D1.5b: Options for integration between PH diversity, biomass, primary production and eutrophication

Arnaud Louchart[1] – Birgit Heyden[2] – Matthew Holland[3] – Lisette Enserink[4] – Abigail McQuatters-Gollop[3] – Fabrice Lizon[1] – Luis Felipe Artigas[1]

[1]Laboratoire d'Océanologie et Geosciences, UMR 8187 LOG, Centre National de la Recherche Scientifique, Université du Littoral Côte d'Opale, Université de Lille, IRD, Wimereux, France

[2]AquaEcology GmbH & Co. KG, Oldenburg, Germany

[3]Marine Conservation Research Group, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, United Kingdom

[4]Ministry of Infrastructure and the Environment/Rijkswaterstaat, Zuiderwagenplein 2, 8224 AD Lelystad, The Netherlands

Acknowledgments:

This report was produced in the frame of the European Maritime and Fisheries Fund project NEA-PANACEA (North-East Atlantic project on biodiversity and eutrophication assessment integration and creation of effective measures). The project was funded by the European Union (EU). Grant No. 110661/2020/839628/SUB/ENV.C.2

List of abbreviations

BDC: Biodiversity Committee

COBAM: Coordination of Biodiversity Assessment and Monitoring

COMPEAT: Common Procedure Eutrophication Assessment Tool

EQR: Ecological Quality Ratio

EcApRHA project: Applying an Ecosystem Approach to (sub) Regional Habitat Assessment

FW: Food Web

GES: Good Environmental Status

HASEC: Hazardous Substances and Eutrophication Committee

IQR: Interquartile Range

LCBD: Local Contribution to Beta Diversity

MSFD: Marine Strategy Framework Directive

OO-AO: One Out – All Out

PH: Pelagic Habitats

RBINS: Royal Belgian Institute for Natural Science

WFD: Water Framework Directive

Table of content

1. Background

The goal of the Marine Strategy Framework Directive (MSFD; 2008/56/EC) was to achieve the Good Environmental Status (GES) of the European marine waters through 11 Descriptors. The most recent assessment of GES in the North-East Atlantic was carried out using a suite of indicators through their single metric approach, however, in the field of environmental policy management the current best practice involves employing the ecosystem approach, to better inform Ecological Status (Magliozzi et al. 2023). This ecosystem approach can be carried out through synthetising several linked indicators into combined multimetric indices which provide additional meaning. Furthermore, for OSPAR there is a need for better integration within and across ecosystem components, as the establishment of quality status for OSPAR relies on a large variety of indicators to assess MSFD criteria across multiple descriptors.

The ecosystem approach can be carried out by cross-linking relevant indicators within and across MSFD Descriptors. This approach is fundamental to the MSFD because it provides a holistic view of the current environmental quality status. For the descriptor D1 Biological Diversity, previous work carried out for the EcApRHA project (Applying an Ecosystem Approach to (sub) Regional Habitat Assessment) established methods for integrating results from the Benthic Habitat indicators and the Food Webs FW4 'Mean Trophic Level' indicator (Eliott et al. 2017). More recently, a working group dedicated to the integration proposed different methods for integrating across D1C1 to D1C5 Marine Strategy Framework directive biodiversity assessments (Dierschke et al. 2021). However, the D1C6 assessment (Pelagic Habitats) remains to be integrated with other relevant indicators. Until now, the ecosystem approach for D1C6 has strictly focused on the integration of biological quality elements (plankton diversity, phytoplankton biomass, plankton lifeforms abundance and total zooplankton abundance) within Pelagic Habitats (Budria et al. 2017) and applied to the OSPAR convention (Holland et al. 2023). The integration of the Pelagic Habitats indicators assumes an understanding of the factors defining and mechanisms affecting plankton communities at multiple levels (species, lifeforms, biomass), including mechanisms which generate direct top-down and bottom-up effects on the plankton community.

The typical view of the pelagic food web places phytoplankton at the base. Phytoplankton absorb nutrients from the marine environment for growth and reproduction, and they provide the food source which supports zooplankton. Phytoplankton primarily require macro-nutrients, such as carbon, nitrogen, and phosphate, to produce organic matter through photosynthesis. Additional nutrients such as silicate are also required by some groups (i.e., diatoms).

These nutrients are part of a natural cycle. Within coastal and estuarine systems, nutrient inputs originate predominantly from river runoff and atmospheric deposition, but also recycling of organic matter and transboundary nutrient transport can play a role in supplying nutrients to pelagic habitats. In temperate regions such as the North-East Atlantic, the annual maximum concentration of nutrients is usually reached at the end of winter. With their large standing stocks, phytoplankton respond quickly, and rapid growth occurs as the day-light period increases. This period of high productivity and fast growing occurs in early spring and is usually marked by the blooming of one generalist species, or of a group of generalist species. After the spring bloom, by late spring and early summer, nutrients are left depleted. This nutrient depleted environment facilitates the expansion of specialist species. During the spring bloom, the intense biological activity results in dying and sinking organisms. Thus, spring bloom promotes recycling of nutrient through the remineralisation of organic matter. Once nutrients become biologically available again, a second phytoplankton bloom principally composed of diatoms can occur in late-summer or in autumn when environmental conditions still allow phytoplankton growing (e.g. high light regime, mixing and declining thermal stratification).

The concentration of nutrients also impacts other aspects of phytoplankton. It has been noted that nutrients usually have contrasting effects on phytoplankton dominance and richness (e.g., Facca et al. 2014). Usually, at low nutrient concentrations, species dominance and richness are both low. With increasing nutrient concentrations, dominance of generalist species increases to reach an asymptotic maximum at high nutrient concentrations. However, intermediate nutrient concentrations promote species coexistence and maximum richness is typically achieved at this level.

Beyond the natural cycle of nutrients, human activities such as agriculture, aquaculture, and waste-water treatment can modify the stoichiometry or balance of nutrients in the marine environment, potentially disrupting the phytoplankton community and their productivity. Artificially elevated phytoplankton productivity because of human-induced nutrient concentrations is referred to as eutrophication.

This report provides elements of integration for Pelagic Habitats with relevant indicators. The Pelagic Habitat diversity and biomass indicators (PH3 and PH2, respectively) from MSFD Descriptor 1 are linked to primary production indicators (FW2) and to some eutrophication indicators, two relevant indicators of D5 (Eutrophication) and D4 (Food Web), all MSFD descriptors directly connected to the Pelagic Habitat indicators. Integration across MSFD Descriptors remains challenging for several reasons. First, it is crucial to understand exactly the relation and the mechanisms acting between the quality elements. This aspect is currently studied in the section *Synergy between Pelagic Habitat Indicators* of the D1.5c of NEA-PANACEA project (Louchart et al. 2023) in which we investigate

relationships between abundance, productivity, and diversity. These indicators may not be directly comparable, as each Descriptor has its own methodology to assess quality status. Nevertheless, here we present two options for integrating the Pelagic Habitat diversity indicator (PH3), the phytoplankton biomass indicator (PH2), the primary production indicator (FW2), and the concentration of chlorophyll-*a* indicator (HASEC) from the simplest to the most elaborate approach to assess the quality status of the environment. In a second section, we investigate a comparison between the phytoplankton biomass indicator (PH2) and concentration of chlorophyll-*a* indicator (HASEC) as discrepancies between the results of these two indicators have been reported.

This deliverable builds on former work from the Water Framework Directive, is strongly supported by the Marine Strategy Framework Directive and the OSPAR commission and represents the continuation of the integration steps developed during the EcApRHA project (Budria et al. 2017; Eliott et al. 2017).

2. Relationships between PH2, PH3, FW2 indicators and winter nutrient concentration indicators

2.1. Description of the indicators

PH2 indicator (BDC)

PH2 is a state indicator. This indicator is based on identification of phytoplankton biomass and zooplankton abundance trends within plankton time-series. Anomalies represent deviations from the assumed natural variability of a time-series. Thus, the greater the magnitude of the anomaly (in terms of absolute value, since anomalies can be positive or negative), the greater the change. An anomaly value of zero indicates no difference from the time-series mean (which must be de-seasonalised). To understand the changes presented (i.e., annual anomalies) and to be most useful for decision makers, annual anomalies are best interpreted with information provided by anomalies on monthly timescales.

Once the data are at a monthly timescale, the time-series analysis can be run. The analysis uses an R-script for both discrete-station data and non-station data, after the pre-analysis steps have been followed. The first step consists in identifying the mean seasonal cycle (which is called seasonality in this assessment) during the whole study period. Removing the seasonality is required to analyse the variations of each plankton compartment (i.e., phytoplankton biomass or zooplankton abundance) beyond their natural cycle. The second step consists in obtaining anomalies by subtracting this seasonality from the original time series. The method used is the seasonal differentiation by the seasonal deviation. Finally, the cumulative sum of these anomalies was produced to detect regime shifts in the time-series for the assessment and comparison periods. A Spearman rank correlation test

is now implemented to test the trend of the cumulative sum of the anomalies of the assessment and comparison periods. The correlation can move towards a significant ($p \leq 0.05$) increase in phytoplankton biomass/zooplankton abundance (0 to 1), no changes (=0) or decrease in phytoplankton biomass/zooplankton abundance (-1 to 0). The results of the Spearman rank correlation provide an indication of changes. A t-test against the cumulative sum of the anomalies of the comparison period and the assessment period informs whether the trends are significantly different or not.

For this indicator, no thresholds were available. For the QSR2023, the attribution of quality status was determined by the application of the One-Out-All-Out principle after linking pressures to the indicator. Further information can be found in the OSPAR PH2 indicator assessment (Louchart et al. 2022a).

PH3 indicator (BDC)

PH3 is a state indicator. PH3 is a complex multi-metric indicator which describes plankton diversity. The method incorporates both α-diversity (i.e., the diversity within a site or sample) and β-diversity, which focuses on the rate of change, or turnover, in species composition (Rombouts et al., 2019). For the QSR2023, we used the α- and β-diversity as consecutive steps to detect the temporal changes in community composition (through the β-diversity) and subsequently to report the state of the community whenever changes were observed (through the α- diversity). First, the β-diversity was computed and significant deviation from the overall composition was flagged. More specifically, the Local Contribution to Beta Diversity (LCBD) shows how much each observation in a time-series contributes to β-diversity; for example, a site with an average species composition would have an LCBD value of 0. Large LCBD values may indicate sampling units (in time) characterised by high conservation value or degraded and species-poor sites in need of restoration (Legendre and De Cáceres, 2013). High values may also correspond to special ecological conditions or may result from the disturbance effect of invasive species (i.e., differing from normal conditions in a positive or a negative way). When significant community composition was detected, the α-diversity was investigated to observe whether the richness and/or the dominance were responsible. Assessment of richness was processed by the Menhinick index. The dominance of phytoplankton was assessed by the Hulburt index while the dominance of zooplankton was assessed by the Patten index. Further explanations on the choice of indices can be found in Louchart et al. (2022b).

For this indicator, no thresholds were available. For the QSR2023, the attribution of quality status was determined through the application of a normalised EQR across assessment units and by the One-Out-All-Out principle (see paragraph 2.2 Methodology) on the normalised EQR-pressures relationship (see NEA-PANACEA D1.4a for detailed methodology). This indicator can also be calculated based on a multi-

metric approach and an attempt to average the results. Further information can be found in the OSPAR PH3 indicator assessment (Louchart et al. 2022b).

FW2 indicator (BDC)

FW2 is a state indicator. This indicator is derived through time-series analysis, based on identification of trends in primary production anomalies. As primary production is closely related to phytoplankton biomass, we employed the same methodology as for the PH2 indicator 'Changes in phytoplankton biomass/zooplankton abundance'. Anomalies in this case represent deviations from the assumed natural variability within a time-series. Thus, the greater the magnitude of an anomaly (in terms of absolute value, since anomalies can be positive or negative), the greater the change. An anomaly value of zero indicates no difference from the time-series mean trend (which must first be de-seasonalised). To understand the changes presented (i.e., annual anomalies) and to be most useful for decision makers, the annual anomalies must be considered using details given by the monthly anomalies (since an early warning indicator should be assessed at the highest temporal resolution available).

For this indicator, no thresholds were available. For the QSR2023, the attribution of quality status was determined through the application of the One-Out-All-Out principle (see paragraph 2.2 Methodology). Further information can be found in the OSPAR FW2 indicator assessment (Louchart et al. 2022c).

Winter nutrient concentrations in OSPAR Regions II, III and IV (HASEC)

The *Winter nutrient concentrations* indicator is a pressure indicator. This indicator is currently implemented together with other eutrophication indicators in the OSPAR "Common Procedure Eutrophication Assessment Tool" (COMPEAT) developed and maintained by ICES to produce a coherent assessment of eutrophication across the OSPAR regions. Winter nutrient concentrations are at the core of the OSPAR strategy to tackle eutrophication through limiting inputs of nutrients and organic matter to levels that do not give rise to adverse effects on the marine environment (Heyden and Leujak, 2023). The indicator focusses on time-series analysis of Dissolved Inorganic Nitrogen (DIN) and Dissolved Inorganic Phosphorous (DIP) concentrations. The temporal changes are tested for significance using the Mann-Kendall trend test function from the R-package *TTAinterfaceTrendAnalysis* (Devreker & Lefebvre, 2014). Further information can be found in OSPAR Winter Nutrient Concentrations in the Greater North Sea, Celtic Seas and Bay of Biscay and Iberian Coast (Heyden and Leujak. 2023).

Concentration of chlorophyll-a in the Greater North Sea, Celtic Seas and Bay of Biscay and Iberian Coast (HASEC)

This indicator for the concentration of chlorophyll-*a* is a proxy for phytoplankton biomass. To produce a coherent assessment of eutrophication in the OSPAR regions, this indicator is integrated amongst

other eutrophication indicators in the COMPEAT tool. The indicator focusses on the temporal analysis of chlorophyll-*a* concentration by assessment unit and subsequently grouped habitat types within OSPAR Regions II (Greater North Sea), III (Celtic Seas) and IV (Bay of Biscay and Iberian Coast). Temporal changes are tested for significance using the Mann-Kendall trend test function from the R package *TTAinterfaceTrendAnalysis* (Devreker & Lefebvre, 2014). Unlike the PH2 indicator, no transformation of data or extraction of chlorophyll-*a* values beyond their natural cycle were necessary for this indicator analyses. Further information can be found in OSPAR Concentrations of Chlorophyll-a in the Greater North Sea, Celtic Seas and Bay of Biscay and Iberian Coast (Prins and Enserink. 2023). For practical reasons, we will refer to this indicator as "eutrophication indicator" in the rest of this document.

2.2. Conceptual view of the integration

To date, the indicators of the Food Web and Pelagic Habitats have been assessed on an individual basis, while the Dissolved Inorganic Nitrogen, Dissolved Inorganic Phosphate and eutrophication indicator are assessed both individually, and as part of the overall integrated eutrophication assessment result. The first step of the integration process across different MSFD descriptors is to understand the connections between the indicators (**Figure 1**). "Winter nutrient concentrations in OSPAR Regions II, III and IV" affect phytoplankton communities in terms of their diversity (PH3 Changes in phytoplankton diversity), biomass (PH2 Changes in phytoplankton biomass) and primary production (FW2 Changes in primary production). These biological elements themselves also directly affect the "winter nutrient concentrations in OSPAR Regions II, III and IV". Furthermore, phytoplankton diversity directly affects phytoplankton biomass and primary production. In turn, primary production generates phytoplankton biomass and *vice versa*. It is important to note that PH3 and PH2 indicators also assess zooplankton, which is itself not directly linked to the "winter concentrations of nutrients". In fact, zooplankton diversity (PH3) and total abundance (PH2) impact and are impacted by phytoplankton diversity (PH3), phytoplankton biomass (PH2), and primary production (FW2). Finally, the eutrophication indicator' indicator and the PH2 'changes in phytoplankton biomass and zooplankton abundance' are closely interrelated since they each investigate the concentration of chlorophyll-*a* in an analogous manner.

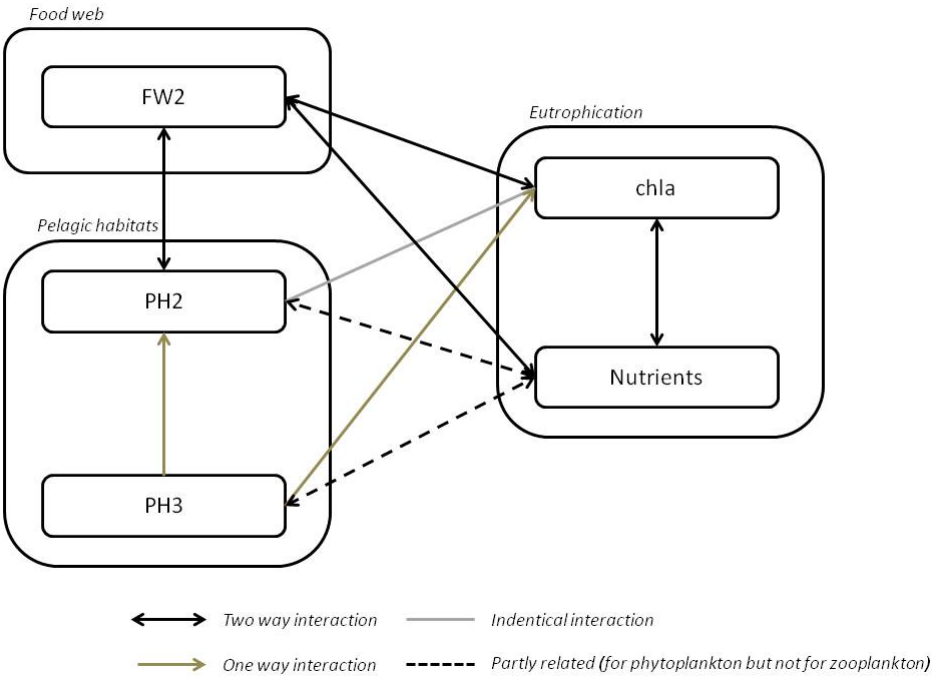

Figure 1: Conceptual view of the interactions between Eutrophication, Pelagic habitats and Food web indicators.

The assessment flow for the descriptor D1 in Magliozzi et al. (2021) provides an interesting view of the requirements for integrating within and across Pelagic Habitats. We reason that this view can be extended for an integration of Pelagic Habitats and Eutrophication and we propose an extended flow for integration across MSFD descriptors.

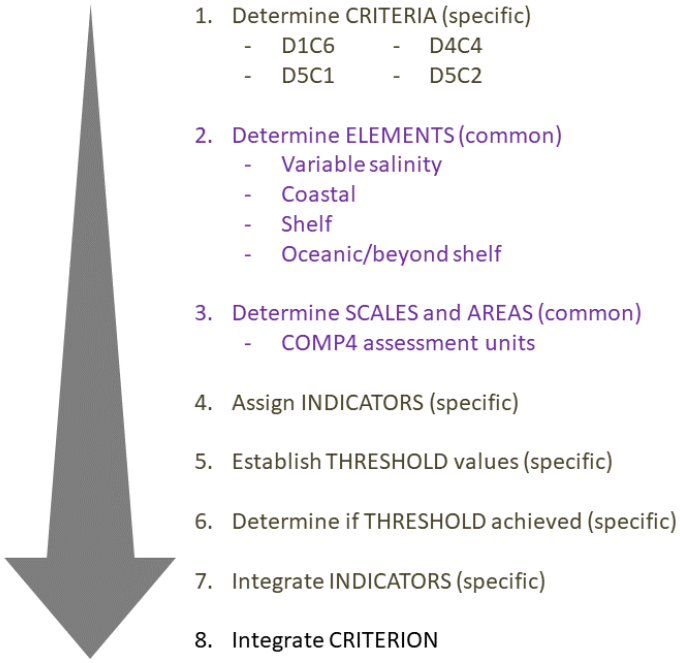

Figure 2: Integration flow between D1C6 (Pelagic Habitats), D5C1 (Dissolved Inorganic Nitrogen and Dissolved Inorganic Phosphate or Concentration of winter concentration of nutrients), D5C2 (concentration of chlorophyll-a in the Greater North Sea, the Celtic Sea and the Bay of Biscay and Iberian Coast or eutrophication indicator) and D4C4 (Productivity of trophic guild). Adapted from Magliozzi et al. (2021).

The first level of consideration for a good integration across MSFD descriptors starts by using the same 'Elements' and 'Scales and areas' (**Figure 2**). According to Magliozzi et al. (2021), the term 'Elements' refers to the 'essential characteristics of the criterion to be evaluated'. For the D1C6 and D5C1, the 'Elements' are common and defined as the four types of habitats, i.e. variable salinity, coastal, shelf, and oceanic/beyond shelf. The term 'Scales and Areas' refers to the 'subdivision of the region or subregion to assess'. For the D1C6 and D5C1, this step is achieved by the OSPAR common procedure (COMP), which aims to define the OSPAR maritime area by a set of ecologically and physically distinct areas or spatial units. This method has been applied by the OSPAR HASEC working group since QSR 2000 for assessing descriptor D5, and since QSR 2023 for D1C6 and D4 by the OSPAR BDC working group. For the current version of COMP (COMP4), the parameters used to differentiate these areas include physical, chemical, and biological factors: depth, salinity, stratification, suspended particulate matter and primary production. Consult Enserink et al. (2019) and OSPAR Agreement 2022-07e for a more detailed description of the procedure used to define assessment units.

After the 'Elements' and the 'Scales and areas' have been defined, integration steps 4 to 7 should be specific to each criterion being assessed. The second step for integration concerns the integration across criterion, as described in subsequent sections of this document.

Two methods have been retained for the integration. The first option is the One Out-All Out approach (OO-AO) and the second option is the weighted Normalised Ecological Quality Ratio (EQRS) approach. Briefly, the OO-AO as used in the Water Framework Directive (WFD), the overall ecological status of a water body is based on the worst status determined by any of its biological quality elements (Heiskanen et al. 2004). The EQRS quantifies the deviation in indicator values between an assessment period and a baseline defined as "reference conditions"

2.3. Data preparation

The results of options 1 (OO-AO approach) and 2 (Normalised EQR approach) rely on existing OSPAR methodologies for indicators and does not need additional preparation. Prior to the analysis for Option 2, we harmonise the timescale across the indicators. Since the PH3 plankton diversity indicator is calculated at the annual scale, we have decided to use all the indicators at this scale. For PH3, we use the annual EQR of the LCBD. For the PH2, the monthly values were aggregated into the interquartile range (IQR) for each year. For the FW2, anomalies are obtained at monthly scale for the indicator but here the annual primary production is derived as the sum of monthly primary production values per year. The eutrophication indicator and the winter nutrient concentration indicators are assessed in the eutrophication assessment for specific seasons on annual level (December to February for winter nutrients and March to September for the growing season).

2.3.1. Option 1: one-out-all-out approach (OO-AO approach)

The One Out-All Out (OO-AO) approach can be considered only when two or more biological quality elements are available. For example, if phytoplankton biomass is in "not good" status (based on three categories "not good", "unknown" and "good") and primary production is in "good" status, the overall quality status is classified as "not good". This approach has three main advantages. It is the simplest way of integrating indicator results which avoids the need to perform calculations. This method supports using different methodologies to assess the separate biological quality elements. Furthermore, this approach avoids calculating averages of multiple indicator metrics. However, the OO-AO approach has been reported to overly downgrade or misclassify quality status in certain cases (Borja et al. 2010), although it is generally the most precautionary approach.

For the integration of Pelagic Habitats and Food Web, the OO-AO approach is supported by a categorisation of quality status (**Table 1**) recently documented by McQuatters-Gollop et al. (2022). This categorisation is based on detecting statistically significant changes in indicator values according

to assessment threshold and / or the influence of anthropogenic pressures and climate change on the indicator's variability. Quality status can only be in one of four categories: "Not good", "Unknown", "Good", or "Not assessed". The OO-AO approach has recently been applied and validated for defining the quality status of Pelagic Habitats to support the OSPAR commission's Quality Status Report 2023 (Holland et al. 2023).

Table 1: Categorisation of quality status and associated narratives for biodiversity and food webs indicators.

Quality status categories	
Not good	Indicator value is below assessment threshold, or change in indicator represents a declining state, or indicator change is linked to increasing impact of anthropogenic pressures (including climate change), or indicator shows no change but state is considered unsatisfactory
Unknown	No assessment threshold and/or unclear if change represents declining or improving state, or indicator shows no change but unknown 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
Not assessed	Indicator was not assessed in a region due to lack of data, lack of expert resource, or lack of policy support.

In the eutrophication working group, quality status is categorised into six classes according to the normalised Ecological Quality Ratio (EQRS; **Table 2**): "bad" (EQRS below 0.2), "poor" (EQRS comprised between 0.2 and 0.4), "moderate" (EQRS comprise between 0.4 and 0.6), "good" (EQRS between 0.6 and 0.8), "high" (EQRS above 0.8) and "not assessed". Despite these two methods using different approaches in terms of vocabulary, number of classes, and assignation, it is possible to integrate them since determination of the final integrated quality status is based on the poorest quality status of both methods.

Table 2: Categorisation of the quality status and their associated narratives for the Eutrophication indicators.

Quality status categories	
Bad	Ecological Quality Ratio is below 0.2
Poor	Ecological Quality Ratio is equal or higher than 0.2 but lower than 0.4
Moderate	Ecological Quality Ratio is equal or higher than 0.4 but lower than 0.6
Good	Ecological Quality Ratio is equal or higher than 0.6 but lower than 0.8
High	Ecological Quality Ratio is equal or higher than 0.8
Not assessed	Indicator was not assessed in a region due to lack of data, lack of expert resource, or lack of policy support.

2.3.2. Option 2: Weighted Averaging Ecological Quality Ratio approach

The second option for integrating pelagic habitats and eutrophication involves applying the ecological quality ratio. The purpose of this approach is to quantify the deviation in indicator values between an assessment period and a baseline defined as "reference conditions". This approach implies knowledge of historical indicator values at the scale of the assessment. These baseline conditions are currently defined for the Eutrophication indicators based on modelled historic scenarios around 1900, while no such reference conditions currently exist for the Pelagic Habitats indicators, due to a lack of historical data to represent negligible impacts from human pressures within the COMP4 assessment units.

There are three possibilities to establish the baseline for the EQR. The first option consists in defining a reference value according to historical time-series (prior to 1930). For this purpose, palaeoecological analysis of long-term time series of historical data can be useful. However, historical phytoplankton records are scarce and often limited to nearshore stations. It is nevertheless possible to determine reference conditions of chl-*a* concentrations using hindcast model predictions for 1900, as is currently also done by HELCOM. The second option is based on literature review which can also help identifying reference conditions if historical data are not available. This second approach requires a large review of historical publications which are not widely applicable to the scale of OSPAR regions. Finally, the third option involves establishing thresholds or reference values based on existing data. With this approach, the reference value is obtained through pooling and averaging the data prior to the assessment period (prior to 2015 in the current OSPAR assessment cycle) more in the sense of a comparison period than actual reference conditions.

Due to the scarcity of historical data required for the first two methods and strong modelling expertise required to conduct historical predictions, the third option is the most appropriate one to evaluate in the current report. We explored this method for phytoplankton biomass and zooplankton abundance (PH2 indicator) as well as primary production (FW2 indicator), since it has already been developed and is currently applied for the PH3 indicator assessment to support the current OSPAR Quality Status Report. Further information can be found in the methodology description for the plankton diversity indicator assessment <PH3 changes in plankton diversity assessment>. For the winter nutrient concentration indicator, historic conditions have been modelled by the OSPAR ICG-EMO group for the Greater North Sea and Celtic Sea and an acceptable deviation of +50% has been used to define area-specific thresholds based on the reference values. The agreed thresholds were applied in the current Quality Status Report. Once thresholds for each biological quality element have been established, values observed during the assessment period can be compared to thresholds to determine the EQR. Integration across descriptors is then conducted by averaging the multiple EQR values using weighted

indicators. Weighting the indicators (in %) is not a mandatory step, and can often be avoided entirely, however, this step tends to overemphasise certain indicators at the expense of others to better reflect the ecosystem approach. Here, we integrate two pressure indicators ('*Winter concentration of nutrient*' and '*eutrophication indicator'*) with three state indicators (PH3 '*Changes in plankton diversity*', PH2 '*changes in phytoplankton biomass and zooplankton abundance*' and FW2 '*Primary production*'). All these indicators are connected, but some have a greater ecological weight than others. For this reason, we established three weighting classes depending on the importance of their contributions to ecosystem function. For example, the nutrient indicator is known to impact diversity, productivity and plankton biomass. Consequently, this indicator is considered as the basis of our approach, but as it only provides information on pressure on the ecosystem, it is given the lowest weight. The other indicators received higher weights. The first class of weight is focused on the '*winter concentration of nutrient*' indicator, which should receive the smallest weighting. The second class concerns the primary production indicator, which should be given moderate weighting. Finally, the most important weighting is given to the third level indicators (PH2, PH3, and the eutrophication indicator). According to these three classes of weighting, we propose the following equation to produce a combined EQR from the integration:

(1) $$EQR_{integrated} = {}^{1}/_{12}\, EQR_{nutrient} + {}^{1}/_{6}\, EQR_{FW2} + {}^{1}/_{4}\, EQR_{CHLA} + {}^{1}/_{4}\, EQR_{PH2} + {}^{1}/_{4}\, EQR_{PH3}$$

Where $EQR_{nutrient}$ is the EQR of the *Winter concentration of nutrient* indicator; EQR_{FW2} is the EQR of the *Primary production* indicator; EQR_{CHLA} is the EQR of the *eutrophication* indicator; EQR_{PH2} is the EQR of the *Changes in phytoplankton biomass and zooplankton abundance* indicator; and EQR_{PH3} is the EQR of the *Changes in plankton diversity indicator*. It is also important to note that phytoplankton and zooplankton both contribute equally for the PH2 and PH3 indicators. When an EQR was not computed for a particular indicator, ratios were reallocated among the three remaining weighting classes described above.

EQR values are bounded between 0 ("not good" or "bad" environmental status) and 1 ("good" or "high" environmental status) for eutrophication and 0 ("far from reference conditions") to 1 ("close to reference conditions") for the Pelagic Habitats and Food web indicators. An R-script is currently in development to establish an EQR for the PH2 and FW2 indicator. This approach follows a similar methodology to the calculation of the EQR for plankton diversity. We use the comparison period (prior to the assessment period) to establish annual mean values of chlorophyll-*a* biomass or zooplankton abundance per assessment unit or fixed station. We present here a case study for the "Channel Well Mixed" COMP4 assessment unit, and extend case study results to OSPAR Regions II, III, and IV. The

selection of assessment unit was arbitrary, although this happens to be an area with high spatial density and temporal frequency of sample collection.

3. Testing of the different options

3.1. Option 1: one-out-all-out approach

In this approach, PH2, PH3, and FW2 were mainly in "not good" status due to changes in plankton communities which were likely linked to pressures with anthropogenic origins (i.e., changes in temperature, pH, nutrient concentration). In this exercise to test different options for combining assessment results of indicators used under different Descriptors (D1, D4 and D5), the eutrophication indicators on chlorophyll-a concentrations and winter nutrient concentrations have been used individually and therefore deviate from the integrated eutrophication assessment results per habitat and OSPAR Region as included in the Eutrophication Thematic Assessment. For the eutrophication indicator, quality status was mostly "high" (OSPAR Region III and IV and Shelf habitat of Region II), except in Plume habitats of Region II which received a "moderate" quality status, and Coastal habitats of OSPAR Region II which received "good" quality status. Dissolved Inorganic Phosphate received "high" quality status in all habitats of the three Regions except in Coastal habitat of Region II where it received a "good" status. Dissolved Inorganic Nitrogen received "poor" status in Plume habitat of Region III, "moderate" status in Plume habitat of Region II, "good" status in Coastal habitats of Regions II and "high" status in the remaining habitats. Dissolved Inorganic Nitrogen and Dissolved Inorganic Phosphate are combined under the wording *winter concentration of nutrient*. Shelf habitats of all three regions received "high" quality status along with Coastal habitats in OSPAR Regions III and IV, and the Oceanic habitats in OSPAR Region IV. Plume habitats of OSPAR Regions II and III and Coastal habitats of Region II received "good" quality status. One tenth of the habitats assessed within the three regions had a "high" or "good" quality status. The Plume habitat of OSPAR region IV was only assessed with the eutrophication indicator. For the remaining habitats, at least one indicator received a "not good" quality status, contributing to a subsequent "not good" integrated quality status for the remaining habitat types within the three OSPAR Regions.

Table 3: Categorisation of the quality status and their associated narratives. Color coding for PH2, PH3, FW2 and Integrated quality status comes from table 1. Color coding for Concentration of chlorophyll-a eutrophication indicator and DIN and DIP (as part of the Winter concentration of nutrient) comes from table 2.

		OSPAR indicator						OO-AO	OO-AO
OSPAR Region	Habitat	PH2	PH3	FW2	eutrophication indicator	Winter concentration of nutrient		Integrated message (habitat)	Integrated message (region)
						DIP	DIN		
RII	Plume								
	Coastal								

	Shelf								
	Oceanic								
RIII	Plume								
	Coastal								
	Shelf								
	Oceanic								
RIV	Plume								
	Coastal								
	Shelf								
	Oceanic								

3.2 Option 2: Averaging Weighted Ecological Quality Ratio approach

In the test area ('Channel Well Mixed'), the reference values calculated are 0.909 (µg L^{-1}), 471 (ind. m^{-3}), 0.03, 0.02 and 0.876 (mg C m^{-2}) for phytoplankton biomass, zooplankton abundance, phytoplankton diversity, zooplankton diversity and primary production, respectively (**Table 4**). In order to create equity among indicators required for the integration, multi-metric indicators are divided by the number of biological elements they describe. For the PH2 indicator, the EQR obtained for phytoplankton biomass ranges between 0.86 and 0.96, representing conditions close to the reference conditions, whereas the EQR calculated for zooplankton abundance ranges between 0.53 in 2015 to 0.84 in 2018 which are close to the reference conditions (except for 2015 which is an "intermediate" year). For the PH3 indicator, values ranged from 0.10 in 2015 and 2016 to 0.47 in 2018 for phytoplankton diversity, and from 0.03 in 2015 to 0.17 in 2019 for zooplankton diversity. In the PH3, the values are far from the reference conditions. For FW2, values ranged from 0.73 to 0.99, being close to the reference conditions. The EQR for Dissolved Inorganic Nitrogen indicator was 0.92 in 2015 and 1 in 2017 while the EQR for Dissolved Inorganic Phosphate was 1 in 2015 and 1 in 2017, representing high environmental status. No values were obtained in 2016, 2018, or 2019 for both indicators. EQR values for the eutrophication indicator ranged from 0.90 (in 2015) to 1 (in 2016, 2017, 2018 and 2019). Regarding the values computed for each indicator and each biological element, the annual weighted average EQR ranged from 0.61 in 2017 to 0.90 in 2015. The year 2015 could be characterised as having "high" quality status, while the 2016 to 2019 period could be characterised as "good" quality status. The status of the entire assessment period could be characterised as "good" as the weighted average EQR was 0.72.

Table 4: Example of annual reference values and EQR obtained for Phytoplankton biomass, zooplankton abundance, Phytoplankton and zooplankton diversity, Primary Production and winter nutrient concentration in the 'Channel Well Mixed' COMP4 assessment unit. The EQR computed for the integration corresponds to the $EQR_{integrated}$ as defined by Equation (1).

Note that Dissolved Inorganic Nitrogen (DIN) and Dissolved Inorganic Phosphate (DIP) formed the 'winter nutrient concentration' indicator. We use the average between DIN and DIP as the value to integrate in average weighted EQR.

Year	Phytoplankton biomass (PH2)		Zooplankton abundance (PH2)		Phytoplankton diversity (PH3)		Zooplankton diversity (PH3)		Primary Production (FW2)		Eutrophication indicator		DIN		DIP		Integration (year)	Integration (assessment period)
	Ref value	EQR	Ref value	EQR	Ref value	EQR	Ref value	EQR	Ref value	EQR	Ref value	EQR	Ref value	EQR	Ref value	EQR	$EQR_{integrated}$	$EQR_{integrated}$
2015	0.909	0.96	471	0.53	0.003	0.10	0.002	0.03	0.768	0.99	1.0	0.90	1.0	0.92	1.0	1	0.90	0.72
2016	0.909	0.86	471	0.71	0.003	0.10	0.002	0.09	0.768	0.73	1.0	1	1.0	-	1.0	-	0.65	
2017	0.909	-	471	0.62	0.003	0.11	0.002	0.07	0.768	-	1.0	1	1.0	1	1.0	1	0.61	
2018	0.909	-	471	0.84	0.003	0.47	0.002	0.04	0.768	-	1.0	1	1.0	-	1.0 -	-	0.76	
2019	0.909	-	471	0.75	0.003	0.34	0.002	0.17	0.768	-	1.0	1.0	1.0	-	1.0 -	-	0.67	

The EQR methodology was then extended to the four pelagic habitat types (plume or variable salinity, coastal, shelf and oceanic/beyond shelf habitat types) within OSPAR Regions II, III and IV (**Table 5**). The EQR values for PH2 ranged between 0.45 for Shelf habitats in Region III and 0.69 in Plume habitats of the Region III. The EQR values of the PH3 indicator ranged between 0.16 (Plume habitats in Region II) and 0.32 (Coastal habitats in Region III). For FW2, the EQR ranged between 0.49 (Oceanic habitats in Region IV) and 0.88 (Plume habitats in Regions II and III). For the eutrophication indicator, values ranged between 0.56 (Plume habitats in Region II) and 1 (Coastal and Oceanic habitats in Region IV). Finally, the EQR for the Dissolved Inorganic Nitrogen indicator ranged between 0.39 (Plume habitats in Region III) and 1 (Shelf habitats in Region IV) while the EQR of Dissolved Inorganic Phosphate ranged between 0.76 (Coastal habitats in Region II) and 1 (Shelf habitats in Region IV). The weighted average EQR of each habitat within OSPAR Regions II, III and IV ranged between 0.55 (Plume habitats in Region II) and 0.88 (Plume habitats in Region III). The integration resulted in 1 habitat with moderate status (Plume habitats in Region II), 7 habitats with good status (Coastal and shelf habitats in Regions II, III and IV and Oceanic habitat in Region IV) and 2 habitats with high status (Plume habitats in Region III and IV). The integrated status per Region was "good" for all three Regions with values comprised between 0.60 and 0.75.

Table 5: EQR obtained for Phytoplankton biomass, Zooplankton abundance, Phytoplankton and zooplankton diversity, Primary Production, eutrophication indicator and Winter nutrient concentration for Plume, Coastal, Shelf and Oceanic habitats within OSPAR Regions II, III and IV for the assessment period (2015-2019). The EQR computed for the integration corresponds to the $EQR_{integrated}$ as defined by Equation (1). The GES of the weighted average EQR is given according to their correspondence to the GES categories given by Eutrophication indicators. Note that Dissolved Inorganic Nitrogen (DIN) and

Dissolved Inorganic Phosphate (DIP) formed the 'winter nutrient concentration' indicator. We use the average between DIN and DIP as the value to integrate in average weighted EQR

		OSPAR indicator						$EQR_{integrated}$	$EQR_{integrated}$
OSPAR Region	Habitat	PH2	PH3	FW2	Eutrophication indicator	DIP	DIN	Integrated message (habitat)	Integrated message (region)
RII	Plume	0.67	0.16	0.88	0.56	0.82	0.59	0.55	0.60
	Coastal	0.56	0.26	0.87	0.78	0.76	0.66	0.61	
	Shelf	0.52	0.27	0.76	0.93	0.89	0.82	0.63	
	Oceanic	-	-	-	-	-	-	-	
RIII	Plume	0.69	-	0.88	0.86	0.95	0.39	0.88	0.71
	Coastal	0.54	0.32	0.87	0.82	0.86	0.86	0.64	
	Shelf	0.45	0.29	0.68	0.98	0.98	0.93	0.62	
	Oceanic	-	-	-	-	-	-	-	
RIV	Plume	-	-	-	0.92	-	-	0.92	0.75
	Coastal	0.46	-	-	1	0.96	0.89	0.76	
	Shelf	0.49	0.25	0.61	1	1	1	0.69	
	Oceanic	0.57	0.27	0.49	1	0.84	0.80	0.61	

4. Discussion

4.1. Integration between D1-D4-D5 indicators

Since the eutrophication, Pelagic Habitats and FW2 'primary production' indicators now all use the COMP4 assessment units for OSPAR assessments, the first two steps of the workflow described in Figure 2 facilitated the integration between the D1C6, D5C1 and D5C2 and D4C4 for the purpose of the QSR2023.

The first attempt to integrate plankton-related indicators across MSFD descriptors D1, D4 and D5 provided a holistic view of the base of the marine food web for the North-East Atlantic. The two integration options presented revealed different results in terms of the overall Environmental Status for each pelagic habitat type within the three OSPAR Regions. While the OO-AO approach resulted in an overall "not good" or "bad" environmental status (independent of the vocabulary used), the weighted average Ecological Quality Ratio resulted in only 9% of the habitats being designated as having "moderate" Environmental Status, and 91% of the remaining habitats in "good" or "high" Environmental Status.

The OO-AO represents a fast and easy integration method and has the advantage that it can be performed even when different methodologies have been used for previous integration steps. It also has the benefit of being relatively simple, since it avoids calculations (**Table 6**). Nevertheless, OO-AO is the most rigid integration approach, which regularly downgrades quality status to be more negative (i.e., frequently reporting not good or poor status) (Borja and Rodriguez, 2010). By contrast, the

weighted average Ecological Quality Ratio approach is less strict and produces a more realistic and complete summary of the state of pelagic ecosystem (**Table 6**). For the example tested in this report, the EQR approach appeared to upgrade, rather than downgrade the results. The weighted average EQR approach does have some drawbacks, including the knowledge required for selecting an appropriate comparison period which exhibits reference values for each biological element of the Pelagic Habitats indicators and corresponding to a period where these habitats are "not adversely affected due to anthropogenic pressures".

Table 6: Pro and cons of the two options explored in the integration between D1, D4 and D5.

	Advantages	Drawbacks
Option 1: One Out-All Out	- Simple - Allows different methodologies prior to the integration - Avoids averaging multiple metrics	- Tendency to downgrade or misclassify quality status
Option 2: Weighted average Ecological Quality Ratio	- More realistic - Easy to interpret	- Requires knowledge and/or data on period with negligeable human impact (defining the baseline date)

The difference in GES results between the two approaches when applied across indicators is likely due to differences in the pressures-indicator relationship within each group of indicators. Within the state indicators, the Pelagic Habitat and Food Webs assessments used this relationship to establish GES (Magliozzi et al. 2021), while the eutrophication indicator did not use this relation to establish GES. Therefore, a change in chlorophyll-*a* concentration in this latter indicator could not be linked to any anthropogenically-induced change, producing contrasting results compared to the PH2 indicator. Further explorations on the origin of discrepancies between these two indicators are provided in part 5 of this report. Changes in chlorophyll-*a* concentration in the eutrophication indicator may also result from natural variability in phytoplankton biomass. Finally, the Dissolved Inorganic nutrient (Nitrogen and Phosphate) indicators could not be linked to any pressure as they are already indicators of pressure.

Within the OO-AO approach, the pressures-indicator relationship tends to produce a highly downgraded integration result, as the GES categories can only be "not good" or "uncertain", as we lack data for historical "conditions that are not adversely affected due to anthropogenic pressures". The observation that the EQR option tends to upgrade integration results is related to the fact that it

does not consider the pressures-indicator relationship for the Pelagic Habitats and Food Web indicators. Nevertheless, since the current definition of GES is established at the level of each descriptor, the inclusion of GES in the integration across descriptors is mostly a policy decision, rather than a technical one.

Currently, the OO-AO approach represents the best option for integration. This method is currently implemented to integrate the Pelagic Habitats assessment within and across indicators (Pelagic Habitat Thematic Assessment and D1.4). However, prior the generalisation of integration across Pelagic Habitats, Food Webs and Eutrophication indicators, some recommendations for Descriptor 1 and Descriptor 4 can be made to increase the robustness of this integration. The OO-AO principle can be refined to integrate partial indicators' results, and weight them according to their relative level of importance (defined as a partial OO-AO principle). As an example, refinement of the OO-AO integration performed in this report would provide more weight to the Pelagic Habitats and eutrophication indicators than it would for the concentration of winter nutrient indicator. In cases where the weighted average EQR option is selected for future integration, improvements can be made by considering the effect of pressures on the biological elements prior to computing the EQR.

4.2. Definition of the baseline

Defining a baseline is a mandatory step for assessing the marine ecosystems within the framework of the Regional Seas convention. At present, two procedures exist for defining baselines in OSPAR assessments. While the eutrophication assessment uses thresholds produced from historical data reconstructed from 1900 to establish a baseline for GES, the Pelagic Habitats and Food Webs assessments use the period prior to the assessment period. The Pelagic Habitats and Food Webs indicator assessments set baselines defined by the initial samples in each time-series, which do not necessarily represent conditions that are not adversely affected by human pressures. For this reason, reconstructions of historical data (as far back as 1900), as currently done for the eutrophication indicator assessments, is probably a more appropriate and realistic method for determining GES. Once a method for defining the baseline has been selected, Borja et al. (2012) recommends as best practice that baseline conditions should be established independently for each assessment unit. For OSPAR assessments, the authors of this report also recommend setting reference conditions independently for each assessment unit, due to the large scale and diverse habitat types contained within each OSPAR region. Further discussion of thresholds and their application to Pelagic Habitat indicators can be found in NEA-PANACEA D1.4a (Holland et al. 2023).

5. Investigation between PH2 phytoplankton biomass indicator from Pelagic Habitat working group and Chlorophyll-*a* concentration indicator from Eutrophication working group.

A. Louchart, B. Heyden, L. Enserink, L.F. Artigas

5.1. Context

For the OSPAR QSR 2023, chlorophyll-*a* concentration was assessed with a Pelagic Habitats indicator (PH2-Changes in phytoplankton biomass) and a Eutrophication indicator (concentration of Chlorophyll-*a* in the Greater North Sea, the Celtic Seas and Bay of Biscay and Iberian Coasts). Even though both indicators report on changes in the same variable, there were several important differences in assessment results. Contrasting results were most frequently detected within Variable salinity and Coastal habitats of the Greater North Sea. Some of these differences may be attributable to differences in data sources. While both indicators used chlorophyll-*a* measurements from satellite datasets derived from remotely sensed ocean colour, for the eutrophication assessments, when chlorophyll-*a* data from satellite were not available, *in situ* data were used and extrapolated to their corresponding assessment unit. The two separate methodologies are described in the following paragraphs.

5.2. Methodology

5.2.1. Concentration of Chlorophyll-*a* in the Greater North Sea, Celtic Seas and Bay of Biscay and Iberian Coast indicator – Eutrophication

The data used for the chlorophyll-*a* concentration Eutrophication indicator spanned from 1998 to 2020 and were provided by the Royal Belgian Institute for Natural Sciences (RBINS) under the name CHL-Gon (and referred in the rest of the document as CHLA-EUT). The model used was a mixture between OLCI and OC5 algorithms (Ocean Color and Land Instrument algorithm and Ocean Color algorithm with 5 estimates of remote sensing reflectance, respectively). The selection of the best algorithm was influenced by the amount of suspended matter in the water. The Gons multiple algorithms was particularly effective for coastal turbid and eutrophic waters, such as waters often encountered in coastal regions of the North Sea, while the OC5 algorithm was better suited for clearer offshore waters. For this indicator, data were not transformed in any way and were analysed on an annual scale. Only chlorophyll-*a* concentration values from within the productive period (March to October) contributed to this analysis.

5.2.2. Pelagic Habitat 2 Changes in phytoplankton biomass indicator – Biodiversity

Data used for the PH2 indicator analysis were also provided by the Royal Belgian Institute of Natural Sciences (RBINS), as well as Plymouth Marine Laboratory (PML). Data provided to the OSPAR Biodiversity Committee (BDC) by RBINS spanned between 2006 and 2019, while PML data spanned

from 1998 to 2016. PML used the OC5 ocean colour algorithm, which is less efficient than GON for more turbid and eutrophic waters. For this indicator, data needed to be temporally resolved to monthly mean values. To achieve this, temporal interpolation was conducted prior to analysis in order to fill in the missing months, particularly at high latitudes where satellite coverage was poor. Temporal interpolation was run independently on each pixel to replace missing values for when there were up to two consecutive months missing and a maximum of four non-consecutive months missing within a calendar year. This case was encountered often for winter months, when dense cloud cover tends to obscure satellite imagery. These data were then spatially aggregated across a grid of 1-degree pixels. Finally, aggregated data were clipped based on their intersection with corresponding assessment units. In cases when there were more than two consecutive months, or four non-consecutives months missing, the entire year was excluded from the time-series analysis.

The first step of the indicator assessment involves logarithmic transformation of data to reduce the relative influence of outliers. Then, the mean annual cycle of chlorophyll-*a* concentration is calculated separately for each assessment unit and so that it can be subtracted from the transformed data. This step resolves the variation in chlorophyll-*a* concentration beyond the natural annual cycle of phytoplankton growth. Finally, two temporal periods are compared, defined as the "comparison period" (period to be compared to) and the "assessment period" (period to compare) was processed. This indicator records changes when there is a statistically significant difference between the comparison and assessment periods.

5.2.3. Hypothesis

An initial comparison of indicator results reveals three potential hypotheses to explain discrepancies between the assessments of chlorophyll-*a*. The first explanation concerns differences in the datasets used for the two assessments. The eutrophication indicator (labelled as CHLA-EUT) preferentially employed Gons model or OC5 algorithm, depending on the concentration of suspended matter in the water column, while PH2 used a mixture of Gons model and OC5. Differences in results are expected to be more pronounced in the vicinity of estuaries, where CHLA-EUT methodology tended to prefer the choice of OCLI algorithm of the Gons method.

The second possibility is that differences in indicator results might stem from differences in integration methodology. While the CHLA-EUT assessment used a relatively fine spatial resolution (i.e., 1 km x 1 km), the PH2 indicator assessment used a coarser spatial resolution of 1° longitude by 0.5° latitude aggregated across a grid of 1° x 1°.

Finally, a third explanation is that differences in assessment results could stem from whether the non-productive period is considered in the analysis or not. For the PH2 assessment, missing winter months

were interpolated, allowing for the whole year to be included in the analysis, whereas the CHLA indicator avoids interpolation of winter values, and focuses only on the productive months (i.e. March to September) which can be too restrictive as blooms can sometimes occur in February, October or November depending on the location.

5.2.4. Case study

This section investigates differences between these two indicators through a case study, by selecting a range of seven out of the 64 COMP4 assessment units representing habitat types which differ in terms of their biological, hydrological and biogeochemical parameters (**Table 7**).

Table 7: List of the seven assessment units, their abiotic characteristics (habitat, mean salinity and mean depth) and the results of OSPAR assessments for the PH2 changes in phytoplankton biomass and the eutrophication indicator.

Assessment unit	Habitat	Salinity	Depth	Trend PH2	EQR eutrophication indicator	Total area (km^2)
Elbe Plume	Variable salinity	30.8	18	0.55	0.43	7 836
Outer Coastal DEDK	Coastal	33.4	27	-0.78	0.47	18 540
Southern North Sea	Coastal	34.3	32	0.78	0.89	61 758
Northern North Sea	Shelf	35.0	121	0.61	0.97	264 253
Eastern North Sea	Shelf	34.8	43	-0.79	0.84	60 634
Dogger Bank	Shelf	35.1	28	0.66	0.95	14 749
Intermittently stratified 2	Shelf	35.1	102	-0.42	0.99	26 517

The assessment units selected to conduct the case study were all located in the Greater North Sea (OSPAR Region II) (**Figure 3**) where the largest discrepancies between results of the two indicators were observed. We selected at least one assessment unit for each pelagic habitat type to ensure representativeness across habitats for this indicator comparison, with one assessment unit in Variable salinity (Elbe Plume), two in Coastal (Outer Coastal DEDK, Southern North Sea), and four in Shelf (Northern North Sea, Eastern North Sea, Dogger Bank and Intermittently stratified 2) habitats. Results of the two indicator assessments are also provided to illustrate the discrepancies (Table 7).

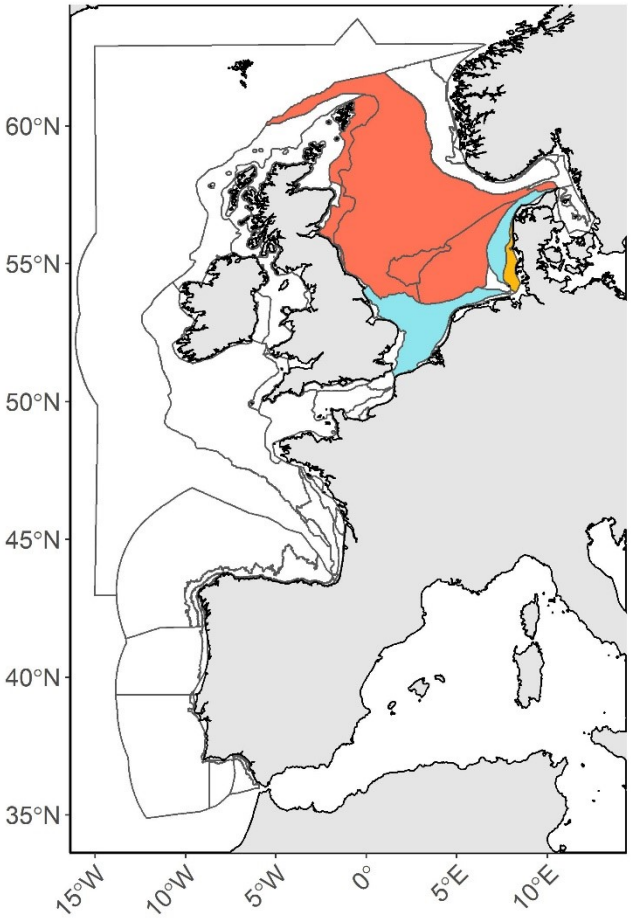

Figure 3: Assessment units selected to conduct the investigation between the PH2 and the eutrophication indicators. Assessment units coloured in orange, blue, and red correspond to variable salinity, coastal and shelf habitats, respectively, as defined by the COMP4 assessment units procedure (Enserink et al. 2019).

5.2.5. Analysis

Prior to analysis, we visually inspected the distribution of chlorophyll-*a* concentration measurements using boxplots for each assessment unit for the separate datasets. Subsequently, each dataset was tested for normality and homogeneity of variance to ensure test assumptions could be met. Since variance was found to not be homogeneous, an ANOVA with 1000 permutations (a more robust option than the Scheirer-Ray-Hare non-parametric test) was applied to investigate whether there were significant differences between datasets within each assessment unit. We conducted a second 1000 permutation ANOVA to understand the difference caused by the choice of the dataset (data used for the eutrophication versus pelagic habitats indicators), the period studied (productive period only versus whole year) considered in the analysis, and to also test for an interaction of these two variables.

We also investigated the importance of the spatial resolution of the gridded data for the estimation of chlorophyll-*a* concentration within each assessment unit. Since the datasets have different spatial resolution (i.e., PML dataset: 1° longitude x 0.5° latitude; RBINS: 1 x 1 km, **Figure 4**), the dataset with the lower resolution had to be modified to match the higher resolution grid. For this, we disaggregated

the resolution of the PML data to match that of the RBINS data. Subsequently, the two datasets were overlayed, and subtraction was performed separately on each cell to produce a map of differences between the two datasets. For this example, we subtracted the RBINS data from the disaggregated PML data. Values close to 0 indicate near-parity between the two datasets. Results greater than 0 indicate overestimation of chlorophyll-*a* concentration for the PML data, relative to the RBINS data, whereas results below 0 indicate underestimation of chlorophyll-*a* concentration for the PML data, relative to the RBINS data. For this step, we only investigated month of August 2012 as a single case study. The choice of month was random and arbitrary in this case. We also investigate for the selected assessment units for the probability of attributing incorrect chlorophyll-*a* concentration measurements with the PML datasets. This probability increases when there are pixels contained within an assessment unit that also intersect adjacent assessment units. This step is sensitive to spatial resolution, as far as the smallest assessment units are expected to contain few pixels.

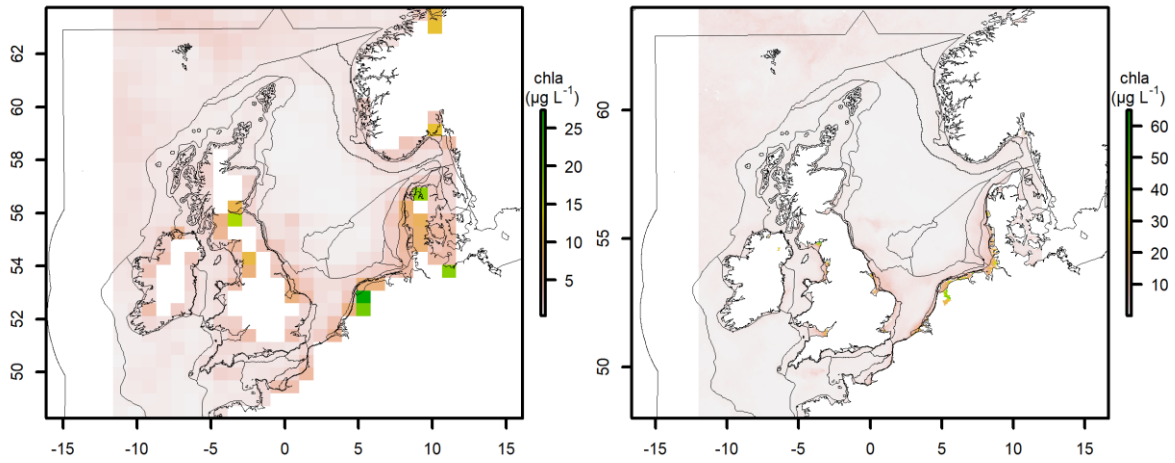

Figure 4: Visualisation of the spatial resolution of chlorophyll-a concentration for August 2012 from PML data (left) and RBINS data (right).

Finally, we investigated differences in chlorophyll-*a* dynamics between the two datasets. First, time-series for annual mean raw chlorophyll-*a* and $\log_{10}$ chlorophyll-*a* were visually inspected. To identify potential bias in the trend analysis, we generated linear models using values from one time-series to predict the other for each assessment unit. This step was intended to compare the slope of the raw chlorophyll-*a* versus the slope of the theoretical relationship (EO_chla = OC5&GON_chla) to highlight whether certain assessment units were more likely subject to biases. In addition, a strict similarity between the two datasets is expected. Therefore, the intercept is expected to be 0. A t-test compare the intercept of each linear model against 0.

5.3. Results

5.3.1. Testing on the importance of considering the winter months

The ANOVA with permutation revealed that the chlorophyll-*a* concentration varied between the two datasets for the three habitat types ($p < 0.05$). The ANOVA also revealed that data used for the Pelagic Habitats indicator assessment had significantly higher mean value than analogous data used for the Eutrophication assessment for 5 out of 7 assessment units (**Figure 5**; $p < 0.05$). There was no difference between datasets for the Elbe plume (Variable salinity habitat) and the Southern North Sea (Coastal habitat) assessment units. In most cases there were no differences between periods considered in the analysis (i.e. productive period only versus full annual cycle), except for the data used for the Pelagic Habitats indicator in the Eastern North Sea, Dogger Bank, and Northern North Sea ($p < 0.05$), and between the productive period and full annual cycle for the dataset used in the Eutrophication assessment of the Dogger Bank.

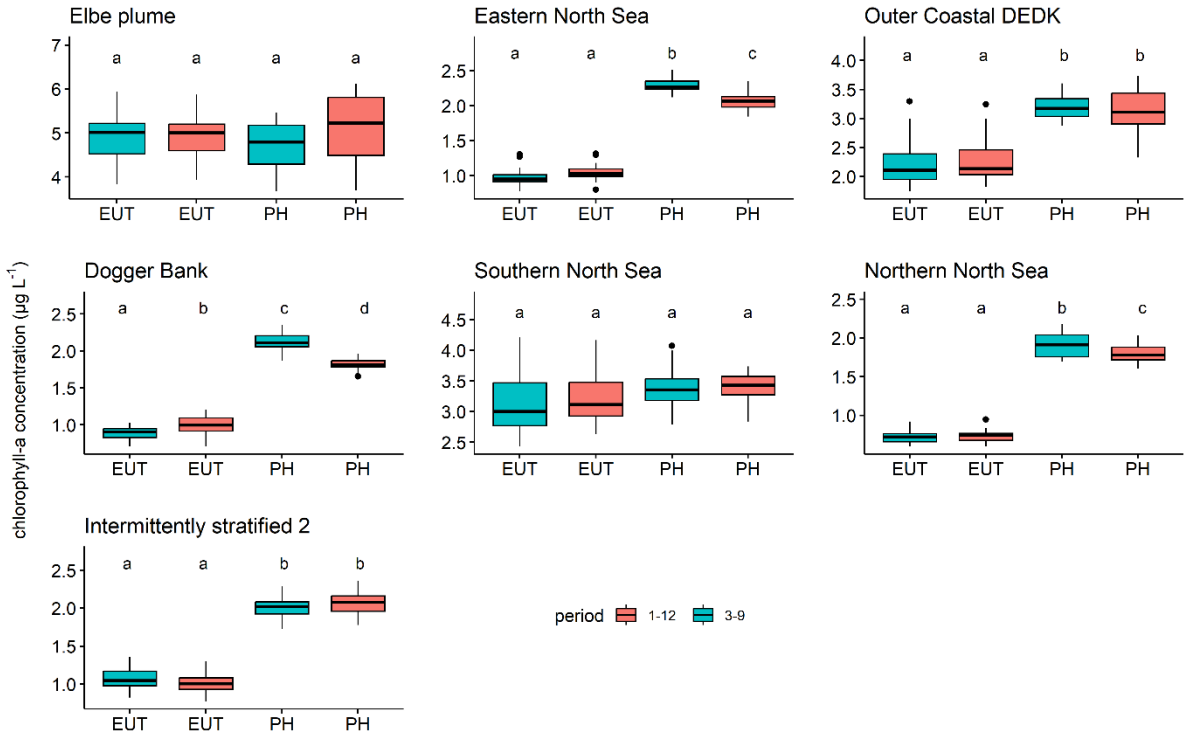

Figure 5: Boxplots of annual averaged concentration of untransformed chlorophyll-a per dataset within each assessment units for the productive period only (March to September) and for the full annual cycle (January to December).

5.3.2. Testing the impact of spatial resolution

Disaggregating the gridded PML data to the same resolution as the RBINS data generated a grid containing 2,599,206 pixels (**Figure 6**)[1]. The difference in chlorophyll-*a* concentration between the two datasets ranged from –61.4 to 19.5 (µg L^{-1}). Positive values represented 74% of total pixels, while negative values represented 26% of total pixels. In other terms, for August 2012 the PML dataset overestimated the concentration of chlorophyll-*a* relative to the RBINS datasets for 74% of the study

area, while it underestimated the concentration of chlorophyll-*a* in 26% of the study area. The PML dataset always provided higher chlorophyll-*a* measurements than the RBINS dataset, except for in Plumes or Variable salinity habitats, in line with results described earlier in section 5.3.1. In Plumes or Variable salinity habitats the RBINS dataset consistently exhibited higher chlorophyll-*a* values than the PML dataset.

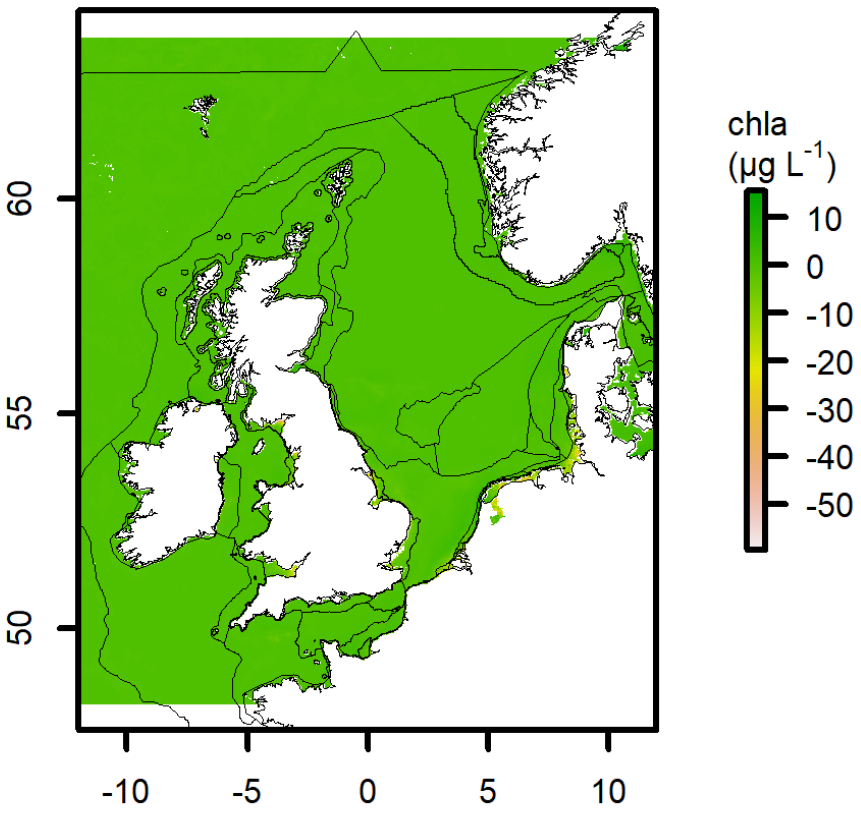

Figure 6: The difference in chlorophyll-a concentration between PML and RBINS datasets for August 2012.

When calculating the mean chlorophyll-*a* concentration for each assessment unit, misattribution can occur through incorporating neighbouring pixels into the calculation, particularly when low resolution gridded data is used. In this section, we quantify the intersection between each of the seven selected assessment units and chlorophyll-*a* pixels as the probability of good (pixel centroid intersects assessment unit) or bad (pixel centroid does not intersect assessment unit) assignment (**Table 8**). The proportion of misclassified pixels represented the probability of bad chlorophyll-*a* assignment. This proportion ranged from 12.2% (Northern North Sea and Eastern North Sea) to 44.5% (Outer Coastal DEDK). Pixels attributed to larger assessment units are less likely misclassified than pixels attributed to smaller assessment units ($\rho = 0.89$; $p<0.05$).

Table 8: Probability of good (IN) and bad (OUT) pixel assignment for each assessment unit and the corresponding area of these groups.

Assessment unit	Habitat	% IN	% OUT	Total Km2	Km^2 IN	Km^2 OUT
Elbe Plume	Plume	75.0	25.0	7 836	5 877	1 959
Outer Coastal DEDK	Coastal	55.5	44.5	18 540	10 290	8 250
Southern North Sea	Coastal	80.2	19.8	61 758	49 530	12 228
Northern North Sea	Shelf	87.8	12.2	264 253	232 014	32 239
Eastern North Sea	Shelf	87.8	12.2	60 633	53 236	7 397
Intermittently Stratified 2	Shelf	58.2	41.8	26 517	15 433	11 084
Dogger Bank	Shelf	75.4	24.6	14 749	11 121	3 628

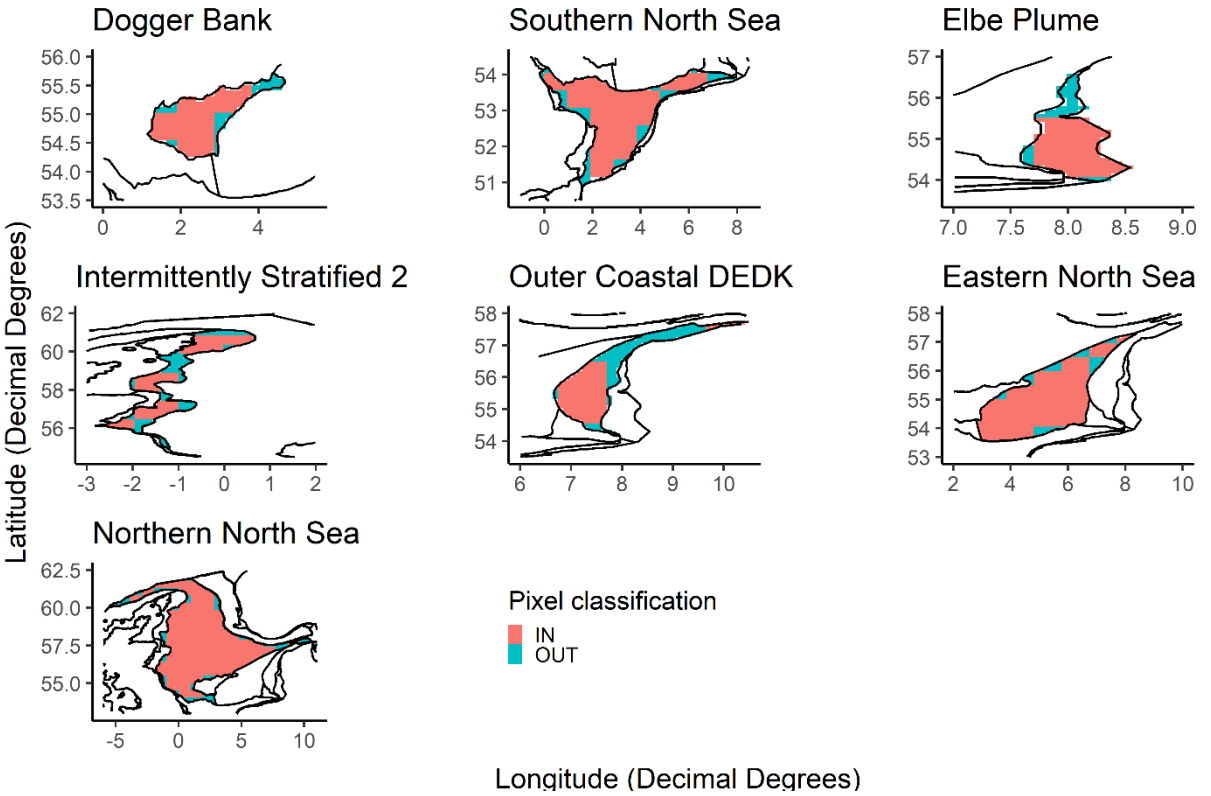

Figure 7: Classification of pixels used for calculation of mean chlorophyll-a concentration for seven assessment units, according to whether each pixel's centroid is located inside (IN) or outside (OUT) the assessment unit.

5.3.3. Comparison of phytoplankton biomass dynamics

Visualisations of long-term chlorophyll-*a* time-series reveal differences in the dynamics of the two datasets, varying by assessment unit (**Figure 8**). While EO-chla (eutrophication indicator) and OC5&GON_chla (PH2 indicator) showed similar dynamics in the Outer Coastal DEDK, Dogger Bank, and Intermittently stratified 2 assessment units (**Figure 8**), there were larger differences observed in the Elbe plume, Eastern North Sea, Southern North Sea, and Northern North Sea assessment units. In other words, we should expect the eutrophication indicator and PH2 indicator will generate contrasting results for Elbe plume, Eastern North Sea, Southern North Sea, and Northern North Sea assessment units.

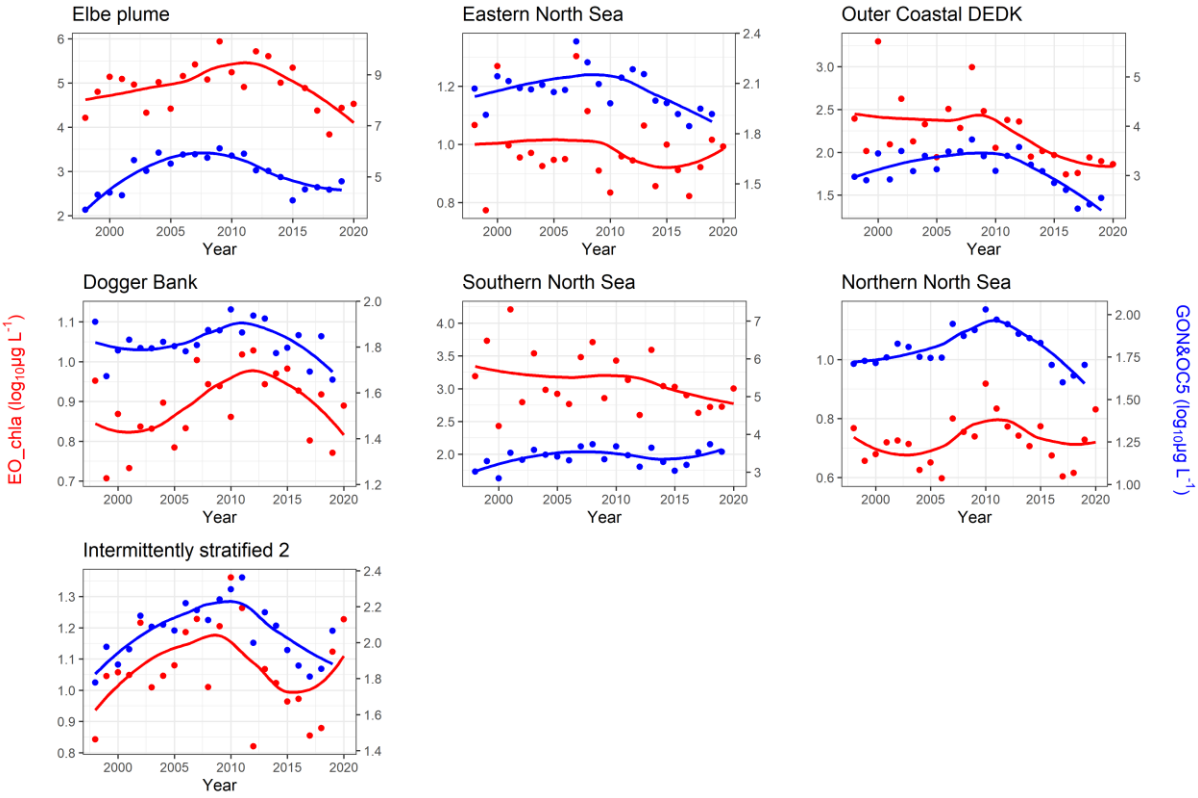

Figure 8: Evolution of annual averaged concentration of untransformed chlorophyll-a.

5.3.4. Evaluation of the relationship

Following the previous section, we examined the linear relationship between the two datasets (**Figure 9**). The theoretical 1:1 relationship was also overlaid in each plot to assist in identifying assessment units likely to be subject to large differences in indicator results.

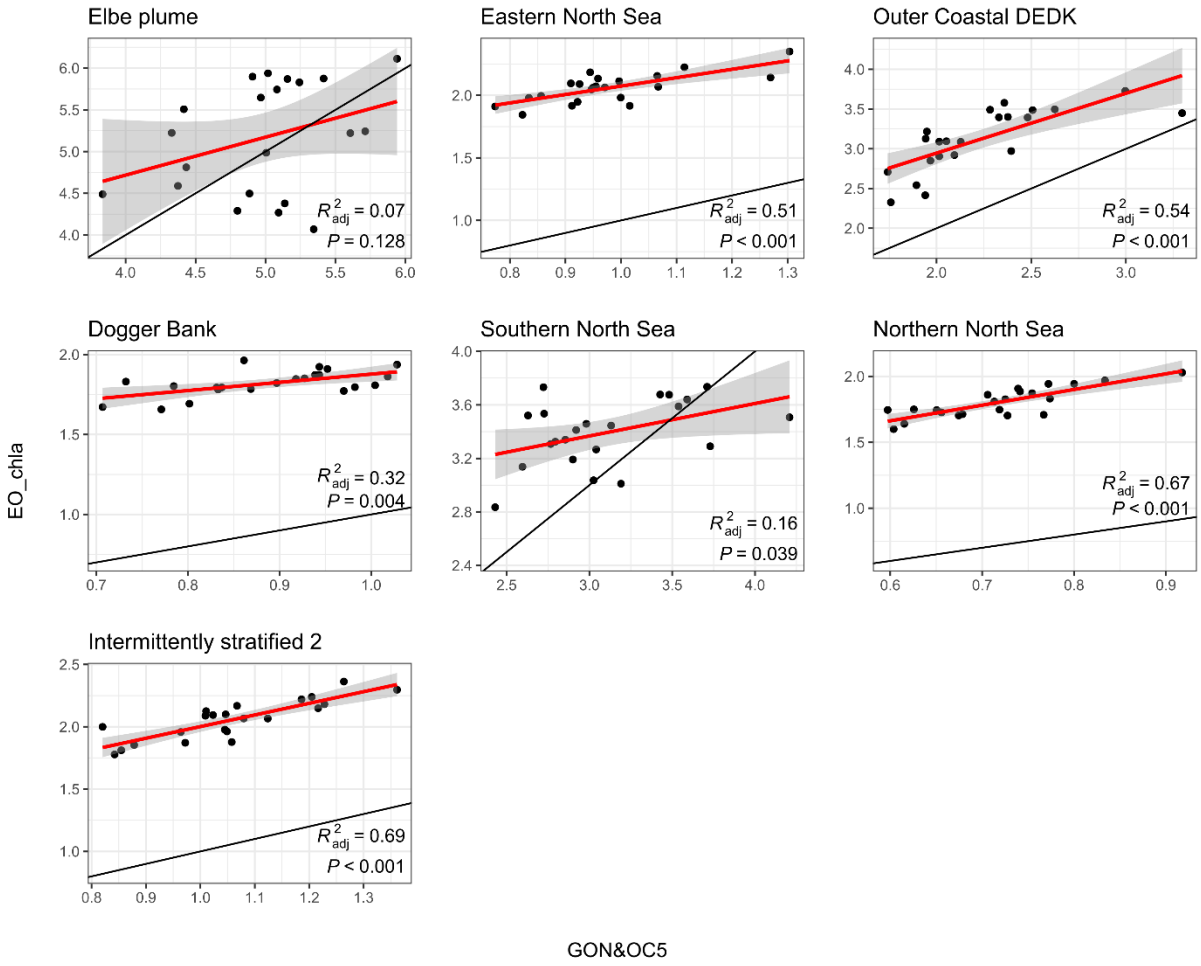

Figure 9: Linear relationship between the EO-chla (Eutrophication indicator) and OC5&GON_chla (PH2 indicator) datasets. The red line represents the linear relationship between the two datasets. The black line represents theoretical relationship (EO_chla = OC5&GON_chla). The grey area represents the standard error between the two datasets.

For 5 out of 7 assessment units (Eastern North Sea, Outer coastal DEDK, Dogger Bank, Northern North Sea and Intermittently Stratified 2), the linear model relationship had a higher y-intercept than the theoretical relationship, revealing that for these 5 assessment unit values from the OC5&GON_chla were overestimated, relative to EO_chla. For the two remaining assessment units (Elbe Plume and Southern North Sea), OC5&GON_chla values were higher than EO_chla, for any value less than 5.30 and 3.45 µg L^{-1} chla respectively, indicating that at values below these breakpoints the OC5&GON_chla dataset overestimated chlorophyll-*a* concentration relative to EO_chla. T-tests comparing the y-intercepts revealed that the overestimation of chlorophyll-*a* concentration by the OC5&GON_chla dataset was significant for all the assessment units except the Elbe Plume (**Table 9**). The high variability of the data around the linear model relationship likely contributed to the non-significant result.

Table 9: Slope and intercept of the linear relationship between the two datasets in the 7 assessment units. The R2 corresponds to the relationship quality between the two datasets. The p-values report whether there is a statistically significant difference between the observed relationship and the ideal theoretical relationship.

Assessment unit	Habitat	Slope	Slope p-value	Intercept	Intercept p-value	Full model R^2	Full model p-value
Elbe Plume	Variable salinity	0.60	0.17	2.13	0.14	0.07	0.128
Eastern North Sea	Shelf	0.67	**<0.05**	1.40	**<0.001**	0.51	**<0.001**
Coastal DEDK	Coastal	0.75	0.11	1.44	**<0.001**	0.54	**0.004**
Dogger Bank	Shelf	0.51	**<0.01**	1.36	**<0.001**	0.32	**0.039**
Southern North Sea	Coastal	0.24	**<0.001**	2.64	**<0.001**	0.16	**<0.001**
Northern North Sea	Shelf	1.19	0.31	0.95	**<0.001**	0.67	**<0.001**
Intermittently Stratified 2	Shelf	0.93	0.63	1.07	**<0.001**	0.69	**<0.001**

Despite high similarity in the slopes of the linear model and the theoretical relationship for 5 of 7 assessment units (**Figure 9**; i.e., Eastern North Sea, Outer coastal DEDK, Dogger Bank, Northern North Sea and Intermittently Stratified 2), t-tests revealed significant differences in slope for the Eastern North Sea, Dogger Bank, and Southern North Sea assessment units. In other words, there were significant differences in measurements of chlorophyll-*a* concentration between the two datasets. Finally, while the weakest relationship between datasets (R^2 = 0.07; full model p-value = 0.128) occurred within a Variable salinity habitat (i.e., Elbe Plume), the strongest relationships occurred mainly in Shelf habitat (R^2 = 0.32 to 0.69; full model p-value < 0.05). The R^2 values for Coastal habitat assessment units were between those of the variable salinity and shelf habitats.

5.4. Discussion

All pelagic habitats, from Variable salinity to Shelf, exhibited differences in results between the eutrophication indicator and the PH2 indicators. There was no evidence that the length of the assessed period (e.g., productive period only or full annual cycle) contributed to differences between datasets. Rather, our results suggest that differences in spatial resolution are likely the main source of differences observed between eutrophication and Pelagic Habitats indicator results. The larger assessment units tended to be less affected by the discrepancies than the smaller and narrower assessment units.

The methodology developed to study the impact of differences in spatial resolution provided a measurement of “spatial confidence” as a percentage of misattributing pixels with their centroid

located outside the polygon for the respective assessment unit. There was high variability between assessment units with misattributed values, accounting for between 12.8 and 44.5% of the total area for each assessment unit. In other words, 12.8 to 44.5% of the total area of each assessment unit is affected by boundary effects, with data from neighbouring assessment units being misattributed due to spatial resolution which is too coarse for this purpose.

Assessment methodology also plays a non-negligeable role. Despite focusing on the same biological element (i.e., phytoplankton biomass), there are important differences between indicators in the establishment of baseline conditions and computation of indicator results (see discussion in Part 4 of this report). Without considering this point, when interpreting the results, one might simply conclude that one of the two indicators is wrong, when these two indicators should realistically be seen as complementary, each having their own methodology and providing different conclusions on GES.

Considering the above explanations, it is possible to provide some further recommendations. First, the methodology used in this report to determine spatial confidence requires further development before being implementated as a part of indicator assessments. Finally, we also recommend for subsequent assessments that the same dataset should be used for the PH2 and concentration of chlorophyll-*a* indicators. A coordinated mutualisation of the data call for the phytoplankton biomass between the Eutrophication and Pelagic Habitats working groups would represent an improvement in this regard.

6. Knowledge gaps

In addition to further development within each indicator (see knowledge gaps section of each indicator for detailed information), further development of this integration is needed, particularly on the following points:

- Improvement of the methodology for integrating the Pelagic Habitats and FW2 indicators with eutrophication indicators.
- Until now, different assessment units have been used for Pelagic Habitats, Food webs and eutrophication hampering an integration of the different indicators. The common use of the new ecologically coherent area classification of COMP4 assessment units under the different MSFD Descriptors (D1, D4 and D5) provides the basis for the integration. The first step towards integration has been to divide the OSPAR regions into assessment units and categorise them into four habitats (variable salinity or plume, coastal, shelf and oceanic or beyond shelf habitats). This step is now commonly used by Pelagic Habitat and HASEC expert groups. Then, each indicator has its own methodology to determine the GES. Pelagic Habitat indicators are state indicators. Their GES is based on the relationship between biological elements and the

impact of pressures on them (D1C6 of the MSFD). Eutrophication indicators are pressure indicators. Their GES is therefore determined directly on the results of each indicator compared with reference values. In the approaches we have explored in this report, the One-Out All-Out approach considers the GES of the indicators, whereas the EQR approach does not consider the GES of the Pelagic Habitat indicators. The EQR approach does no longer fully comply with the MSFD D1C6 criterion. As the determination of the GES is approached at different levels in each MSFD descriptor, we need to clarify matters for future integrations. It therefore seems necessary to define the appropriate level of analysis for integration in order to determine the GES of the integration as accurately as possible.

- More coherence of data availability between Pelagic Habitats and Eutrophication indicators. OSPAR QSR 2023 highlighted discrepancies between PH2 and Concentration of Chlorophyll-*a* in the Greater North Sea, Celtic Sea and Bay of Biscay and Iberian Coast indicators due to different spatial resolution of the datasets. HASEC and COBAM expert groups identified that same data products provided for OSPAR QSR 2023 had different temporal length. COBAM group received chlorophyll-*a* data from RBINS from 2009 to 2020 while HASEC group received chlorophyll-*a* data from RBINS from 1998 to 2020. Recommendations to strengthen the coherence of data collection and use for future assessments were made to HASEC and COBAM expert groups.
- Incorporate spatial and temporal confidence to weight the determination of the GES. Non-station data are very often unevenly distributed in time and space. An example of spatio-temporal heterogeneity in the Northeast Atlantic is shown with the sampling distribution of the Continuous Plankton Recorder (Holland et al. 2023), which is particularly weak in the Plume habitat. Satellite data are also affected by this spatio-temporal heterogeneity, as during winter months, primary production and chlorophyll-*a* concentration at high latitudes are poorly estimated. It is therefore advisable to modify the weight of each indicator for the integration in each assessment unit, considering the spatio-temporal confidence of these data and make a robust and reliable integration.

7. Conclusion

Integrating across MSFD descriptors provides a holistic view of marine ecosystems for the purpose of marine management. Recent improvement within the NEA-PANACEA project have now made it possible to integrate within and across Pelagic Habitats indicators, within and across Food Webs indicators, and within and across eutrophication indicators. This report provided two options for integrating related plankton indicators under descriptors D1, D4, and D5, and represented a synoptic view of GES for the first trophic level of the marine ecosystem under MSFD legislation. Although each

option has its own advantages and disadvantages, they provide simple information on GES which is suitable for a range of stakeholder audiences, from experts to policy makers and the general public.

Additionally, we explored mechanisms contributing to differences between two indicators which describe phytoplankton biomass. We identified that the two indicators define GES at a different level and thus this may lead to differences in the key message they produce. While the PH2 indicator (change in phytoplankton biomass) defines GES by "the condition of the habitat type that is not adversely affected due to anthropogenic pressures", the concentration of chlorophyll-*a* indicator defines GES as "human-induced eutrophication is minimised, especially adverse effects thereof, such as losses in biodiversity, ecosystem degradation, harmful algae blooms, and oxygen deficiency in bottom waters". In addition to differences in the definition of GES, there were also fundamental differences between the datasets used. While the time period considered did not impact the quality of the results, spatial resolution of the data source was found to be a major contributor to differences between results of the two indicators. A mutualisation of data calls and other databases and data products used (e.g., ICES, RBINS satellite data) involving both the Eutrophication and Pelagic Habitats expert groups has been identified as a priority for improving consistency between related indicators. It is hoped that this collaboration between working groups will enhance the relevance of the integration between plankton indicators under Descriptors 1, 4, and 5.

8. References

Borja, A., Rodríguez, J.G., 2010. Problems associated with the 'one-out, all-out' principle, when using multiple ecosystem components in assessing the ecological status of marine waters. Marine Pollution Bulletin, 60, 1143-1146

Borja, Á., Dauer, D. M., & Grémare, A. (2012). The importance of setting targets and reference conditions in assessing marine ecosystem quality. Ecological Indicators, 12, 1-7.

Budria, A., Aubert, A., Rombouts, I., Ostle, C., Atkinson, A., Widdicombe, C., Goberville, E., Artigas, F., Johns, D., Padegimas, B., Corcoran, E., McQuatters-Gollop, A., 2017. Cross-linking plankton indicators to better define GES of pelagic habitats. EcApRHA Deliverable WP1.4. 54 pp.

Dierschke V., A. Kreutle, N. Häubner, C. Magliozzi, S. Bennecke, L. Bergström, A. Borja, S. T. Boschetti, A. Cheilari, D. Connor, F. Haas, M. Hauswirth, S. Koschinski, C. Liquete, J. Olsson, D. Schönberg-Alm, F. Somma, H. Wennhage, A. Palialexis, 2021. Integration methods for Marine Strategy Framework Directive's biodiversity assessments, EUR 30656 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-79-33990-8, doi:10.2760/4751, JRC124613

Eliott, S.A.M., Arroyo, N.L., Safi, G., Ostle, C., Guérin, L., McQuatters-Gollop, A., Aubert, A., Artigas, F., Budria, A., Rombouts, I., Pesch, R., Schmitt, P., Vina-Herbon, C., Meakins, B., Gonzales-Irusta, J.M., Preciado, I., Lopez-Lopez, L., Punzon, A., Torriente, A., Serrano, A., Haraldsson, M., Capuzzo, E., Claquin, P., Kromkamp, J., Niquil, N., Judd, A., Padegimas, B., Corcoran, E., 2017. Proposed approaches for indicator integration. EcApRHA Deliverable WP 4.1. 46 pp.

Enserink, L., Blauw, A., van der Zande, D., Markager, S., 2019. Summary report of the EU project 'Joint monioring programme of the eutrophication of the North Sea with satellite data' (Ref: DG ENV/MSFD Second Cycle/2016). 21 pp.

Facca, C., Bernardi Aubry, F., Socal, G., Ponis, E., Acri, F., Bianchi, F., Giovanardi, F., and Sfriso, A., 2014. Description of a Multiparametric Phytoplankton Index (MPI) for the assessment of transitional waters. Marine Pollution Bulletin, 79, 145-154.

Heiskanen, A. S., Van de Bund, W., Cardoso, A. C., and Noges, P., 2004. Towards good ecological status of surface waters in Europe-interpretation and harmonisation of the concept. Water science and technology, 49, 169-177.

Heyden, B., Leujak, W., 2022. Winter nutrient concentrations in the Greater North Sea, Celtic Seas and Bay of Biscay and Iberian Coast. In: OSPAR, 2023: The 2023 Quality Status Report for the Northeast Atlantic. OSPAR Commission, London.

Holland, M., Louchart, A., Artigas, L.F., and McQuatters-Gollop, A., 2023. Integration within and across Pelagic Habitats indicators. NEA PANACEA Deliverable 1.4a. 31 pp.

Legendre, P., De Cáceres, M., 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecology letters, 16, 951-963.

Louchart, A., Artigas, L. F., Holland, M., McQuatters-Gollop, A., 2022a. *Change in phytoplankton biomass and zooplankton abundance*. In: OSPAR, 2023: The 2023 Quality Status Report for the Northeast Atlantic. OSPAR Commission, London.

Louchart, A., Artigas, L. F., Holland, M., McQuatters-Gollop, A., 2022b. *Change in plankton diversity*. In: OSPAR, 2023: The 2023 Quality Status Report for the Northeast Atlantic. OSPAR Commission, London.

Louchart, A., Lizon, F., Claquin, P., Artigas, L. F., 2022c. *Pilot assessment on primary production*. In: OSPAR, 2023: The 2023 Quality Status Report for the Northeast Atlantic. OSPAR Commission, London.

Louchart, A., Holland, M., Schückel, U., Raabe, T., McQuatters-Gollop, A., and Artigas, L.F., 2023. Options for integration of Pelagic Habitats indicators with Benthic Habitats, birds, mammals and Food web. NEA PANACEA Deliverable 1.5b. 30 pp.

Magliozzi, C., Palma, M., Druon, J. N., Palialexis, A., McQuatters-Gollop, A., Varkitzi, I., González-Quirós, R., Gorokhova, E., Heyden, B., Boicenco, L., and Artigas, L.F., 2023. Status of pelagic habitats within the EU-Marine Strategy Framework Directive: Proposals for improving consistency and representativeness of the assessment. Marine Policy, 148, 105467.

Magliozzi, C., Druon, J.-N., Palialexis, A., Artigas, L. F., Boicenco, L., González-Quirós, R., Gorokhova, E., Heyden, B., McQuatters-Gollop, A., Varkitzi, I., 2021. Pelagic habitats under MSFD D1: current approaches and priorities, EUR 30619 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-30988-8, doi:10.2760/942589, JRC123960

McQuatters-Gollop, A., Guérin, L., Arroyo, N.L., Artigas, L.F., Bedford, J., Corcoran, E., Dierschke, V., Elliott, S.A.M., Geelhoed, S.C.V., Gilles, A., Gonzalez-Irusta, J.M., Haelters, J., Johansen, M., Le Loc'h, F., Lynam, C.P., Niquil, N., Meakins, B., Mitchell, I., Padegimas, B., Pesch, R., Preciado, I., Rombouts, I., Safi, G., Schmitt, P., Schückel, U., Serrano, A., Stebbing, P., De la Torriente, A., Vina-Herbon, C. 2022. Assessing the state of marine biodiversity in the Northeast Atlantic. Ecological Indicators, 141, 109148.

Rombouts, I., Simon, N., Aubert, A., Cariou, T., Feunteun, E., Guérin, L., Hoebeke, M., McQuatters-Gollop, A., Rigaut-Jalabert, F., Artigas, L.F., 2019. Changes in marine phytoplankton diversity: Assessment under the Marine Strategy Framework Directive. Ecological Indicators, 102, 265-277.