NEA-PANACEA Activity 2 - Eutrophication and physical conditions informing MSFD D1, D4 and D6 assessments

Task 2.4: Towards coherent threshold value setting methods

Task Lead: Lisette Enserink (RWS, NL)
Other involved: Thomas Raabe (AquaEcology, DE), Anouk Blauw (Deltares, NL), Laurent Guérin (OFB, FR), Abigail McQuatters-Gollop (UoP, UK), Ian Mitchell and Stefano Marra (JNCC, UK)
Other Activities involved: Activity 1, 3, 4 and 5 (Tasks 1.4, 2.1, 3.1, 3.4, 3.6, 4.1, 5.2)

Acknowledgments
This report was produced in the frame of the European Maritime and Fisheries Fund project NEA PANACEA. The project was funded by the European Union (EU). Grant No. 110661/2020/839628/SUB/ENV.C.2

NEA PANACEA

North East Atlantic project on biodiversity and eutrophication assessment integration and creation of effective measures

Table of content

Summary

The main aim of Task 2.4 is to identify options to improve coherence between threshold value (TV) setting methods used in OSPAR's Quality Status Report (QSR) 2023 indicator assessments. These assessments contribute to MSFD Article 8 reporting for EU member states that are also OSPAR contracting parties. We considered all indicator assessments where threshold values have been applied, albeit at different levels of development, in the area of biodiversity and eutrophication. In addition to the ecosystem components covered in the NEA PANACEA project (pelagic habitats, foodwebs, benthic habitats, birds and eutrophication) the remaining indicator assessments developed under the OSPAR Biodiversity Committee were also investigated, namely those addressing marine mammals, fish and non-indigenous species. In total 29 indicator assessments were analysed, of which 21 used threshold values to assess the status of these indicators in the QSR2023 assessment period (usually 2015-2020).

In order to compare the TV setting methods, we used categories developed for the MSFD common implementation strategy (CIS). This resulted in an overview of the level of coherence within ecosystem components, usually using similar approaches. However, it was clear that full coherence is not possible, due to the lack of historic information/time series for many indicators, understanding of what 'good' status looks like in a changing world, and also due to the nature of the indicator, *eg*. bycatch of birds or marine mammals in fishing gear is approached in a different way than population abundance and distribution. We also made an attempt to compare threshold value setting approaches between indicators that are linked to each other though pressure-state relationships or foodweb interactions. This was further investigated in two examples, connecting eutrophication indicators with indicators for benthic and pelagic habitats. Understanding the level of (in)coherence really requires in-depth comparison, taking into account the underlying data, assessment methods and spatial and temporal patterns in the assessment outcomes.

The next – and most important – question is whether incoherence is a problem, for instance when assessment outcomes of two indicators lead to conflicting management measures. As an example, nutrient concentrations can be an important driver of phytoplankton productivity. A threshold value for nutrient concentrations should therefore be compatible with those for plankton assessments, to avoid situations where one indicator reports good status (no management action needed) and the other indicates non-satisfactory status (implying management action may be in order). A simple decision tree was developed to identify situations where action is needed to improve the coherence. Potential steps to do so are suggested, including improved collaboration between different expert groups in OSPAR and elsewhere and between experts and policy makers. Threshold value setting is a sensitive process that requires thorough discussions between experts and policy makers as well as insight in any consequences in terms of management action to facilitate that discussion. The present approach and analyses can support such discussions in OSPAR and other Regional Sea Conventions and in the MSFD CIS process.

1. Introduction

1.1. The NEA PANACEA project

NEA PANACEA is an EU-funded project in which 8 partners from 5 OSPAR Contracting Parties (Germany, France, the United Kingdom, Spain and the Netherlands) collaborate to deliver biodiversity assessments for OSPAR's Quality Status Report (QSR) 2023.

The project focus lies specifically on pelagic habitats, benthic habitats, food webs and marine birds' assessments in the framework of the OSPAR Quality Status Report (QSR) 2023. These assessments can be used by EU member states in the North East Atlantic region to inform their reporting to the EU for the Marine Strategy Framework Directive (MSFD). In addition, Activity 2 uses the newly developed and coherent approaches applied in OSPAR's eutrophication assessment (COMP4, based on the EU funded project JMP EUNOSAT). It provides tools to link pressure and state indicators to achieve more informative assessments of pelagic and benthic habitats and food webs, that will increase our understanding of ecosystem functioning in a changing climate and enable more effective management. Furthermore, Activity 2 - Task 2.4 investigates feasible options for better coherence between baseline and threshold setting methods for eutrophication, pelagic and benthic habitats, food webs, marine birds, marine mammals, fish and non-indigenous species (NIS) indicator assessments.

NEA PANACEA also aims to have value for those members of the OSPAR family that are not directly involved. For this reason, the investigation of baseline and threshold value (TV) setting methods has been extended to also cover the indicator assessments of marine mammals, fish and non-indigenous species that have been developed outside of the project for the QSR 2023 and the MSFD Art 8 assessments.

1.2. Activity 2 Task 2.4 - Towards coherent threshold value setting methods

A major challenge in the implementation of the MSFD is the development of baselines and threshold values in a coherent manner. OSPAR has a good track record of common assessment levels in some themes (hazardous substances (Environmental Assessment Concentrations (EACs)) and biodiversity (Ecological Quality Objectives (EcoQOs)) and has recently improved coherence in the assessment of eutrophication. This coherence so far addresses comparability between subregions and Contracting Parties and within Descriptors, for instance D5. Coherence *across Descriptors* is not yet addressed and discussions are only starting.

For a coherent assessment framework and where relevant (*eg.* pressure-state relationships) the MSFD and also OSPAR asks us to look into the type of narrative (or philosophy) that is used to set the threshold value. If these narratives conflict we may implement measures that steer in opposite directions. For instance, if we take measures to reduce nutrient levels to reach good status for eutrophication (currently: historic pre-eutrophication level + 50%) and at the same time we aim for large populations of a specific species that flourishes in high productivity environments, this will be conflicting. This issue was discussed in an EU-level MSFD workshop called Horizontal Issues – Threshold Values (30 September 2020) and findings were summarized in a number of workshop documents, which have been used as input for the present Task.

This Task aims to investigate and compare the threshold value setting experiences in OSPAR for D1/D6 (pelagic and benthic habitats, birds, mammals, fish and NIS), D4 and D5. Since eutrophication is one of the pressures affecting biodiversity, it is important to ensure that measures taken to reduce eutrophication are also beneficial to biodiversity and do not contradict measures to achieve GES for biodiversity indicators. This Task uses the background documents and outcomes of the MSFD Workshop on Horizontal issues.

The following steps have been taken:

1. Sharing the principles and MSFD language for baseline and threshold setting methods as in the background document for the Horizontal issues workshop to achieve a common language for further discussion;
2. Inventory and comparison of current methods used or under development in D1, D4 and D5 in OSPAR, investigation of why these methods have been chosen. This inventory also takes into account the limitations often faced with threshold setting, eg. limited time series. It also records the stage of development of a threshold value, which can range from not available/not started to policy acceptance.

3. Define functional links between criteria/indicators within and between these Descriptors to understand where coherence in baseline and threshold value methods is really crucial, so as to avoid potentially ineffective management measures;
4. Investigate inconsistencies between methods, and whether these are a problem and whether inconsistencies can be solved, given the circumstances, *eg*. limitations in data sets;
5. Identify options for next steps to improve the coherence.

Due to time constraints and the focus of OSPAR's Biodiversity Committee (BDC) and Hazardous Substances and Eutrophication Committee (HASEC) on delivering the QSR 2023 products in the 2022-2023 meeting cycle, the discussion on Task 4 results in these Committees has not yet taken place and will be organised after the NEA PANACEA project has ended (*ie*. in the 2023-2024 meeting cycle). The same situation is valid for the communication with the MSFD CIS level work in this area (follow up of the Workshop on Horizontal Issues), where the NEA PANACEA coordination team expects to present the project's results in a meeting of the Working Group GES (date to be determined).

Comparisons with approaches to GES assessments in other European Sea Regions has been limited so far to the workshop under Task 4.4 (*cf*. NEA PANACEA Final Report Annex AN) that focused on the assessments of marine birds to identify regional synergies and differences and to define an action plan detailing priorities for future co-working and establishing best practice for assessment.

2. Methods

Step 1. Sharing the principles and MSFD language for baseline and threshold setting methods

In the NEA PANACEA SuperCOBAM workshop (20-22 October 2021, Utrecht) a session was dedicated to this Step 1. The session was prepared by distributing a set of pre-read documents (*cf.* Annex 1) in order to familiarize participants with the language used in the MSFD CIS process and facilitate the discussion during the workshop.

Description of the categories of narratives and some examples as used in the Step 1 session:

1. *Acceptable deviation from historic or pristine state*
    - Similar to Water Framework Directive and the Habitats and Birds Directive, whereby TVs are set in relation to natural characteristics, such as the distributional range of a species, the extent of a habitat or the condition of its biological community.
    - Example 1: OSPAR 50% deviation from background concentrations for eutrophication parameters such as nutrient concentrations.
    - Example 2: Changes in occupancy rate and shifts in distribution from OSPAR's Changes in Harbour Seal and Grey Seal Distribution. Changes in seal distribution assessed between assessment period, against the baseline distribution, which is the potential distribution area or observed area in the baseline period.
2. *Non-deterioration*
    - To maintain good status (Art. 1 MSFD).
    - If scientific knowledge to set TVs is lacking, in combination with an improving trend (Art. 4(2) Commission Decision (EU) 2017/848).
3. *Points-of-no-return and tipping points*
    - Points-of-no-return are system condition parameter values that indicate a level, which, when surpassed, will lead to irreversible alterations in system conditions. A point-of-no-return might be reached if, due to a pressure, an ecosystem component declines beyond recovery. This could also include declines as a result of cumulative effects of multiple pressures.
    - Tipping points are system condition parameter values that indicate a level, which, when surpassed, will alter system conditions drastically.
4. *Removal and conservation targets*
    - Removal targets are TVs based on "unacceptable mortality levels" caused by human activities for the indicator species.
    - Example: HELCOM indicator on drowned mammals and waterbirds in fishing gear.
    - Conservation targets are TVs relating to the state of biological management units (i.e. stocks or populations). A limit value for a safe human-induced mortality of marine

species is usually the outcome of a simulation over a certain time period using a suitable population dynamic model. During the time period, the conservation target for the stock size is to be reached with a given certainty in a predefined fraction of the simulation time (e.g. at least 95% likelihood of reaching at least 80% of carrying capacity within 100 years).
    - Example: FMSY (Fishing mortality under the overall aim of Maximum Sustainable Yield) in the management of commercial fish species.
5. *Limit reference level*
    - Approach for defining TVs based on targeted estimated "optimal", "favourable" or "acceptable" condition.
    - Example: the Habitats Directive's Favourable Reference Range and Favourable Reference Area: The threshold value indicates how much habitat is needed to maintain its specialised species in viable populations.
6. *Trend-based approaches*
    - In the absence of knowledge of historic baseline and reference conditions and historical time-series, future state and trend-based approaches can be used individually or in combination with a baseline.
    - Example: OSPARs intermediate assessment for Harbour Seal and Grey Seal Abundance, and OSPAR's grey seal pup production.
7. *Other*

We used the same categories as in the MSFD workshop on threshold values in order to facilitate future interaction between the OSPAR and MSFD networks and the potential use of the NEA PANACEA outcomes in the MSFD CIS process.

Step 2. Inventory and comparison of current methods used or under development in D1, D4 and D5 in OSPAR and investigation of why these methods have been chosen

The NEA PANACEA SuperCOBAM workshop was also used for an initial inventory of methods for defining reference levels and threshold as used in the indicator assessments developed under NEA PANACEA. As mentioned above, also indicator lead authors outside of the project (notably authors of the QSR/MSFD assessments related to marine mammals, fish and NIS) were invited to contribute, either in person (when present in the workshop) or in a written procedure. Methods were categorized using the descriptions of Step 1.

Furthermore, a joint effort was made to compare the development stage of the threshold values, ranging from 'not started' to 'policy acceptance':

- P*olicy acceptance*: threshold values agreed and used by most countries in reporting under legal frameworks, such as MSFD and WFD in the case of EU member states;
- *Policy consequences clear*: threshold values understood by policy makers, including (potential) consequences (ecological but also societal) of reaching or not reaching the TV and whether there is a need for management measures. Threshold values may also be used by some countries in reporting under legal frameworks;
- *TV calculated*: expert agreement on threshold values, but not yet political acceptance. Could also be used in situations where experts accept that the current threshold values are based on current knowledge, but there is uncertainty whether these indicate 'good' status. This is the minimum level of development for application of the threshold values in an indicator assessment for the QSR 2023;
- *Narrative/method decided:* expert agreement on the threshold value setting method, but the values have not yet been calculated and/or agreed. No threshold values have been used in the assessment;
- *Not started/under development:* development of threshold values has not yet started or is at an early stage;

During the workshop participants also exchanged experiences related to the limitations often faced in threshold setting, *eg*. limited availability of time series and lack of information on non-impacted or 'good' state.

After the indicator assessments were finalised and made available on the OSPAR QSR 2023 SharePoint the inventory was updated and completed by the Task lead. An Excel matrix (Appendix 2) was produced that includes information on:

- the parameters used;
- season (if applicable/known);
- the OSPAR Region the indicator applies;
- the threshold value setting method used;
- the category it fits in;
- whether threshold values indeed have been used and how 'good' or 'not good' status has been defined;
- whether these threshold values are area-specific;
- (where relevant) comments on coherence between indicators.

Step 3. Define functional links between criteria/indicators within and between Descriptors D1, D4 and D5
One of the aims of Activity 2 was to stimulate the use of methods developed for D5 assessment in the indicator assessments for D1 and D4. This was supported by the programmes of the NEA PANACEA meetings (Kick-off, SuperCOBAM, UltraCOBAM, Final meeting and a workshop called MiniCOBAM which was organised by OSPAR's ICG-COBAM), where dedicated sessions were held to facilitate cross-cutting work between expert groups. Part of the discussions were on coherence in threshold setting methods, notably between the indicator assessments related to eutrophication, food webs, pelagic habitats and some benthic habitats indicators. There is a common understanding that coherence in baseline and threshold value methods is important, to avoid potentially ineffective management measures. However, limitations in data availability (time series) and understanding of what 'good' status looks like for a specific indicator hamper harmonisation of approaches.

Step 4. Investigate inconsistencies between methods
In the Final meeting of NEA PANACEA a session was dedicated to discussing the updated inventory in Step 2 and identifying examples where narratives of related indicators are not consistent and may lead to measures that would improve the assessment outcome of one indicator while at the same time compromising the assessment outcome of another indicator. These examples have been further elaborated in the present document.

Step 5. Identify options for next steps to improve the coherence.
For this purpose a decision tree was developed that helps investigating to which extent an apparent inconsistency is a real problem that should be solved, and to which extent it is at all possible to solve these inconsistencies, given the circumstances, *eg*. limitations in data sets. Based on the examples and discussions in the project, options for next steps, to be taken in OSPAR and in the MSFD CIS framework, are identified.

3. Results

3.1. Inventory of threshold setting approaches

The majority of the biodiversity indicators assess ecological state, and some assess a combination of the intensity of a pressure and the condition of an ecosystem component (*eg.* seafloor impacting pressures acting on benthic communities). Unlike threshold value setting for pressure-related indicators, where basic policy principles can be applied such as stand still or the precautionary approach, it is in many cases intrinsically difficult to do so for ecosystem components. This is primarily due to a lack of (monitoring) data on unimpacted (either historical of present-day) areas or populations, and also to the uncertainty of current and future changes in ecosystem structure and functioning caused by climate change. In addition, there are many uncertainties around the behaviour of a perturbated ecosystem and whether it can ever return to the previous unimpacted state, even when human pressures are removed (tipping point theory, Scheffer et al, 2001). Next to the biodiversity indicators the eutrophication indicators were also considered.

In the MSFD workshop Horizontal issues – Threshold values (30 September 2020) two types of pressures were distinguished: naturally occurring pressures and pressures solely made or introduced by man. Eutrophication, as one of the few pressures considered in this report, is by definition the result of excess input of nutrients by human activities, on top of naturally occurring nutrient inputs, from sources on land or at sea. For this type of pressure indicators the challenge is to define the acceptable level of a pressure on top of the natural level. In the case of eutrophication this requires understanding of ecosystem functioning at current levels in comparison with natural levels. Here, the link with biodiversity and foodweb indicator assessments becomes apparent and also the need for some level of coherence between the eutrophication and biodiversity threshold value setting approaches.

OSPAR biodiversity expert groups, covering all ecosystem components addressed in OSPAR, developed an interpretative assessment scheme applicable to each biodiversity indicator to classify indicator change within the wider ecosystem context (McQuatters-Gollop *et al.*, 2022). This categorical assessment uses expert interpretation of indicator change with respect to assessment thresholds (where available), links to pressures, and knowledge of indicator state to categorise indicators as in poor, uncertain, or good biodiversity status (Table 1). For a number of ecosystem components (pelagic habitats, foodwebs) it was felt that threshold values won't work and the interpretative scheme allows for an informal assessment of 'good' or 'not good' status, with no legal implications. Either way, the narrative is key in linking drivers of change to indicator change.

Table 1. Biodiversity status categories and colours used for the interpretation, by expert judgement, of indicator biodiversity state.

Poor	Indicator value is below assessment threshold, **or** change in indicator represents a declining state, **or** indicator change is linked to increasing effect of anthropogenic pressures (including climate change), **or** indicator shows no change but state is considered unsatisfactory
Uncertain	No assessment threshold **and/or** unclear if change represents declining **or** improving state, **or** indicator shows no change but uncertain if state represented is satisfactory
Good	Indicator value is above assessment threshold, **or** indicator represents improving state, **or** indicator shows no change but state is satisfactory
Unassessed	Indicator was not assessed in a region due to lack of data, lack of expert resource, or lack of policy support.

Threshold setting methods in biodiversity and eutrophication indicator assessments were investigated, using 29 indicator assessments developed for the OSPAR QSR 2023.

Table 2 summarizes the indicator assessments investigated, the type/category of thresholds, whether threshold values have been used in this assessment cycle, and the level of development and acceptance as indicated in October 2022, when these assessments were still under development. In some cases the judgement on the level of development has been adjusted by the author [in square brackets] according to the use of threshold values in the final version of the indicator assessment.

Table 2. Overview of threshold value setting methods used in biodiversity and eutrophication indicator assessments. Colours in the first column indicate assessments performed under the NEA PANACEA project (Tasks 1, 3 and 4). Grey cells indicate assessments carried out in OSPAR outside of the project. All assessments contribute to the QSR 2023 and – for EU member states – to MSFD 2024 reporting. Square brackets in the last column indicate an adjustment or interpretation of the level of development by the author.

Group	Indicator code	Indicator name	TV category	TV used in assessment	Level of development
Pelagic habitats	PH1 (FW5)	Changes in Phytoplankton and Zooplankton Communities	trend-based	No	Narrative/method decided
	PH2	Changes in Phytoplankton Biomass and Zooplankton Abundance	trend-based	No	Narrative/method decided
	PH3	Changes in Plankton Diversity in the Celtic Seas (common), Greater North Sea, and Bay of Biscay and Iberian Coast (candidate indicator)	trend-based	No	Narrative/method decided
Foodwebs	FW2	Pilot assessment primary productivity	trend-based	No	Narrative/method decided
	FW3	Size composition in fish communities	NA	No	Not started/under development
	FW4	Changes in average trophic level of marine predators in the Bay of Biscay (cf MTI)	NA	No	Not started/under development

Group	Indicator code	Indicator name	TV category	TV used in assessment	Level of development
	FW7	Pilot Assessment of Feeding Guild	NA	No	Not started/under development
	FW9	Pilot Assessment of Ecological Network Analysis Indices in Region II (Bay of Seine, Elbe Plume and Kattegat) en V (Azores)	NA	No	Not started/under development
Eutrophication	EU1*	Winter Nutrient Concentrations in the Greater North Sea, Celtic Seas and Bay of Biscay and Iberian Coast	Acceptable deviation from historic or pristine state	Yes	Policy consequences clear
	EU2*	concentrations of chlorophyll a in the Greater North Sea, Celtic Seas and Bay of Biscay and Iberian Coast	Acceptable deviation from historic or pristine state	Yes	Policy consequences clear
	EU3*	Concentrations of Dissolved Oxygen Near the Seafloor in the Greater North Sea, Celtic Seas and Bay of Biscay and Iberian Coast	Points-of-no-return and tipping points	Yes	[Policy consequences clear]
Benthic habitats	BH1	Sentinels of the Seabed	Points-of-no-return and tipping points	No	Narrative/method decided
	BH2a	Condition of benthic habitat communities: assessment of some coastal habitats in relation to nutrient and/or organic enrichment (WFD)	Acceptable deviation from historic or pristine state	Yes	Policy acceptance
	BH2b	Condition of Benthic Habitat Communities: Margalef diversity in Region II (Greater North Sea)	Acceptable deviation from historic or pristine state	No	Not started/under development
	BH3a	Extent of Physical Disturbance to Benthic Habitats (BH3a): Fisheries with mobile bottom-contacting gears	Limit reference level (extent physical damage) and 'other' (condition/disturbance based on sensitivity). LE: should 'other' also be limit reference level?	No	Narrative/method decided
	BH3b	Extent of Physical Disturbance to Benthic Habitats (BH3b): Aggregate Extraction	Limit reference level (extent physical damage) and 'other' (condition/disturbance based on sensitivity). LE: should 'other' also be limit reference level?	No	Narrative/method decided
	BH4	Area of habitat loss	NA	No	Not started/under development
Birds	B1	Marine bird abundance	Limit reference level	Yes	Policy acceptance
	B3	Marine Bird Breeding Productivity	Limit reference level	Yes	TV calculated
	B5	Marine bird bycatch (pilot)	Removal and conservation targets	Yes	Policy consequences clear
	B7	Marine bird habitat quality (pilot)	Acceptable deviation from historic or pristine state	No	[Not started/under development]
Marine mammals	M3	Seal abundance and distribution	Trend-based approaches	Yes	[TV calculated]
	M4	Abundance and Distribution of Cetaceans	Trend-based approaches	Yes	[TV calculated]
	M5	Grey seal pup production	Trend-based approaches	Yes	[TV calculated]
	M6	Marine mammal bycatch	Removal and conservation targets	Yes	[TV calculated]
Fish	FC1	Recovery of sensitive fish species	Trend-based approaches	Yes	[TV calculated]
	FC2	Proportion of large fish (Large Fish Index, LFI)	Trend-based approaches	Yes	[TV calculated]
	FC3	Mean maximum length of fish (candidate)	NA	No	[Not started/under development]
Non-indigenous species	NIS3	Trends in New Records of Non-Indigenous Species (NIS) Introduced by Human Activities	Trend-based approach	Yes	TV calculated

*Abbreviation for the purpose of this Task, not used in OSPAR.

Below, a more extensive description is given of the threshold setting methods and narratives used for each indicator for which threshold values have been applied in the assessment or where the narrative/method has been decided, but threshold values were not defined or not used. The entire result of the investigation is in Appendix 2 (Excel file), that also includes information on *eg*. the parameters used, season (if applicable/known), to which OSPAR Region the indicator applies and, where relevant, comments on coherence between indicators.

3.1.1. Pelagic habitats

As mentioned above the categorisation reflected in Table 1 is used for the determination of the status of the pelagic habitats indicators. A trend-based approach was used for all indicators, comparing the trend in the assessment period against the available time series over previous years, *ie*. baseline period (Table 3). It was not possible to only select the oldest part of the time series, since too much interpretative information and context would then be lost. A set of 16 environmental variables (modelled or observed), some of which influenced by human activities (such as nutrient concentrations and sea surface temperature), was considered to identify their importance for the observed spatial and temporal patterns in the pelagic indicators. This is an important part of the assessment, since it informs where and how human pressures may have caused a change in the pelagic system. However, it is currently unknown what a healthy pelagic habitat in 'good environmental status' looks like and it was therefore considered inappropriate to determine quantitative/fixed threshold values. The assessments primarily detect change, which is considered a warning signal according to the reasoning in Table 1.

These indicator assessments are area specific and they used the ecologically coherent assessment areas developed for the eutrophication assessment, see also Task 2.3 (NEA PANACEA Final Report Annex Q).

Table 3. Threshold value setting methods used in pelagic habitat indicator assessments. Source: All Indicator Assessments - OSPAR-OAP (Prod)

Indicator code&name	TV narrative category	Reasoning	TV setting or assessment narrative/methodology	Assessment methodology	Threshold Value	TV area specific?	Area/scale
PH1 (FW5) Changes in Phytoplankton and Zooplankton Communities	trend-based	We have this approach because we have time-series of all different lengths. If we take the oldest period available across all of them to use as our baseline (2008-2014) we lose way too much interpretive information and context. Right now we feel that TVs probably won't work for our pelagic indicators but we have developed categories, that will still allow us to determine GES or not GES. Either way, the narrative is key in linking drivers of change to indicator change.	comparing significant change in annual mean abundance values from assessment period (2015-2019), to preceding years (1960-2014, depending on dataset) AND whether these changes can be linked to human pressure	See Table 1	abundance in present assessment period compared to 1960-2014. Not a TV	Y	COMP4 areas, also grouped into pelagic habitat categories: variable salinity, coastal, shelf, and oceanic / beyond shelf
PH2 Changes in Phytoplankton Biomass and Zooplankton Abundance	trend-based	Same as above	comparing significant change in de-seasonalised data (monthly anomalies) from assessment period (2015-2020), to preceeding years (1997-2014)	See Table 1	biomass/abundance in present assessment period compared to 1997-2014. Not a TV	Y	Using 1° longitude by 0.5° latitude blocks aggregated across a grid of 1° x 1°, and to COMP4 assessment areas, also grouped into pelagic habitat categories: variable salinity, coastal, shelf, and oceanic / beyond shelf
PH3 Changes in Plankton Diversity in the Celtic Seas (common), Greater North Sea, and Bay of Biscay and Iberian Coast (candidate indicator)	trend-based	Same as above	comparing significant change in EQR from assessment period (2015-2019), to (all?) preceding years. EQR=index value at a given year/reference value, where reference value is taken from preceding years.	*assessment only done for Celtic Seas (common indicator)* See Table 1	indicator value in present assessment period compared against preceding years. Not a TV	Y?	COMP4 areas, also grouped into pelagic habitat categories: variable salinity, coastal, shelf, and oceanic / beyond shelf

3.1.2. Foodwebs

For only one food web indicator assessment a threshold value narrative was used, although it was considered inappropriate to determine quantitative/fixed threshold values, as in the pelagic habitats assessments. Instead, the categories in Table 1 were used and also the set of environmental variables to identify potential drivers, see under pelagic habitats. In line with the pelagic habitats assessment a trend-based approach was used to compare the trends in the current assessment period against trends observed in previous years (Table 4).

These indicator assessments are area specific and they also used the ecologically coherent assessment areas developed for the eutrophication assessment, see also Task 2.3.

Table 4. Threshold value setting methods used in foodweb indicator assessments. Source: [All Indicator Assessments - OSPAR-OAP (Prod)]

Indicator code&name	TV narrative category	Reasoning	TV setting or assessment narrative/methodology	Assessment methodology	Threshold Value	TV area specific?	Area/scale
FW2 Pilot assessment primary productivity	trend-based	We have this approach because we have time-series of all different lengths. If we take the oldest period available across all of them to use as our baseline (2008-2014) we lose way too much interpretive information and context. Right now we feel that TVs probably won't work for our pelagic indicators but we have developed categories, that will still allow us to determine GES or not GES. Either way, the narrative is key in linking drivers of change to indicator change.	comparing significant change in de-seasonalised data (monthly anomalies) from assessment period (2015-2020 for station data, 2015-2016 for satellite data), to preceeding years (1992-2014, depending on data set)	See Table 1	primary production in present assessment period compared to 1992-2014. Not a TV	Y	Using 1° longitude by 0.5° latitude blocks aggregated across a grid of 1° x 1°, and to COMP4 assessment areas, also grouped into pelagic habitat categories: variable salinity, coastal, shelf, and oceanic / beyond shelf

3.1.3. Eutrophication

The three common indicators used to assess eutrophication have threshold values, which are area-specific and reflect an acceptable deviation from historic (pre-eutrophic) state in two indicators. The third indicator has a generic TV that intends to avoid drastic changes in benthic communities caused by hypoxia (tipping point) (Table 5).

Recently, the threshold values for nutrient concentrations and chlorophyll-a have been harmonized between OSPAR contracting parties. The threshold setting narrative was retained, but the reference conditions were recalculated and harmonized. In addition, new assessment areas were defined, using ecologically relevant characteristics rather than international boundaries (COMP4 areas, see OSPAR COMP agreement – OSPAR, 2022). The acceptable deviation (50%) from the reference conditions was politically agreed in OSPAR in the 1980's, when eutrophication in the North-East Atlantic, in particular the North Sea and the Irish Sea, was more prominent than today. The policy acceptance of the newly developed threshold values varies among contracting parties, however among contracting parties that are also EU member states there is a general intention to use – where possible – the outcomes of the indicator assessments for MSFD reporting. Although rooted in past policy agreement, it is not clear to which extent the threshold value setting method of 50% acceptable deviation represents a healthy ecosystem, in terms of ecosystem functioning. Therefore, the NEA PANACEA project included comparisons with related indicator assessment, for instance under pelagic habitats and foodwebs.

Table 5. Threshold value setting methods used in eutrophication indicator assessments. Source: [All Indicator Assessments - OSPAR-OAP (Prod)]

Indicator code&name*	TV narrative category	Reasoning	TV setting or assessment narrative/methodology	Assessment methodology	Threshold Value	TV area specific?	Area/scale
EU1 Winter nutrient concentrations (DIN and DIP)	Acceptable deviation from historic or pristine state	Nutrient concentrations in the pre-eutrophic period are modelled and area-specific. 50% acceptable deviation has been agreed in OSPAR in the 1980's.	concentrations should be below area-specific TV, which is pre-eutrophic conditions plus acceptable deviation of 50%	**Not good**: indicator value is above assessment threshold; **Good:** indicator value is at or below assessment threshold.	pre-eutrophic (1900) conditions plus acceptable deviation of 50%	Y	COMP4 areas
EU2 Concentration chlorophyll a (in situ and satellite)	Acceptable deviation from historic or pristine state	Chlorophyll- a concentrations in the pre-eutrophic period are modelled and area-specific. 50% acceptable deviation has been agreed in OSPAR in the 1980's.	concentrations should be below area-specific TV, which is pre-eutrophic conditions plus acceptable deviation of 50%	**Not good**: indicator value is above assessment threshold; **Good:** indicator value is at or below assessment threshold.	pre-eutrophic (1900) conditions plus acceptable deviation of 50%	Y	COMP4 areas
EU3 Concentration dissolved oxygen near seafloor	Points-of-no-return and tipping points	Below the TV adverse impacts of anoxia on benthic communities can occur. These impacts also depend on the duration of the low oxygen level period.	concentrations should be at or above 6 mg dissolved oxygen/l to avoid hypoxia impacts on benthic organisms	**Not good**: indicator value is below assessment threshold; **Good:** indicator value is at or above assessment threshold.	6 mg dissolved oxygen/l	N	COMP4 areas

*The indicator codes EU1, 2 and 3 have been introduced for this Task 2.4. They are not used in OSPAR.

3.1.4. Benthic habitats

The four benthic habitat indicator assessments that include an assessment of the state of the indicator use different narratives for TV setting: Points-of-no-return and tipping points (BH1), Acceptable deviation from historic or pristine state (BH2a), and Limit reference level (extent physical damage) and 'other' (condition/disturbance based on sensitivity) for BH 3 a and b (Table 6). BH 1, 3a and 3b focus on the extent and severity of physical disturbance and use disturbance gradients based on distance to the source (*eg.* demersal fishing or aggregate extraction) in combination with impacts on sensitive species and information on dose-response relationships.

As a well-developed example, BH1 defines the degradation point of sentinel species (species characteristic of the habitat and sensitive to a given pressure) proportion per habitat type based on pressure response curves across a pressure gradient, starting at undisturbed (reference) conditions. The condition threshold is established as a percentile of the distance between the origin of the curve and the degradation point. Depending on the sensitivity of the habitat the TV can be standard, precautionary, or tolerant, where (i) the standard corresponds with the middle point (0,5) between the beginning of the curve and the tipping point, (ii) the precautionary located in the first third (0,33) of that range and (iii) the tolerant threshold located in the second third (0,66) of that range (Figure 1). The degradation point is the point at which the habitat has lost most of its quality. At this point, the pressure-state curves change their trend, decreasing the rate at which the reduction in the habitat state is observed. Currently, the method relies on the 45 degrees slope of the tangent to the curve.

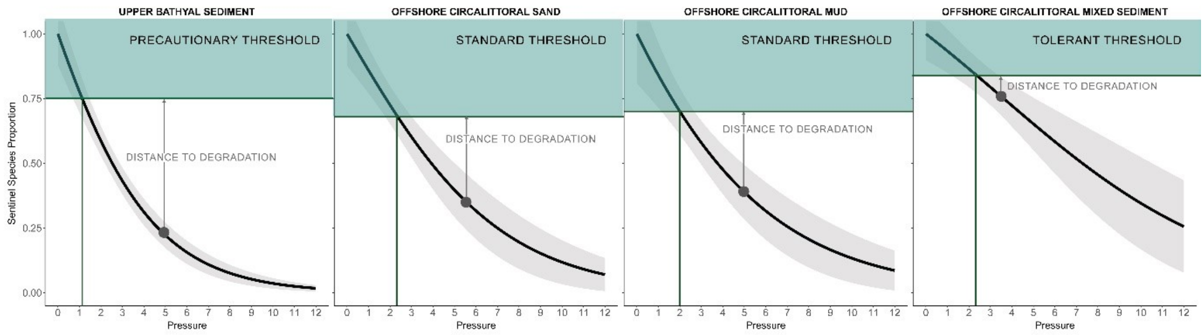

Figure 1. Distance to degradation approach methodology for setting thresholds to evaluate disturbance on seabed habitats. The four pressure-state curves show the four BBHTs from North Iberian Atlantic with different sensitivities, from more sensitive (sensitivity 4) to less sensitive (sensitivity 2) (In: BH1 assessment)

BH2a relates to chemical disturbance (eutrophication) in WFD coastal water bodies. Here, reference conditions and acceptable deviation vary between countries, which hamper comparison of assessment outcomes between countries. However, since this indicator is assessed and reported under the WFD, the level of policy acceptance is high. This is not the case for the other BH indicators, where TV setting has been a politically sensitive process in OSPAR and no agreement could be reached on the formal use of the TV setting methods. The BH1, BH3a and b assessment outcomes inform about the level of disturbance (expressed in disturbance categories) in the assessment areas, and refrain from an assessment in terms of 'good' or 'not good'.

All BH indicator assessments are area specific since they use (broad) habitat types as assessment areas. The eutrophication (COMP4) assessment areas were considered less useful, as these do not include information about seafloor substrate.

Table 6. Threshold value setting methods used in benthic habitat indicator assessments. Source: [All Indicator Assessments - OSPAR-OAP (Prod)]

Indicator code&name	TV narrative category	Reasoning	TV setting or assessment narrative/methodology	Assessment methodology	Threshold Value	TV area specific?	Area/scale
BH1 Sentinels of the Seabed	Points-of-no-return and tipping points	Combines information on extent and intensity of (physical) disturbance caused by human activities, such as demersal trawling, with dose-response information of sentinel species. See explanation in text above.	degradation point of sentinel species proportion per habitat type based on pressure response curves across a pressure gradient, starting at reference conditions.	The areas were classified as follows: **no pressure**, the value of the pressure on the area is zero, **low disturbance** when the proportion of sentinel species was higher than the threshold, even after removing the standard error; **high disturbance** when the proportion of sentinel species was lower than the threshold, even after adding the standard error and **Moderate disturbance** areas when the position (higher or lower) of the proportion of sentinel species related to the threshold changes after adding/removing the standard error.	degradation point of sentinel species proportion per habitat type, giving the most sensitive habitats the highest distance to degradation. Precautionary for sensitive habitats (sensitivity value 4), standard or tolerant for less sensitive habitats (sensitivity values 3 or 2 respectively).	Y	broad habitat types submitted to trawling effort in Region IV
BH2a Condition of benthic habitat communities: assessment of some coastal habitats in relation to nutrient and/or organic enrichment	Acceptable deviation from historic or pristine state	Focus on the effects of nutrient and/or organic enrichment, but indices may also respond to other pressures.	Acceptable deviation from reference condition; CPs use various reference conditions. Ratio between current and reference condition expressed as EQRS.	The values of the boundaries between classes of Ecological Quality Ratio of the CPs/Member States are established per country and per habitat type: usually **High-Good** is approximately 0,8 and **Good-Moderate** is approximately 0,6 at the EQRS scale from 1 to 0, for benthic invertebrate fauna and Angiosperms and macroalgae.	varies between MS, approximately ≥ 0,8 high; ≥ 0,6 good	Y	WFD areas and in Scotland partly overlapping with COMP4
BH3a Extent of Physical Disturbance to Benthic Habitats (BH3a): Fisheries with mobile bottom-contacting gears	Limit reference level (extent physical damage) and 'other' (condition/disturbance based on sensitivity).	Assesses the extent and intensity of disturbance by bottom contacting fishing gear types in combination with habitat sensitivity	Habitat sensitivity is quantified in 5 classes as a combination of resistance (structural and functional) and resilience/recovery time. Pressure extent (Swept Area Ratio) is also quantified in 5 classes as a combination of %grid	Disturbance categories were summarised into three groups ('Low' = disturbance categories 1-4, 'Moderate' = disturbance categories 5-7, and 'High' = disturbance	These groupings are not representative of thresholds and should be used for comparative interpretations of	Y	Central North Sea, Southern North Sea, Channel, Norwegian Trench, Kattegat, Northern Celtic Sea, Southern Celtic Sea, Gulf

			cell swept and frequency of fishing. Disturbance matrix combines extent of pressure and habitat sensitivity in 9 classes. Focuses on pressures associated with bottom-contacting fishing.	categories 8 and 9) derived from the 1-9 disturbance scale.	disturbance outputs across the OSPAR Maritime Area only.		of Biscay, North-Iberian Atlantic, South-Iberian Atlantic, and Gulf of Cadiz. For fishing pressure: 0,05x0,05 degree grid cells. For disturbance: broad habitat types subdivisions up to EUNIS level 6.
BH3b Extent of Physical Disturbance to Benthic Habitats (BH3b): Aggregate Extraction	Limit reference level (extent physical damage) and 'other' (condition/disturbance based on sensitivity).	Assesses the extent and intensity of disturbance by aggregate extraction for commercial purposes (*eg*. sand and gravel) in combination with habitat sensitivity	Habitat sensitivity is quantified in 5 classes as a combination of resistance (structural and functional) and resilience/recovery time. Pressure extent is also quantified in 5 classes as a combination of %grid cell swept and duration of the dredging. Disturbance matrix combines extent of pressure and habitat sensitivity in 9 classes. Focuses on pressures associated with commercial aggregate extraction.	Disturbance categories were summarised into three groups ('Low' = disturbance categories 1-4, 'Moderate' = disturbance categories 5-7, and 'High' = disturbance categories 8 and 9) derived from the 1-9 disturbance scale.	These groupings are not representative of thresholds and should be used for comparative interpretations of disturbance outputs across the OSPAR Maritime Area only.	Y	Central North Sea, Southern North Sea, Channel, Norwegian Trench, Kattegat, Northern Celtic Sea, Southern Celtic Sea, Gulf of Biscay, North-Iberian Atlantic, South-Iberian Atlantic, and Gulf of Cadiz. For aggregate extraction: 50mx50m grid cells. For disturbance: broad habitat types.

3.1.5. Seabirds

Two bird indicator assessments use TVs as limit reference levels (B1 and B3), where for B5 Removal and conservation targets are defined (Table 7). All TV setting methods aim to prevent (long-term) decline of bird populations and applied per species. Where possible (B1 and B3) IUCN red-list criteria have been used, which facilitates uptake in legally binding assessment frameworks. For indicator B1 a baseline of 1991-2000 is used and the resilience of a species, as indicated by the number of eggs in a clutch, is taken into account. Both B1 and B3 assessments integrate the assessment outcomes into functional groups and for that integration a TV of 75% of the species in 'good status' is used.

For B5 three methods have been applied per species, depending on data availability (Figure 2),:

1. (default) assess threat to protection and conservation of birds: Population Viability Analysis (PVA) to be applied where data allow;
2. minimise/eliminate incidental by-catches of birds: by-catch level ≤1% of total annual adult mortality as an approximation of zero by-catch, which acknowledges that small numbers of birds will probably still be caught even when the most effective mitigation measures are deployed;
3. identify risk areas by investigating spatio-temporal overlap in distributions of marine bird species and fisheries causing by-catch. This is a precautionary method only applied to OSPAR threatened and/or declining species to prevent further threatening of such species.

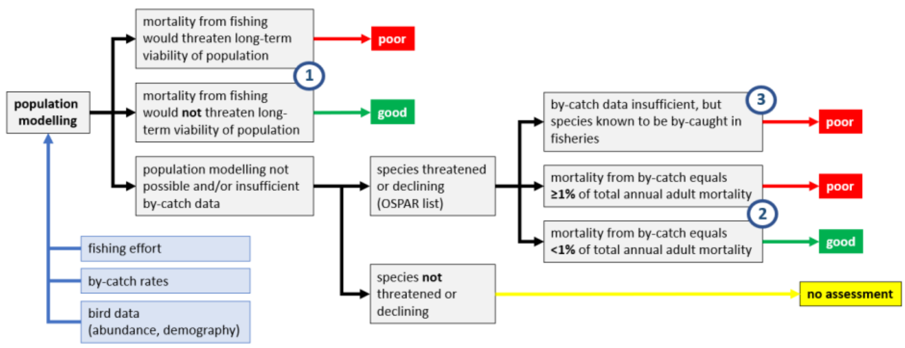

Figure 2: Workflow for the candidate indicator B5 Marine Bird By-catch. Numbers indicate where the Assessment Methods 1, 2 and 3 are applied (in: B5 assessment).

The TVs used in B1 and B3 are generic across Regions. Step 3 in the B5 assessment is area-dependent. The level of development of the TVs used for B1 and B3 is high; whereas B5 is a candidate indicator applied to a restricted part of the OSPAR area.

Table 7. Threshold value setting methods used in bird indicator assessments. Source: [All Indicator Assessments - OSPAR-OAP (Prod)]

Indicator code&name	TV narrative category	Reasoning	TV setting or assessment narrative/methodology	Assessment methodology	Threshold Value	TV area specific?	Area/scale
B1 Marine bird abundance	Limit reference level	In the lack of information of abundance level in pristine areas, the indicator uses the starting point levels from long term time series as baseline. However, data scarcity prior to 1991 prevents using an earlier baseline. Considering the length of the time series the broad assumption is that pressures affecting bird abundance were less intense in the first ten years of the time series, albeit unlikely at pristine levels.	abundance per species should be above 70 (more than 1 egg) - 80% (1 egg) of 1991-2000 baseline populations. These thresholds are derived from statistical properties of normal distribution (70% of observation will fall within one standard deviation of the mean) The lower TV for species laying more than 1 egg account of the greater resilience in terms of reproductive capacity. The baseline is calculated from regression analysis on the first ten years of the time series (or using the mean 1991 to 2000 if no significant regression).	Threshold values: Good: >0,8 (i.e., 80% of the baseline) – for species that lay one egg; or >0,7 (i.e., 70% of the baseline) – for species that lay more than one egg. If 75% or more of species assessed exceed their individual threshold values, an assemblage of bird species is considered to be healthy.	75% of all species in a functional group above individual species threshold	N	OSPAR regions and sub-divisions of Greater North Sea and Norwegian part of Arctic Waters
B3 Marine Bird Breeding Productivity	Limit reference level	The indicator does not use a traditional baseline. The indicator uses the population growth rate which, if sustained, would lead to a decline in population size of under 30% over three generations, this TV keeps the species above their IUCN level for 'vulnerable' designed for red list species.	population growth rate should be high enough to prevent a 30% decline over 3 generations (calculated per species with population models).	Threshold values: Good: < 30% decline in population growth rate over three generations. A stable population has a growth rate of 1, a growing or increasing population has a growth rate of greater than 1 and a declining population has a growth rate of less than 1. If 75% or more of species assessed is below their individual threshold values, an assemblage of bird species is considered to be healthy.	75% of all species in a functional group below individual species threshold	N	OSPAR regions

		Common approach between OSPAR, ICES and HELCOM and consistent with IUCN.					
B5 Marine bird bycatch	Removal and conservation targets (candidate)	3 assessment methods, applied per species, see description and Figure 2 above.	Method 1: The annual loss of individuals from by-catch in fisheries is to be assessed against the threshold that the long-term viability of a population, using Population Viability Analysis (PVA), is not threatened. Method 2: the number of individuals reflecting 1% of the annual adult mortality of a species (or population). Method 3: the distribution of a bird population in a given assessment unit does not overlap spatially and temporally with the exercise of a fishing method which is known to cause by-catch in that species (only applies to threatened/declining species).	Integration of Assessment methods 1 (first step), 2 and 3 (second step) determine whether the indicator is assessed as good or poor.	Method 1: TV species-dependent, using PVA; Method 2: by-catch level ≤1% of total annual adult mortality; Method 3: no spatio-temporal overlap in distributions of marine bird species and fisheries causing by-catch	Y	OSPAR Regions

3.1.6. Marine mammals

As in the birds indicators, all marine mammal TVs essentially aim to conserve or restore populations. For indicators M3, M4 and M5 trend-based approaches were used (Table 8). The baseline (1992) is used by some Member States for reporting under the European Union Habitats Directive (Council Directive 92/43/EEC). Because of the long generation time of marine mammals two TVs are used (short- and long term) to avoid the 'shifting baseline syndrome'. With a shifting baseline, each successive assessment is comparing slightly different sets of consecutive data points. This could allow an indicator to continually decline at a slower rate than the threshold value, but after many years the population may have declined substantially without actually being below the threshold value. In the M4 assessment the annual threshold rate depends on the generation time, which varies between species, *e.g.* the shorter the generation time the higher the annual threshold rate.

Similar to the birds bycatch indicator B5, the marine mammal bycatch indicator M6 uses TVs that are classified as Removal and conservation targets, using a limit to human-caused mortality. This limit is described as a maximum tolerable deviation from the carrying capacity. No baseline is needed in this case. The conservation objectives differ between species groups, *ie.* 80% and 50% of the carrying capacity for cetaceans and seals respectively. Taking these values into account, species- and area specific TVs for anthropogenic removals (number of animals) were calculated in an OSPAR/HELCOM workshop (2019).

In all of the four indicators assessments TVs have been used to determine 'good' or 'not good' status and the threshold setting methodologies have been agreed in OSPAR's Biodiversity Committee.

Table 8. Threshold value setting methods used in marine mammal indicator assessments. Source: [All Indicator Assessments - OSPAR-OAP (Prod)]

Indicator code&name	TV narrative category	Reasoning	TV setting or assessment narrative/methodology	Assessment methodology	Threshold Value	TV area specific?	Area/scale
M3 Seal abundance and distribution	Trend-based approaches	The 25% (long-term trend) currently approximates to 1% a year since 1992. Testing shows that there is sufficient monitoring to assess against this assessment value with confidence. Where a shorter timescale is assessed, the 25% decline since the baseline is not equivalent to those AUs where data do extend to 1992 (for example, a 25% decline since 2003 describes a more rapid contraction in the population than a 25% decline since 1992). Two assessment thresholds were used to address the issue known as 'shifting baselines'.	Baseline set to 1992 (or start of the data series). Two assessment thresholds were used to address the issue known as 'shifting baselines'. To avoid the problem of shifting baselines when using the rolling baseline applied in assessment value 1, an assessment value relating to a fixed baseline is needed (assessment value 2).	**Good:** Threshold value 1 (short term): "No decline in seal abundance of > 1% per year in the previous 6-year period (this is approximately 6% over 6 years)."; Threshold value 2 (long term): "No decline in seal abundance of >25% since the fixed baseline in 1992 (or closest value)."	decline ≤1%/year in short term; decline ≤25% since 1992 (or closest value) in long term.	Y	specific assessment units, reflecting a balance between population structure evidence (e.g. telemetry and genetics) and feasible monitoring sites. Entire area for grey seal abundance.
M4 Abundance and Distribution of Cetaceans	Trend-based approaches	The rate of decline is assessed for each species-specific assessment unit by comparing the latest abundance estimates with the baseline, equal to the earliest reliable population estimate for the assessment unit (e.g. from SCANS or SCANS II/CODA.) The generation time varies between species which causes the different thresholds per species; e.g. the shorter the generation time the higher the annual threshold rate. Importantly, although thresholds are based on the IUCN criterion of a 30% decline over three generations, it is not necessary to wait for three generations for the assessment.	Baseline set to 1992 (or start of the data series); The proposed trend-based threshold is species specific and has two parts: (1) no absolute decrease and is relevant irrespective of a time period. (2) allows to compare an annual trend: i.e. thresholds are an annual rate of decline in abundance that must not be exceeded. These annual rates of decline, if sustained over three generations, will lead to 30% decline in abundance.	**Good:** For each assessment unit: threshold value (1) maintain [insert species name] population size at or above baseline levels with no absolute decrease of >30% AND treshold value (2) a rate of decrease no greater than 30% over three generations.	(1) no absolute decrease of >30% AND (2) a rate of decrease no greater than 30% over three generations.	Y	species specific assessment units
M5 Grey seal pup production	Trend-based approaches	The 25% (long-term trend) currently approximates to 1% a year since 1992. Testing shows that there is sufficient monitoring to assess against this assessment value with confidence. Where a shorter timescale is assessed, the 25% decline since the baseline is not equivalent to those AUs	Baseline set to 1992 (or start of the data series). Trends in pup production informs on drivers of change (pressures) and where measures should be taken to manage these pressures. Two assessment thresholds were used to address the issue known as 'shifting baselines'. To avoid the	**Good:** Threshold value 1 (short term): " No decline in grey seal pup production of > 1% per year in the previous 6-year period (this is approximately 6% over 6 years)." Threshold value 2 (long term): "No decline in grey seal pup production of	decline ≤1%/year in short term; decline ≤25% since 1992 (or closest value) in long term.	Y	25 coastal assessment units in Regions I, II and III, based on the behaviour of mature grey seals of both sexes that are usually faithful to

		where data do extend to 1992 (for example, a 25% decline since 2003 describes a more rapid contraction in the population than a 25% decline since 1992). Two assessment thresholds were used to address the issue known as 'shifting baselines'.	problem of shifting baselines when using the rolling baseline applied in assessment value 1, an assessment value relating to a fixed baseline is needed (assessment value 2).	>25% since the fixed baseline in 1992 (or closest value)."			particular breeding sites and may return to within 10–100 m of individual breeding locations
M6 Marine mammal bycatch	Removal and conservation targets	Baseline not needed because the approach consists, for common marine mammal species, in setting the conservation objective as a maximum tolerable deviation from the carrying capacity (depletion). The threshold describes a limit to human-caused mortality. The threshold setting method is model-based and incorporates life-history and demographic parameters specific to the species and population assessed. The threshold has undergone thorough testing to ensure robustness against uncertainties and possible biases in the data.	For cetaceans - Conservation objective: "A "population" should be able to recover to or be maintained at 80% of carrying capacity, with 80% probability, within a 100-year period." For seals - Conservation objective (follows the US Potential Biological Removal PBR approach): "A population will remain at, or recover to, its maximum net productivity level MNPL (typically 50% of the populations carrying capacity), with 95% probability, within a 100-year period."	**Good:** The mortality rate from incidental catches is below levels which threaten any protected species, such that their long-term viability is ensured. OSPAR/HELCOM workshop 2019: translated in maximum number of by-catch per species per region	TVs for anthropogenic removals species and area specific	Y	species specific assessment units

3.1.7. Fish

The TVs used in the two fish indicator assessments are trend-based (Table 9). The FC1 assessment of sensitive fish species compares the presence of these species in survey catches in the current period to the previous assessment period, assuming that the population abundance and occurrences of each sensitive species sampled by each survey is assumed to have declined as a result of past human activities. Thus, achieving acceptable status for these sensitive species will require population recovery (primary TV), or halting the decline in occurrence (secondary TV).

For the assessment of FC2 (proportion of large fish) a length threshold as well as a proportion threshold is used. Fishing mortality as well as climate change affects this indicator. In the absence of reliable historic data (the longest running survey programmes started in the 1980's, and half of the programmes started after 2000) time series have been analysed to identify relatively stable periods that are used as baselines.

The TVs in both FC1 and FC2 are species and survey (and area) specific.

FC1 has been used in the (integrated) assessment of the overall status of fish in the OSPAR area, together with assessments of commercial fish stocks produced by ICES and ICCAT. FC2 has not been used for that purpose.

Table 9. Threshold value setting methods used in fish indicator assessments. Source: [All Indicator Assessments - OSPAR-OAP (Prod)]

Indicator code&name	TV narrative category	Reasoning	TV setting or assessment narrative/methodology	Assessment methodology	Threshold Value	TV area specific?	Area/scale
FC1 Recovery of sensitive fish species	Trend-based approaches	Each survey assessed has a differing start year, meaning that assessments of long-term change are not necessarily directly comparable between surveys or regions (CEMP Guideline). A temporally coherent measure of change that is comparable across the whole of the OSPAR Maritime Area is captured by the assessment of short-term change, which can highlight where signs of recovery or ongoing depletion are evident. Each species is assessed separately for each survey and both the primary and the secondary threshold is considered when data allows. Species are classified as "not assessed" if no data exists in the available surveys but the species is known to be present in the area. If fewer than five records are reported then the assessment outcome is classified as "unknown" (CEMP Guideline). Spatial integration within species across each Region.	By virtue of their sensitivity to additional human-related mortality, the population abundance and occurrences of each sensitive species sampled by each survey is assumed to have declined as a result of past human activities. Thus, achieving acceptable status for these sensitive species will require population recovery (primary TV), or halting the decline in occurrence (secondary TV).	Primary: **Good:** population is recovering, indicated by a significant increase in occurrence of a sensitive species between assessment periods in the area sampled; Secondary: **Good:** population is stable, ie. no decline in occurrence between assessment periods. Additional binomial test to determine whether a population was recovering among a significant fraction of the surveys available within each OSPAR Region or, for deep-sea species, across the whole OSPAR Maritime Area.	occurrence in previous assessment period (per species, per survey)	Y	Regions
FC2 Proportion of large fish (Large Fish Index, LFI)	Trend-based approaches	Fishing mortality constrains the age structure of fish communities, reducing the proportion of larger and older individuals of species impacted. Climate change (warming) also is expected to decrease the size of fish. Long-term trend is modelled. Subsequent breakpoint analyses define stable underlying periods and determines significant change in the time series state over time, namely whether the recent period is significantly different from the historically observed period. The method avoids the arbitrary choice of reference periods for assessment (i.e., how many years to use to calculate an average) which can lead to subjective assessments. The analysis uses two statistical approaches:	length threshold (LT) defining 'large fish' is used to calculate the ratio of large fish in the demersal fish community. LT and TV for the proportion of large fish are specific for each survey programme/gear type and hence the area where the survey (otter or beam trawl) is performed. TVs taken from time series and periods with relatively high and stable LFI values (at a specific value of LT) or periods where ICES stock	**Good:** proportion of fish ≥ LT is at or above some (stable) historic assessment value. **Not good:** proportion of fish ≥ LT is below some (stable) historic assessment value.	threshold value LT is specific for each survey programme/gear type and hence the area where the survey (otter or beam trawl) is performed. LT is between 30 and 50 cm in this assessment. For only 3 surveys (in Regions II and III) a TV for the proportion of large fish was available (range: 0,165-0,508).	Y	subregion (scale of the surveys), in a second step grouped for Region-wide assessment

		First applying the 'supremum F test' to establish whether a non-stationary time series or a constant period for the entire time series is more suitable. If the former, then breakpoint analysis is applied to find periods of at least six years duration.	assessments indicate stocks not unduly depleted. These periods are defined using modelling of trends and breakpoint analysis.				

3.1.8. Non-indigenous species

The management of non-indigenous species mainly focuses on reducing the number of new introductions. This is reflected in the trend-based assessment of indicator NIS3 (Table 10). The number of new introductions in the current assessment period is compared to the previous 6-year period and a downward trend is considered 'good'. Time series started in 2003.

The assessment is performed at the level of countries and integrated in OSPAR Regions.

Table 10. Threshold value setting methods used in non-indigenous species indicator assessments. Source: [All Indicator Assessments - OSPAR-OAP (Prod)]

Indicator code&name	TV narrative category	Reasoning	TV setting or assessment narrative/methodology	Assessment methodology	Threshold Value	TV area specific?	Area/scale
NIS3 Trends in New Records of Non-Indigenous Species (NIS) Introduced by Human Activities	Trend-based approach	focus on the approach applied in the IA2017. Basically this compares the number of new arrivals, ie. first records in each Region assessed, in separate 6 year periods from 2003 to 2020. In addition to this, we have ongoing analysis to investigate thresholds based on absolute vs relative changes.	The "New Introductions" parameter was chosen because (a) preventing NIS introductions is the most cost-effective management approach and (b) these records are the only reliable data available. While it was originally suggested to apply a threshold of zero new NIS introductions per six-year period, most Contracting Parties and EU Member States (JRC) support a trend-based approach.	Not good: no trend or upward trend. Good/improving: downward trend;	number of new introductions in previous 6-year assessment period	Y	Regions

3.2. Comparison of threshold setting methods in relation to functional links between criteria/indicators

3.2.1. Within ecosystem components

The level of coherence in threshold setting methods within sets of indicators applied to a specific ecosystem component varies and depends on the characteristics and the drivers of change considered in the indicator assessment. Below, some observations are given, restricted to the use of threshold values for the OSPAR QSR2023. No comparison was made with approaches used in other Regional Sea Conventions.

The *pelagic habitats* and *foodweb* indicators all use the same assessment areas and trend-based approaches, comparing the status and trends therein in the current assessment against previous years. No formal TVs have yet been identified, since it is unknown what a 'good' status looks like for these indicators. The coherence in threshold value setting methods between these indicators is high.

The assessment areas used in the pelagic habitats and food web assessments were developed for the *eutrophication* indicators (so-called COMP4 areas) and applied to all three common eutrophication indicators. The TV narratives are coherent for two indicators (acceptable deviation from historic pre-eutrophication state) but not for the third indicator (oxygen concentration near the seafloor), where a tipping point TV is used. Measures applied to achieve good status for the first two indicators may not lead to good status for the third.

The indicators used for the assessment of *benthic habitats* can be separated in those assessing the impact of physical pressures (demersal fishing and aggregate extraction) and one focusing on chemical pressures (eutrophication). The TVs for the indicators focusing on physical pressures were identified as tipping points and limit reference levels, but they all use a combination of information on the extent and intensity of the pressure and the sensitivity of the habitat or species therein. Whether there is sufficient coherence in TV setting methods is currently unknown, as for various reasons (methodological, data, policy support) no integration of indicator assessments was done. The indicator focusing on eutrophication (BH2a) has only been applied to WFD areas. The TV is based on acceptable deviation from historic or pristine state, the latter may well be represented by current unaffected areas, which then resembles the approach taken for the other benthic indicators (gradient in pressure intensity). However, the choice of reference conditions varies between countries (EU member states) and may well be incoherent with the other benthic TVs (BH1, BH3a and BH3b) in certain areas.

The two *seabird* indicators assessing population status and trends use the same type of TV setting method (limit reference levels), although B1 uses a 'traditional' past state baseline, while B3 looks at the future consequences of trends in population growth rate. These indicators are combined in the integrated assessment for each functional group (one out-all out). The bird bycatch indicator uses another type of TV setting method, looking specifically at additional bycatch mortality. All bird indicators aim at long-term conservation of bird populations, which is an aspect of coherence when it comes to management measures. However, measures to reduce pressures may affect these bird indicators in different ways, and even benefit some bird species while negatively affecting others.

Three out of four *marine mammal* indicators are trend-based, using the same 1992 baseline, which contributes to coherence. Indicators M3 and M5 consider seals, while M4 addresses cetaceans. The bycatch indicator M5 addresses all marine mammals bycaught in fishing nets, using another TV setting method (removal target, as in seabirds). All indicators relevant to the species of species group are considered in the integrated assessment for these groups (one out-all out approach). As in the case of birds, all marine mammal indicators aim at long-term conservation of populations, which is also an element contributing to coherence.

The TVs used for the two *fish* indicator are trend-based, but the baselines are not the same, and vary between survey programmes in both indicators. Both indicators aim at recovery of species sensitive to human pressures, using groundfish surveys, looking at abundance of fish species (FC1) or size-structure of fish communities (FC2). Whether there is sufficient coherence in TV setting methods is currently unknown; FC1, not FC2, is used in the integrated assessment.

Only one indicator for *non-indigenous species* comes with a TV, and analysis of coherence is therefore irrelevant.

3.2.2. Between ecosystem components and between state and pressure indicators

The NEA PANACEA project stimulated and facilitated interactions between expert groups, which helped the Activity 2 team to investigate relationships between eutrophication and biodiversity indicator assessments. The eutrophication indicators usually are regarded as pressure indicators, but they also contain information on the basis of the food web (biomass of algae) and habitat quality (oxygen level near the seafloor).
The majority of the connections was found between the PH and FW indicators on the one hand and eutrophication indicators on the other hand, sometimes even using the same parameter (*ie*. chlorophyll-*a* in PH2 and EU2). However, the PH and FW assessments also took into account other pressures, notably parameters linked to climate change. The connection with the benthic assessments was much weaker, since the majority of the assessments focused on physical pressures and only one (BH2a) on the impacts of nutrient and organic matter enrichment. Connection with the bird assessments was considered limited, although productivity of primary producers and hypoxic conditions of the seafloor may impact the assessment of the bird indicators, which is however not reflected in the assessments. Figure 3 is an initial attempt to map connections between the indicators for pelagic habitats, benthic habitats, eutrophication and birds.

Comparison between *pelagic habitats, foodweb and eutrophication* assessments has become much easier and more relevant with the harmonisation of assessment areas for all of these indicators. This is a significant step forward. With regard to the TV setting methods, there are two main differences between eutrophication and pelagic habitat and food web indicators: trend-based approaches (PH and FW) versus acceptable deviation from historic conditions (EU) and the choice of baseline/reference conditions, using pre-eutrophic (around 1900) state in EU1 and 2, and more modern conditions for PH and FW. The latter also include relatively high eutrophication status in some areas in the 1980s to early 2000s, due to excess inputs of nutrients in those years.

The TV setting approaches used in the *benthic* BH2a assessment are much more coherent with the *eutrophication* EU1 and 2 assessments, *ie*. acceptable deviation from historic or pristine state. There are differences between the two indicator groups with regard to the interpretation of historic or pristine, but the narratives are the same.

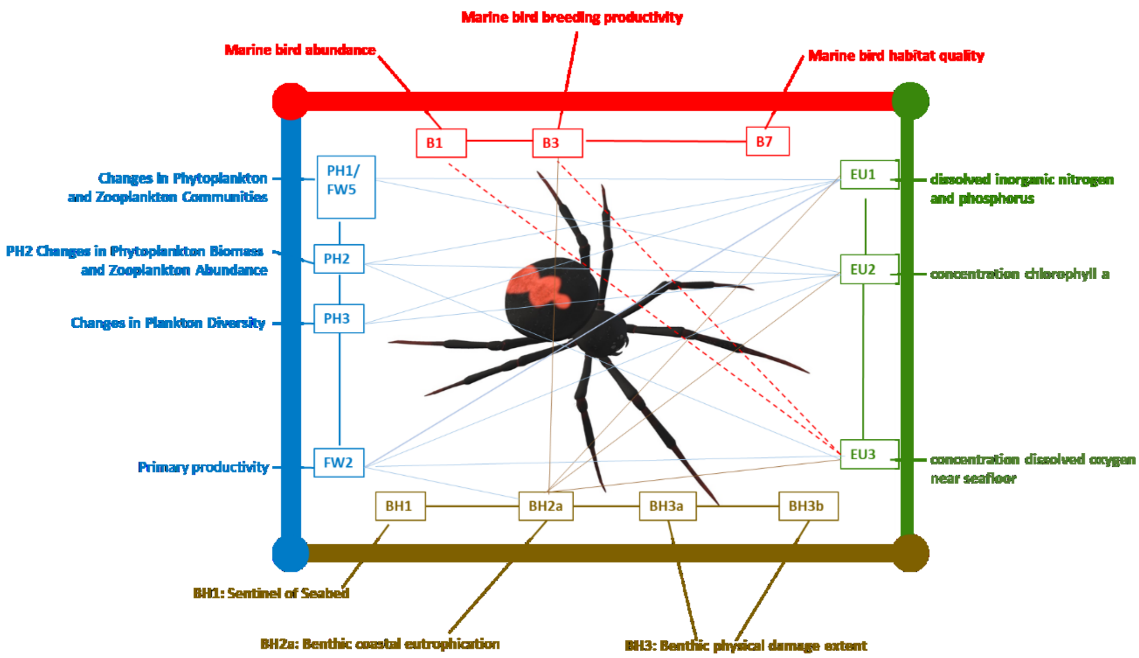

Figure 3. Relationships between biodiversity and eutrophication indicators (modified from Laurent's spider web)

Investigation of coherences between the eutrophication and PH, FW and the one benthic indicator assessment (BH2a) also involves comparison of the outcome of the assessments. This requires a more detailed analysis, taking into account space and time, in addition to the comparison between TV setting methods. This is illustrated in the examples below. The main issue is whether differences in approaches lead to inconsistencies in policy and management measures.

3.2.3. Examples of (in)coherence between threshold value setting approaches in related indicators

Example 1: BH2a vs EU1/2 - comparable TV setting methods and assessment outcomes in neighbouring areas

This example compares the benthic habitat indicator BH2a - *effects of nutrients on benthic vegetation and invertebrates* with two eutrophication indicators, *ie.* EU1 - *Winter nutrient concentrations* and EU2 - *Chlorophyll a.* The assessment areas do not overlap but can be neighbours, since BH2a is applied to WFD coastal waters, whereas EU1 and 2 apply to COMP4 assessment areas, that also contain coastal waters, but exclude WFD areas as advised in the MSFD Commission Decision (2017). The BH2a indicator is taken from the WFD 2016 reporting, which covers the period 2010 to 2015. This overlaps with the OSPAR COMP3 period from 2006 to 2014.

The TV setting methods are comparable:

- Eutrophication TVs 50% above pre-eutrophication condition (around 1900);
- Benthic assessment assesses deviation (defined by EU member states) from a reference condition with negligible impacts.

Figure 4 below depicts the assessment outcomes for BH2a Vegetation (first panel from the left), BH2a Invertebrates (second panel), EU1 Winter concentrations of dissolved inorganic nitrogen (DIN, third panel) and EU2 growing season chlorophyll-a concentrations (fourth panel). The highlighted areas in the Wadden Sea, around Denmark and in the Kattegat show assessment outcomes classified as moderate to poor/bad status, which are roughly in agreement. Since the main driver here is nutrient enrichment, measures taken to improve the status can be expected to benefit both BH2 and EU1 and 2 in these areas.

However, it is well-known that TVs used for WFD water coastal waters are not entirely coherent across countries and with neighbouring OSPAR COMP4 areas. This may cause conflicts, for instance when in neighbouring areas and in the case of nutrient inputs from land the more landward area is in good status, while the more offshore area is not in good status. Or when national boundaries determine the outcome of the assessment rather than ecological or pressure gradients. These types of issues have been the subject of WFD intercalibration efforts and projects such as the INTERREG Deutschland-Nederland project 'Wasserqualität – Waterkwaliteit'[1] and needs further consideration in future work.

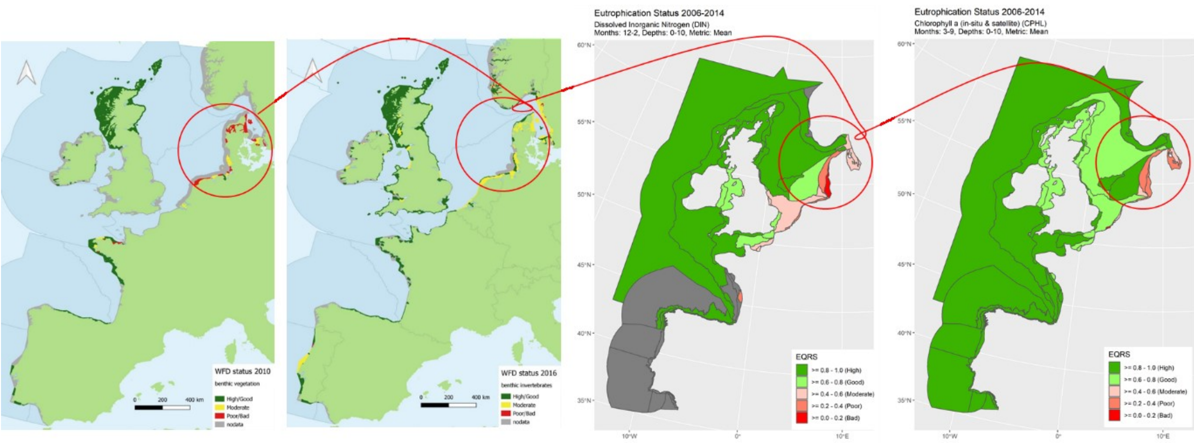

Figure 4. Comparison between assessment outcomes of BH2a Vegetation (1st panel), Benthic invertebrates (2nd panel), EU1 Winter concentrations of dissolved inorganic nitrogen (DIN, 3rd panel). Red circles highlight areas classified as moderate or poor/bad. In: BH2a, Winter nutrient concentrations and Chlorophyll-a concentrations indicator assessments.

Example 2: PH2 vs EU2 – different TV setting methods and assessment outcomes in the same areas

The second example compares the pelagic habitats indicator PH2 - *Changes in Phytoplankton Biomass and Zooplankton Abundance* with the eutrophication indicator EU2 - *Concentrations of chlorophyll a.* Phytoplankton biomass in PH2 is measured as concentration of chlorophyll-*a*, so essentially the same parameter is used in both assessments. The comparison is further facilitated by the use of the same (COMP4) assessment areas. The assessment periods largely overlap: 2015-2019 in PH2 and 2015-2020 in EU2.

The TV setting narratives however are incoherent:

- Eutrophication TVs are set at 50% above pre-eutrophication condition (around 1900);
- Pelagic Habitats compare significant change in annual mean abundance values from assessment period (2015-2019), to preceding years (1960-2014, depending on dataset).

[1] https://interregv.deutschland-nederland.eu/en/project/wasserqualitat-waterkwaliteit/

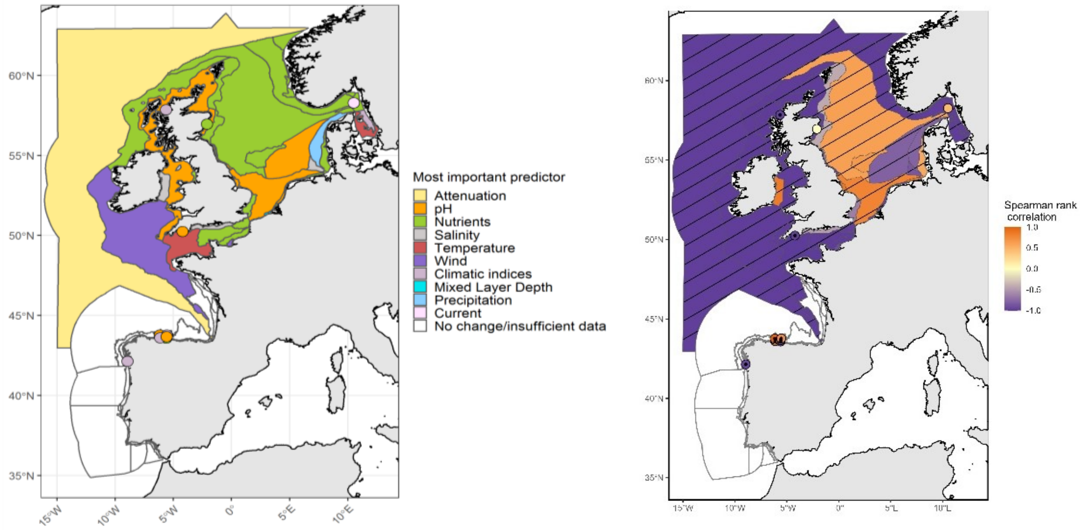

Figure 5. PH2 assessment outcomes for phytoplankton abundance. Left panel: most important variables addressing changes in phytoplankton abundance within the COMP4 assessment areas. Right panel: trend in phytoplankton biomass anomalies between the assessment period (2015–2019) and the comparison period (station data: 1992–2014; non-station data: 1997–2014). Hatched areas were characterised by significant changes (p≤0,05) in phytoplankton biomass between the comparison and the assessment periods. White areas indicate no data or insufficient data to assess the area. In: PH2 indicator assessment.

Whether this is problematic depends on the main driver for change, where the eutrophication assessment focuses solely on human-induced inputs of nutrients (via rivers, direct inputs and atmospheric deposition), while the PH2 indicator considers multiple drivers, including parameters connected to climate change.

Figure 5 left panel depicts the predominant pressure per assessment area for phytoplankton biomass. The colour green indicates nutrients as the main driver for change. The right panel shows the direction of change, which indicates significantly increasing chlorophyll-*a* concentrations, calculated from *monthly anomalies over the entire year*, in *eg.* the Elbe Plume, Northern North Sea and Dogger Bank (orange), and decreasing concentrations in the Channel areas (purple). In the PH2 assessment nutrient concentrations, notably the ration between nitrogen and phosphorous concentrations (N:P ratio) was considered not in balance and driving changes in chlorophyll-*a* concentrations, therefore assessment areas in the North Sea were classified as 'not good'.

The EU2 chlorophyll-*a* assessment classified fewer areas as 'not good', primarily located along the North Sea East coast (Figure 6). This relates to inputs of nutrients from land. 'Not good' is defined by the mean *growing season* chlorophyll-a concentration exceeding the area-specific TV. The chlorophyll-*a* indicator assessment also contains trend information, calculated as the linear trend in the entire period from 1990-2020. In the majority of areas no significant long-term trend was found, although in some areas displaying elevated chlorophyll-a concentrations in the beginning of the time series a downward trend was seen in the 1990's and early 2000's, owing to measures to reduce the nutrient inputs, which has levelled off in more recent years.

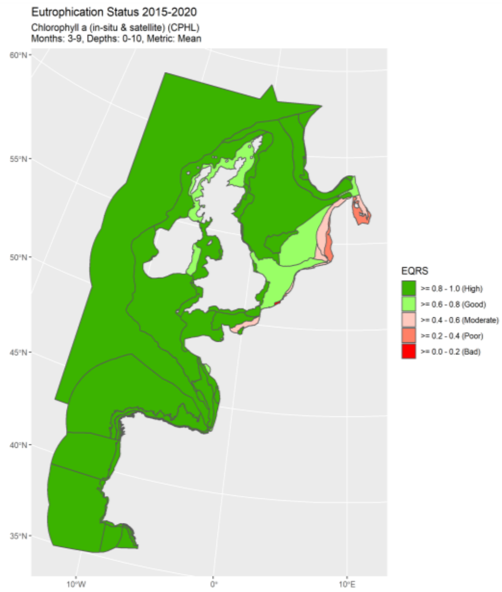

Figure 6. Results of the COMP4 assessment of chlorophyll growing season mean, based on the combination of in situ and satellite data, for OSPAR Regions II, III and IV. In: Chlorophyll-a indicator assessment.

Table 11 lists all assessment areas affected by nutrients (green coloured areas in Figure 5 left panel) with their assessment outcomes according to the PH2 assessments. The EU2 assessment outcomes are included as well for comparison. There is limited coherence between the two assessments; in only one assessment area (Coastal FR Channel) the status (not good) and the trend (downwards) are similar. In all of the other assessment areas in Table 11 differences occur in the status classification and/or the direction of the trend. In addition, the interpretation of these outcomes also differ. In the case of EU2 a downward trend in chlorophyll-*a* concentrations in an area classified as 'not good' is considered a positive signal, indicating that measures to reduce nutrient loads to the sea are being effective. In contrast, the PH2 assessment considers significant increases or decreases a cause for concern for the future of 'traditional' pelagic food webs, wherein blooms of phytoplankton feed copepods, which support commercially exploited fish.

Table 11. Assessment outcomes (status and trends) in the PH2 and EU2 assessments of chlorophyll-a in the areas identified as influenced by nutrients in the PH2 assessment (green coloured areas in Figure 6 left panel). Trend colours are the same as in Figure 6 right panel. The area names and codes are explained in both the PH2 and EU2 assessments.

Area code	Area name	PH2 status	PH2 trend	EU2 status	EU2 trend
CNOR1,2,3	Coastal NOR 1, 2, 3	Not good	↓	Good	↔
NT	Norwegian Trench	Not good	↓	Good	↔
NNS	Northern North Sea	Not good	↑	Good	↔
IS1	Intermittently Stratified 1	Not good	↓	Good	↔
IS2	Intermittently Stratified 2	Not good	↑	Good	↔
DB	Dogger Bank	Not good	↑	Good	↔
ELPM	Elbe Plume	Not good	↑	Not good	↔
THPM	Thames Plume	Not good	↓	Good	↔
CUKC	Coastal UK Channel	Not good	↓	Good	↔
CCTI	Channel Coastal shelf tidal influenced	Not good	↓	Good	↔
CFR	Coastal FR Channel	Not good	↓	Not good	↓

Part of these inconsistences in assessment outcomes are caused by methodological differences, for instance the use of the entire year in PH2 versus the Summer growing season in EU2. These differences have been investigated in Activity 1 and reflected in Chapter 5 of the D1.5b report '*Options for integration between PH diversity, biomass, primary production and eutrophication*' (NEA PANACEA Final Report, Annex L). This chapter includes a list of actions to improve the comparability between the two approaches.

From the perspective of improving coherence between threshold value setting methods the following actions to be taken by the pelagic habitats, foodweb and eutrophication experts in OSPAR are proposed:

- Further investigation of TV narratives and values, involving pelagic habitats, foodwebs and eutrophication experts in OSPAR;
- Investigation of potential issues with conflicting messages to policy makers and areas where these occur;
- In areas where a management decision needs to be taken in the short term: decide which indicator guides these management measures, using transparent criteria;
- Pragmatic approach:
    - continue implementation of reduction measures to improve the status in coastal eutrophication problem areas. Reduction of nutrient input takes time and change will be slow and limited, especially in more offshore areas.
    - develop an easy to understand explanation of the differences between the PH2 phytoplankton biomass and EU2 eutrophication chlorophyll-*a* assessment, that can be shared with the Hazardous Substances and Eutrophication Committee and the Biodiversity Committee in OSPAR.

4. Options to improve coherence

The main aim of Task 2.4 was to identify feasible options to improve coherence between threshold setting methods used in various indicator assessments. The two examples above, connecting eutrophication indicators with indicators for benthic and pelagic habitats respectively, show that understanding the level of (in)coherence really requires in-depth comparison, taking into account the underlying data, assessment methods and spatial and temporal patterns in the assessment outcomes.

The next – and most important – question is whether incoherence is a problem, for instance when assessment outcomes of two indicators lead to conflicting management measures. A simple and pragmatic decision tree was developed to identify situations where action is needed to improve the coherence (Figure 7), either for management or for scientific rigour of the assessment framework.

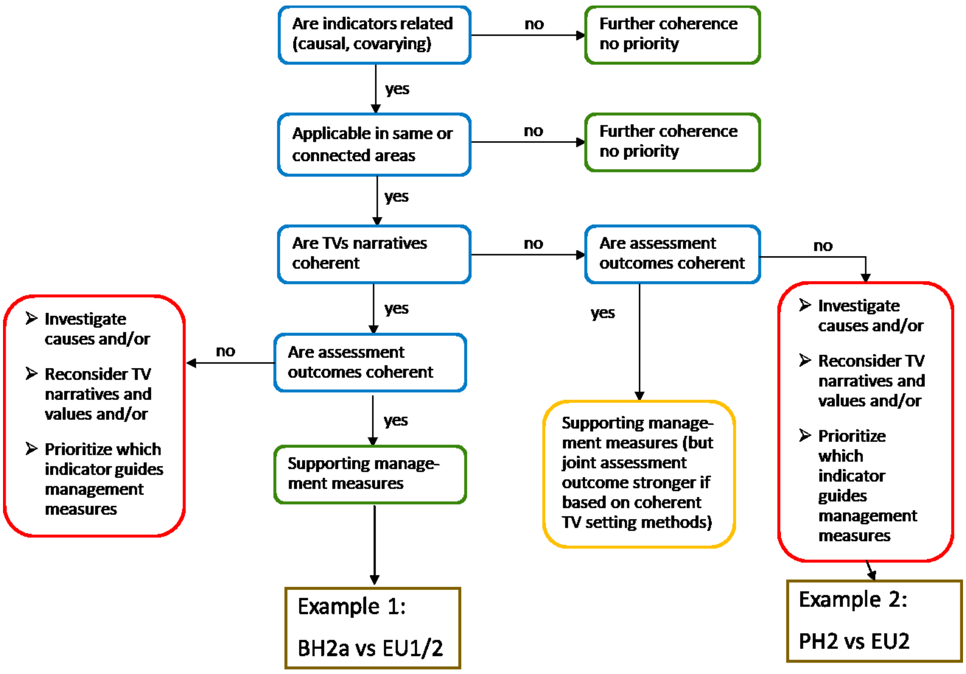

Figure 7. (In)coherence between TV setting narratives or assessment outcomes - problem or not? Successive questions (blue boxes) lead to either acceptable situations (green boxes) or need for further work to improve the coherence (red and orange boxes). The examples refer to section 3.2.3.

The decision tree helps to prioritize – in a pragmatic way – situations where incoherence occurs and whether it is important (or not!) to take action. Given the difficulties already encountered in (biodiversity) indicator assessments it is considered unrealistic to tackle all incoherencies, hence prioritisation is realistic. The red boxes in Figure 7 identify situations that call for priority action. Further action is also advised when the outcomes of assessments of related indicators, applicable in the same or connected areas are coherent, although based on different threshold value setting methods. Here, coherent assessment outcomes might be coincidental and therefore not a firm basis for management measures. This situation is indicated by the orange box.

If action is needed, suggested options are:

- Further investigation of TV narratives and values, involving relevant expert groups in OSPAR or other Regional Sea Conventions, ICES and JRC expert networks. This may lead to adjustments and improvements;
- Investigation of potential issues with conflicting messages to policy makers and areas where these occur;
- In areas where a management decision needs to be taken in the short term: decide which indicator guides these management measures, using transparent criteria;
- Pragmatic approach:
    - In case measures are already in place, continue these measures to improve the status of at least one of the indicators and where possible monitor the effects on related indicators;
    - Be transparent about the (potential) inconsistencies to policy makers in OSPAR, and – where applicable – MSFD CIS;
    - In the meantime develop a plan to address important incoherencies, involving relevant expert groups and policy makers (OSPAR Committees).

Initial discussions already occur in the OSPAR working groups on biodiversity (ICG-COBAM) and eutrophication (ICG-Eut). The NEA PANACEA project has fuelled these discussions and experts active in OSPAR will ensure that the issue of coherence between threshold value setting methods remain on their respective work programmes. We also advise that this type of cross-cutting work needs better collaboration between the expert groups mentioned above, after the NEA PANACEA project has ended. Furthermore, as a follow up of this project in combination with the MSFD Workshop Horizontal issues – threshold values, we will invite the D1, D4 and D5 expert networks and the Working Group on Good Environmental Status under the MSFD CIS process to continue this work and build on the findings in this Task 2.4 report.

5. References

McQuatters-Gollop, A., L. Guérin, N.L. Arroyo, A. Aubert *et al*. (2022) Assessing the state of marine biodiversity in the Northeast Atlantic. Ecological Indicators 141. https://doi.org/10.1016/j.ecolind.2022.109148

OSPAR (2022). Revision of the Common Procedure for the Identification of the Eutrophication Status of the OSPAR Maritime Area. OSPAR Agreement 2022-07.

OSPAR-HELCOM, 2019. Outcome of the OSPAR-HELCOM workshop to examine possibilities for developing indicators for incidental by-catch of birds and marine mammals. 2019 Copenhagen, Denmark. Available: https://portal.helcom.fi/meetings/Incidental%20bycatch%20WS%201-2019-647/MeetingDocuments/Outcome%20OSPAR-HELCOM%20incidental%20by-catch%20indicator%20workshop_final.pdf

Scheffer, M., S. Carpenter, J.A. Foley, C. Folke and B. Walker (2001) Catastrophic shifts in ecosystems. Nature 413, p. 591-596. https://doi.org/10.1038/35098000

Appendix 1. NEA PANACEA SuperCOBAM workshop (20-22 October 2021, Utrecht) pre-read documents

In advance of the workshop the following MSFD related documents were sent to participants in order to prepare the session on threshold value setting methods:

- **JRC review report on Species Threshold Values methods**
- **EU MSFD Horizontal Issues: Threshold Values workshop pre-read document**
  This document contained several types of narratives for setting TVs that were considered relevant for the type of assessments developed under NEA PANACEA and was used for a common language to discuss and compare baseline and threshold setting methods.
- **Report of the EU MSFD Horizontal Issues: Threshold Values workshop**
- **Tsiamis K, et al (2021), Marine Strategy Framework Directive- Descriptor 2, Non-Indigenous Species, Delivering solid recommendations for setting threshold values for non-indigenous species pressure on European seas**
- (Submitted to) BDC 2019: Setting regional threshold values for Non-Indigenous Species primary criteria: pros, cons and how it could be achieved.
  Note: This Word file is attached to this report and can be accessed through the attachments tab (paperclip symbol). This will not work when this report has been opened in your browser, a pdf reader is required.

Appendix 2. Inventory of threshold value setting methods used in biodiversity and eutrophication indicators

Excel matrix containing results of the inventory of threshold value setting methods in 29 indicators developed for OSPAR's Quality Status Report 2023 and MSFD Art 8 reporting.

Note: The Excel file is attached to this report and can be accessed through the attachments tab (paperclip symbol). This will not work when this report has been opened in your browser, a pdf reader is required.