OSPAR CEMP Guideline

Common indicator: PH1/FW5 Change in plankton communities

Adopted by BDC(2) 2022
OSPAR Agreement 2018-07[1]

This OSPAR biodiversity indicator has been further developed from its initial use in the Intermediate Assessment 2017. As a result of iteration and learning, it is anticipated that there will continue to be evolution of the methods and approaches documented in the CEMP guidelines. Version updates will be clearly indicated and will be managed in a phased approach via ICG-COBAM through its expert groups and with the oversight and steer of BDC.

Contents

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

1 Introduction

Indicators based on plankton lifeforms have been used to assess community response to sewage pollution (Charvet et al. 1998; Tett et al. 2008), anoxia (Rakocinski 2012), fishing (Bremner et al. 2004), eutrophication (HELCOM 2012), climate change (Beaugrand 2005; Bedford et al. 2020; McQuatters-Gollop et al. 2019), and ocean acidification (Keys et al. 2018). Indicators based on functional groups have been proven relevant for the description of the community's structure and biodiversity and are more easily inter-compared than other indicators based on taxonomy (Estrada et al. 2004; Gallego et al. 2012; Garmendia et al. 2012; Mouillot et al. 2006).

In practice, it is often preferable to aggregate species with similar traits into functional groups, such as lifeforms, rather than assessing the dynamics of individual species. Measures of species abundance are frequently subject to large interannual and regional variation, often due to natural physical dynamics and habitat preferences rather than anthropogenic stressors (de Jonge 2007). Functional group abundance is often less variable because variability in the abundances of the group's constituent species averages out. Cryptic speciation (species with near-identical appearance) within the plankton community, alongside the limitations of identifying plankton using routine light microscopy techniques, make it difficult to generate accurate counts at a species or genus level. Functional group abundance is more reliable as many plankton lifeforms are easily identified, making comparisons between different laboratories and institutes feasible. Both abundance and biomass data can be used to inform lifeform time-series, depending on the lifeform in question and data availability from monitoring programmes.

In addition to studying change in individual lifeforms, these concepts can be extended to investigating changes in ecologically relevant pairs of lifeforms in the form of a lifeform pairs index. The precise combination of lifeforms composing the pairs will depend on the habitat and the objective of the indicator, e.g. as a measure for change in pelagic habitats, food webs, seafloor integrity or eutrophication. Change in the abundance of ecologically linked lifeforms over time can also provide an indication of changes in various aspects of ecosystem function. These include, for example: the transfer of energy from primary to secondary producers (changes in phytoplankton and zooplankton); the pathway of energy flow and top predators (changes in gelatinous zooplankton and fish larvae); benthic/pelagic coupling, i.e. changes in holoplankton (fully planktonic) and meroplankton (only part of the lifecycle is planktonic, the remainder is benthic) (see also Gowen et al. 2011; McQuatters-Gollop et al. 2019).

While studying lifeform pairs can be helpful for detecting changes in the annual cycle of ecologically linked lifeforms, assessing lifeforms separately is more suitable for evaluating gradual change over time. It is simpler to interpret results of an analysis of correlation between lifeform abundance and environmental pressures, than between an index of lifeform pairs and environmental pressures (Bedford et al. 2020). A method has been developed to detect potential links between environmental pressures and change in lifeform abundance over time. The main features of the method are: (i) the grouping of planktonic taxa into functional types or lifeforms; (ii) spatially division of plankton samples to construct distinct time-series of lifeform abundance; (iii) using a robust nonparametric test to quantify long-term changes in lifeform abundance; (iv) relating change in lifeform abundance to trends in environmental pressures and climate indices.

For pelagic habitats, three common OSPAR indicators have been identified based on species (PH3), functional groups (PH1) and community abundance/biomass (PH2). Each indicator will provide specific and complementary information on the state/change and functioning of plankton communities. Currently,

these indicators are developed in parallel but for a robust assessment of the pelagic habitat, these should be considered simultaneously to 1) understand changes and dynamics within the community, 2) reduce the uncertainty in the assessment and 3) to understand the links with anthropogenic pressures. Moreover, plankton strongly depend on physico-chemical properties in the environment so these should be integrated in the assessment; that way, if environmental status must be improved, measures can be taken at the habitat rather than the biological community level.

2 Monitoring

2.1 Purpose

What is the objective of the assessing the indicator; only status of the environment, or also to support identification of pressures and programmes of measures?

- PH1/FW5 is a state indicator which can now be used to quantify links to environmental pressures. PH1/FW5 can be used for the identification of deviation from natural variability for D1, D4 and D6, and will provide information supporting D2, D3, and D5.

2.2 Quantitative Objectives

Plankton sampling collects data which can be used for not only pelagic indicators, but also for food web indicators as well. One plankton sample can be used to inform PH1/FW5, PH2 and PH3. Data collected can also be used to inform D2, D3 and D5. Therefore, one set of monitoring data can be used in multiple ways.

Which parameter needs to be measured?

- Plankton abundance or biomass (per species/genera/taxa)

2.3 Monitoring Strategy

Plankton abundance and biomass must be monitored. PH1/FW5 has been developed using existing datasets which are required for informing the indicator.
Several protocols can be used. It's most cost effective to go with what we already have than to get all CPs to use the same methodology; this also enables establishments of baselines through use of historical data. Integration of existing time-series is key – the pelagic team has considered this practical approach throughout.

- As of the QSR2023 assessment, three types of data are used (**Table 1**):
    - 1) The Continuous Plankton Recorder (CPR) survey (Marine Biological Association (MBA)), a regional monitoring programme at the European scale, including offshore areas; the CPR's regional scale is scientifically critical for understanding plankton dynamics. The CPR survey is funded by the UK with limited funding from some other CPs but is not funded by the EU.
    - 2) Fixed point time-series datasets form the UK and Germany: (Plymouth Marine Laboratory (PML), Marine Scotland Science (MSS), Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN), and Instituto Espanol de Oceanografia (IEO, Spain).
        - The data submitted by The Scottish Association of Marine Science (SAMS) are distributed data from a small area of freshwater influence, however, these data were aggregated and treated as fixed-point time-series.
    - 3) Spatially distributed data collected as part of research cruises, collected in a scattered distribution around coastal regions, or collected at several distributed stations within close proximity to one another (all other datasets).

Table 1: Contracting parties and institutes that provided the datasets used for the indicator assessment.

Contracting Party	Institute	Dataset name	Date range
Belgium (BE)	Vlaams Instituut voor de Zee (VLIZ)	VLIZ_LW_zoo	2014-2020
Denmark (DK)	Aarhus University (AU)	NOVANA phytoplankton	1985-2020
Germany (DE)	Landesamt für Landwirtschaft, Umwelt und ländliche Räume des Landes Schleswig-Holstein (LLUR)	OSPAR_LLUR-SH_2010-2020	2010-2020
	Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten und Naturschutz (NLWKN)	OSPAR_NLWKN_1999-19_phyto	1999-2019
Netherlands (NL)	Rijkswaterstaat (RWS)	RWS_Fpzout_2000-2019_phyto	2000-2019
Portugal (PT)	Instituto Portugues do Mar e da Atmosfera (IPMA)	PseudoNitzschia vs Dinophysis_IPMA	2002-2020
Spain (ES)	Instituto Espanol de Oceanografia (IEO)	IEO_RADIALES_Phyto	1989-2016
		IEO_RADIALES_Zoo	1991-2018
Sweden (SE)	Swedish Meteorological and Hydrological Institute (SMHI)	National data_SMHI_Kattegat-Dnr: S/Gbg-2021_116_phyto	1989-2021
		National data_SMHI_Kattegat-Dnr: S/Gbg-2021_116_zoo	1996-2020
		National data_SMHI_Skagerrak-Dnr: S/Gbg-2021_116_phyto	1986-2020
		National data_SMHI_Skagerrak-Dnr: S/Gbg-2021_116_zoo	1996-2020
United Kingdom (UK)	Centre for Environment, Fisheries and Aquaculture Science (Cefas)	Cefas SmartBuoy Marine Observational Network - UK Waters Phytoplankton Data 2001-2019	2001-2019
	Environment Agency (EA)	EA PHYTO 2000-2020	2000-2020
	Marine Biological Association (MBA)	CPR dataset 1960-2019	1960-2019
	Marine Scotland Science (MSS)	MSS Scalloway Phytoplankton dataset	2000-2018
		MSS Loch Ewe Phytoplankton	2000-2020
		MSS Loch Ewe zooplankton	2002-2017
		MSS Scapa Phytoplankton dataset	2000-2020
		MSS Stonehaven Phytoplankton	2000-2020
		MSS Stonehaven zooplankton	1999-2020
	Plymouth Marine Laboratory (PML)	PML_L4 phytoplankton	1992-2020
		PML_L4 zooplankton	1988-2020
	Scottish Association for Marine Science (SAMS)	SAMS-LPO-Phyto-Dec2021	1970-1981, 2000-2017

2.4 Sampling Strategy

PH1/FW5 is important at the regional level and will be assessed at the scale of COMP4 assessment units (Enserink et al. 2019) which receive regular plankton sampling where possible (**Table 2**).

Table 2: Minimum sampling strategy

	Coastal	Shelf	Offshore
Frequency of data collection*	Monthly	Monthly	Monthly
Monitoring method	In situ	In situ	In situ
Who is responsible for monitoring?	Member state	Member state	Member state
Freq of indicator update and assessment	Annual update	Annual update	Annual update
Minimal amt of monitoring locations	Monitoring must cover all COMP4 assessment units.	Monitoring must cover all COMP4 assessment units.	Monitoring must cover all COMP4 assessment units.

Current data availability	Single point stations exist mainly in coastal waters but there are gaps in some regions.	The CPR is a European scale plankton monitoring programme, focusing on the shelf and offshore regions. Institutional datasets also contain regular samples from research cruises within national jurisdictions.	The CPR is a European scale plankton monitoring programme, focusing on the shelf and offshore regions. Regular (monthly) fisheries and research cruises could be used for sample collection.
*A complementary need exists for both long-term time-series as well as high frequency monitoring, particularly in habitats considerably influenced by anthropogenic pressures.			

2.5 Quality assurance/ Quality Control

Belgium - Vlaams Instituut voor de Zee (VLIZ)

Through regular sampling surveys, the Flanders Marine Institute (VLIZ) is generating a long-term data series for the Belgian coastal water and sandbank system, a designated site in the Long Term Ecological Research (LTER) network. The data series is built from sampling activities initiated in 2012 in the framework of the LifeWatch marine observatory, a Flemish contribution to the LifeWatch ESFRI by VLIZ. Nine nearshore stations are sampled monthly, with an additional eight offshore stations sampled seasonally. Zooplankton densities and size measurements are measured, using a ZooScan plankton imaging device together with the ZooProcess and Plankton Identifier software packages. The entire methodology and QA/QC practices are described in Mortelmans et al. (2019)(https://doi.org/10.1002/gdj3.68) whereas associated metadata and simultaneous water quality parameters (e.g., nutrient and pigment concentrations; salinity; temperature) are described in Flanders Marine Institute (2022)(https://doi.org/10.14284/543).

Denmark – Aarhus University (AU)

The analysts of the Danish samples do intercalibrations every five years. The analysts are taxonomically trained continuously and updated with taxonomic changes. The data are stored in a database and undergo screening for outliers that are further investigated, followed by annual reporting that include scientific screening for deviation in the data

Germany - Landesamt für Landwirtschaft, Umwelt und ländliche Räume des Landes Schleswig-Holstein (LLUR)

The analyst of the German samples from Schleswig-Holstein regularly participates in internationally, e.g. within HELCOM, and nationally available phytoplankton ring tests. Their analyst has been continuously analysing the plankton samples for more than 20 years. Their knowledge of taxonomic changes is continuously updated.

Germany - Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten und Naturschutz (NLWKN)

The phytoplankton samples of NLWKN monitoring have been analysed by an independent lab (AquaEcology GmbH & Co. KG, Oldenburg Germany) since 2006. AquaEcology operates a quality management system according to DIN EN ISO/IEC 17025 and is a member of the biological and chemical quality assurance group of the Federal Environment Agency in Germany. The processing of phytoplankton samples is conducted according to the "Standard operation procedure for laboratories of the German Marine Monitoring Programme - Phytoplankton investigations of coastal surface waters (qualitative and quantitative)" (SOP-BLMP-PP 2009).

Plankton analysts and technicians of AquaEcology regularly participate in national and international biological intercalibrations and ring tests, e.g. using the services of IPI or HELCOM. The analysts are taxonomically trained continuously and updated with taxonomic changes.

Netherlands – Rijkswaterstaat (RWS)
Analysis of the Rijkswaterstaat samples is carried out in accredited laboratory facilities following NEN-EN-ISO/IEC 17025:2017. Senior analysts take part in yearly intercalibration exercises via the IPI programme since 2009.

Portugal - Instituto Portugues do Mar e da Atmosfera (IPMA)
The Portuguese laboratory has been working with marine microalgae for more than 30 years, having a significant focus on the portuguese national program of HAB monitoring. For each evaluation period (2 years), all analysts perform at least one international laboratory intercomparison (IPI-IOC) or at least one Training Course on Identification of Harmful Marine Algae (IOC-UNESCO-Univ. Copenhagen). The analysts are continuously updated with taxonomic information.

Spain - Instituto Español de Oceanografía (IEO)
Both phyto- and zooplankton datasets include data from three different IEO-CSIC centres (Vigo, A Coruña and Gijón), where the samples are analysed by their well-trained taxonomic specialists. Data from the different centres are compiled and processed to unify the format and the nomenclature.

Sweden – Swedish Meteorological and Hydrological Institute (SMHI)
The analysts of the Swedish samples do yearly intercalibrations using either the service of IPI or HELCOM. The analysts are taxonomically trained continuously and updated with taxonomic changes.

United Kingdom - Centre for Environment, Fisheries and Aquaculture Science (Cefas)
Cefas analysts have undertaken IPI (formerly BEQUALM) ringtrials since 2006 and the lab holds ISO 17025 accreditation for Phytoplankton analysis. Analysis is undertaken to meet IOC UNESCO Manual on Phytoplankton Analysis methods, and includes internal QC on 1 in 30 samples, and external QC with a partner lab on an additional 1 sample in 80. Senior staff members have undertaken Freshwater Species ID courses.

United Kingdom – Environment Agency (EA)
APEM, the lab that analyses the UK Environment Agency samples, has gained UKAS accreditation to ISO/IEC17025:2005 for marine and freshwater phytoplankton analysis, the method is fully compliant with CEN standard EN – 15204 'Guidance on the enumeration of phytoplankton using inverted microscopy (Utermöhl technique)', and the quality management system is a ISO 9001:2015 certified. Internal QC analysis is undertaken routinely on 1 in 30 samples and consists of original analysis, original repeat QC, and subsample repeat QC. All analysts have a degree in a relevant discipline and routinely complete internationally recognised external quality assurance schemes including IPI (formerly known as BEQUALM) for marine and the External Quality Assessment Trials Phytoplankton (EQAT) for freshwater. In addition, APEM also participates in an interlab AQC with a partner lab.

United Kingdom – Marine Biological Association (MBA)
The CPR has a QA/QC method which has remained virtually unchanged since 1948. MBA procedures are documented, plankton analysts have International Phytoplankton Intercomparison (IPI; formerly known as BEQUALM) qualifications and MBA chairs the NE Atlantic Marine Biological Analytical Quality Control (NMBAQC) scheme which is working to develop first a standard and then quality control scheme for zooplankton analysis.

United Kingdom – Marine Scotland Science (MSS)
MSS phytoplankton data follows the MSS joint code of practice for data quality. Phytoplankton analysts participate in the International Phytoplankton Intercomparison/BEQUALM exercise annually since it started

as well as in annual internal sample QC and validation exercises. Zooplankton quality assurance follows the MSS joint code of practice and analysts participate in the NE Atlantic Marine Biological Analytical Quality Control (NMBAQC) external identification ring trials every 2 years.

United Kingdom – Plymouth Marine Laboratory (PML)
The main Plymouth Marine Laboratory (PML) analyst for L4 was trained by a skilled MBA CPR analyst and holds NMBAQC qualifications.

United Kingdom – Scottish Association of Marine Science (SAMS)
Because SAMS also does commercial work for Food Standards Scotland (FSS), monitoring algae that are potentially contaminants of commercially grown or harvested shellfish, we use the SOP specified by the United Kingdom National Reference Laboratory for marine biotoxins based at Agri-Food and Biosciences Institute (AFBI) in Belfast (https://www.afbini.gov.uk/articles/nrl-marine-biotoxins-procedures-and-links). Additionally, SAMS uses either SOP written for laboratory Health & safety purposes or methods unchanged since 1970 and described by Tett, P. (1987). Plankton. In: Biological Survey of Estuaries and Coasts. J. Baker and W. J. Wolff (eds), Cambridge University Press: 280-341.

2.6 Data reporting, handling, and management

- *Reporting format (Available via a link in the CEMP Appendices)*
- *Data metadata schema (Link to ODIMS, INSPIRE compliant)*
    - Each dataset is responsible for its own metadata
- *Confidence levels in data*
    - Confidence levels in the datasets used for this analysis are established by the quality control measures described in **2.5 Quality assurance/ Quality Control**.
    - Information on QA/QC of other datasets will be included as those datasets are incorporated.
- *Data flows described (Additional to information in CEMP Appendix)*
    - Individual datasets submitted to OSPAR are forwarded to DASSH and given a unique Digital Object Identifier (DOI) which can be used to link assessment outputs to specific dataset versions
    - New datasets are scrutinised to ensure they do not contain any missing (blank, NA or NaN) values or Aphia IDs and to ensure that data conforms to the formatting requirements issued with the data call. Any issues identified at this stage need to be rectified with the data provider before proceeding further.
    - New datasets are also scrutinised to ensure that their spatial and temporal distribution are appropriate for the current assessment (i.e. samples collected within OSPAR Regions II, III, IV jurisdictions; samples collected within the assessment period and prior to the assessment period). Fixed-point time-series data collected within transitional waters do not technically fall within MSFD jurisdiction (e.g. MSS, SAMS), but can still be included in assessment if this is made clear.
    - Datasets which do not satisfy the checks described above cannot be used for assessment but can still be hosted on DASSH for other researchers to use.
    - Once a dataset has been received by DASSH and has passed initial scrutiny, new species will be ingested into the Master Taxa List which assigns biological traits to taxa, allowing them to be sorted into lifeforms for PH1/FW5.

    - The new dataset's species list will be compared with the Master Taxa List via Aphia IDs, and any new species not currently represented on the list will be identified. This process ensures that each species is only entered in the database once.
    - Any new species will be manually assigned functional traits by obtaining expert advice directly from data providers, Pelagic Habitats Expert Groups, and by searching the literature. Data providers may be approached to provide more information if there are taxa in their dataset which are not currently represented on the Master Taxa List.
    - Once traits have been assigned for new species, they can be added to the Master Taxa List. Missing trait information leads to identified taxa not being included in the lifeform analysis.
    - New datasets can then be uploaded to DASSH, along with the updated version of the Master Taxa List with functional traits for all new taxa added.
    - New datasets with suitable spatial and temporal distribution can now be used for assessment by using a set of six sequential R-scripts described in section **3.2 Preparation of data**.
- *Data storage*
    - Plankton abundance data is quality controlled to ensure consistency. Data is stored on DASSH and documents are stored on the OSPAR SharePoint for the QSR.
    - Plankton datasets stored on DASSH are publicly accessible through the Plankton Lifeform Extraction Tool (PLET; Ostle et al. 2021): https://www.dassh.ac.uk/lifeforms/.
    - The PLET uses the Master Taxa List to extract abundance time series of plankton functional groups, or "lifeforms", according to sets of shared biological traits. The purpose of the PLET is to make complex plankton datasets accessible and meaningful for policy, public interest, and scientific discovery.
    - Visit https://github.com/hollam2/PH1_PLET_tool to download an R-script tool which calculates the PH1/FW5 indicator directly from a CSV file output from the PLET.

3 Assessment

3.1 Data acquisition

- *How you extract the data specifically for your assessment question*
    - Data were extracted from individual plankton abundance datasets collected and maintained by the institutes described in **Table 1**. Additional data have been submitted via the pelagic data call which came out in 2021, but those data have not been incorporated into the assessment due primarily to insufficient temporal extent.

3.2 Preparation of data

- *Normalisation of data (If it has come from different monitoring methods)*
    - Because the data are from different sources, it is important that they are kept separate and not directly combined.
- *Aggregation and integration of data acquired*
    - Data are aggregated into means for each calendar month (e.g. January 1960, February 1960, etc).
    - Across the Greater North Sea, Celtic Seas, Bay of Biscay and Iberian Coast (OSPAR Regions II, III, IV), data are analysed at the scale of COMP4 assessment units (Enserink et al. 2019).

- The Master Taxa List should be used to create plankton lifeforms, which are defined based on common functional traits (Ostle et al. 2021). This is maintained/updated by the COBAM pelagic group and a copy is held by OSPAR. The most current version is available from DASSH: https://doi.org/10.17031/1709.

- The R-code (six scripts to be run sequentially) for data processing can be found at:

    - https://github.com/hollam2/PH1-FW5_Change_in_plankton_communities
    - Read the associated README file in the GitHub repository
    - Download the compressed zip file from the "Code" drop-down menu for the GitHub repository. This file contains the raw data, R-scripts and directory structure required for this analysis.
    - Extract the compressed files which
    - Run the six scripts sequentially, as each script generate interim outputs which are required by subsequent scripts
    - Note: To run the full analysis across all datasets may take several hours
    - The purpose of each R-code script is described below:
        - **1-RAW_CPR_loader**: This script reads raw CPR data and prepares it to be in list format
            - CPR taxa IDs are converted into Aphia IDs using an included CSV file.
            - Aphia IDs are used to download consistent taxon names from WoRMS.
        - **2-RAW_plankton_loader**: This script reads raw plankton data from institutional datasets and outputs aggregated lifeform abundances.
            - Aphia IDs are used to download consistent taxon names from WoRMS.
            - This data is combined with the CPR data from the previous script.
            - Lifeforms are identified from a Master Taxa List so that lifeform abundance values can be calculated.
            - The lifeform designations for each taxon (a taxon can belong to multiple lifeforms) are used to tag the raw data to construct time-series for individual taxa within lifeforms later in the dataflow.
        - **3-PROCESSED_spatial**: This script reads processed plankton lifeform abundance data and applies a shapefile to extract lifeform time-series for each polygon.
            - Mean lifeform abundance values per month are calculated separately for samples falling within each polygon. Fixed-point time series are calculated in the same manner.
            - Inverse distance weighted interpolation is used to generate two-dimensional surfaces representing monthly lifeform abundance.
            - Interpolated data is extracted by calculating a mean value within each polygon. These values are used to fill gaps in lifeform time-series wherever possible.
            - The raw plankton data are tagged with IDs for the polygons containing each sample.
            - Spatial and temporal confidence is estimated for each assessment unit
        - **4-PH1_indicator**: This script reads processed and spatially referenced plankton abundance time-series data and calculates the indicator results (lifeform pairs indicator and Kendall statistic)
            - Years from each time-series with less than 8 months represented are excluded.
            - Linear interpolation is used to fill gaps of 3 months or less in each time-series. This step and the previous step are used to ensure that mean annual abundance values are not seasonally biased.

- The PH1/FW5 indicators (lifeform pairs indicator and Kendall statistic) are then calculated for each lifeform time-series, generating a set of figures in an "output" folder.
- The indicator results (lifeform pairs indicator and Kendall statistic) are applied across the polygons of a shapefile to provide a spatial context for change in lifeform abundance. Whenever there are multiple datasets representing the same polygon, the dataset with the most unique months of samples is displayed.
- The tagged raw data are also used to construct time-series for individual taxa for each polygon/fixed-point station and lifeform, providing a means to identify which taxa are driving change in lifeform abundance. The taxon time-series are also used to generate figures in the output folder.

▪ **5-Environmental_drivers**: This script is used to identify potential links between environmental variables (e.g. nutrient concentration, climate indices, sea surface temperature) and plankton lifeform abundance.
- Environmental variables have been pre-extracted from monthly gridded NC files, using the same shapefile which was used to extract lifeform time-series.
- A lifeform-specific subset of environmental variables is selected to model the abundance of each lifeform. Not all environmental variables are used for each lifeform to avoid potentially spurious correlations (e.g. using nutrient concentration to predict crustacean abundance).
- Monthly lifeform abundance and environmental data are separated into training and testing sets for cross-validation.
- Missing values for environmental variables within the training and testing sets are imputed.
- A 12-month moving window smoother is applied to training and testing sets to remove seasonality and extract the long-term trends.
- The Boruta algorithm, a wrapper around random forest, is applied to rank the relative importance of a subset of environmental variables to predict variation in lifeform abundance separately for each polygon and fixed-point time-series.
- Random forest models are generated from the environmental variables selected in the previous step and models are validated by predicting on the testing set.

▪ **6-Integration**: This script loads the results from the previous script and integrates them with the Kendall statistics for the plankton lifeform abundance time-series.
- A matrix is generated containing the variable importance scores for each lifeform time-series.
- Important predictors are reported for time-series with significant positive correlation between predicted and observed values ($p \leqslant 0.05$).
- Indicator results are collated for pelagic habitat types within each OSPAR Region (II, III and IV).
- Indicator results and confidence are reported for each lifeform based on the majority trend for each habitat type within an OSPAR Region.

3.3 Assessment criteria

- *Defining assessment unit/scale (Temporal and spatial)*
  - For the Greater North Sea, Celtic Seas, Bay of Biscay and Iberian Coast we report state according to COMP4 assessment units (Enserink et al. 2019).
- *Baseline/ reference condition / assessment value*

- The starting conditions will be set as the start of each time-series and the entire time-series will be evaluated to identify any long-term trends in abundance. The assessment value will be evaluated as "absence of a significant increasing or decreasing trend".
- In addition to evaluating long-term change over time, samples from the assessment period will be used to define a contemporary reference envelope to represent conditions in the assessment period; for the QSR2023 the period 2015-2019 was selected. A comparison was made with all previous months in each time-series (for example, this was 1960-2014 for the CPR dataset and 1988-2014 for the PML zooplankton dataset). In accordance with our proposed assessment value the absence of a significant trend in an indicator or lack of a significant correlation between the indicator trend and the trend in a human pressure will be used as evidence that the assessment value has been met (for that criterion and the plankton community as a whole). However, this presupposes that the starting point of the time-series represented Good Environmental Status. This may not be the case. Where data exist, it will be necessary to use this to determine the current status of the plankton at those locations but more than 10 years of data will have to be collected to characterise the status of the plankton. If, however, existing data sets can be used to characterise Good Environmental Status for plankton communities (using ecological theory, modelling, the absence of obvious human pressure and expert opinion), it may be possible to use such data as reference conditions for new monitoring sites and existing sites at which the status of the plankton does not meet GES.

- *Proposed assessment value*
  - Plankton community is not experiencing long-term change. If it is experiencing long-term change, it is not significantly influenced by anthropogenic pressure.

3.4 Spatial Analysis and / or trend analysis

- *Statistical analysis (e.g. Method for trend analysis, Establishment of confidence limits)*

Previous assessment

The previous assessment (Intermediate Assessment 2017; IA2017) was based largely on state space theory (see below) and examined differences in the relative abundances of ecologically relevant lifeform pairs between a past comparison period (2004-2008) and a contemporary assessment period (2009-2014). While this method is suitable for detecting changes in the annual cycle of ecologically linked lifeforms, it is less suited to evaluating gradual change over time. Further, stronger statistical links can be drawn by examining relationships between the abundances of distinct planktonic lifeforms and environmental pressures, than between an index of lifeform pairs and environmental pressures (Bedford et al. 2020). For the current assessment a simplified approach was developed, based on measuring long-term changes in the abundance of key lifeforms and evaluating how environmental pressures co-varied with changes in lifeform abundance. A long-term perspective helped improve confidence in establishing links to environmental pressures. Short-term changes within the assessment period were also evaluated to identify emerging trends. The PH1 indicator as originally defined in IA2017 was still evaluated for this assessment and is described under the **Lifeform pairs indicator approach** section of **Section 3.4**.

Kendall trend test and statistic

Abundance trends for planktonic lifeforms can be evaluated over time by applying the Kendall trend test to annual mean abundance values (Bedford et al. 2020; Desmit et al. 2020). The test was performed on annual $\log_{10}$ transformed mean abundance values, rather than monthly or seasonal values, to remove the seasonal variation typical of plankton time-series data (**Figure 1**). This

nonparametric test generates a statistic which is derived by comparing each value in a time-series with each of the values preceding it. If a latter value is greater than a previous one, the pairwise comparison is assigned a value of 1. If it is lower it is assigned a value of -1, with 0 assigned to cases when there is no difference between values. The sum of the pairwise comparisons for the time-series produces Kendall's S-statistic. The variance in the S-statistic is used to derive a Z-score with an approximately normal distribution; thus, confidence in this statistic can be assessed with an associated p-value, with $p \leq 0.05$ generally accepted as statistically significant change. The sign of the test statistic reveals the direction of the trend, with a positive statistic indicating an increasing trend and a negative statistic indicating a decreasing trend. The magnitude of the statistic is proportional to the strength of the trend. A great benefit of this nonparametric test is that it yields identical results irrespective of the data transformation method and is not sensitive either to gaps in data or to non-linear or irregular trends. It also has the advantage that abundance trends are comparable across datasets, lifeforms, and assessment units.

Time-series plots (**Figure 1**) indicate variation in lifeform abundances through time for a single assessment unit, with the blue lines displaying monthly variability (thinner line), and annual mean abundance (thicker line) values used to derive the Kendall statistic. Data correspond to the assessment unit representing the western entrance to the English Channel ('Channel well mixed' see **Figure a**) and contains the fixed-point station 'L4', consistently monitored by the Plymouth Marine Laboratory since 1988.

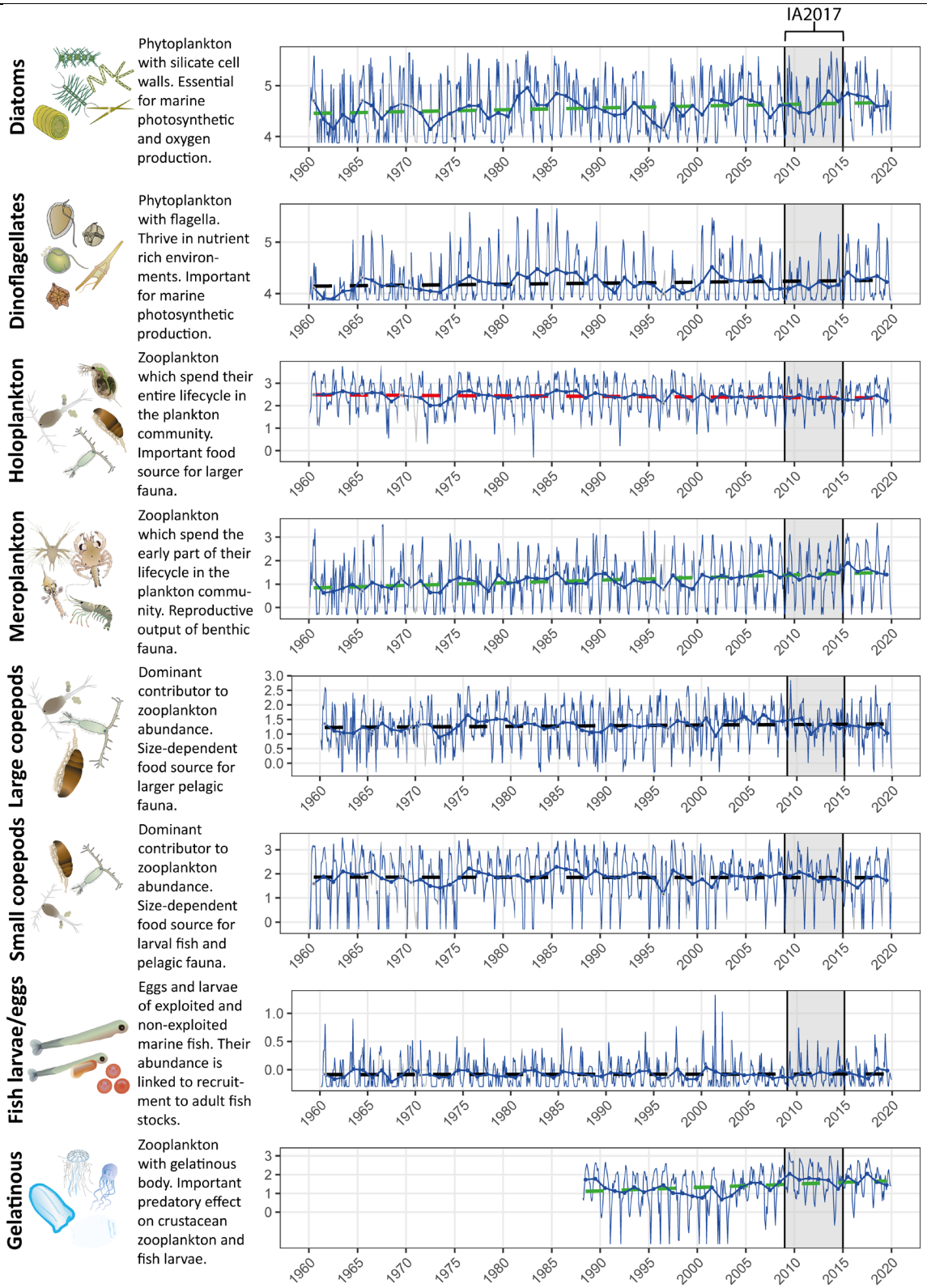

Figure 1: Long-term monthly and annual log_{10} transformed abundance time-series for eight plankton lifeforms in the Western Channel. Blue lines display monthly variability (thinner line), and annual mean abundance (thicker line) values. Dashed lines indicate linear trend lines in annual abundance without any inference on statistical significance. The Kendall trend test is used to infer significance of trends, with red: decreasing trend, green: increasing trend, and black: no trend. Data obtained from the Continuous Plankton Recorder (CPR) survey and Plymouth Marine Laboratory (for gelatinous zooplankton only due to non-quantitative capture of gelatinous taxa by CPR). The shaded region represents the time-period of IA2017. Plankton images courtesy of the Integration and Application Network, University of Maryland Center for Environmental Science (ian.umces.edu/symbols/).

The ability to aggregate complex plankton data into ecologically relevant lifeforms is a useful tool for assessment because it can help identify where changes are occurring (McQuatters-Gollop et al. 2019). The ability to understand which taxa are driving changes in a lifeform can also be valuable. By calculating the Kendall statistic for each taxon, additional information can be gleaned to assist in interpreting lifeform time-series trends. For example, **Figure 2** displays change in the abundance of diatoms for the 'Channel well mixed' COMP4 assessment unit described in **Figure 1**. The abundance of diatoms has been increasing since 1960 ($z = 2.66$, $p \leq 0.05$). Through examining the annual abundance time-series for individual diatom taxa it becomes possible to also conclude that overall trend of increasing diatom abundance is driven by positive abundance trends in 11 out of 54 diatom taxa. Increasing trends and greater absolute abundances for *Proboscia alata* and *Thalassiosira spp* indicate that they are predominantly responsible for the increasing abundance trend in diatoms. Additionally, while the net trend is increasing, four diatom taxa (*Guinardia striata*, *Skeletonema costatum*, *Thalassionema nitzschioides*, and *Odontella sinensis*) are actually decreasing in abundance.

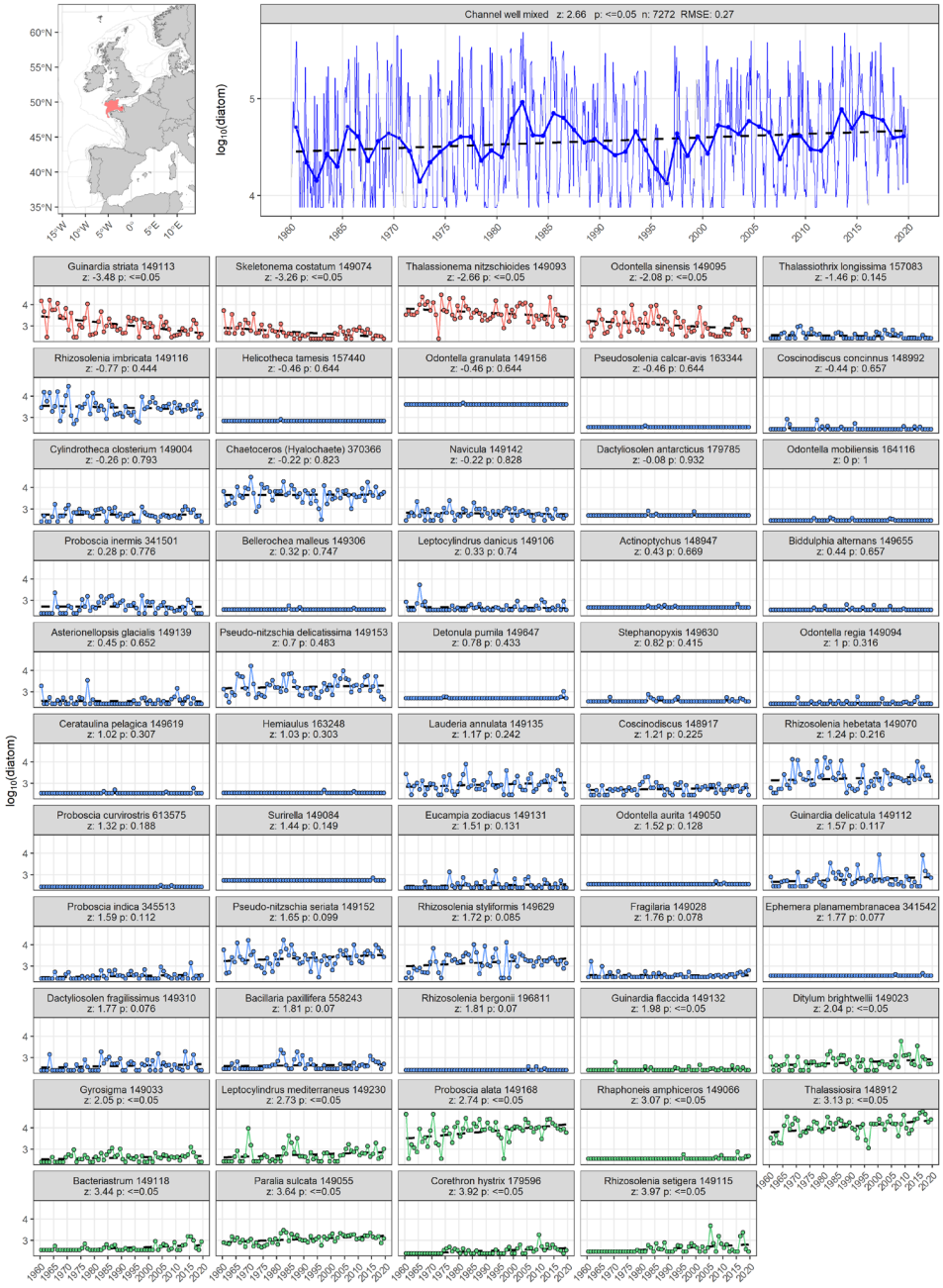

Figure 2: Long-term log_{10} transformed abundance time series for diatoms in the 'Channel well mixed' COMP4 assessment unit. Data obtained from the Continuous Plankton Recorder (CPR) survey. Individual diatom taxa abundance time series are displayed as facets. Facet labels indicate taxon scientific name and Aphia ID, z: derived from Kendall's S-statistic, and p: Kendall p-value. Taxon time-series are coloured according to the outcome of the Kendall trend test, with red: decreasing trend and $p \leq 0.05$, blue: $p > 0.05$ (no trend), green: increasing trend and $p \leq 0.05$.

Since this assessment was based on data from multiple sources, the Kendall statistic was calculated independently for each combination of dataset, assessment unit, and lifeform. Since the datasets used also varied in duration, to avoid discarding data the full duration of each dataset was assessed up until the end of 2019. For a dataset to be included in this analysis, it needed to contain samples collected within the assessment period (2015-2019) and prior to the assessment period.

Spatial Scale

Because plankton community composition, distribution, and dynamics are closely linked to their environment, the analysis was performed at the scale of the 'COMP4 assessment units' (COMP4 v8a; **Figure 3**, **Table 1**). Assessment units within the Greater North Sea and Celtic Sea (OSPAR Regions II and III, respectively) were initially developed by Deltares and partner institutes as part of the EU 'Joint Monitoring Programme of the Eutrophication of the North Sea with Satellite data' (JMP-EUNOSAT; Enserink et al. 2019) and further refined in the revision process of the eutrophication assessment by OSPAR expert groups ICG-EMO and TG-COMP. Assessment units with similar phytoplankton dynamics were derived from cluster analysis of satellite data for chlorophyll a and primary production. Boundaries between assessment units were derived by relating clustering results to the best-matching gradients in environmental variables obtained from the three-dimensional hydrodynamic Dutch Continental Shelf model version 6 (DCSMv6 FM). The variables which best matched the divisions highlighted by clustering were depth, salinity, and stratification regime. Additional geographic areas were added such as the Channel, Irish Sea, and Kattegat. These assessment units are a geographical representation of the conditions most likely to drive plankton distribution, dynamics, and community composition. The assessment units are being regularly updated and may not perfectly reflect the true state of the pelagic environment in their present form.

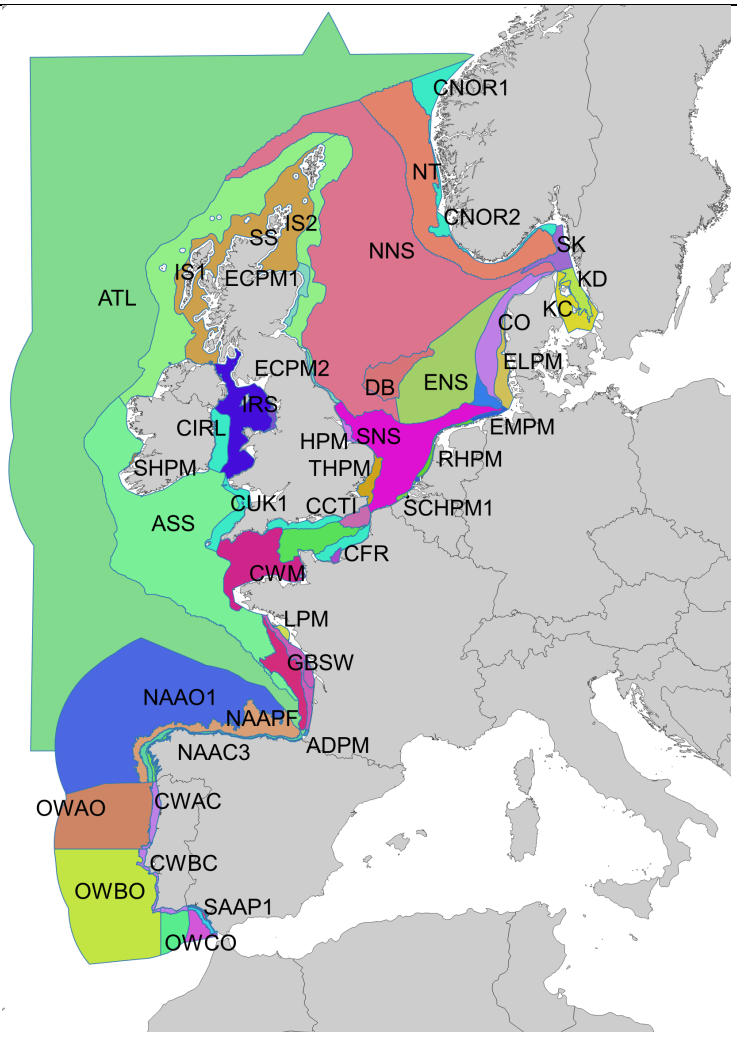

Figure 3: COMP4 assessment units developed by JMP-EUNOSAT (Enserink et al. 2019) and OSPAR. Assessment unit codes displayed in this figure are referred to in **Table 1**.

Table 1: COMP4 assessment unit definitions and categorisation according to pelagic habitat type (variable salinity, coastal, shelf or oceanic / beyond shelf) and OSPAR Region. Additional fixed-point stations which were not part of COMP4 but were evaluated in this indicator analysis are described at the end of the table.

COMP 4 assessment units categories	Unit code	Unit name	Salinity (surface mean)	Depth (mean)	OSPAR Region
Variable salinity assessment units	ADPM	Adour plume	34.4	87	IV
	ELPM	Elbe plume	30.8	18	II
	EMPM	Ems plume	31.4	19	II
	GDPM	Gironde plume	33.5	34	IV
	HPM	Humber plume	33.5	16	II
	LBPM	Liverpool Bay plume	30.6	15	III
	LPM	Loire plume	33.8	38	IV
	MPM	Meuse plume	29.3	16	II
	RHPM	Rhine plume	31.0	17	II
	SCHPM1	Scheldt plume 1	31.4	13	II
	SCHPM2	Scheldt plume 2	30.9	15	II
	SHPM	Shannon plume	34.1	61	III
	SPM	Seine plume	31.8	25	II
	THPM	Thames plume	34.4	22	II
Coastal assessment units	CFR	Coastal FR channel	34.2	33	II
	CIRL	Coastal IRL 3	34.0	65	III
	CNOR1	Coastal NOR 1	34.3	190	II
	CNOR2	Coastal NOR 2	34.0	217	II
	CNOR3	Coastal NOR 3	32.4	171	II

	CUK1	Coastal UK 1	34.5	60	III
	CUKC	Coastal UK channel	34.8	37	II
	CWAC	Coastal Waters AC	No information	No information	IV
	CWBC	Coastal Waters BC	No information	No information	IV
	CWCC	Coastal Waters CC	No information	No information	IV
	ECPM1	East Coast (permanently mixed) 1	34.8	73	II
	ECPM2	East Coast (permanently mixed) 2	34.5	43	II
	GBC	German Bight central	33.4	39	II
	IRS	Irish Sea	33.7	65	III
	KC	Kattegat Coastal	25.7	21	II
	KD	Kattegat Deep	27.6	50	II
	NAAC1A	Noratlantic Area NOR-NorC1	No information	No information	IV
	NAAC1B	Noratlantic Area NOR-NorC1	No information	No information	IV
	NAAC1C	Noratlantic Area NOR-NorC1	No information	No information	IV
	NAAC1D	Noratlantic Area NOR-NorC1	No information	No information	IV
	NAAC2	Noratlantic Area NOR-NorC2	No information	No information	IV
	NAAC3	Noratlantic Area NOR-NorC3	No information	No information	IV
	OC	Outer Coastal DEDK	33.4	27	II
	SAAC1	Sudatlantic Area SUD-C1	No information	No information	IV
	SAAC2	Sudatlantic Area SUD-C2	No information	No information	IV
	SAAP2	Sudatlantic Area SUD-P2	No information	No information	IV
	SNS	Southern North Sea	34.3	32	II
Shelf assessment units	ASS	Atlantic Seasonally Stratified	35.2	134	III, IV
	CCTI	Channel coastal shelf tidal influenced	34.8	40	II
	CWM	Channel well mixed	35.1	77	II, III
	CWMTI	Channel well mixed tidal influenced	35.0	59	II
	DB	Dogger Bank	35.1	28	II
	ENS	Eastern North Sea	34.8	43	II
	GBCW	Gulf of Biscay coastal waters	34.6	53	IV
	GBSW	Gulf of Biscay shelf waters	34.9	107	IV
	IS1	Intermittently Stratified 1	35.3	138	II, III
	IS2	Intermittently Stratified 2	35.1	102	II
	NAAP2	Noratlantic Area NOR-NorP2	No information	No information	IV
	NAAPF	Noratlantic Area NOR-Plataforma	No information	No information	IV
	NNS	Northern North Sea	35.0	121	II
	NT	Norwegian Trench	34.1	349	II
	SAAP1	Sudatlantic Area SUD-P1	No information	No information	IV
	SK	Skagerrak	31.8	134	II
	SS	Scottish Sea	35.1	89	II, III
Oceanic / beyond shelf assessment units	ATL	Atlantic	35.3	2291	II, IV, V
	NAAO1	Noratlantic Area NOR-NorO1	No information	No information	IV
	OWAO	Ocean Waters AO	No information	No information	IV
	OWBO	Ocean Waters BO	No information	No information	IV
	OWCO	Ocean Waters CO	No information	No information	IV
	SAAOC	Sudatlantic Area SUD-OCEAN	No information	No information	IV

Fixed-point station categories	**Unit code**	**Data provider**	**Salinity (surface mean)**	**Depth (mean)**	**OSPAR Region**
Variable salinity fixed-point stations	Norderney	NLWKN	No information	No information	II
	LPO	Scottish Association of Marine Science	No information	No information	III
	Loch Ewe	Marine Scotland Science	No information	No information	III
Coastal fixed-point stations	L4	Plymouth Marine Laboratory	No information	No information	II
	NOVANA	Aarhus University	No information	No information	II
	Scalloway	Marine Scotland Science	No information	No information	II
	Scapa	Marine Scotland Science	No information	No information	II
	Stonehaven	Marine Scotland Science	No information	No information	II

	RADIALES	Instituto Espanol de Oceanografia	No information	No information	IV

Because the Bay of Biscay and Iberian Coast (OSPAR Region IV) extended beyond the boundaries of the DCSMv6 FM, assessment units within this region were developed using a different methodology, based on phytoplankton dynamics (Spain) and salinity dynamics (Portugal). To delineate assessment units for the Spanish coast, a polygon was created to extend from the coast to the exclusive economic zone (EEZ) boundary. Daily NASA Aqua/MODIS Level-2 satellite images (https://doi.org/10.5067/AQUA/MODIS/L2/OC/2018) were used to calculate climatological mean values of chlorophyll *a* for each pixel. K-means clustering was then used to group pixels with similar dynamics, resulting in six distinct groupings within the main Spanish polygon. Portugal's three Water Framework Directive assessment units were extended to the boundaries of the Portuguese EEZ. These assessment units were further divided longitudinally to separate pelagic waters from coastal waters more subject to eutrophication from river influence by applying a salinity threshold, followed by a bathymetry threshold.

COMP4 assessment units and fixed-point stations were categorised according to the OSPAR Regions they intersected as well as the habitat type they represented (**Table 1**, **Figure 4**). The four habitat types considered were: variable salinity (representing areas of freshwater influence where estuarine plumes extend beyond waters designated as Transitional Waters under Directive 2000/60/EC), coastal (representing areas with mean salinity < 34.5 psu), shelf (representing areas with mean salinity > 34.5 and mean depth < 200 m) and oceanic / beyond shelf (representing areas with mean salinity > 34.5 and mean depth > 200 m).

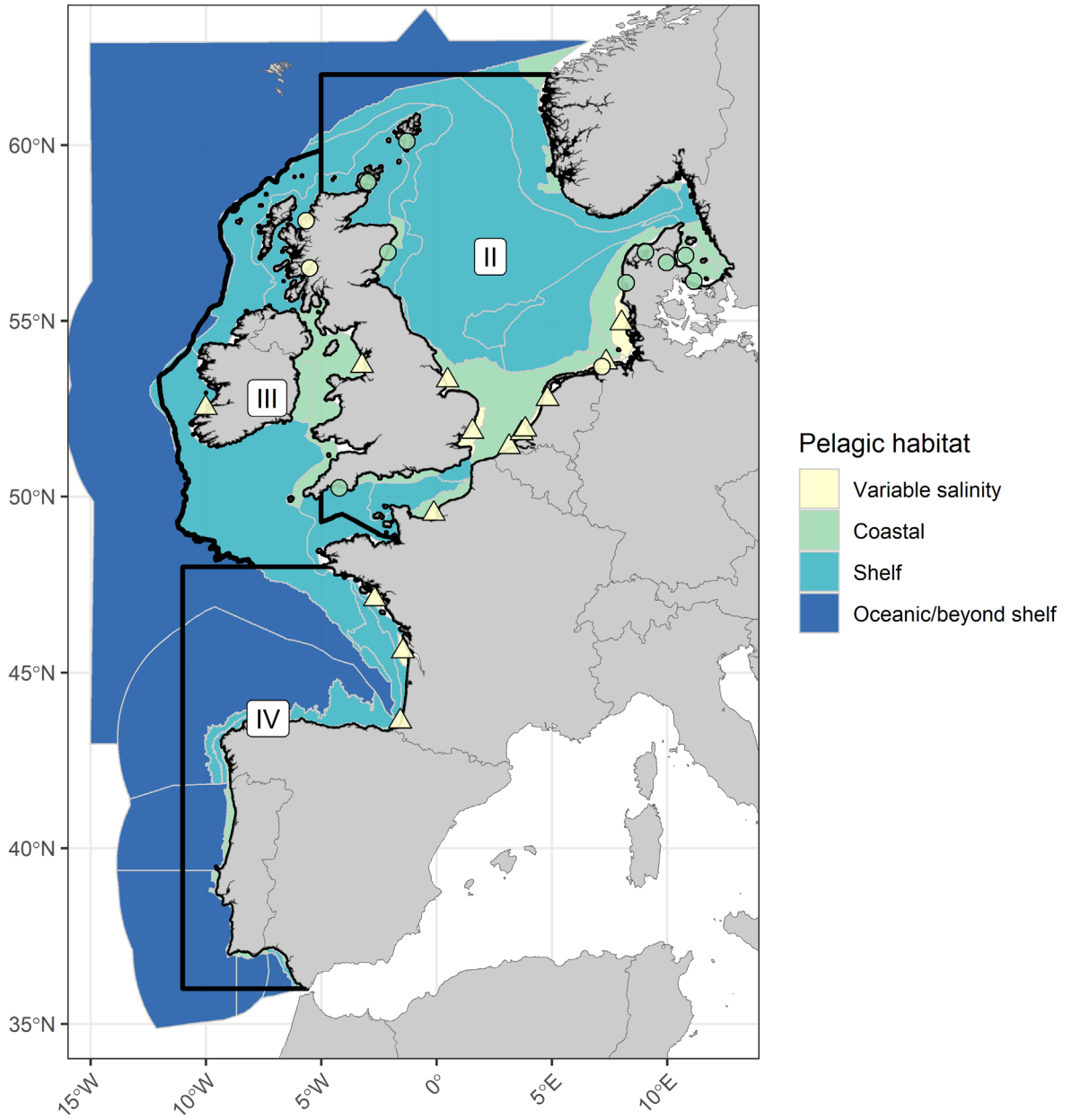

Figure 4: The categorisation of COMP4 assessment units and fixed-point stations according to pelagic habitat type. Three assessed OSPAR Regions (I, II, and III) indicated by black outlines. River plumes are indicated by triangle symbols and fixed-point stations are indicated by circles.

Plankton Data

The assessment has been carried out using 24 phytoplankton and zooplankton datasets from 14 sources (**Table 2**, **Figure 5**). Other datasets were provided but they were out of the scope of this assessment (e.g. time-series with comparison period duration shorter than assessment period duration, time-series ending prior to the assessment period, time-series commencing during assessment period, distributed sampling locations restricted to transitional waters only). These additional datasets were submitted by: Vlaams Instituut voor de Zee (BE), Bundesamt für Seeschifffahrt und Hydrographie (DE), Centre for Environment, Fisheries and Aquaculture Science (UK), Newcastle University (UK), and Natural Resources Wales (UK).

Table 2: Contracting parties and institutes that provided the datasets used for the indicator assessment.

Contracting Party	Institute	Dataset name	Date range

Belgium (BE)	Vlaams Instituut voor de Zee (VLIZ)	LW_VLIZ_zoo	2014-2020
Denmark (DK)	Aarhus University (AU)	NOVANA phytoplankton	1985-2020
Germany (DE)	Landesamt für Landwirtschaft, Umwelt und ländliche Räume des Landes Schleswig-Holstein (LLUR)	OSPAR_LLUR-SH_2010-2020	2010-2020
	Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten und Naturschutz (NLWKN)	OSPAR_NLWKN_1999-19_phyto	1999-2019
Netherlands (NL)	Rijkswaterstaat (RWS)	RWS_Fpzout_2000-2019_phyto	2000-2019
Portugal (PT)	Instituto Portuguêes do Mar e da Atmosfera (IPMA)	PseudoNitzschia vs Dinophysis_IPMA	2002-2020
Spain (ES)	Instituto Español de Oceanografía (IEO)	IEO_RADIALES_Phyto	1989-2016
		IEO_RADIALES_Zoo	1991-2018
Sweden (SE)	Swedish Meteorological and Hydrological Institute (SMHI)	National data_SMHI_Kattegat-Dnr: S/Gbg-2021_116_phyto	1989-2021
		National data_SMHI_Kattegat-Dnr: S/Gbg-2021_116_zoo	1996-2020
		National data_SMHI_Skagerrak-Dnr: S/Gbg-2021_116_phyto	1986-2020
		National data_SMHI_Skagerrak-Dnr: S/Gbg-2021_116_zoo	1996-2020
United Kingdom (UK)	Centre for Environment, Fisheries and Aquaculture Science (Cefas)	Cefas SmartBuoy Marine Observational Network - UK Waters Phytoplankton Data 2001-2019	2001-2019
	Environment Agency (EA)	EA PHYTO 2000-2020	2000-2020
	Marine Biological Association (MBA)	CPR dataset 1960-2019	1960-2019
	Marine Scotland Science (MSS)	MSS Scalloway Phytoplankton dataset	2000-2018
		MSS Loch Ewe Phytoplankton	2000-2020
		MSS Loch Ewe zooplankton	2002-2017
		MSS Scapa Phytoplankton dataset	2000-2020
		MSS Stonehaven Phytoplankton	2000-2020
		MSS Stonehaven zooplankton	1999-2020
	Plymouth Marine Laboratory (PML)	PML_L4 phytoplankton	1992-2020
		PML_L4 zooplankton	1988-2020
	Scottish Association for Marine Science (SAMS)	SAMS-LPO-Phyto-Dec2021	1970-1981, 2000-2017

The data submitted by AU (DK), PML (UK), MSS (UK), NLWKN (DE) and IEO (ES) were from discrete fixed-point stations which were evaluated independently. The data submitted by SAMS (UK) were from multiple stations within a small area of freshwater influence. These data were aggregated and treated as a single fixed-point time-series.

Data from the Continuous Plankton Recorder (CPR) survey, collected by the Marine Biological Association (MBA, UK), consisted of spatially distributed data collected along transects. The CPR survey, provides offshore open ocean data at a broad spatial scale using ships-of-opportunity (Reid et al. 2003). Due to the distributed nature of CPR data, CPR samples are typically aggregated across a grid or a set of polygons at monthly temporal resolution (Bedford et al. 2020). The remaining fixed-point datasets, including VLIZ (BE), LLUR (DE), RWS (NL), IPMA (PT), SMHI (SE), Cefas (UK), and EA (UK), were also treated in this manner since samples were collected at several distributed stations within close proximity to one another. The dataset from IPMA (PT) only records abundances for the genera *Pseudo-nitzschia* (diatom) and *Dinophysis* (dinoflagellate), thus results from the Portuguese data should only be considered as a proxy for diatom and dinoflagellate lifeform abundances.

Data from the different providers were not combined for analysis due to differences in sampling, plankton analysis and enumeration methods. Instead, the datasets were analysed separately. Each dataset has internal QA/QC procedures to ensure consistency and accuracy of the data. Before total lifeform abundance values were $\log_{10}$ transformed, a nominal value equivalent to half the minimum non-zero observed value for each time-series was added to each sample. All spatially distributed data (e.g. excluding AU, PML, MSS, IEO, NLWKN and SAMS) were averaged per month within each assessment unit. In a few cases, this resulted in localised groupings of samples being extrapolated across much larger assessment units (see distribution of CPR samples in **Figure 5 (i)**), as was the case for CPR data along the west of Scotland (Intermittently Stratified 1) and Ireland (Atlantic Seasonally Stratified). For months when there were no samples within an assessment unit, gaps in the time series were filled by extracting a mean value from an inverse distance weighted interpolated surface generated from samples adjacent to the assessment unit (<250 nmi) from the same dataset. Years containing less than eight months of averaged and spatially interpolated values were discarded. Remaining gaps of three months or less were filled via linear interpolation.

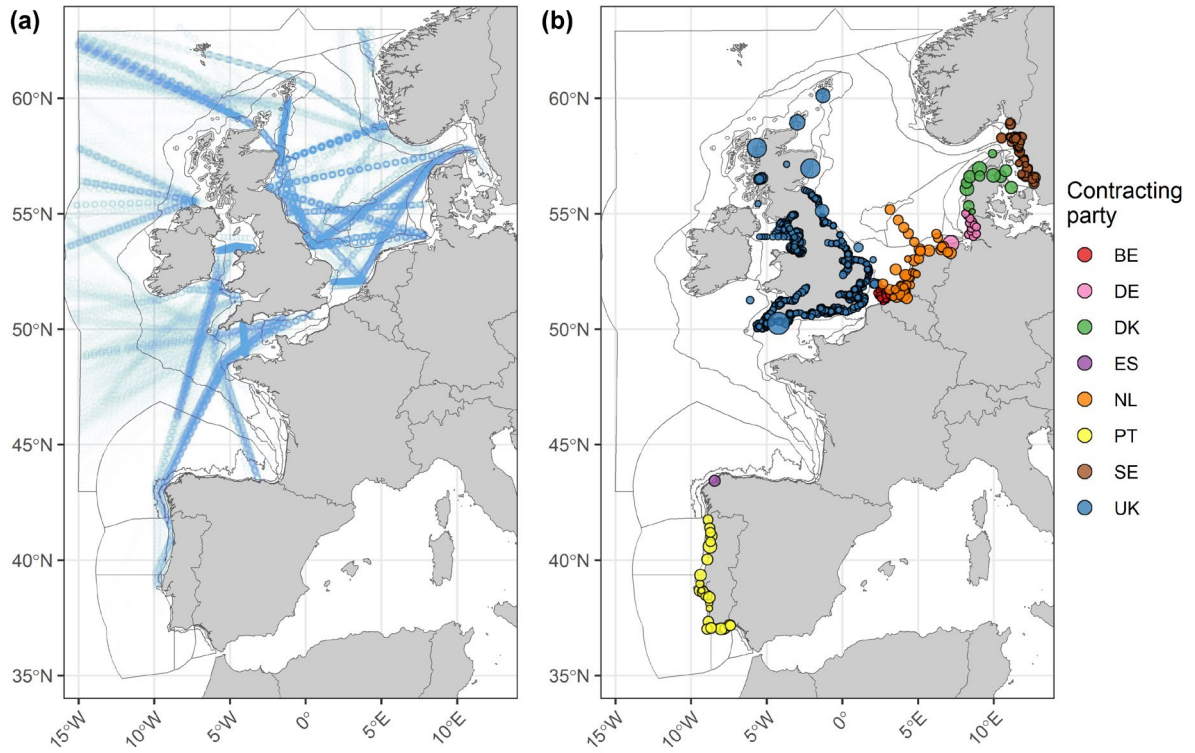

Figure 5: CPR tracks (a) and the locations of samples from all other datasets (b) used in the assessment. For (b), point size is proportional to the number of samples taken at each location. Points are coloured according to contracting party.

Confidence scoring

A confidence scoring methodology, based on an approach developed by ICG-Eut to validate the output from their COMPEAT Tool, was applied to evaluate the robustness of reported trends for each plankton dataset for each assessment unit it intersected. For each assessment unit or fixed-point station temporal confidence was evaluated using two metrics, which we refer to here as specific temporal confidence (STC) and general temporal confidence (GTC).

STC was calculated on an annual basis as the proportion of months when samples were reported out of a total of 12 possible months. For example, if a dataset reported samples within an assessment unit only for June, July and August in a particular year, the STC for that year would be 0.25. GTC was

calculated as the proportion of unique months with recorded samples out of the total number of potential months across the time-series. For example, a dataset commencing in January 2015 and ending in December 2019 contains a total of 60 potential months. If samples were reported for 30 of those months, the GTC would be 0.5. An overall temporal confidence score was then determined by calculating both the mean STC across all years and subsequently the mean of this value with the GTC.

Spatial confidence was also evaluated for distributed datasets such as the CPR. However, for fixed-point stations spatial confidence was not considered since there was no spatial element to the data. For each COMP4 assessment unit spatial confidence was calculated using two metrics, which we similarly refer to here as specific spatial confidence (SSC) and general spatial confidence (GSC).

SSC was calculated on an annual basis based on the proportion of longitudinal and latitudinal ranges of an assessment unit covered by the spatial extent of samples. For example, if an assessment unit spanned longitude from 1 to 5° E, and the distribution of samples collected within that assessment unit within a particular year ranged from 2 to 4° E, the proportion of longitudinal range covered by samples would be 0.5. The same procedure is also applied to the latitudinal range. The SSC for a particular year would be the mean of the longitudinal and latitudinal proportions covered by the spatial extent of samples.

GSC was derived using the same approach as SSC, but across all samples in a dataset that intersect with an assessment unit. An overall spatial confidence score was then determined by calculating the mean SSC across all years and subsequently calculating the mean of this value with the GSC.

Finally, an overall confidence score was calculated as the mean of temporal and spatial confidence. For fixed-point stations, spatial confidence was not considered, therefore only temporal confidence was used to represent the overall confidence of fixed-point station data. For the reporting of trends and linking to pressures, in cases where multiple suitable datasets contained samples from within the same assessment unit, the trends displayed were based on the dataset with the highest confidence score.

Lifeform construction

The eight plankton lifeforms highlighted in this assessment were chosen due to their ecological-relevance and owing to the high confidence in their classification (**Table 3**). All lifeforms were defined based on common functional traits. The rationale for selecting the lifeforms and additional criteria containing supplementary information on lifeforms is listed in **Table 4**.

Table 3: Confidence in the definition of each lifeform and reasoning for low confidence where applicable (Ostle et al. 2021).

Lifeform	Confidence	Reason for confidence (where not 'high')
(micro)Phytoplankton	High	
Large microphytoplankton	Medium	Can reliably identify individual plankton species size class but cannot always reliably assign the size trait if the group counted spans taxa that are both larger and smaller than 20 microns.
Small microphytoplankton	Medium	
Diatoms	High	
Dinoflagellates	High	
Autotrophic and mixotrophic dinoflagellates	Medium	Can identify taxa, but assigning feeding mechanism trait is not always clear (see e.g.: discussion in Flynn et al., 2019)
Pelagic diatoms	High	
Tychopelagic diatoms	High	

Potentially toxic and nuisance diatoms	Low	Designation of some algal blooms as "harmful" (i.e.: Harmful Algal Blooms, 'HABS'), relates more to societal assessment than plankton traits, these 'lifeforms' are therefore not currently recommended for use though they are defined in the traits list and will be the focus of future development work. Specific issues include: • The toxin producing diatom genus Pseudo-nitzschia contains both amnesic shellfish toxin-producers which can render shellfish unfit for human consumption and potentially negatively impact the health of marine mammals, and non toxin-producing species/individuals. It is not possible to identify these cells to species level using routine light microscopy; some toxin and non-toxin producing species are morphologically identical. • The genus *Alexandrium* contains both paralytic shellfish toxin- and non-toxin-producing species/strains and it is not possible to distinguish these using routine light microscopy; some toxin and non-toxin producing species are morphologically identical. • The taxon *Karenia mikimotoi* forms high biomass blooms which strips the water column of oxygen which can be fatal to other lifeforms (Barnes et al. 2015). • Not all datasets included in PLET reliably record key species (e.g.: CPR does not record *Alexandrium*)
Potentially toxic and nuisance dinoflagellates	Low	
Ciliates	Low	• Ecological function can be duplicated by mixotrophic dinoflagellates. Ciliates do not preserve well in the standard 0.5% Lugol's iodine preservative used to preserve phytoplankton samples and some (but not all) are too small to be well sampled by many of the datasets currently in PLET.
Holoplankton	Medium	• May not identify taxa specifically enough to determine traits. • Some of the rarer species are resuspended from the seabed and definition of their holo- or meroplanktonic status is difficult
Meroplankton	Medium	
Gelatinous zooplankton	High	
Fish larvae/eggs	High	
Carnivorous zooplankton	Medium	Can identify taxa, but assigning diet trait is unclear, especially at different life stages.
Non-carnivorous zooplankton	Medium	
Crustaceans	High	
Small copepods	High	
Large copepods	High	

Table 4: Each lifeform comprises organisms with particular traits. A query is then used to assign individual species to lifeforms (McQuatters-Gollop et al. 2019).

Lifeform	**Traits**	**Criteria**
Diatoms	'Diatom' only	PhytoplanktonType=Diatom
Dinoflagellates	'Dinoflagellate' only	PhytoplanktonType=Dinoflagellate
Gelatinous zooplankton	'Gelatinous' only	PlanktonType=Zooplankton AND Gelatinous=Y
Fish larvae/eggs	'Fish' only	ZooType=Fish
Carnivorous zooplankton	'Carnivore' only	PlanktonType=Zooplankton AND Diet=Carnivore
Non-carnivorous zooplankton	'Zooplankton' AND either 'Herbivore', 'Omnivore', OR 'Ambiguous'	PlanktonType=Zooplankton AND (Diet=Herbivore OR Omnivore OR Ambiguous)
Crustaceans	'Crustacean' only	Crustacean=Y
Large phytoplankton	'Phytoplankton' AND 'Lg'	PlanktonType=Phytoplankton AND PhytoplanktonSize=Lg
Small phytoplankton	'Phytoplankton' AND 'Sm'	PlanktonType=Phytoplankton AND PhytoplanktonSize=Sm
Phytoplankton	'Phytoplankton' only	PlanktonType=Phytoplankton
Autotrophic and mixotrophic dinoflagellates	'Dinoflagellate' AND either 'Auto' OR 'Auto/Mixo'	PhytoplanktonType=Dinoflagellate AND (FeedingMech=Auto OR Auto/Mixo)
Pelagic diatoms	'Diatom' AND 'Pelagic'	PhytoplanktonType=Diatom AND DiatomDepth=Pelagic
Tychopelagic diatoms	'Diatom' AND 'Tychopelagic'	PhytoplanktonType=Diatom AND DiatomDepth=Tychopelagic

Nuisance and toxin-producing diatoms	'Diatom' AND either 'Toxic' OR 'Nuisance'	PhytoplanktonType=Diatom AND (HAB = Toxic)
Nuisance and toxin-producing dinoflagellates	'Dinoflagellate' AND either 'Toxic' OR 'Nuisance'	PhytoplanktonType=Dinoflagellate AND (HAB = Toxic)
Holoplankton	'Holoplankton' only	Habitat=Holoplankton
Meroplankton	'Meroplankton' only	Habitat=Meroplankton
Large copepods	'Copepod' AND 'Lg'	Copepod=Y AND ZooSize=Lg
Small copepods	'Copepod' AND 'Sm'	Copepod=Y AND ZooSize=Sm
Ciliates	'Ciliate' only	PhytoplanktonType=Ciliate
Microflagellates	'Dinoflagellate' AND 'Sm'	PhytoplanktonType=Dinoflagellate AND PhytoplanktonSize=Sm

The database master species list (Ostle et al. 2021) was built by assigning functional traits to each species for a dataset and then adding additional new datasets to expand the master species list (**Figure 6)**. The species list for each new dataset was assigned a unique Aphia ID via the World Register of Marine Species (WoRMS). The new dataset's species list was then compared with the plankton database's master species list via Aphia IDs and any new species were identified. This process ensures that each species is only entered in the database once. The new species were then manually assigned functional traits by searching the literature; fields were left blank where functional traits for species were unknown. Once traits were assigned, the new species were added to the master species list. Queries were constructed (**Table 6**) to build lifeforms from the functional trait information (McQuatters-Gollop et al. 2019).

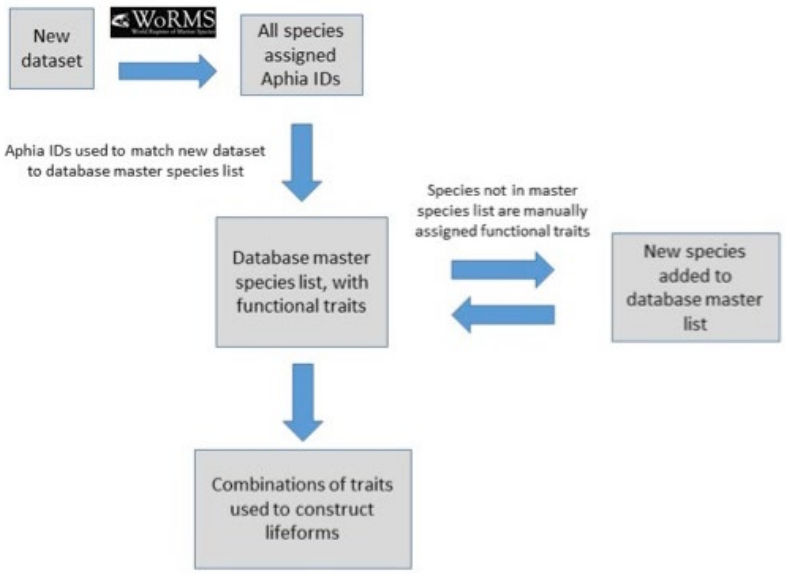

Figure 6: Schematic illustrating the process undertaken to assign functional traits to species, then species to lifeforms. Each species must first be assigned a unique Aphia ID to determine whether it is already present in the master species list (McQuatters-Gollop et al. 2019).

A simple method of confidence assessment was applied for each lifeform (**Table 5**). Using expert opinion, each lifeform was evaluated on two characteristics: the ability to identify and speciate organisms in that lifeform using light microscopy and the understanding of the accuracy of determining traits assigned to the lifeform. For example, medium confidence is assigned to the lifeforms 'autotrophic dinoflagellates' and 'mixotrophic dinoflagellates' because the mode of nutrition of many dinoflagellate species is currently uncertain (Flynn et al. 2019). Thus, the accuracy of assigning the lifeform category is medium. Likewise, the lifeform 'non-carnivorous zooplankton' has a medium confidence designation since the feeding habits of many abundant and common zooplankton species remain not strictly defined. Further investigation must also be conducted to decide whether both harmful-bloom-forming algae and

potentially toxin-producing lifeforms should be considered in future assessments of this indicator. Work is ongoing to increase confidence in lifeforms, however, this work is resource dependent.

Table 5: Matrix used to determine the confidence in each lifeform. Only pairs with high confidence were used for this assessment (Ostle et al. 2021).

	Easy to ID/speciate	**Difficult to ID/speciate**
Known traits	High	Low
Unknown traits	Low	Low

Assessing links to pressures

The Boruta algorithm (Kursa and Rudnicki 2010) was applied to evaluate which environmental variables were the best predictors of lifeform abundance. The Boruta algorithm is a tree-based permutational variable importance method which uses a wrapper around random forest to evaluate the predictive performance of each variable being tested against a shuffled set of the same predictor variables. If a variable achieves better predictive accuracy than the highest-performing shuffled variable, it is determined to be important and is assigned a score based on the mean decrease in predictive accuracy when the variable is excluded from the model.

For each unique combination of dataset, assessment unit, and lifeform whenever a significant Kendall statistic (i.e. a significant change in lifeform abundance over the time-period assessed) was observed, a separate permutational variable importance process based on random forest (Boruta algorithm) was conducted to evaluate relationships between lifeform abundance and a set of 16 environmental variables described in **Table 6**. Based on the results of the previous assessment, a table of implicating factors was developed, linking lifeforms with a set of environmental pressures with plausible links to variation in lifeform abundance (**Table 7**). Only variables indicated in **Table 7** were tested for each lifeform.

Table 6: Environmental variables and their descriptions and sources which were evaluated in the random forest permutational variable importance process.

Variable name	**Variable description**	**Source**
Sea surface temperature	Temperature of surface layer, as measured by satellite	International Comprehensive Ocean-Atmosphere Data Set (ICOADS): https://psl.noaa.gov/data/gridded/data.coads.1deg.html
Salinity	Salinity of the surface layer	European Regional Seas Ecosystem Model (ERSEM; NWSHELF_MULTIYEAR_PHY_004_009): https://doi.org/10.48670/moi-00059);
Total oxidised nitrogen	Total oxidised nitrogen concentration in surface layer	In situ data from Marine Scotland Science (MSS): https://doi.org/10.7489/1881-1 and Aarhus University (Svendsen et al. 2005)
Nitrate	Nitrate concentration in surface layer	European Regional Seas Ecosystem Model (ERSEM; NWSHELF_MULTIYEAR_BGC_004_011): https://doi.org/10.48670/moi-00058); In situ data from Plymouth Marine Laboratory (PML): https://www.westernchannelobservatory.org.uk/l4_nutrients.php
Phosphate	Dissolved inorganic phosphate concentration in surface layer	European Regional Seas Ecosystem Model (ERSEM; NWSHELF_MULTIYEAR_BGC_004_011; https://doi.org/10.48670/moi-00058); In situ data from Plymouth Marine Laboratory (PML): https://www.westernchannelobservatory.org.uk/l4_nutrients.php and Marine Scotland Science (MSS): https://doi.org/10.7489/1881-1
Total phosphorous	Total dissolved phosphorous concentration in surface layer	In situ data from Aarhus University (Svendsen et al. 2005)
N:P ratio	The ratio of molar nitrogen concentration to molar phosphorous concentration	Calculated from European Regional Seas Ecosystem Model (ERSEM; NWSHELF_MULTIYEAR_BGC_004_011): https://doi.org/10.48670/moi-00058); In situ data from Plymouth Marine Laboratory (PML): https://www.westernchannelobservatory.org.uk/l4_nutrients.php, Marine

		Scotland Science (MSS): https://doi.org/10.7489/1881-1 and Aarhus University (Svendsen et al. 2005)
Silicate	Dissolved silicate concentration in surface layer	In situ data from Marine Scotland Science (MSS): https://doi.org/10.7489/1881-1, Plymouth Marine Laboratory (PML): https://www.westernchannelobservatory.org.uk/l4_nutrients.php and Aarhus University (Svendsen et al. 2005)
Wind speed	Wind speed (a proxy for turbulence)	International Comprehensive Ocean-Atmosphere Data Set (ICOADS); https://psl.noaa.gov/data/gridded/data.coads.1deg.html
Mixed layer depth	The depth below the surface where the steepest change in density occurs (the thickness of the surface layer in which photosynthesis can occur)	European Regional Seas Ecosystem Model (ERSEM; NWSHELF_MULTIYEAR_PHY_004_009): https://doi.org/10.48670/moi-00059)
Light attenuation	The extinction coefficient for visible light, or the decrease in the intensity of solar radiation with depth (a proxy for the opacity of the water column)	European Regional Seas Ecosystem Model (ERSEM): NWSHELF_MULTIYEAR_BGC_004_011; https://doi.org/10.48670/moi-00058)
Precipitation	Rate of precipitation (a proxy for freshwater input)	International Comprehensive Ocean-Atmosphere Data Set (ICOADS); https://psl.noaa.gov/data/gridded/data.coads.1deg.html
Current velocity	Current velocity in surface layer (a proxy for horizontal transport which affects how quickly organisms are advected to adjacent systems)	European Regional Seas Ecosystem Model (ERSEM; NWSHELF_MULTIYEAR_PHY_004_009; https://doi.org/10.48670/moi-00059)
Atlantic Multidecadal Oscillation (AMO)	The Atlantic Multidecadal Oscillation, also known as Atlantic Multidecadal Variability, is the theorised variability of the sea surface temperature of the North Atlantic Ocean on the timescale of several decades	National Oceanic and Atmospheric Administration (NOAA); https://psl.noaa.gov/data/timeseries/AMO/
North Atlantic Oscillation (NAO)	The North Atlantic Oscillation is a weather phenomenon over the North Atlantic Ocean of fluctuations in the difference of atmospheric pressure at sea level between the Icelandic Low and the Azores High	National Oceanic and Atmospheric Administration (NOAA): https://www.ncdc.noaa.gov/teleconnections/nao/
pH	Sea water pH reported on total scale	European Regional Seas Ecosystem Model (ERSEM; NWSHELF_MULTIYEAR_BGC_004_011): https://doi.org/10.48670/moi-00058)

Table 7: Each lifeform evaluated for this assessment, along with the relevant environmental variables (✓) tested for each lifeform where data was available.

Variable name	Diatoms	Dinoflagellates	Holoplankton	Meroplankton	Large copepods	Small copepods	Fish larvae/eggs	Gelatinous
Sea surface temperature	✓	✓	✓	✓	✓	✓	✓	✓
Salinity	✓	✓	✓	✓			✓	✓
Total nitrogen	✓	✓						
Nitrate	✓	✓						
N:P ratio	✓	✓						
Phosphate	✓	✓						
Total phosphorous	✓	✓						

Silicate	✓	✓						
Wind speed	✓	✓	✓	✓			✓	✓
Mixed layer depth	✓	✓	✓	✓			✓	✓
Light attenuation	✓	✓					✓	✓
Precipitation	✓	✓	✓	✓	✓	✓	✓	✓
Current velocity	✓	✓	✓	✓	✓	✓	✓	✓
Atlantic Multidecadal Oscillation	✓	✓	✓	✓	✓	✓	✓	✓
North Atlantic Oscillation	✓	✓	✓	✓	✓	✓	✓	✓
pH	✓	✓	✓	✓				

Prior to any manipulation of data, lifeform time-series were divided into separate training and testing sets. The training data were used for variable selection and for generating the random forest models, while the testing data were used to validate model predictive accuracy. For the purposes of this analysis, the training data were limited to all months prior to the assessment period (e.g. 1960 to 2014 for CPR data) and the testing data were restricted to the assessment period itself (e.g. 2015 to 2019).

This analysis covered the full temporal extent of each plankton dataset (1960 to 2019 in the case of CPR data). While gridded data from NOAA were available from 1960 to 2019, the longest duration modelled dataset for nutrients only spanned from 1993 to 2019, with in-situ nutrient datasets commencing as early as 1980. To evaluate long-term links to pressures and avoid excluding the first several decades of many plankton time-series from our analysis we used the 'missRanger' package for R (Mayer and Mayer 2019) to perform multiple imputation by chained random forests. This method uses multiple random forest regressions to impute missing values based on collinearities among observed values in the predictors. For each variable containing missing values the algorithm generates a separate regression model based on all the other predictors. To improve imputation performance, a numeric variable representing 'month' was included in this step to better predict the consistent seasonal patterns in some variables. Missing values in the predictors were imputed separately for the training and testing datasets. It is important to note that no imputation of lifeform abundance time-series was performed.

All variables including lifeform abundance and excluding AMO (Atlantic Multidecadal Oscillation; AMO data product already has seasonality removed) were smoothed to remove seasonality and uncover long-term trends by calculating the mean value across a 12-month moving window (**Figure 7)**. This step required the first and last six months of data from training and testing sets to be excluded from the analysis.

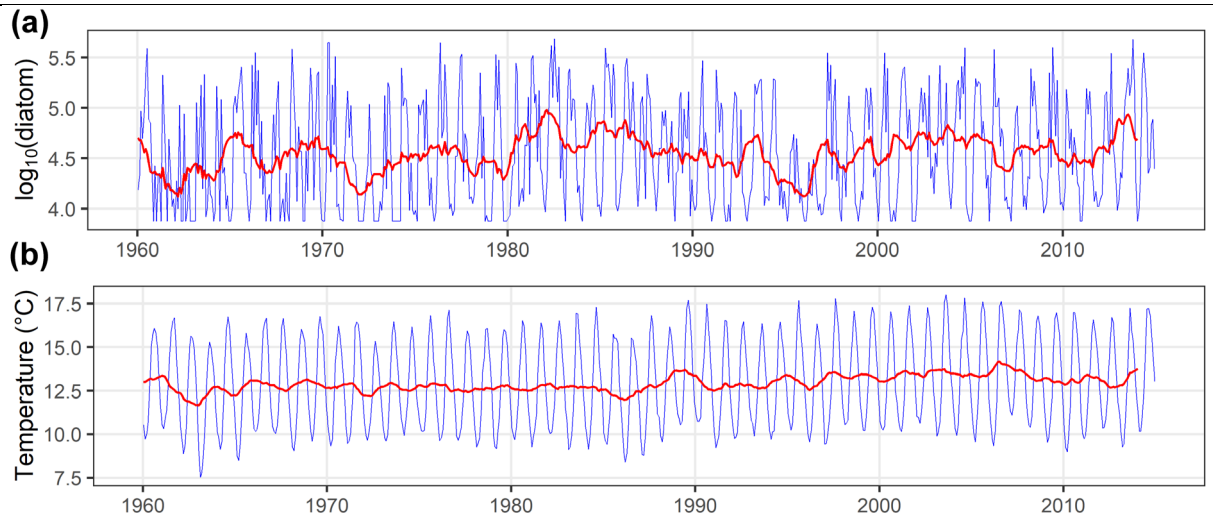

Figure 7: Long-term time-series (1960 to 2019) for (i) diatom abundance and (ii) sea surface temperature for the 'Channel well mixed' COMP4 assessment unit. Data obtained from the Continuous Plankton Recorder (CPR) survey. The blue line on each plot displays monthly mean values and the red line displays the 12-month moving average or monthly long-term trend in each time-series.

Values for each variable were calculated as the mean of monthly mean gridded values (modelled and remotely sensed; **Table 6**) intersecting each COMP4 assessment unit. For fixed-point stations, mean values were calculated from all measurements within a 5-nautical mile radius of the station. Where in-situ data were available (total nitrogen, nitrate, phosphate, total phosphorous, silicate) they were evaluated instead of the modelled environmental variables. For Atlantic Multidecadal Oscillation (AMO) and North Atlantic Oscillation (NAO), monthly values were applied identically across all assessment units since these variables have basin-scale influence likely to cover the entire assessment region.

The most important predictor variable for each model, or the predictor variable which resulted in the greatest mean decrease in accuracy when it was removed, were reported across a map. Instances when the validation step with the testing data indicated significant positive correlation ($p \leq 0.05$) between predicted and observed lifeform abundance values were also indicated in this map to provide greater confidence in the variable importance results and to avoid reporting spurious relationships.

It is important to note that observing high importance of any environmental variable is simply an indication that it co-varied with lifeform abundance. Our approach cannot imply causation since it is impossible to hold all other potential influences on lifeform abundance constant, as would be the case in a controlled experimental setting.

Integration of indicator results

A primary objective of this indicator assessment was to integrate results to facilitate an understanding of changes occurring across pelagic habitat types within OSPAR Regions II, III and IV. This required indicator results for each OSPAR Region to be integrated according to the following pelagic habitat categories: variable salinity, coastal, shelf, and oceanic / beyond shelf. This categorisation of COMP4 assessment units and fixed-point stations is described in **Table 1.** To meet this objective, we focused on the primary direction of change detected across assessment units and fixed-point stations within each pelagic habitat category for each of the 8 high confidence plankton lifeforms highlighted in this assessment. We then reported the mean confidence, spatial representativeness, and most likely links to environmental pressures.

As an example, changes in dinoflagellate abundance were assessed across 4 COMP4 areas and 4 fixed-point stations representing variable salinity habitats within OSPAR Region II. If 1 decreasing trend, 4 increasing trends, and 3 instances of no trend were detected across these locations, we would report an increasing net trend and the proportion of assessment units studied where this trend was detected, in this case 0.5.

To report the level of confidence in this result we calculated the mean confidence score for locations where dinoflagellate abundance was increasing. We considered COMP4 assessment units and fixed-point stations as equivalent for this integration.

To report the spatial representativeness of the result, we calculated the proportion of the total number of COMP4 assessment units considered in the analysis, in this case 8, out of the total number of possible COMP4 assessment units representing variable salinity habitats within the OSPAR Region, in this case 16. Therefore, the spatial representativeness of the result would be 0.5. Note that fixed-point station datasets do not contribute to this score.

Finally, to report links to environmental pressures which can drive changes in lifeform abundance for the net trend, we ranked environmental variables for each location based on their relative variable importance, with 1 assigned to the variable with highest importance, 2 to the variable with second highest importance and so on. For locations where the net trend was increasing, we calculated the mean rank of each environmental variable and reported the variable with the lowest mean rank.

Assessing the status of pelagic habitats

To assign a designation of assessment status to pelagic habitats based on the integration of indicator results we applied a semi-quantitative methodology described in McQuatters-Gollop et al. (2022), which was developed from the lessons gained from the previous OSPAR assessment (IA 2017). In this case, the status of a habitat type can be designated as either “Good”, “Unknown”, “Not good”, or “Not assessed” (**Table 8**). Following the criteria outlined in this study, if a pelagic habitat has been assessed, it should by default be considered as either “Unknown” or “Not Good”. At this stage it is not realistic to assign ‘Good’ status to pelagic habitats, since it is difficult to develop meaningful assessment thresholds for plankton and generally not possible to determine whether a particular state is desirable or undesirable, except under specific circumstances such as eutrophication. Following this logic, the status of pelagic habitats should be considered “Unknown” by default. In cases when change has been detected and that change can be confidently linked to the impact of an anthropogenic pressure, the status of this habitat is “Not good”.

Table 8: Biodiversity status categories and colours used for the interpretation, by expert judgement, of indicator biodiversity state (McQuatters-Gollop et al. 2022).

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 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
Not assessed	Indicator was not assessed in a region due to lack of data, lack of expert resource, or lack of policy support.

We considered status at the level of each of the 8 high confidence lifeforms within each pelagic habitat within OSPAR Regions and integrated the results for multiple lifeforms to assign a single quality

status designation for the pelagic habitat type. For the status of a lifeform to be shifted from "Unknown" to "Not good" the results of the integration had to meet certain criteria:

- The net trend should be either increasing or decreasing and must be present in at least 0.5 of the locations assessed.
- The mean confidence for locations considered for the determination of the net trend must be at least 0.5.
- The spatial representativeness of the assessed locations out of the total number of possible locations for that habitat type must be at least 0.5.
- The environmental pressure with the greatest mean rank for locations expressing the net trend must represent either sea surface temperature, pH, or nutrients.
- The mean rank of the most important environmental pressure must be ≤ 3, indicating that across all locations the variable ranks in the top 3 most important for predicting the abundance of the lifeform.

If all the above criteria are met, the lifeform is assigned a status of "Not good". If 25% or more of assessed lifeforms within a pelagic habitat type are assigned a status of "Not good" then the whole habitat type is also assigned a status of "Not good". Otherwise, the habitat type is assigned a status of "Unknown".

Lifeform pairs indicator approach:

Although this assessment focused primarily on long-term trends in lifeform abundance due to their stronger links with environmental pressures at large spatial scales, the lifeform pairs indicator was also evaluated for the current assessment (which constitutes the PH1 indicator as originally defined in IA2017).

Tett et al. (2008) proposed to track changes in the state of the plankton community by means of plots in a state space and calculating a Plankton Community Index, referred to here as a Plankton Index (PI). The conceptual framework is that ecosystems can be viewed as systems with an instantaneous state defined by values of a set of system state variables which are attributes of the system that change with time in response to each other and external conditions. Building on this approach and plotting plankton lifeform abundance in a multi-dimensional state space provides a means of monitoring changes in the structure of plankton communities. A state can be defined as a single point in state space, with coordinates provided by the values of the set of state variables, in this case two lifeform abundances, which together are used as a pair.

The plankton community index is calculated by initially mapping relative plankton lifeform abundances from an assessment period in a state space. The distribution of lifeform abundances within the state space is then used to define an envelope, representing the prevailing conditions for the assessment period. Finally, a set of abundance values from a comparison period are mapped onto the state space to evaluate the plankton community index, which represents the proportional similarity of the two periods.

Mapping lifeforms in a state space

In the example illustrated in **Figure 8** for diatoms and dinoflagellates, the axes of the two-dimensional space are the abundance values of the two lifeforms. Each point represents the state of the ecosystem in terms of the two lifeforms at the time of sample collection. Subsequent samples yield additional pairs of abundance values that can be mapped onto the lifeform state space.

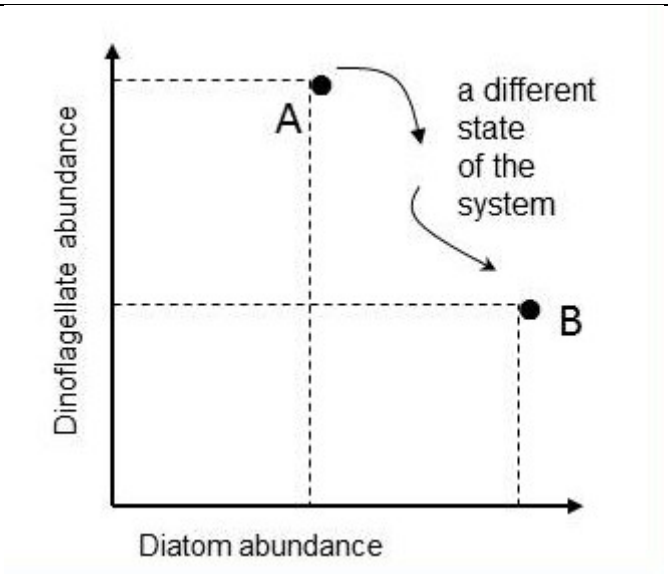

Figure 8: Mapping the abundance of two lifeforms in state-space

Point A is the ecosystem state at the instant a water sample was taken and is characterised by the abundance values of two lifeforms. Another sample, taken in the same location, yields abundance values that map to a different point in the diatom-dinoflagellate state space (point B).

The path between the two states is called a trajectory, and the plankton condition is defined by the trajectory drawn in the state space by a set of points. Such trajectories reflect: (i) cyclic and medium-term variability (the higher order consistencies in the plankton that result from seasonal cycles, species succession and interannual variability); and (ii) long-term variability that might result from environmental pressure. The seasonal nature of plankton production and the succession of species in seasonally stratifying seas, result in this trajectory tending in a certain direction (as plankton growth increases in the spring and declines during autumn), such that the trajectory tends towards its starting point. Given roughly constant external pressures, the data collected from a particular location over a period of years form a cloud of points in state space that can be referred to as a regime. Long-term variability may show a persistent trend of movement away from a starting point in state space.

Approach for defining the envelope

To define a regime, an envelope can be drawn about this group of points using a convex hull method. Because of theoretical arguments that the envelope should be doughnut-shaped with a central hole (Tett et al. 2008), bounding curves can be fitted outside and inside the cloud of points (**Figure 9**). The data are from the CPR dataset for the 'Channel well mixed' COMP4 assessment unit. The colour of each point corresponds to the season within which it was sampled; blue = winter, green = spring, yellow = summer, red = autumn.

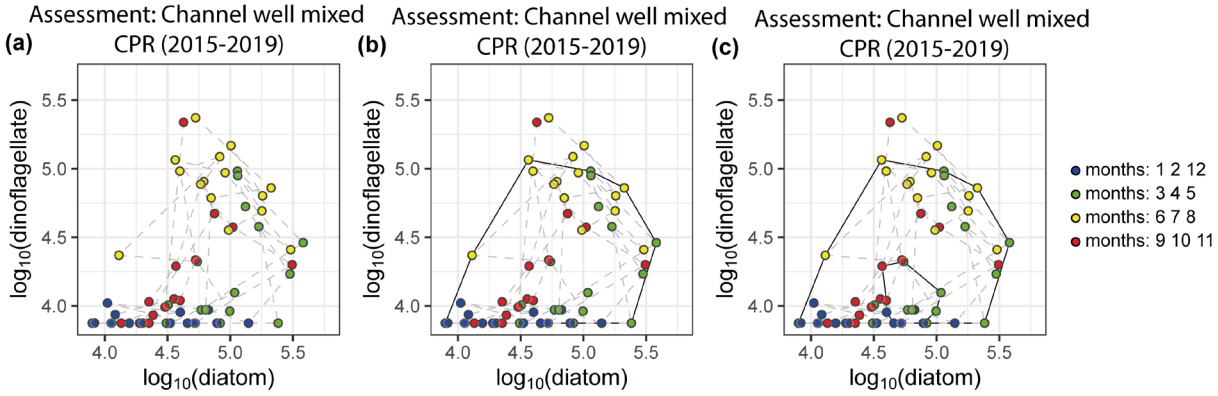

Figure 9: Creating the envelope in three steps, from left (a) to right (c): An example of a regime defined by the envelope drawn by the convex hull method

The size and shape of the envelope are sensitive to sampling frequency and the total number of samples. Envelopes are made larger by including extreme outer or inner points, and the larger the envelope, the less sensitive it will be to change in the distribution of points in state space and therefore to detect a change in condition. Conversely, if too many points are excluded the envelope will be small and even minor changes will result in a statistically significant difference. It is therefore desirable to exclude a proportion (p) of points, to eliminate these extremes, and so the 90-percentile was used. Envelopes are therefore drawn around the cloud of points to include a proportion (p=0.9) of the points: with the 5% of points that were most distant from the cloud's centre and the 5% of points that were closest to the centre, excluded.

For a Plankton Index to be calculated, it is necessary to establish a set of conditions as the basis for making comparisons. For this assessment the Plankton Index was calculated by using the assessment period (2015-2019) to define the envelope. For each dataset, all measurements collected prior to the assessment period were used for comparison. For the CPR data this meant the envelope was defined with samples from 2015-2019 and samples from 1960-2014 were used for comparison. This made it possible to maintain the same range of years to define the envelope across multiple datasets with different temporal coverage. Like the Kendall statistic, the PI was calculated independently for each combination of dataset, assessment unit and lifeform pair.

In the example shown in **Figure 10,** CPR data collected from the 'Channel well mixed' COMP4 assessment unit during the assessment period (2015-2019), were used to create an envelope. The envelope, thus drawn (**Figure 10 (a)**) defines a domain in state space that contains a set of trajectories of the diatom-dinoflagellate component of the pelagic ecosystem.

Calculating the plankton community index

The next step is to map a new set of data onto the same state space for comparison (**Figure 10 (a)**). The value of the PI is the proportion of new points that fall between the inner and outer envelopes. In this example, 30% or 198 of the 660 new points lie outside, and the PI is 0.7 (**Figure 10 (b)**). A value of 1.0 would indicate no change, and a value of zero would show complete change. The envelope was made by excluding 10% of points, so 66 (10% of 660) points are expected to fall outside. The exact probability of getting 198 by chance alone when only 66 are expected, can be calculated using a chi-square calculation (with 1 df and a 1-tail test). The value of 0.7 is significantly less than the expected value of 0.9, and so the difference between the two periods is statistically significant ($p < 0.01$).

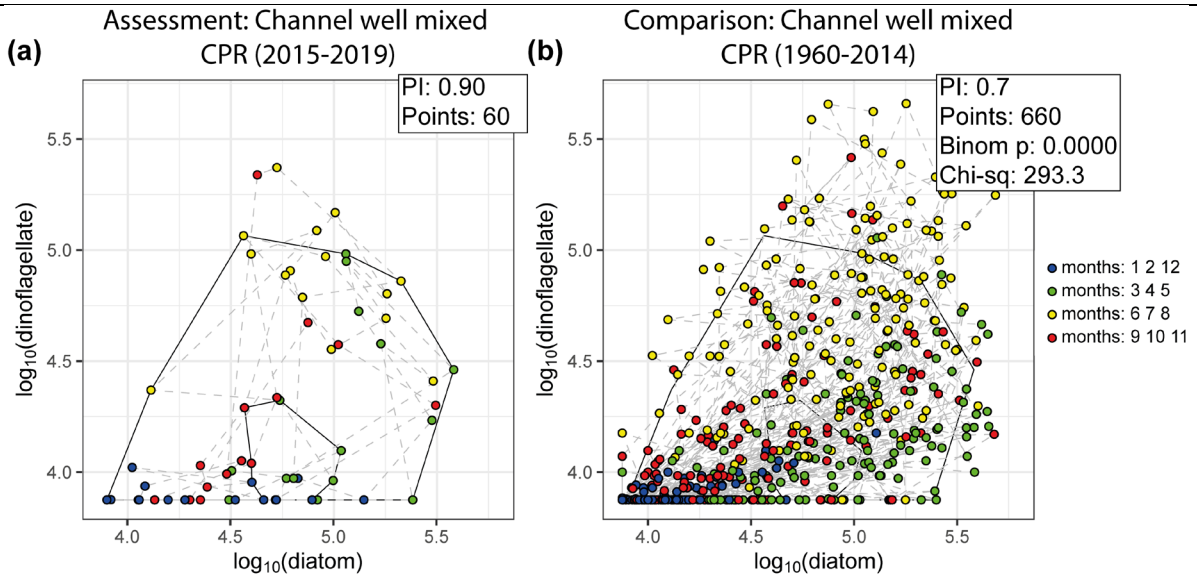

Figure 10: An example for the comparison between assessment and comparison periods within the 'Channel well mixed' COMP4 assessment unit, with (a) representing the assessment period of 2015-2019 and (b) representing the comparison period of 1960-2019.

3.5 Presentation of assessment results

- *Consideration of target audience and appropriate communication style*
- *Assessment metadata schema (link to ODIMS)*

The common indicator assessment is published on the OSPAR Assessment Portal https://oap.ospar.org/en/ospar-assessments/intermediate-assessment-2017/biodiversity-status/habitats/changes-phytoplankton-and-zooplankton-communities/

4 Change Management

- *Responsibility for follow up of assessment (e.g. if the monitoring is not adequate)*
    - Pelagic subgroup of ICG-COBAM
    - BDC

5 References

Barnes, M. K., G. H. Tilstone, T. J. Smyth, C. E. Widdicombe, J. Gloël, C. Robinson, J. Kaiser, and D. J. Suggett. 2015. Drivers and effects of *Karenia mikimotoi* blooms in the western English Channel. Progress in Oceanography **137:** 456-469.

Beaugrand, G. 2005. Monitoring pelagic ecosystems using plankton indicators. ICES Journal of Marine Science **62:** 333-338.

Bedford, J., C. Ostle, D. G. Johns, A. Atkinson, M. Best, E. Bresnan, M. Machairopoulou, C. A. Graves, M. Devlin, and A. Milligan. 2020. Lifeform indicators reveal large-scale shifts in plankton across the North-West European shelf. Global Change Biology **26:** 3482-3497.

Bremner, J., C. L. J. Frid, and S. I. Rogers. 2004. Biological traits of the North Sea benthos – Does fishing affect benthic ecosystem function? , p. 477-489. *In* P. Barnes, Thomas, J. [ed.], Benthic habitats and the effects of fishing. American Fisheries Society.

Charvet, S., M. C. Roger, B. Faessel, and M. Lafont. 1998. Biomonitoring of freshwater ecosystems by the use of biological traits. Annals of Limnology – International Journal of Limnology **34:** 455-464.

de Jonge, V. N. 2007. Toward the application of ecological concepts in EU coastal water management. Marine Pollution Bulletin **55:** 407-414.

Desmit, X., A. Nohe, A. V. Borges, T. Prins, K. De Cauwer, R. Lagring, D. Van der Zande, and K. Sabbe. 2020. Changes in chlorophyll concentration and phenology in the North Sea in relation to de-eutrophication and sea surface warming. Limnology and Oceanography **65:** 828-847.

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

Estrada, M., P. Henriksen, J. M. Gasol, E. O. Casamayor, and C. Pedrós-Alió. 2004. Diversity of Planktonic Photoautotrophic Microorganisms Along a Salinity Gradient as Depicted by Microscopy, Flow Cytometry, Pigment Analysis and DNA-based Methods. FEMS Microbiology Ecology **49:** 281-293.

Flynn, K. J., A. Mitra, K. Anestis, A. A. Anschütz, A. Calbet, G. D. Ferreira, N. Gypens, P. J. Hansen, U. John, and J. L. Martin. 2019. Mixotrophic protists and a new paradigm for marine ecology: where does plankton research go now? Journal of Plankton Research **41:** 375-391.

Gallego, I., T. A. Davidson, E. Jeppesen, C. Perez-Martinez, P. Sanchez-Castillo, M. Juan, F. Fuentes-Rodriguez, D. Leon, P. Penalver, J. Toja, and J. J. Casas. 2012. Taxonomic or ecological approaches? Searching for phytoplankton surrogates in the determination of richness and assemblage composition in ponds. Ecological Indicators **18:** 575-585.

Garmendia, M., Á. Borja, J. Franco, and M. Revilla. 2012. Phytoplankton composition indicators for the assessment of eutrophication in marine waters: Present state and challenges within the European directives. Marine Pollution Bulletin **66:** 7-16.

Gowen, R. J., A. McQuatters-Gollop, P. Tett, M. Best, E. Bresnan, C. Castellani, K. Cook, R. Forster, C. Scherer, and A. Mckinney. 2011. The Development of UK Pelagic (Plankton) Indicators and Targets for the MSFD. Advice to Defra, p. 41 pp.

HELCOM. 2012. Development of the HELCOM core-set indicators Part B, GES 8/2012/7b. HELCOM.

Keys, M., G. Tilstone, H. S. Findlay, C. E. Widdicombe, and T. Lawson. 2018. Effects of elevated CO 2 and temperature on phytoplankton community biomass, species composition and photosynthesis during an experimentally induced autumn bloom in the western English Channel. Biogeosciences **15:** 3203-3222.

Kursa, M. B., and W. R. Rudnicki. 2010. Feature selection with the Boruta package. Journal of statistical software **36:** 1-13.

Mayer, M., and M. M. Mayer. 2019. Package 'missRanger'. R Package.

McQuatters-Gollop, A., A. Atkinson, A. Aubert, J. Bedford, M. Best, E. Bresnan, K. Cook, M. Devlin, R. Gowen, and D. G. Johns. 2019. Plankton lifeforms as a biodiversity indicator for regional-scale assessment of pelagic habitats for policy. Ecological Indicators **101:** 913-925.

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

Mouillot, D., S. Spatharis, S. Reizopoulou, T. Laugier, L. Sabetta, A. Basset, and T. D. Chi. 2006. Alternatives to taxonomic-based approaches to assess changes in transitional water communities. Aquatic Conservation-Marine and Freshwater Ecosystems **16:** 469-482.

Ostle, C., K. Paxman, C. A. Graves, M. Arnold, L. F. Artigas, A. Atkinson, A. Aubert, M. Baptie, B. Bear, and J. Bedford. 2021. The Plankton Lifeform Extraction Tool: a digital tool to increase the discoverability and usability of plankton time-series data. Earth System Science Data **13:** 5617-5642.

Rakocinski, C. F. 2012. Evaluating macrobenthic process indicators in relation to organic enrichment and hypoxia. Ecological Indicators **13:** 1-12.

Reid, P. C., J. Colebrook, J. Matthews, J. Aiken, and C. P. R. Team. 2003. The Continuous Plankton Recorder: concepts and history, from Plankton Indicator to undulating recorders. Progress in Oceanography **58:** 117-173.

Svendsen, L., L. van der Bijl, S. Boutrup, and B. Norup. 2005. NOVANA. National Monitoring and Assessment Programme for the Aquatic and Terrestrial. and number: NERI Technical Report.

Tett, P., C. Carreira, D. K. Mills, S. van Leeuwen, J. Foden, E. Bresnan, and R. J. Gowen. 2008. Use of a phytoplankton community index to assess the health of coastal waters. ICES Journal of Marine Science **65:** 1475-1482.