marine Science, 65, 1475-1482.

Tett P., Gowen R., Painting S., Elliott M., Foster R., Mills D., Bresnan E., Capuzzo E., Fernandes T., Foden J., Geider R., Gilpin L., Huxham M., McQuatters-Gollop A., Malcolm S., Saux-Picart S., Platt T., Racault M.F., Sathyendranath S., VanderMolen J., Wilkinson M., 2013. A framework for understanding marine ecosystem health. Marine Ecology Progress Series, 494, 1-27. DOI: 10.3354/meps10539.

Walvoort D.W.W. and Van Loon W.M.G.M., 2016. BENMMI user manual. http://publicaties.minienm.nl/documenten/user-manual-benmmi-3-0.

Van Hoey G., Drent J., Ysebaert T., Herman P., 2007. The Benthic Ecosystem Quality Index (BEQI), intercalibration and assessment of Dutch coastal and transitional waters for the Water Framework Directive, Final report, NIOO.

Van Loon W.M.G.M., Verschoor A.J., Gittenberger A., 2012. Benthic Ecosystem Quality Index 2: Design and Calibration of the BEQI-2 WFD metric for marine benthos in transitional waters. Report, RWS Waterdienst, 35 pp

Van Loon W.M.G.M., Walvoort D.J.J., Van Hoey G., Vina-Herbon C., Blandon A., Pesch R., Schmitt P., Scholle J., Heyer K., Lavaleye M., Phillips G., Duineveld G.C.A., Blomqvist M. A regional benthic fauna assessment method for the Southern North Sea using Margalef diversity and reference value modelling. Ecological Indicators, in press, 2017.Wells E., Wilkinson M., Wood P., Scanlan C., 2007. The use of macroalgal species richness and composition on intertidal rocky seashores in the assessment of ecological quality under the European Water Framework Directive. Mar Poll Bull 55: 151-161

Annex 1 Appendix 1:

Summary description of the existing (mostly WFD) biotic indices using multi-metric calculation. *Update was made through associated research programmes (Southern North Sea, BenthoVAL and EcApRHA): See Annex 2 for details of WFD indices and methods used in OSPAR maritime area.*

COASTAL MACROALGAE		
INDICATOR	**CP APPLYING for WFD**	**BRIEF DESCRIPTION AND REFERENCES**
RSL	IE, UK, NO, ES	"Reduced species list" is an indicator that includes 5 components: species richness (from a reduced list of between 68 and 70 species) /proportion of Rhodophyta/ proportion of Chlorophyta/ proportion of opportunistic species/ratio of perennial to annual forms. Wells et al. 2007: Mar. Pollution. Bull. 55: 151–161. Recently adapted to Southern Spain (Bermejo et al, 2012: Ecol. Ind, 12: 46-57)
CFR	ES	"Quality of rocky bottoms". This indicator includes: percentage cover of characteristic macroalgae/ macroalgae population richness / percentage cover of opportunistic species. (Juanes et al, 2008. Ecol. Ind, 8: 351-359)
p-MarMAT	PT	"Portuguese marine macroalgae assessment tool". This indicator contemplates all the components of the RSL method, including an additional parameter of % coverage of opportunistic species. Neto et al. 2011, Ecol. Ind: in press.
MAB	IE, UK, DE	"Macroalgae blooming", which measures 5 components: % cover of algae from the intertidal area / total extent of algal blooms/ biomass of algae/ biomass of algae over the affected bloom area/ presence of entrained algae. Scanlan et al, 2007: Marine Poll. Bull., 55: 162-171
RICQI	ES	"Rocky Intertidal Community Quality Index". It includes: ESS: ecological status similarity, PC: presence of Cystoseira; Morphologically Complex Algae; Species Richness and Faunal cover. Díez I., M. et al 2012. Ecol. Ind*,* 12: 58-71
Subtidal algae	NO, SE, DK	This indicator incorporates: depth extension of selected perennial macroalgae species / cover of macroalgae along depth gradients / macroalgae composition in rocky littoral and sublittoral zones (based on the integration of multiple parameters). Krause-Jensen et al, 2008. Ecol. Indicat. 8: 515-529
COASTAL ANGIOSPERMS		
INDICATOR	**CP APPLYING**	**BRIEF DESCRIPTION AND REFERENCES**
Intertidal *Zostera*	UK, IE, NL DE (only bed extent), FR	Multimetric indicator for *Zostera* communities. It includes three parameters: 1: Species composition; 2. Seagrass abundance (acreage/bed extent) and 3. Seagrass abundance (coverage/density)
BENTHIC INVERTEBRATE FAUNA		
INDICATOR	**CP APPLYING**	**BRIEF DESCRIPTION AND REFERENCES**
BEQI	BE	Benthic ecosystem quality index. Indicators included: density, biomass, species richness, species composition changes (Van Hoeij et al., 2007. Report NIOO/NIOZ).
BEQI-2	NL	Benthic ecosystem quality index 2. Indicators included: species richness, Shannon index and AMBI; univariate calibration (Van Loon et al., Report, 2011)
M-AMBI	ES, FR, DE	Multimetric-AMBI: Species richness, Shannon diversity, and AMBI, multi-variate calibration (Muxika et al., Marine Poll. Bull., 2007. Mar. Poll. Bull., 55: 16-29).
IQI	IE, UK	Infaunal quality index (IQI) is comprised of three indicators: species richness, Simpson diversity and AMBI (Borja et al., Marine Poll. Bull., 2007)
NQI	NO	Norwegian quality index: includes SN-diversity, total abundance and AMBI (Josefson et al., Marine Poll. Bull., 2009).
P-BAT	PT	Portuguese benthic assessment tool. Indicators included: Margalef species richness, Shannon diversity and AMBI (Teixeira et al., Marine Poll. Bull., 2009).

BQI	SE	Benthic Quality index. Indicators included: species richness and total abundance (Josefson et al., 2009).
DKI	DK	Danish quality index. Indicators included: AMBI, Shannon's diversity, total abundance and species richness (Josefson et al., Marine Poll. Bull., 2009).
BOPA	ES	Benthic Opportunistic Annelida Amphipoda Index /Benthic Opportunistic PolychaeteAmphipoda Index. It uses the ratio between opportunistic polychaete and amphipods frequencies. Dauvin, J.C., T. Ruellet, 2007. Mar. Pollut. Bull, 55(1-6), 215-224.
BHQ	SE	Benthic Habitat Quality index. Used in analyze of sediment profile image. Tested against organic enrichment, oxygen deficiency, physical disturbance (e.g. trawl fishery). Nilsson H.C. and Rosenberg R., 1997. Marine System, 11, 249-264.

ANNEX 2

OSPAR – Common Biodiversity Indicators

Condition of benthic habitat communities (BH2-A): Some coastal habitats in relation to nutrients and/or organic enrichment

Guidelines for Coordinated Environmental Monitoring Programme (CEMP)

Last submission to OSPAR Secretariat and BDC (March 2022), OSPAR Benthic Habitat Expert Group (November 2022), COBAM (November 2022). All comments received to date are addressed in this updated version.

Adoption (BDC 2017) = OSPAR Agreement 2018-06

Adoption of this update: BDC (December 2022)

Main authors: Laurent Guérin (FR), Anna Lizińska (FR) and Petra Schmitt (DE)

The CEMP guidelines are a tool for clearly documenting the detailed methodology for the monitoring and/or assessment of the individual parameters/indicators, to ensure transparency, coherence and consistency in the implementation of such monitoring and assessment. *These guidelines should outline themonitoring and assessment requirements for the components outlined within the CEMP and should be developed for each element (i.e. common indicators)*

1 Introduction

Coastal waters represent the interface between land and ocean and are defined under Water Framework Directive (WFD) as extending one nautical mile from a baseline defined by the land points where territorial waters are measured. “Water bodies” have been defined by European Union Member States by sub-dividing these coastal waters, along the coast, according to their hydrological characteristics. Some coastal waters of OSPAR Contracting Parties have, to a varying degree, been affected by nutrient and organic enrichment (either as a direct or indirect effect of nutrient inputs or of organic matter enrichment) and this may lead to nuisance and toxic algal blooms, loss of benthic habitats by shading out benthic vegetation andmodification of benthic fauna communities due to specific sensitivities. “Coastal habitats”, are defined here according to the Marine Strategy Framework Directive (MSFD) terminology, as the seabed and subsoil on the landward side of a line, every point of which is at a distance of one nautical mile on the seaward side from the nearest point of the baseline from which the breadth of territorial waters is measured.

From Solheim et al. (2012):

“The biological quality elements required for the integrated assessment of ecological status (WFD Annex V)of coastal waters are phytoplankton, macroalgae, angiosperms and benthic fauna.”

The WFD requires that standardised methods are used for the monitoring of the quality elements and that the good status class boundaries for each biological quality element are intercalibrated across the Member States sharing similar types of water bodies. As a result of the intercalibration exercises, the national methods and their boundary values, for each Member State monitoring system classifications, were established by the European Commision (2018) decision.

The WFD requires that the overall ecological status of a water body is determined by the results for the biological or physicochemical quality element with the worst class determined by any of the biological quality elements. This is called the “one out - all out” principle. The rationale of this principle is to avoid averaging the impacts on different quality elements due to different pressures and therefore overlook some significant pressures, and also to provide sufficient protection of the most sensitive quality elementto significant pressure.”

By definition, the WFD aims to assess the ecological quality status of water bodies, reflected by quality elements such as benthic invertebrates or macroflora. Benthic habitat communities are thus mainly assessed for the WFD to reflect the main pressures in the coastal zone, such as direct or indirect effects of nutrient and organic enrichment, notably due to nutrient inputs (as terrestrial or freshwater sources of pressure) in the OSPAR Maritime Area. Contaminants are generally assessed in sediment. For these reasons, WFD monitoring was generally designed to avoid other pressures, and assess only the effects of nutrient and organic enrichment (including for reference stations). This does not mean that the indices used are not sensitive, per se, to other pressure types, but currently, several statistical indices used, notablythose including a species sensitivity classification, have been developed and calibrated to assess nutrient inputs or organic matter enrichment. MSFD’s requirements lead to a reconsideration/ adaptation of monitoring, to take into account and assess other pressure types (both inshore and offshore).

2 Monitoring

Specific monitoring programmes have been implemented for Water Framework Directive, and the Water Environment Regulations (WER) in [15]UK waters, to provide data for this specific assessment. For benthic habitats, specific habitat sub-types sampled, sampling methods and reference conditions used, are detailed in the following chapters.

[15] This refers to the Water Environment (Water Framework Directive) (England and Wales) Regulations 2017, The Water Environment (Water Framework Directive) Regulations (Northern Ireland) 2017, Water Environment and Water Services (Scotland) Act 2003

Benthic Invertebrates

Sampling methods used for the benthic multimetric indices, habitat types sampled and the source to derivereference conditions are summarised in Table 1. Mostly a Van Veen grab or an equivalent is used for takingsublittoral samples and a corer for intertidal habitats. The number of spatial replicates, sampling occasions and the total area surveyed for the classification of a water body varies by country. All of the indices are designed to assess subtidal soft-bottom sediments and in some member states also intertidal habitats. The reference conditions are derived from historical data or sites with the least disturbed conditions, when available, supported by expert knowledge and in some cases also statistical modelling.

Table 1: Sampling method, habitat types sampled and reference conditions for the assessment of benthic invertebrates (updates from Van Hoey et al. 2019)

Contracting Party	Index	Sampling Method	Habitat type	Reference conditions
BE	BEQI	Van Veen grab (0.1 m²), 18-20 replicates, yearly	subtidal soft-bottom (muddy sediments, fine muddy sand, clean sands)	expert knowledge, historical data, least disturbed conditions; data period 1994-2012, habitat specific
DE	M-AMBI[1]	Van Veen grab (0.1 m²) or corer (diameter 9-15 cm), 6-10 replicates per ecotope, once or twice per year	subtidal and intertidal soft-bottom	expert knowledge, historical data, least disturbed conditions, reference time: 1959 up to now, habitat-specific
DK	DKI	Van Veen grab (0.1 m²) or Haps-corer (0.0143 m²), 6 (grab) or 40 (corer) replicates, yearly	subtidal soft-bottom (sand - mud)	least disturbed conditions, recent data from least impacted sites, surface water type-specific
ES	M-AMBI	Van Veen grab (0.1 m²), 3 replicates, 2-6 locations per water body, yearly	subtidal and intertidal soft bottom	expert knowledge, historical data, modelling (extrapolating model results), period 1995-2005, habitat-specific
ES	BOPA	Van Veen grab (0.1 m²), 3 replicates, yearly	subtidal soft-bottom	least disturbed conditions, habitat-specific
FR	M-AMBI[2]	Van Veen grab / Day grab / Smith-McIntyre grab (0.1 m²), 3 locations with 3 replicates each, once per sampling season	subtidal and intertidal soft-bottom	expert knowledge, historical data, least disturbed conditions; data period 1995-2006, habitat specific
IE, UK	IQI	Van Veen grab / Day grab (0.1 m²), hand corer (0.01 m²), replicates and sampling occasions variable according to habitats, methodology and required confidence	subtidal and intertidal soft-bottom	expert knowledge, least disturbed conditions and modelling (extrapolating model results), data from 1979 to 2003, habitat-specific
NL	BEQI2	corer tube, box corer (e.g. Reineck box corer), at least once per year	subtidal and intertidal, all present habitats	historical data for 1991-2006, statistical modelling

NO	NQI1	Van Veen grab (0.1 m²), 4 replicates, yearly	subtidal soft-bottom	recent data from least impacted sites
PT	BAT	Van Veen grab (0.1 m²) or equivalent, replicates variable according to habitats and required confidence, once per sampling season	subtidal soft-bottom (sand - mud)	existing near-natural reference sites, expert knowledge, historical data, least disturbed conditions, habitat-specific
SE	BQI	Van Veen grab, Smith-McIntyre grab (0.1 m²), at least 5 locations per water body, yearly	subtidal soft-bottom	expert knowledge, least disturbed conditions

[1]m-AMBI method, but other reference and boundary settings. (Van Hoey et al. 2019)
[2]m-AMBI method, but other reference settings. (Van Hoey et al. 2019)

Macroalgae

The sampling method applied to assess macroalgae depends on the location of habitats (Table 2). Opportunistic intertidal macroalgae are often surveyed by aerial imagery. Intertidal and subtidal rocky substrates are assessed by field surveys and diver observations along transects. The sampling procedure, number of spatial replicates, total area surveyed and sampling occasions vary considerably across countries and reflect the methodological diversity of indices used. Reference conditions are mostly based on expert knowledge and existing near-natural reference sites. For several countries, historical data is also available.

Table 2: Sampling method, habitat types sampled and reference conditions for the assessment of macroalgae

Contracting Party	**Index**	**Sampling Method**	**Habitat type**	**Reference conditions**
DE	Opportunistic Macroalgae	aerial mapping (extent) of total area, 3 - 5 occasions persampling season	intertidal, all habitats	expert knowledge, historical data
DE	HPI	field survey (% cover, taxonomic composition, RSL-Index) along transects and on fixed stations, diver observations along transects (depth limits), sampling occasions variable according to methodology	intertidal and subtidal hard-bottom	expert knowledge

DK	Total Macroalgal Cover	diver observations (% cover, total and individual species) along a depth gradient (transect), 2-5 transects per area, 3 replicates each about 25 m², once every 2 years	subtidal hard-bottom habitats	Modelling (extrapolating model results), historical data (nitrogen loads)
ES	CFR	field survey, diver observations, 3 replicates (10-100 m²) per assessed level, once per sampling season	intertidal and subtidal hard-bottom	existing near-natural reference sites, expert knowledge
ES	RSL	field survey (taxonomic composition, cover), ten replicates, yearly	intertidal hard-bottom	existing near-natural reference sites, expert knowledge
FR	QI Sub Mac Fr	diver observations along a transect, identification and counting of macroalgae, 10 quadrates (2.5 m²) in upper and 8 quadrates (2 m²) in lower infralittoral of each algal belt, once per sampling season	infralittoral hard-bottom	expert knowledge, historical data, least disturbed conditions, existing near-natural reference sites
FR	CCO	field survey, 3 permanent spotsper level/belt marked by a frame (1.65 x 1.65 m), onceevery three years	intertidal hard-bottom	existing near-natural reference sites & least impacted sites, expert knowledge
IE	Opportunistic Green Macroalgal Abundance	field survey, transects across algal beds, at least 5 quadrates (0.25 m²) per transect, once per sampling season	intertidal sediments suitable for green algae	existing near-natural reference sites, expert knowledge, historical data
IE	RSL	field survey, at least 3 sites per water body with 4-5 shore heights and 5 quadrats at each shore height, twice in 6 years	intertidal hard-bottom	existing near-natural reference sites, expert knowledge, historical data
NO	MSMDI	diver observations along transects (% cover, depth limits), at least two transectsper water body, once per sampling season		expert knowledge, historical data, least disturbed conditions
NO	RSLA	field survey, divers observations, 10-15 m wide transects	intertidal and subtidal hard-bottom	existing near-natural reference sites, expert knowledge

PT	PMarMAT	field survey along transects, seven stations per transect with 3 replicates each (quadrates 0.2x0.2 m), 1-3 transects per site, 1-3 sites perwater body, once per sampling season	intertidal hard-bottom	existing near-natural reference sites, expert knowledge, historical data, least disturbed conditions
SE	MSMDI	diver observations along transects (% cover, depth limits), at least three transects per waterbody, once per year	subtidal, all suitable habitats	historical data, expert knowledge, least disturbed conditions
UK	RSL	field survey along transects, 100-300 m shore length, minimum of 3 sites per water body, once per sampling season	intertidal hard-bottom	expert knowledge, historical data, least disturbed conditions
UK	Macroalgal Bloom Assessment	field survey, aerial imagery, multiple quadrats (0.25 m²), replicates vary with extent, patchiness and abundance of macroalgal blooms, once per sampling season	intertidal soft sediment, mussel beds	expert knowledge, historical data

Angiosperms

The methods applied for the assessment of angiosperms depend on the location of the habitat (Table 3).Intertidal habitats are assessed either with aerial imagery or by field surveys. Subtidal habitats are sampled by diving surveys along transects. Preferably, all existing seagrass beds shall be assessed. Reference conditions are derived from historical data, mostly supported by expert knowledge. Additionally, in some countries, existing near-natural reference sites or habitats with the least disturbed conditions are considered.

Table 3: Sampling method, habitat types sampled and reference conditions for the assessment ofangiosperms

Contracting Party	Sampling Method	Habitat type	Reference conditions
DE - saltmarsh	aerial mapping (extent) in combination with field mapping (extent and zonation), once in six years, some selected sites every 2-3 years	intertidal, all available habitats	existing near-natural reference sites, expert knowledge, historical data, modelling (extrapolating model results)

DE - seagrass	aerial mapping (extent) in combination with field mapping (species composition and density along transects), all seagrass beds, once in six years, some selected sites every year	intertidal, all available habitats	expert knowledge, historical data, modelling (extrapolating model results)
DK	diver observations (% cover) along with a depth gradient (transect), about 5 sitesper area, yearly	subtidal soft-bottom	historical data
FR	aerial photography, field survey, once in 3 years	intertidal and subtidal	historical data, expert knowledge
IE	field survey of all seagrass beds in a waterbody, transects across each bed, one occasion over several years	intertidal seagrass beds	existing near-natural reference sites, expert knowledge, historical data
SE	diving surveys, investigating changes along transects, (cover % of phytobenthic species) together with quantitative sampling, at least three transects per waterbody, yearly	subtidal, all available habitats	expert knowledge, historical data, least disturbed conditions
UK	field survey with quadrat, aerial photography, all seagrass beds, once per sampling season	intertidal, all available habitats	expert knowledge, historical data, least disturbed conditions

3 Assessment

The first (2017) OSPAR assessment results for the Biological Quality Elements 'Benthic Invertebrates' and 'Macroalgae and Angiosperms' in coastal waters were requested through a formal data call to OSPAR Contracting Parties and notably WFD national contacts. In addition, information on the indices used and the monitoring stations in coastal water bodies was requested. Information on the methodology of the various benthic indices (sampling strategy, data evaluation, reference conditions, detected pressures) was mainly derived from the WISER methods database and intercalibration reports (Birk et al. 2010, van Hoey et al. 2015).

The current (2023) OSPAR assessment results were obtained in early 2022, from WISE Water Framework Directive Database[16], which includes information for the countries that have reported under WFD 2010-2016 reporting cycle. A formal data call was done by OSPAR to encourage Contracting Parties to make their data available under the WISE database. Data from the next reporting cycle (2017-2022) were not available from any Contracting Parties. The information about coastal water bodies and monitoring stations were taken from WISE WFD reference spatial data sets[17]. Information on the methodology of the various benthic indices was mainly derived from intercalibration reports (Birk et al., 2010; Van Hoey et al., 2015;

[16] https://www.eea.europa.eu/data-and-maps/data/wise-wfd-4

[17] https://www.eea.europa.eu/data-and-maps/data/wise-wfd-spatial-3

Neto et al., 2018; Wilkes et al., 2018; Van Hoey et al., 2019; Salas Herrero et al., 2020). The intercalibration of the ecological quality status and assessment methods were carried out for benthic invertebrates (Van Hoey et al., 2019), opportunistic macroalgae (Wilkes et al., 2018) and seagrasses (Neto et al., 2018) in the coastal waters of the North-East Atlantic. As a result, the values of the boundaries between classes of Ecological Quality Ratio of the Member States were established (European Commission, 2018). In the UK the approach is set out as part of the Water Environment Regulations, which is not part of WFD implementation nor reporting.

In 2022, WFD assessment results for benthic invertebrates were provided from all of the eleven OSPAR Contracting Parties, with coastal water bodies in the three OSPAR regions (II – Greater North Sea, III – Celtic Seas, IV – Bay of Biscay & Iberian coast) where this indicator was adopted as common (Table 4). The quality element 'other aquatic flora' (QE1-2) contains four sub-elements QE1-2-1 – Macroalgae, QE1-2-2 – Angiosperms, QE1-2-3 – Macrophytes and QE1-2-4 – Phytobenthos. Two first mentioned are relevant to coastal water bodies. The assessment results for macroalgae and angiosperms were available from nine countries. For the Netherlands and Belgium, this quality element is not present, thus not relevant or their coastal waters. Germany, Spain, Ireland, Netherlands and Portugal reported the status of some coastal water bodies with regards to 'other aquatic flora' using joint normative definitions for macroalgae and angiosperms (WFD 2003, Annex V 1.2.4).

Table 4: Data availability on WFD assessment results for benthic invertebrates and macroalgae/ angiosperms / other aquatic flora

Contracting Party	BE	DE	DK	ES	FR	IE	NL	NO	PT	SE	UK
Benthic invertebrates	x	x	x	x	x	x	x	x	x	x	x
Macroalgae	*na*	x	*na*	x	x	x	*na*	x	x	x	x
Angiosperms	*na*	x	x	x	x	x	*na*	x	x	*na*	x
Other aquatic flora	*na*	x	*na*	x	*na*	x	x	*na*	x	*na*	*na*

x – reported and assessed, na – reported not assessed; BE = Belgium, DE = Germany, DK = Denmark, ES = Spain, FR = France, IE = the Republic of Ireland, NL = the Netherlands, NO = Norway, PT = Portugal, SE = Sweden, UK = the United Kingdom

Detailed assessment methods and intercalibration exercises are available at:

https://ec.europa.eu/environment/water/water-framework/facts_figures/guidance_docs_en.htm

https://www.eea.europa.eu/themes/water/european-waters/water-quality-and-water-assessment/water-assessments

Details on metrics and indices used by Contracting Parties are described in the following chapters.

Benthic Invertebrates

The quality status of the benthic fauna is generally assessed with multimetric indices, as the WFD requires the inclusion of the metrics taxonomic composition, diversity, abundance and taxa sensitive to disturbance. In the OSPAR region, eleven different benthic assessment approaches are applied (Table 5). Except for the Spanish BOPA (Benthic Opportunistic Polychaete Amphipoda Index), all approaches comply with the WFD requirements. Most of the countries use a similar approach with the AMBI (AZTI Marine Biotic Index) as a metric indicating anthropogenic disturbance, a diversity index (Shannon index, Simpson index or Margalef) and the number of species. The Belgian BEQI with the metric taxonomic composition (Bray-Curtis similarity) has a different approach where the assessment is performed on habitat level instead of the sample level. The BOPA used in Andalusia (Spain) is based on the relative abundance of opportunistic polychaetes and amphipods but lacks a metric for diversity (Van Hoey et al. 2015, Van Hoey et al. 2019). The BOPA index is also the only assessment approach that has not been intercalibrated within the North East Atlantic intercalibration exercise (Van Hoey et al. 2015). However, the BOPA intercalibration exercise was presented in a separate intercalibration document (Van Hoey et al. 2019).

In the UK and Norway, additional indices are applied for the assessment of benthic fauna in coastal waters. The Vas Deferens Sequence Index (VDSI) is an imposex tool used in the UK as a subcomponent for the statusassessment of invertebrates. Norway assesses some coastal water bodies with univariate diversity indices (Table 6). It should be noted that the UK benthic invertebrate ecological status is not only targeted at organic enrichment pressures.

Table 5: Multimetric indices used for the assessment of benthic invertebrates

Contracting Party	**Index**	**Metrics**	**Reference**	**Assessed pressure (Borja et al. 2015)**
BE	BEQI Benthic Ecosystem QualityIndex	number of species, abundance, biomass,Bray-Curtis similarity	Van Hoey et al. 2007), http://www.beqi.eu	multipressure
DE	M-AMBI Multivariate-AZTI MarineBiotic Index	factor analysis: number of species,Shannon index (H'), AMBI	Borja et al., (2004), Muxika et al. 2007, http://ambi.azti.es	aquaculture, change coast, climate, dredging, harbor, multipressure, organic enrichment, sewage
DK	DKI Danish Quality Index	number of species, abundance, Shannonindex (H'), AMBI	Borja et al. 2007	Chemical pollution, climate, multipressure
ES	M-AMBI Multivariate-AZTI MarineBiotic Index	factor analysis: number of species,Shannon index (H'), AMBI	Borja et al., 2004, Muxikaet al. 2007, http://ambi.azti.e	aquaculture, change coast, climate, dredging, harbor, multipressure, organic enrichment, sewage

			s	
ES	BOPA Benthic Opportunistic Polychaete Amphipoda Index	relative abundance of opportunisticpolychaetes and amphipods	Dauvin & Ruellet 2007	Aquaculture, multipressure, oil expraction, physical alteration, sewage
FR	M-AMBI Multivariate-AZTI MarineBiotic Index	factor analysis: number of species,Shannon index (H'), AMBI	Borja et al.2004, Muxika et al. 2007, http://ambi.azti.es	aquaculture, change coast, climate, dredging, harbor, multipressure, organic enrichment, sewage
IE, UK	IQI Infaunal Quality Index	number of species, Simpson index, AMBI	Phillips et al. 2014	Climate, Eutrophication, multipressure
NL	BEQI2 Benthic Ecosystem QualityIndex 2	number of species, Shannon index (H'),AMBI	Van Loon et al. 2015	Multipressure Van Loon et al. (2015)
NO	NQI1 Norwegian Quality Index	number of species, AMBI	Rygg 1985, 2002	multipressure
PT	BAT Benthic Assessment Tool	factor analysis: Margalef (D), Shannonindex (H'), AMBI	Teixeira et al. 2009, Marques et al. 2009	Dredging, multipressure
SE	BQI Benthic Quality Index	number of species, abundance, relativeabundance of sensitive and tolerant species	Leonardsson et al. 2009	Aquaculture, chemical polutions, multipresures, organic enrichment

Table 6: Additional indices for the assessment of benthic invertebrates

Contracting Party	**Index**	**Method**	**Reference**
NO	Density Index (DI)	univariate	Rygg 2014
NO	Shannon Index (H’)	univariate	Shannon 1949
NO	Hurlbert Index (ES100)	univariate	Hurlbert 1971
UK	Vas deferens sequence index (VDSI)	imposex tool	Gibbs et al. 1987

Macroalgae

The indices for the assessment of the subcomponent macroalgae apply either to intertidal and/or subtidal macroalgae on rocky shores or to blooming opportunistic macroalgae. Some countries have developed several indices (e.g. DE, FR, UK). Information on 15 approaches used in nine OSPAR's Contracting Parties has been compiled (Table 7). Opportunistic macroalgae are assessed in Germany, Ireland, France and the United Kingdom. These indices are designed to survey intertidal soft-bottom habitats suitable for green algae growth and to map the extent and density of algal beds (Table 2). Various indices have been developed for the classification of macroalgae's quality status on rocky habitats. Metrics used are species richness, coverage, depth limits or proportion of opportunistic and sensitive species. An index adopted by several countries is the Reduce Species List (RSL) developed by Wells et al. (2007). The RSL is a multimetric index based on the principlethat species richness has been shown to remain constant in the absence of anthropogenic disturbance. Metrics incorporated in this index are the number of species, the proportion or number of red and green algae, the proportion of opportunistic species and the ratio of perennial forms (ecological status group 1) to annual or ephemeral forms (ecological status group 2). The intercalibration exercise for opportunistic macroalgae ecological assessment methods was carried out between France and Germany (Willes et al. 2018) and common class boundaries have been established (Wilkes et al. 2018).

Table 7: Indices for the assessment of macroalgae

Contracting Party	**Index**	**Metrics**	**Reference**	**Assessed pressure** Høgslund et al. 2022 and others
DE	Opportunistic Macroalgae	spread and density of opportunistic macroalgae	Kolbe 2007	
DE	HPI Helgoland Phytobenhic Index	taxonomic composition, cover, density, depthlimits	Kuhlenkam p &Bartsch 2007 Kuhlenkamp et al. 2011	euthrophication
DK	Total macroalgal cover	total macroalgal cover (%) per water depth (notyet fully approved)	Carstensen et al.2008	
ES	CFR Quality of Rocky Bottoms	coverage, fraction and richness of characteristicmacroalgae	Juanes et al. 2008,Guinda et al. 2008	euthrophication
ES	RSL Reduced Species List	number of species, the proportion of green algae, number of red algae, proportion of ESG 1, the proportion of opportunistic species, shore description	Wells et al. 2007, Bermejo et al. 2012	euthrophication

FR	QI Sub Mac Fr Quality index of subtidal macroalgae of French Channel and Atlantic coast	depth of infralittoral belts, the density of structuring species, number of characteristic species, densitiesof opportunistic species, presence of "good ecological status species", biodiversity, stipe length and surface of epibionts of *Laminaria hyperborea*	Derrien-Courtel & LeGal 2009, Le Gal & Derrien-Courtel 2015	turbidity, sediment inputs, eutrophisation
FR	CCO Cover, Characteristic species, Opportunistic species on intertidal rocky bottoms	contribution of each macroalgal community (belt)to the total cover of a given intertidal rocky shore,number of characteristic species, cover of opportunistic species in every belt	Ar Gall & Le Duff2014	euthrophication
FR	Macroalgal Bloom Assessment (Opportunistic Green macroalgae) - CWOGA	maximum % of the colonisable area covered by ulva, average % of colonisable area covered by ulva, frequency of bloom of colonisable area covered by ulva, frequency of blooms	Wilkes et al. 2018	anthropogenically elevated nutrient levels
IE	Opportunistic Green Macroalgal Abundance	total cover, the total affected area, biomass	Scanlan et al. 2007	euthrophication
IE	RSL Rocky Intertidal Macroalgae - Reduced Species List	number of species (green, brown, red algae, opportunists, ecological status groups 1 and 2),shore description	Wilkinson et al. 2007	euthrophication
NO	MSMDI Multi-Species MaximumDepth Index	depth limits of 3 to 9 disturbance sensitive species	Kautsky et al. 2006	euthrophication
NO	RSLA Rocky Shore Reduced Species List with Abundance	number of species (green, brown, red algae, opportunists, ecological status groups 1 and 2),abundance, shore description	Wilkinson et al.2007	euthrophication
PT	PMarMAT Marine Macroalgae Assessment Tool	taxa richness, proportion of green algae, numberof red algae, ESG ratio, proportion and cover of opportunists, shore description	Marques et al. 2009, Neto et al. 2011	euthrophication
SE	MSMDI Multi-Species MaximumDepth Index	depth limits of 3 to 9 disturbance sensitive species	Kautsky et al. 2006	eutrophication

UK	RSL Rocky Shore Reduced Species List	normalised number of macroalgal taxa (normalised to shore diversity), proportion of green algae, red algae and of opportunistic taxa, ecological status group ratio	Wells 2006, Wells etal. 2007	euthrophication
UK	Macroalgal Bloom Assessment	total extent of macroalgal bed, cover of availableintertidal habitat, biomass of opportunistic macroalgal mats, biomass over the available intertidal habitat, proportion of entrained algae	Scanlan et al. 2007,Wells et al. 2007	

Angiosperms

Two intercalibration exercises were conducted for angiosperms. The first one was focused on ecological assessment methods of seagrasses (Neto et al 2018), and the second one considers saltmarshes ecological assessment methods (Neto et al. 2019). Indices for the ecological status assessment of seagrass have been developed by nine countries (Table 8). Metrics used by most member states are the spatial extent of seagrass beds, number of species and shoot density. Denmark determines the depth limit of seagrass by measuring the percentage of seagrass cover along transects.

For the saltmarsh intercalibration exercise was made between Germany, Ireland, the Netherlands and the United Kingdom. The participating countries considered saltmarshes as an ecologically meaningful Biological Quality Element (BQE) and use it in the assessment of the ecological quality of waterbodies (Neto et al. 2019)

The main method of assessment was based on national expert judgement and the intercalibration was not possible however the national methods were accepted.

Table 8: Indices for the assessment of angiosperms (updates after Neto et al. 2018, 2019)

Contracting Party	**Index**	**Metrics**	**Reference**	**Assessed pressure**
DE	SG = Assessment tool for intertidal seagrass in coastal and transitional waters	proportion of intertidal area covered by seagrass,density within seagrass beds, number of species	Dolch et al. 2008,Kolbe 2007	
DE	EM = Assessment of saltmarsh vegetation in coastal and transitional	the extent of saltmarsh area (percentage of saltmarsh area of the whole water body) compared to historical	Adolph & Arens, 2011	

	waters	references and on the relative extent of vegetation zones (percentage of zones of the whole saltmarsh area). The overall EQR value is obtained by calculating the mean of the mentioned metrics without any additional weighting		
DK	Depth limit of eelgrass	eelgrass cover (%) per water depth	Krause-Jensen et al. 2005, Duarte et al.2007	
FR	SBQ = Seagrass beds quality in coastal and transitional water bodies (same method for CW & TW)	taxonomic composition, spatial extent, shootdensity, trends in abundance, number of taxa (trends)	Auby et al. 2010	
IE	SG = Seagrass Intertidal tool	spatial extent, shoot density, trends in abundance, number of species	Foden & de Jong 2007, Foden & Brazier 2007	
IE	SMAATIE = Saltmarsh Angiosperm Assessment Tool for Ireland	saltmarsh zonation (taxonomic composition), saltmarsh extent (angiosperm abundance) and presence of halophytes (disturbance-sensitive taxa) The overall EQR is calculated with the attribution of different weightings for combining the metrics.	Devaney &Perrin 2015	
SE	Assessment of Biological Quality Elements - macro-vegetation	seagrass shoot density		
UK	SG = Seagrass Intertidal tool	taxonomic composition, trends in abundance, spatial extent, shootdensity	Foden & de Jong 2007, Foden & Brazier 2007,	

			UKTAG 2014	
UK	SM = UK Saltmarsh Tool		UKTAG 2014	
PT	SQI = Seagrass quality index	taxonomic composition, spatial extent, shootdensity	Neto et al. 2013	
ES	AQI = Angiosperms Quality Index	Relative coverage of estuarine habitats (relative deviations from optimal coverage); Variations in the surface area of natural tidal habitats.	García et al. 2009	Mixed ecological-pressure
NL	SG = Monitoring beds of SG per waterbody using aerial photographs, ground truth and specifying surface & density per species	spatial extent, shoot density, trends in abundance, number of species	Foden & de Jong 2007	
NL	TSM = WFD-metrics for natural water types: tidal salt marsh	condition acreage (area) and condition quality (zonation). The overall EQR value is obtained through the calculation of the mean between quality and quantity metrics, without weighting	Dijkema et al. 2005	

4 Change Management

Since changes of this method often require a more general level of discussion and policy decision making, it is proposed to discuss proposed changes in ICG-COBAM, in link with ICG-C and ICG-EUT, and to make decisions on proposed changes in ICG-COBAM and BDC.

The similarities and differences between different assessment approaches of benthic habitats were discussed during the 29th Meeting of the European Union Marine Strategy Coordination Group (MSCG) to discuss a document from TG Seabed (TG Seabed, 2021). Three options of applying WFD and HD assessments to MSFD obligations were described. The document highlights that WFD indices have been successfully established to reflect the condition of benthic communities, in particular in relation to organic and nutrient enrichment for Member States waters. This should be considered in OSPAR process and thematic assessment.

References

Adolph, W. & Arens, S. (2011). Assessment of reeds, salt- and brackish marshes in the transitional and coastal waters of Lower Saxony (Germany). Report to Niedersächsicher Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, DE.

Ar Gall, E. & Le Duff, M. (2014). Development of a quality index to evaluate the structure of macroalgal communities. Est. Coast. Shelf Sci. 139, 99-109.

Arens S. (2009). Erfassung und Bewertung der Röhrichte, Brack- und Salzmarschen (Makrophyten/Angiospermen) im Rahmen eines Praxistests zur Umsetzung der EG-WRRL in den Übergangsgewässern von Weser und Ems. Berichte des NLWKN 2009

Auby I., Oger-Jeanneret H., Sauriau P.-G., Hily C. & Barille L. (2010). Angiospermes des côtes françaises Manche-Atlantique. Propositions pour un indicateur DCE et premières estimations de la qualité. http://archimer.ifremer.fr/doc/00032/14358/

Bermejo R., Vergara J. J. & Hernández I. (2012). Application and reassessment of the reduced species list index for macroalgae to assess the ecological status under the Water Framework Directive in the Atlantic coast of Southern Spain. Ecological Indicators, 12: 46-57

Birk S., Strackbein J. & D. Hering, D. (2010). WISER methods database. Version: March 2011. Available at http://www.wiser.eu/results/method-database

Borja A., Franco F., Valencia, V., Bald, J., Muxika, I., Belzunce, M.J., et al., (2004). Implementation of the European Water Framework Directive from the Basque country (northern Spain): a methodological approach. Marine Pollution Bulletin 48 (3-4), 209-218

Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., Rodriguez, J.G., Rygg, B. (2007). An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. Marine Pollution Bulletin 55, 42–52.

Borja, A., Marín, S.L., Muxika, I., Pino, L. & Rodríguez, J.G. (2015). Is there a possibility of ranking benthic quality assessment indices to select the most responsive to different human pressures? Marine Pollution Bulletin 97 (2015) 85–94 http://dx.doi.org/10.1016/j.marpolbul.2015.06.030Carstensen, J., Krause-Jensen D., Dahl, K. & Henriksen, P. (2008). Macroalgae and phytoplankton asindicators of ecological status of Danish coastal waters. National Environmental Research Institute, University of Aarhus. 90 pp. - NERI Technical Report No. 683. http://www.dmu.dk/Pub/FR683.pdf

Dauvin, J.C. & Ruellet, T. (2007). Polychaete/amphipod ratio revisited: Implementation of the Water Framework Directive in European marine waters. Marine Pollution Bulletin 55 (1-6): 215-224.

Derrien-Courtel, S. & le Gal, A. (2009). Mise au point du protocole de suivi des macroalgues subtidales pourla façade manche-atlantique. Contrat IFREMER-MNHN, July 2009, 30p.

Devaney, F.M. & Perrin, P.M. (2015). Saltmarsh Angiosperm Assessment Tool for Ireland (SMAATIE). End of Project Report (2013-WS-D-10). Prepared for the Environmental Protection Agency, Ireland.

Dijkema, K.S., de Jong, D.J., Vreeken-Buijs, M.J. & van Duin, W.E. (2005). Saltmarshes in the Water Framework Directive. Development of Potential Reference Conditions and of Potential Good Ecological Statuses. Report to Rijkwaterstaat, NL.

Dolch, T., Buschbaum, C. & Reise, K. (2008). Seegras-Monitoring im Schleswig-Holsteinischen Wattenmeer. AWI, Sylt.

Duarte, C.M., Marbà, N., Krause-Jensen, D. & Sánchez-Camacho, M. (2007). Publications using data on eelgrass depth limits from Danish coastal waters: Testing the Predictive Power of Seagrass Depth Limit Models. Estuaries and Coasts 30 (4): 652-656.

European Commission (2018). Commission Decision (EU) 2018/229 of 12 February 2018 establishing, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, the values of the Member State monitoring system classifications as a result of the intercalibration exercise and repealing Commission Decision 2013/480/EU (notified under document C(2018) 696)

Foden, J. & D.P. Brazier (2007). Seagrass within the EU Water Framework Directive; a UK perspective. MarinePollution Bulletin 55: 181?195.

Foden, J. & de Jong, D.J. (2007). Assessment metrics for littoral seagrass under the European Water Framework Directive; outcomes of UK intercalibration with the Netherlands. Hydrobiologia 583 (1): 187- 197.

García, P., Zapico, A., Colubi, A. (2009). An angiosperm quality index (AQI) for Cantabrian estuaries. Ecological indicators, 856-865.

Gibbs, P.E., Bryan, G.W., Pascoe, P.L. & Burt, G.R. (1987). The use of the dogwhelk, Nucella lapillus, as an indicator of tributyltin (TBT) contamination. Journal of the Marine Biological Association of the United Kingdom, 67, pp507-523.

Guinda, X., Juanes, J.A., Puente, A. & Revilla, J.A. (2008). Comparison of two methods for quality assessment ofmacroalgae assemblages, under different pollution types. Ecological Indicators 8 (5): 743-753.

Høgslund, S., Krause-Jensen, D. & Carstensen, J. (2022). Macroalgae indicators for assessing ecological status in the Baltic and North-East Atlantic. Aarhus University, DCE - Danish Centre for Environment and Energy, 26 pp. – Scientific briefing no. 2022|31

https://dce.au.dk/fileadmin/dce.au.dk/Udgivelser/Notater_2022/N2022_31.pdf

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

Juanes, J.A., Guinda, X., Puente A. & Revilla J.A. (2008). Macroalgae, a suitable indicator of the ecological status of coastal rocky communities in the NE Atlantic. Ecological Indicators 8 (4): 351-359.

Kautsky L., Wibjörn, C. & Kautsky, H. (2006). Förslag till och vidareutveckling av bedömningsgrunder för kust och hav enligt krav i ramdirektivet vatten – Makroalger och några gömfröiga vattenväxter [In Swedish].

Kolbe, K. (2007). Intercalibration Report (NEA GIG). Assessment of German Coastal Waters (NEA1/26, NEA3/4) and Transitional Waters (NEA11) by Macroalgae and Angiosperms. NLWKN Wilhelmshaven.

Krause-Jensen, D., Greve, T.M. & Nielsen, K. (2005). Eelgrass as a Bioindicator under the European Water Framework Directive. Water Resources Management 19: 63-75.

Kuhlenkamp R. & Bartsch, I. (2007). Benthosbewertung Helgoland, Teil I Phytobenthos. Bericht für das LLUR-SH

Kuhlenkamp, R., Schubert P. & Bartsch, I. (2011). Water Framework Directive monitoring component macrophytobenthos N5 Helgoland EQR Evaluation 2010. MMH-Report 17.

Le Gal, A. & Derrien-Courtel, S. (2015). Quality index of subtidal macroalgae (QISubMac): A suitable tool for ecological quality status assessment under the scope of the European Water Framework Directive. Marine Pollution Bulletin, 101(1), 334-348. doi:10.1016/j.marpolbul.2015.10.053

Leonardsson, K., Blomqvist, M. & Rosenberg, R. (2009). Theoretical and practical aspects on benthic quality assessment according to the EU-Water Framework Directive - examples from Swedish waters. Marine Pollution Bulletin 58: 1286-1296.

Marques, J.C., Salas, F., Patricio, J., Teixeira H. and Neto, J.M. (2009). Ecological Indicators for Coastal and Estuarine Environmental Assessment. A user guide. WIT Press, UK. ISBN: 978-1-84564-209-9. 208 pp.

Muxika, I., Borja, A. & Bald, J. (2007). Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive. Marine Pollution Bulletin 55: 16-29.

Neto, J.M., Gaspar, R., Pereira, L. & Marques, J.C. (2011). Marine Macroalgae Assessment Tool (MarMAT) for intertidal rocky shores. Quality assessment under the scope of the European Water Framework Directive. Ecological Indicators, 19:39-47.

Neto J.M, Barroso D.V. & Barría P. (2013). Seagrass Quality Index (SQI), a Water Framework Directive compliant tool for the assessment of transitional and coastal intertidal areas. Ecological Indicators 30, 130-137

Neto, J.M., Salas Herrero, F., Best, M., Buchet, R., Heiber, W., Juanes, J.A., Kolbe, K., Recio, M., Ruiter, H., Scanlan, C. & Wilkes, R. (2018). Coastal and Transitional waters North East Atlantic Geographic Intercalibration Group. Seagrasses ecological assessment methods, EUR 29591 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-79-98477-8, doi:10.2760/86847, JRC115082

Neto, J.M., Salas Herrero, F., Best, M., Buchet, R., Heiber, W., Juane,s J.A., Kolbe, K., Recio, M., Ruiter, H., Scanlan, C., Duarte, B. & Wilkes, R. (2019). Coastal and Transitional waters North East Atlantic Geographic Intercalibration Group. Saltmarshes ecological assessment methods, EUR 29658 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-00026-6, doi:10.2760/450257, JRC115608

Phillips, G.R., Anwar, A., Brooks, L., Martina, L.J., Miles, A.C. & Prior, A. (2014). Infaunal Quality Index: Water Framework Directive Classification Scheme for Marine Benthic Invertebrates Environment Agency (UK) R&DTechnical Report, No SC080016."

Rygg, B. (1985). Distribution of species along pollution-induced diversity gradients in benthic ommunities in Norwegian fjords. Marine Pollution Bulletin 16:469-474

Rygg, B. (2002). Indicator species index for assessing ecological quality in marine waters of Norway. In: NIVA report 4548.

Rygg, B. (2014). DI (density index). Density of marine macroinvertebrates as indicator of environmental status. NIVA-rapport 6654-2014

Salas Herrero F., Aráujo R., Claussen U., Leujak W., Boughaba J., Dellsaea J., Somma F. & Poikane, S. (2020). Physico-chemical supporting elements in coastal waters: Links between Water and Marine Strategy Framework Directives and Regional Sea Conventions. EUR 30383 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-22418-1, doi:10.2760/444388, JRC121759.

Scanlan, C., Foden, J., Wells, E. & Best, M.A. (2007). The monitoring of opportunistic macroalgal blooms forthe Water Framework Directive. Marine Pollution Bulletin 55: 162-171.

Shannon, C.E. (1949). A Mathematical Theory of Communication. In: Bell System Technical Journal. Short Hills N.J. 27.1949, p. 379–423, 623–656.

Solheim, A.L., Peterlin, M., Kodeš, V., Spiteri, C., Collins, R.P., et al. (2012). Ecological and chemical status and pressures in European Waters – Thematic assessment for EEA Water 2012 Report. EEA/NSV/10/002. 146p.

Teixeira, H., Neto, J.M., Patrício, J., Veríssimo, H., Pinto, R., Salas, F. & Marques, J.C. (2009). Quality assessment of benthic macroinvertebrates under the scope of WFD using BAT, the Benthic Assessment Tool. Marine Pollution Bulletin, 58: 1477-1486. (doi:10.1016/j.marpolbul.2009.06.006).

TG Seabed (2021). Review of relevant methods for assessing habitat status under other policies, Marine Strategy Framework Directive (MSFD), Common Implementation Strategy, 29th Meeting of the Marine Strategy Coordination Group (MSCG)

UKTAG (2014). Transitional and Coastal Water Assessment Method, Angiosperms, Intertidal Seagrass Tool; ISBN: 978-1-906934-36-1

Van Hoey, G., Drent, J., Ysebaert, T. & Herman, P. (2007). The Benthic Ecosystem Quality Index (BEQI), intercalibration and assessment of Dutch coastal and transitional waters for the Water Framework Directive: Final report. NIOO Rapporten, 2007-02. NIOO. 244 pp.

Van Hoey, G., Bonne, W. & Salas Herrero, F. (2015). Intercalibration report for benthic invertebrate fauna of the North East Atlantic Geographical intercalibration group for Coastal Waters. (NEA 1/26), ILVO MEDEDELING 191, ISSN 1784-3197

Van Hoey, G., Bonne, W., Muxica, I., Josefson, A., Borgersen, G., Rygg, B., G., Borja, A., Phillips, G., Miles, A.l., Dubois, S., Desroy, N., Buchet, R., Ximenes, MC., O'Beirn, F., Witt, J., Heyer, K., van Loon, W., Ruiter, H., Neto, J., Marques, J.C., Garcia, E., Puente, A. & Salas Herrero., F. (2019). Coastal waters North East Atlantic Geographic Intercalibration Group. Benthic invertebrate fauna ecological assessment methods; EUR 29640 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-79-99232-2, doi:10.2760/16318, JRC115475

Van Loon, W.M.G.M., Boon A.R., Gittenberger, A., Walvoort, D.J.J., Lavaleye, M., Duineveld, G.C.A. & Verschoor A.J. (2015). Application of the Benthic Ecosystem Quality Index 2 to benthos in Dutch transitional and coastal waters. Journal of Sea Research 103, 1-13

Wells, E. (2006). Water Framework Directive Marine Plants Task Team Tools paper, Intertidal Coastal Waters macroalgae. Rocky Shore Tool. Draft verion 3. Paper No. MPTT/MAT01.

Wells E., Wood, P., Wilkinson, M. & Scanlan, C. (2007a). The use of macroalgal species richness and composition on intertidal rocky seashores in the assessment of ecologcial quality under the European Water Framework Directive. Marine Pollution Bulletin. 55, 151 – 161.

Wells, E., Best, M., Scanlan, C., Holt, S. & Foden, J. (2007b). WFD Tools Paper, Draft v2. OpportunisticMacroalgae, Abundance. MPTT/MAT01.

WFD (2003) COMMON IMPLEMENTATION STRATEGY FOR THE WATER FRAMEWORK DIRECTIVE (2000/60/EC) Guidance Document No 5 Transitional and Coastal Waters – Typology, Reference Conditions and Classification Systems Produced by Working Group 2.4 – COAST https://circabc.europa.eu/sd/a/85912f96-4dca-432e-84d6-a4dded785da5/Guidance%20No%205%20-%20characterisation%20of%20coastal%20waters%20-%20COAST%20(WG%202.4).pdf

Wilkes, R., Best, M., Scanlan, C., Rossi, N., Kolbe, K. & Salas Herrero, F. (2018). Coastal and Transitional waters North East Atlantic Geographic Intercalibration Group. Opportunistic macroalgae ecological assessment methods, EUR 29594, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-98481-5, doi:10.2760/167718, JRC115074

Wilkinson, M., Wood, P., Wells, E. & Scanlan, C. (2007). Using attached macroalgae to assess ecological statusof British estuaries for the European Water Framework Directive. Marine Pollution Bulletin 55: 136 -150.

ANNEX 3

OSPAR – Common Biodiversity Indicators

Condition of benthic habitat communities (BH2-B): Subtidal Habitats of the Southern North Sea

Guidelines for Coordinated Environmental Monitoring Programme (CEMP)

Last submission to OSPAR Secretariat and BDC (March 2022), OSPAR Benthic Habitat Expert Group (November 2022), COBAM (November 2022). All comments received to date are addressed in this updated version.

Adoption (BDC 2017) = OSPAR Agreement 2018-06

Adoption of this update: BDC (December 2022)

Main author: Willem Van Loon M.G.M. (NL) 2018 *revised (2022) by Wijnhoven, S. (NL), Walvoort, D.W.W. (NL) & Schilder, J. (NL)*

The CEMP guidelines are a tool for clearly documenting the detailed methodology for the monitoring and/or assessment of the individual parameters/indicators so as to ensure transparency, coherence and consistency in the implementation of such monitoring and assessment. *These guidelines should outline themonitoring and assessment requirements for the components outlined within the CEMP and should be developed for each element (i.e. common indicators)*

1 Introduction

In the OSPAR Biodiversity Committee meeting of November 2016, for the Common indicator BH2-B, an assessment sheet containing the Margalef diversity method for benthic communities to assess benthic habitat quality status, and its assessment results for the Southern North Sea have been accepted (OSPAR 2016). This assessment sheet (Annex 3) builds upon the BH2 common approach CEMP appendix A (OSPAR 2016), initially applied to the specific context of the Southern North Sea subtidal sediments. With the current (2022) revision the application is extended to the entire OSPAR region II (and application in other regions according to the same methodology is an option). The initial methodology as described in the CEMP accepted in 2018 and the results of application for the Intermediate Assessment (IA2017; see Assessment sheet BDC17/D116) have been published (Van Loon et al., 2018) and did make use of the BENMMI software, specifically developed for OSPAR Common indicator BH2. The software amongst others, facilitates the automated analysis and reporting of the quality status of benthic habitats on basis of Margalef diversity (Walvoort and Van Loon 2017). The Margalef diversity indicator is further developed and now indicated as 'Relative Margalef diversity', which builds upon the traditional Margalef diversity, however, includes standardization by taking the case specific reference diversity into account. Also, procedures, calculations and standard output (tables and maps) have been included in a standardized script (Relative Margalef diversity –

script; Walvoort et al, 2022) that can be run in case input data are according to a fixed format. With the current revision, the methodology has been adapted to the application at the level of Broad Habitat Types (BHTs) within Assessment Units (AUs), with the definition of case-specific references for the Margalef diversity and has been extended to the other AUs (subregions) of Region II (the Greater North Sea region). The methodology makes use of case specific reference values for Margalef diversity (Margalef assessed is divided by the reference) to make results (presented as Relative Margalef diversity on a scale from 0-1 after truncation) comparable between cases. Regardless various standardization procedures, results appeared to some extent be related to data acquisition procedures. Therefore, besides that specific references are derived at the level of sampling techniques, also differences between countries are taken into account. Application at BHT level improves general applicability and repeatability, facilitates thematic assessment as other indicators (BH1, BH3, BH4) make use of BHTs as well and connects to the systematics in use for EU MSFD assessments (Article 8 MSFD Assessment Guidance) improving or allowing exchangeability of results.

2 Assessment methodology

As an indicator of the benthic habitat quality status the community diversity is assessed using the Margalef diversity index. 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 is 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. It has to be noticed that 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 data sets. The value for D_{ref} is estimated based on low pressure observations. Initially D_M is calculated at the level of samples. Results are combined with pressure mapping after which the D_{ref} is achieved as a percentile value from a case specific selection of low-pressure data, therefore taking 'natural' (and potentially other sources of) variability into account (see paragraph 3.2). For Margalef diversity, the bad ecological state value equals 0. Hence the D_M' calculation can be simplified to:

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

Before, the step of taking a reference into account was referred to as normalization (Van Loon et al., 2018; Annex 3 of 18-06e_cemp_guideline_bh2). Here the term ‘relative Margalef diversity’ is preferred, as strictly spoken, normalization implicates dividing by the maximum value to achieve a value on the scale from 0-1. Here, theoretically the D_M’ can transgress a value of ‘1’ although truncation at this value is suggested for the presentation of assessment results, where all values for D_M’ above 1 are considered relative high diversity; habitat in good quality with no further difference in the relative quality.

3 Monitoring

3.1 Purpose

The purpose of the benthos monitoring in the North Sea region is:

- To establish the current state (condition) of the benthic communities in selected assessment units., as an assessment of benthic habitat quality status. (For EU countries the specific request with regards to the MSFD (Article 8 MSFD Assessment Guidance) is to do this (at least) at the level of (MSFD) broad habitat types, which is taken into account here as well so that assessment results can be used for both purposes (OSPAR QSR and EU MSFD reporting).
- To detect possible trends in the benthic habitat condition, and to evaluate effectivity of specific management measures to improve the benthic habitat quality status.
- To use the available benthos data to estimate suitable reference values for specific cases to achieve comparability of assessment results.

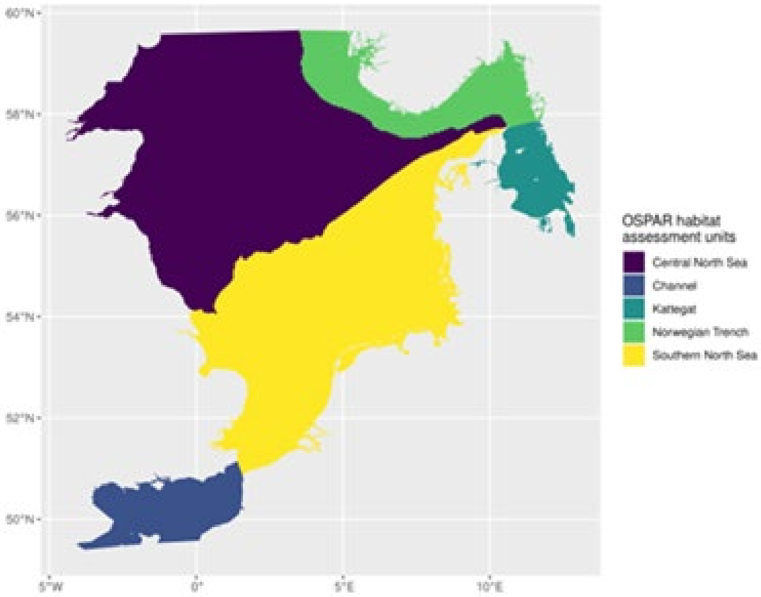

Figure 1. Division of OSPAR region II (Greater North Sea region) in OSPAR assessment units.

3.2 Quantitative Objectives

Although the index of Margalef diversity is specifically developed to minimize the impact of possible size variability (sampled surface area or volume) of samples on assessment results, sampling methodologies should be comparable to a certain extent. Therefore, different categories of sampling methodology are defined. At the moment three sampling categories are suggested for application of BH2b in region II (of which in the end actually only the assessment results of two of these categories are used for the QSR2023, as application showed that assessment results for the category 'Dredge-Trawl' were strongly related to the specific methodology used, more than the expected (fishing) pressure level. Distinguished categories are:

- Grab-Core: Sampled surface area around 0,1 m^2 and mesh size of 0,1 cm
- Dredge-Trawl: Sampled surface area >3 m^2 and mesh sizes in range 0,5-2,2 cm
- Small core: Sampled surface area around 0,0143 m^2 and mesh size of 0,1 cm

Sampled surface area for the 'Grab-Core' category can be 50% smaller or larger than 0,1 m^2 and there is large variation in the sampled surfaces for the 'Dredge-Trawl' category from 3 up to 1000s of square meters. At least there is a certain standardization that reduces possible variability in assessment results due differences in methodology. Working with stricter standards for data that can be included reduces data availability and potentials to compare benthic quality status with BH2b for different habitats, sub-regions and countries. The selection of these categories already means that not all potentially available data provided by OSPAR CPs are used. This does not mean that these data cannot be used to calculate relative Margalef diversity at all; it is just that now sampling methodologies are restricted to certain cases (often applied in only one of the countries) and deviating to much from other monitoring to be included. Running There are potentials (e.g. in case of application of certain techniques in various parts of region II) to apply BH2b on such techniques in the future as well (e.g. video transect monitoring seems to become more common practice and is potentially suitable for BH2b application).

- In principle, relative Margalef diversity (D_M') can be calculated at the level of individual samples once a case-specific reference value for Margalef diversity (D_M') is available. However, to obtain a reliable assessment of the benthic habitat quality status based on D_M', at least 10 benthos samples per case (Assessment Unit x Broad Habitat Type x country = AU x BHT x country) for the sampling methodology of concern and the assessment period to be evaluated, need to be available. Assessment of D_M' always comes with relative confidence estimation. Confidence is determined by the total number of samples, number of years covered by samples (see Table 1) and additionally spatial relative representativity of sampling which should also consider coverage of different abiotic conditions/gradients, pressure levels and management in case the monitoring design cannot be considered random.
- For reference value estimation at least 20 samples or more must be available for a certain case (AU x BHT x country). This corresponds to a quality code 3 as defined in Table 1. With more data available (at least 30 or at least 50 samples) a higher quality code of 2 or 1 is reached in case respectively at least 2 or 3 years are covered by the data. Quality codes combined with (low-)pressure levels (in this case fishing pressure levels distinguishing year average Swept Area Ratios (SAR) of SAR 0-0,1, SAR 0,1-0,5 and SAR 0,5-1) determine the percentile value to be selected as the reference (Table 1). In case of 2 or 3 reference value estimations for a certain case (i.e. based on 2 or 3 pressure classes of low fishing pressure as indicated) the highest value is accepted.
- In region II fishing pressure (and physical disturbance of the seafloor in particular) is expected to be by far the most important pressure determining the quality of the benthic habitats (ICES, 2021a,b). Nevertheless, there will be impact from other pressures on the benthic habitat quality status as well. Ideally, pressure layers of all possible pressures are generated and combined, and references are only extracted from low pressure areas with regards to all pressures. In absence of such detailed information on all types of pressures, the focus of current assessment is on the most important pressure. BH2b assessments therefore use percentile values as it is known that although areas selected to extract reference values for Margalef diversity have low fishing pressure but might include differences in fishing pressure at detailed scale. Chance of sampling exactly in moderate to high physically disturbed areas due to fisheries is low in low fishing pressure areas. However, areas might and generally will include disturbed communities due to other pressures as well. Therefore, a relative

high percentile value is selected for the D_{ref} to reflect good quality status. Moreover, if for certain cases (AU x BHT x country) impacts of other types of pressures are common and/or on average substantial, this will be reflected in a relatively low D_{ref} for that case. Other pressure impacts are than part of the observed 'natural' variability or should be considered typical background disturbances for a certain case. As the indicator basically responds in a similar way to all kind of pressures with impact on benthic communities, D_M' results are expected to be representative for an overall quality status (independent of the type of disturbance), if reference areas are not severally disturbed by other pressures. Therefore, it is advised not to consider samples from areas with known substantial other disturbances. It would be wise to exclude such areas (like aggregate extraction sites, areas with point sources of pollution or eutrophication or known hypoxia problems, or areas surrounding recent construction works, when widely distributed. Percentile values as suggested related to disturbance classes (75% percentile value for average SAR 0-0,1, 95% percentile value for average SAR 0,1-0,5, 99% percentile value for average SAR 0,5-1) as D_{ref} has been proven realistic for a relatively good quality status (not clear at the moment how it compares to a pristine reference), as obtained values for different pressure classes are generally well in line (comparable). Nevertheless, the highest obtained value for D_{ref} per case (AU x BHT x country per monitoring technique) is adopted. It is realized that for certain cases (e.g. BHTs typically common in shallow coastal areas of for instance the Southern North Sea and Channel) the assessed quality status might neglect a generally prevalent background of increased nutrients and certain pollutant levels. As a result, the good quality status based on solely the D_M' might be a slight underestimation in these cases and/or it could be that part of observed quality developments are not related to developments in fishing pressures. Therefore, it is important to consider other thematic assessments as well.

- Quality codes for data sets (cases) also indicate the relative level of confidence for the assessment results, with higher confidence in case of more samples and/or covering more years (typically per period as for instance the years 2016-2021 identified as the years for which the quality status for the QSR2023 is estimated). In case of lack of data, older data can be used to do assessment of the current quality status on best available data, however with lower confidence as indicated in Table 1. (Comparatively, the quality status of the 2009-2015 period can be estimated using older 1998-2008 data in case of lack of data for 2009-2015, however with a (one category) lower confidence as indicated in Table).

Table 1. Systematics of division of subsets of data per case (Assessment Unit x Broad Habitat Type x country, specific for a sampling method) into quality codes to select suitable subsets for reference value derivation and to indicate relative confidence of assessment results (based on these data).

Quality code	5	4	3	2	1
Samples (n)	<10	≥10	≥20	≥30	≥50
Covering n years				≥2	≥3
Percentile value to derive D_{ref}:					
SAR 0-0,1			99%	95%	75%
SAR 0,1-0,5				99%	95%
SAR 0,5-1					99%
Final D_{ref} is the highest value as obtained from up to three estimations based on distinguished low SAR classes (here SAR 0-0,1, SAR 0,1-0,5 and SAR 0,5-1).					
Confidence of assessment of current status:					
Based on 2016-2021 data	Poor	Low	Sufficient	Good	High
Based on 2009-2015 data	Poor	Poor	Low	Sufficient	Good
Based on 1998-2008 data	Poor	Poor	Poor	Low	Sufficient

3.3 Monitoring Strategy

- For the assessment of benthic community condition (indicator of benthic habitat quality) it is essential that species or alternative taxonomic levels are identified to the highest taxonomic level for all samples. At least this should be done in a standardized way by all participating countries, and/or for all datasets included. This includes the way how occurrences of agreed taxa are recorded (i.e. abundances in densities, presence/absence recordings; as an alternative also biomass recordings can potentially be used). As there is no joined or agreed standard (OSPAR) protocol among the participating countries yet, the highest level of taxonomic detail that can be reached in a standardized way is determined by the dataset(s) with the lowest level of taxonomic detail. Options are either accepting a lower level of taxonomic detail or not taking the dataset (with limited detail) into account. Rules for taxonomic level of detail and the type of occurrence recording can differ between types of sampling methodologies for which assessments are done separately. Although it might be expected that assessment results are to a certain level comparable for different monitoring techniques, the focus of techniques is on different types of species (and can differ with the taxa included); therefore, assessment results are expected to differ in detail. Assessments based on different techniques are typically complementary focusing on different aspects of community diversity and/or the possible impact of different pressures on benthic communities.
- To come to optimal taxonomic standards to be accepted for the assessments; sampling and laboratory procedures with regards to the recording of different taxa have been

compared and discussed among participating countries of the Greater North Sea region for the QSR2023. The agreed standards for species/taxa recordings that are in line with provided data (for the QSR2023) are presented in Appendix 3a. Either Those standards are proposed as the lowest acceptable data standards for future data provision and assessments as well and can be a step towards a joint monitoring protocol as well. One can decide to opt for other standard levels of identification in other regions, for other monitoring techniques or with regards to assessments at sub-regional scale. But it has to be considered that other choices might lead to (slight) deviations in assessment results and have consequences with regards to comparability as the focus might be on other parts of the benthic diversity.

The internationally standardized WoRMS taxa name list (www.marinespecies.org) has to be used. With the list standardized names and synonym names can be distinguished and all taxa can be linked to the agreed standardized names. All species and taxa recordings are linked to accepted taxa names, and it is indicted whether:

- Densities (D) or

- Presence (P) indicated as a value of ‘1’, compared to ‘absence’ indicated with a ‘0’ (for the QSR2023, no biomass recordings are used; in case no densities are recorded, each biomass recording >0 results in P=1), are recorded or if an

- Alternative (A) (more detailed) level of taxonomy should be used.

- ‘0’ means not included in the assessment at all.

Potentially, relative Margalef diversity can be calculated based on biomass recordings as well (taking into account that biomass specific reference values for Margalef diversity should be estimated in that case as well), although biomass-based diversity calculations are not common in scientific literature.

In case a sub-selection of specimens is identified at a less detailed taxonomic level (this is often the case for juvenile or incomplete specimens that could not be identified to a more detailed level) whereas another more detailed taxonomic level (e.g. species) should be recorded, these specimens are divided according to the observed occurrence ratios for the more detailed taxonomic levels for the dataset of concern.

3.4 Sampling Strategy

In principle the relative Margalef diversity can be calculated based on each sampling method taking method specific references into account. As indicated, specific references will be defined at the level of AUs x BHTs x countries and for different classes of sampling methodologies (i.e. ‘Grab-Core’, ‘Small core' and ‘Dredge-Trawl’, with further subdivision into 3 classes with different mesh sizes for the last category). For assessments at the level of individual samples, a larger sampled surface area will provide a better representativity. By using sampling methodology specific references and optimal taxonomic standardization within identified categories it is expected that results for D_M’ based on different techniques, are rather comparable. Deviations in details can however be expected as each sampling technique is focused on a specific part of the benthic community diversity. Specific monitoring techniques for instance particularly focus on infauna or epifauna (typically smaller grabs and cores) or especially the mobile species (techniques sampling larger surface areas and using nets in the water column as well) and certain techniques are more representative for common widely distributed species (like grabs and cores) than for larger species naturally occurring in low densities (trawls and dredges could be a better option).

- Basically, any method of sample treatment can be used if specific reference values are defined. It is proposed to submit benthos monitoring data for BH2b assessments in line with one of the 3 presented categories of sampling methodologies. Therefore, samples should be sieved over 1 mm mesh for grabs and cores (including small cores), and mesh sizes or netting of 0,5, 1 or 2,2 cm are allowed for benthic dredge and trawl samples. Optionally other sampling methods can be suggested for future assessments (e.g. video transect recordings), but it is highly preferred that joint OSPAR standards are developed for such techniques and that standards for grabs, cores, dredges and trawls are further tuned among participants. Joint monitoring, at least as a baseline to compare results according to national procedures with, would be ideal (Van Hoey et al., 2022). Aspects with regards to the used procedure during sampling are used mesh size (as indicated), fixation of retained organisms using formaldehyde (a final concentration of 4-6%) or ethanol (a final concentration of 70%), and identification of samples within 6 months to not loose specimens and/or crucial characteristics of species to be identified.
- Assessments take place at the level of AU x BHT x country. Preferably a representative monitoring is in place for each combination (at least when substantial surface areas are covered). Representative monitoring can be achieved by a stratified random sampling design over the BHTs within Assessment Units and countries. Ideally fixed sampling locations are used for quality assessments, as such a design best supports trend analysis (by reducing internal natural variation). Alternatively, random sampling is expected to provide sufficient representative data to do trend analysis, although the power of the design will be lower. In case sampling sites are largely determined by project monitoring, there is an increased risk that these are not entirely representative for broad-scale areas as assessed for OSPAR. Project monitoring is often determined by specific interventions, activities, and management of the system. In that case it should be analyzed for what part of the AU x BHT x country such data are representative and if a sub selection of samples (possibly with recurrent sampling) can be indicative for developments of parts of the area.
- In case of the availability of replicates (as part of a national monitoring programme), information from all replicates is included by taking the median of calculated D_M values per replicate. (Summing would actually increase the sampled surface area at the site and therewith artificially increase D_M due to expected increase of species richness just by chance).

3.5 Quality control and exploratory data analyses

The following quality control procedures must be used. These quality control procedures are partly facilitated by using the ‘Relative Margalef diversity scripts’, which are an update and elaboration of one of the indicators formerly part of the BENMMI software. Scripts and examples (from the QSR2023) are provided in 'step-1-data-prep.html' (OSPAR BH2b indicator - data preparation and cleaning) and 'step-2-eda.html' (OSPAR BH2b indicator: exploratory data analysis (EDA)).

- Benthos data

Confirm that the dataset used is the validated/correct/official one. For OSPAR assessments there will be official data calls according to prescribed data standards. Using the format for data submission is a first step to gather essential information and reduces the risk on large amounts of incomplete data (data not sufficient for BH2b application).

- OSPAR_BH1-BH2b-reporting_format_210330
- OSPAR_BH1-BH2b-BH2a-reporting_format_guidance_210330

Essential data are 'Region', 'Country', 'meta_id', 'sample_id', 'data', 'year', 'mesh_size', 'sampled_area', 'gear_type', 'assess_cat_gear', 'assess_cat_period', 'species_name_assessment', 'density', 'bbht' (Broad Habitat Type), 'subregion', 'sar' (year average swept area ratio), 'lon' (longitude), 'lat'(latitude) to be completed in correct format.

- Gear type

Check whether data fit in one of the proposed methodological categories or be sure that an additional methodology can be used with comparable data for all areas to be included in the assessment. (It can be a choice to select additional methods to be analyzed for certain areas although ideally results for prescribed methodological categories are also available for parts of the areas for comparison. Be aware that in that case own standardization rules of taxonomy must be defined, and no Margalef reference values are available in advance; reference estimations must be part of the calculations in that case).

- Time series

Sampling years and distribution over assessment periods ('1998-2008', '2009-2015', '2016-2021') is visualized per country. Older data can potentially be used to get an indication of the current quality status, however with consequences towards confidence (as clarified later).

Now there is no standardization of sampling season, which means that natural variability due to seasonality is introduced. The season effect might be particularly strong as juveniles are not excluded from the data. Juveniles are included as they were treated differently in the recordings of different data sets, so that they cannot easily be filtered out. At least it is good to be aware of possible effects of seasonality on the assessment results, for which data distributions are visualized for different countries.

- Check the locations of the samples within the Assessment areas using a plot. With the 'Relative Margalef diversity' software plots are produced where those sample sites located outside the region of concern (Region II in case of the QSR2023) should be excluded. Check whether geographically outlying samples are not the results of typing errors or mistakenly swapped coordinates.
- Make connection of samples with Assessment Units and Broad Habitat Types (potentially other habitat clarifications can be used as well). In case no BHTs are distinguished for certain

samples, these are excluded from the analyses (be aware that at the moment the EU MSFD Broad Scale Habitats map (update September 2021; EMODnet, 2021) does not distinguish intertidal habitats, so quality status of these is not assessed). As with regards to the QSR2023, no monitoring techniques have been deployed that can sample hard substrate communities. It is therefore clear that those sample sites indicated as 'rock and biogenic reef' habitats are likely misestimations of the BHT (as largescale biogenic reefs are not expected either). The samples indicated as being 'rock and biogenic reef' have therefore been added to the 'mixed sediment' categories, only distinguishing the depth/photonic classes of infralittoral -, circalittoral - and offshore circalittoral mixed sediment classes.

- Check for clear outliers in the average total abundance per sample. It is however known that aggregations of certain species in banks or reefs occur, or that densities can be huge in case of presence of juveniles. Values due to typing errors should be corrected, however large numbers that might be realistic are maintained. The impact of large numbers on assessment results is to a certain level minimized by using the logarithmic value for the densities considered in calculations.
  Doublecheck whether indeed densities are used in the calculations (so that recorded values reported as numbers per sample, and possibly confusingly called 'densities' are indeed divided by sampled surface area before calculation of indexes.
- For the estimation of reference values, samples are related to average fishing pressure distribution mapping at the level of c-squares. Average fishing pressure in 'Swept Area Ratio (SAR) per year for the years 2013-2018 (ICES EUTRADE data product; ICES, 2021b) is used for the QSR2023. The data product distinguishes average SAR classes of which for estimation of Margalef reference values (D_{ref}), the low fishing pressure classes with year average SAR from '0-0.1', '0.1-0.5' and '0.5-1' are used. In case of no information available (indicated as 'NA') related samples are not used to estimate the reference as there might be fisheries after all of which we are not aware. (It has to be noticed that at the basis of average SAR distribution mapping are VMS (Vessel Monitoring System) data, in general only in use by fishing vessels larger than 12 meters.

3.6 Data quality and confidence of assessments

- In principle also with regards to the BH2b assessments the confidence guidance for the QSR2023 from OSPAR (OSPAR agreement 2019-02, updated version 2021 of Guidance document QSR (2023)) is leading. Amongst others the guidance makes use of a relative classification in confidence classes with regards to data availability and spatial and temporal resolution. Basically, the expected representativity of data is estimated there. As indicated before, also the specific methodology for BH2b makes use of a relative estimation of the quality of the data by indicating a quality code (Table 1). This specifically to estimate reliable reference values for D_M' on sufficient representative data. However, the same quality coding can be used to indicate the expected confidence of assessments based on subsets of data (i.e. for specific cases; combinations of AU x BHT x country) as well. Quality coding can easily be translated into the proposed confidence levels for the QSR2023 as provided by OSPAR as well.

4 Assessment

4.1 Data acquisition

Data are gathered according to an official OSPAR data call, in this case a combined data call for BH1 and BH2. In this way, the first steps with regards to standardization and exchangeability of data and results between indicators (e.g. of specific value with regards to thematic assessments and exchange of data with other themes) have already been made. Additionally, storage (within ODIMS), use and access (whether publicly available or restricted) is immediately arranged via the OSPAR secretary. It would be beneficiary to have recurrent data calls (e.g. every year or every 2nd year), to ensure that recent data are available (even in case a deadline is missed and in that way providing the data to OSPAR might become common practice).

4.2 Estimation of reference values in case of low data availability

- As indicated in Table 1, D_{ref} can be calculated in case sufficient data from low fishing pressure cases are available. Dependent of the fishing pressure class, data subsets with quality code 1-3 can be used. In case 2 or 3 estimations can be done, the highest achieved reference value is accepted. In case data availability in poorer (quality codes 2-5 depending on the level of fishing pressure) or in case (almost) no low fishing pressure data are available, a D_{ref} can be obtained using multivariate statistics. The procedure is integrated in the BH2b script and considers 4 variable characteristics of the data subset: Assessment Unit, Country, Depth/photonic class, and Sediment type. Each data subset (case) is characterized by the four so that the average of already obtained average D_{ref} values can be calculated as the best estimate. In case no references are available for one of the variables yet, the arithmetic average of the options can be taken (i.e. use average value of country A, B and C in case no D_{ref} for country D is available yet).
- Quality coding of the data subsets and amongst others spatial representativity (according to OSPAR Guidance document QSR2023) determines the level of confidence of the assessment results based on limited data availability.
- Using case specific references is a way of standardization allowing comparison of assessment results for different cases and/or development in D_M' in time.
- The methodology suggests the availability of to a certain extent representative monitoring data. To maintain a representative benthos monitoring programme to assess the quality status of benthic habitats including all aspects distinguished (e.g. BHTs) is of national concern, although international joint monitoring could improve monitoring efficiency. Representative monitoring does not necessarily mean that all parts of national waters are covered (with the same monitoring efforts). If designs are not random or stratified random in essence (either or not fixed after initial selection),

care has to be taken that monitoring is not biased by focus on management or activities. Such reduced representativity should be reflected in the confidence or results should solely be presented for subareas of proposed assessment units.

4.3 Assessment criteria

- The assessment scale is the national part of a BHT at the level of an Assessment Unit. Reporting in principle takes place at the level of an Assessment Unit, possibly subdivided into BHTs. For abstracting purposes, it is proposed to present the median value for D_M' for an assessment period (accompanied by confidence estimation). As abstracting might cover important details it is suggested to present the assessment results as a spatial representation of values (or quality categories) at the level of individual sample sites (as well).
- As it is expected that recurrent frequent sampling of fixed locations is limited, it is suggested to present assessment results per assessment period (e.g. period QSR2023 = 2016-2021) potentially comparing results with former periods (i.e. IA2017 – 2009-2015 and QSR2010 - 1998-2008). In case of lack of data, to present the best possible assessment of the current quality status possible, optionally results for former periods can be presented as the current quality status, however in that case with lower confidence (see Table 1).
- For certain areas assessment at the level of BHTs is not specific enough. Especially the estimation of a reference D_M might be hampered by (known) other disturbances besides fishing pressure and/or 'natural' gradients might be present in abiotic conditions with large impact on benthic communities. In case of other pressures, these can be considered in spatial pressure mapping, and typically known high pressure areas (e.g. eutrophic or hypoxic areas or sediment extraction sites) can be excluded from potential low pressure areas. In case of known gradients these can be considered by defining condition to D_{ref} value regressions and estimate D_{ref} at sample site level. It is suggested in that case to present and compare the two (using median D_{ref} and site specific D_{ref}) to enable discussion on possible impacts on comparability of assessment results at larger scale.

4.4 Spatial Analysis and / or trend analysis

- The BH2b relative Margalef diversity indicator primarily uses spatial median values per Broad Habitat Type per Assessment Unit per assessment period. A single sample assessment is not very informative for a local quality status. Visualization of spatial patterns in sample site specific quality status recordings (in a map) is however very informative with regards to spatial variability and specific patterns. Groups or clusters of samples (at least 30 samples covering at least 2 years) for a representative area can however be very informative with regards to the local quality status. A lower number of samples and years in line with the confidence estimation of Table 1 can be of use as well.
- For abstracting purposes, it is proposed to indicate case specific (UA x BHT, optionally per country) temporal developments based on median values of D_M' per assessment

period, taking representativity of data into account. Trend analysis and testing for significance of developments with the focus on long-term (1998-2021), mid-term (12-year periods) and short-term (6-year periods) using individual D_M' values per year is an option. In case of doubt of relatedness of developments to (physical disturbance by fishing) pressure levels, evaluation of patterns in developments (trend analysis) for selections of re-sampled stations (fixed monitoring design) is proposed. At the moment comparing relative Margalef diversity levels for just three periods, two-sided independent t-testing at $p<0,05$ is applied for all combinations.

4.5 Presentation of assessment results

- An abstracting list (table) of Assessment Units x Broad Habitat Types and median relative Margalef diversity scores for the most actual assessment period (preferably 2016-2021) is presented. In case of lack of most recent data, the score for a former period is adopted. Results with regards to overall quality conditions of benthic habitats based on D_M' are accentuated by presenting D_M' scores ≥0,8 in green, scores in range ≥0,6-0,8 in orange, and scores <0,6 in red. This qualification definitely is not related to threshold values (TVs) for Good Environmental Status (GES) yet. When TVs come available or are defined within OSPAR, coloring can be adjusted to proposed TVs. At the moment proposals for TVs are being developed in EU MSFD CIS TG-Seabed. As OSPAR indicators are considered and a large part of OSPAR region II will be assessed under EU MSFD as well (moreover, OSPAR results will be used in several cases for MSFD reporting as well), those proposals might be informative for OSPAR as well. However, future applicability in OSPAR should be further investigated when TV proposals are available. Results are linked to a similar table indicating relative confidence based on data density, temporal coverage (both integrated in quality code). Spatial representativity should be considered as well, it is for the time being restricted to a description in the ain text of the document. In the D_M' abstracted results table with median values per case also developments in the benthic habitats' quality status based on analyses of developments comparing current status with former periods (where possible) can be indicated using rrows in case of significant increases or decreases and/or question marks in case data availability is too limited for testing (low sample number, no former data available or current status based on former period (with reduced confidence level). Abstracting of developments is also based on trend analyses for fixed re-sampled stations if relevant (applicable) for the AU x BHT combination of concern. The last might differentiate at the level of countries as well. In principle results are presented for 'grab-core' samples or 'small cores' (if these are the common procedure for the AU x BHT combination), but can be the median of the two.
- Abstracting results are accompanied with a map showing results at the level of separate samples (sites), created for different periods and different monitoring techniques; for which results can be compared, and more detailed spatial patterns can be extracted.
- More detailed results will be presented at the level of separate countries and for the different categories of monitoring techniques, including trend analyses and where relevant trend analyses for groups/aggregations of fixed stations with re-current sampling. Possible

difference among monitoring techniques (‘grab-core’, ‘small core’ and ‘dredge-trawl’ optionally separated into samples with mesh size 0,5, 1 and 2,2 cm) and in more small-scale spatial patterns will be discussed. This might include analyses of the use of adaptive reference values for D_M along depth- or salinity gradients (D_{ref} on basis of regression) where relevant (e.g. Kattegat).

4.6 ‘Relative Margalef diversity’ assessment software

- Most of the data analysis and assessment steps described in this CEMP can be performed automatically and standardized using the ‘Relative Margalef diversity’ software. This software has been specifically developed for this indicator for application in Region II for the QSR2023 and is based on earlier BENMMI software developed for the IA2017. In principle same methodology and software can be applied on comparable benthos data for other regions and/or alternative purposes.
- ‘Relative Margalef diversity’ assessment software gives accurate error messages in case the input is not yet correct or as expected. These error-messages facilitate correction of the input data (see ‘Relative Margalef diversity’ assessment software user manual). Scripts and (examples of) results are presented in ‘step-3-assessment-nosar.html’ (OSPAR BH2b indicator: Assessment (implicit SAR)).

5 Change Management

Since changes to this method often require a more general level of discussion and policy decision making, itis proposed to discuss proposed changes in ICG-COBAM, and to make decisions on proposed changes in ICG-COBAM and BDC.

References

Article 8 MSFD Assessment Guidance. Reference to Draft Article 8 MSFD Assessment Guidance. Final version 1.4 – 2 May 2022 as submitted to the Marine Strategy Coordination Group (MSCG), document MSCG_30-2022-04 (https://circabc.europa.eu/ui/group/326ae5ac-0419-4167-83ca-e3c210534a69/library/68d25427-2f85-4e2b-89d2-765a9cf8048e/details).

EMODnet (2021). EUSeaMap 2021 Broad-scale predictive habitat map for Europe (updated September 2021; EUSeaMap 2021).

ICES (2021a). A series of two Workshops to develop a suite of management options to reduce the impacts of bottom fishing on seabed habitats and undertake analysis of the trade-offs between overall benefit to seabed habitats and loss of fisheries revenue/contribution margin for these options (WKTRADE3). ICES Scientific Reports. 3:61. 100 pp. http://doi.org/10.17895/ices.pub.8206.

ICES (2021b). 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.

OSPAR Intermediate Assessment 2017 - Common indicator. Condition of benthic communities: subtidalhabitats of the Southern North Sea.

OSPAR 2016 – Common Biodiversity Indicators. Condition of benthic habitat communities (BH2). Generic guidelines for Coordinated Environmental Monitoring Programme (CEMP), Annex A.

Van Hoey, G., de Backer, A., Wijnhoven, S. (2022). Roadmap for integrated benthic monitoring in Greater North Sea. ILVO report in cooperation with Ecoauthor, commissioned by Rijkswaterstaat within the frame of OSPAR.

Van Loon, W.M.G.M., Walvoort D.J.J., Van Hoey, G., Vina-Herbon, C., Blandon, A., Pesch, R., Schmitt, P., Scholle, J., Heyer, K., Lavaleye, M., Phillips, G., Duineveld, G.C.A., Blomqvist, M. (2018). A regional benthic fauna assessment method for the Southern North Sea using Margalef diversity and reference value modelling. Ecological Indicators 89, 667-679.

Walvoort, D.J.J., Wijnhoven, S., schilder, J.C. (2022). 'Relative Margalef diversity script'. Assessment software OSPAR BH2b indicator, consisting of three parts: 1) Data preparation and cleaning (step-1-data-prep.html); 2) Exploratory data analysis (EDA) (step-2-eda.html) and 3) Assessment (implicit SAR) (step-3-assessment-nosar.html).