

Using Dynamic Ocean Color Provinces to Elucidate Drivers of North Sea Hydrography and Ecology



Key Points:

- Four ocean color types are defined for the North Sea, which are associated with bio-optical, physical, and zooplankton community covariates
- Ocean color province dynamics relate to variability in local wind-forced currents and large-scale variations in source waters
- Atlantic-associated ocean color types show highest coverage in the North Sea during periods of reduced wind-forced circulation

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. H. Taylor,
marc.taylor@thuenen.de

Citation:

Taylor, M. H., Akimova, A., Bracher, A., Kempf, A., Kühn, B., & Hélaouët, P. (2021). Using dynamic ocean color provinces to elucidate drivers of North Sea hydrography and ecology. *Journal of Geophysical Research: Oceans*, 126, e2021JC017686. <https://doi.org/10.1029/2021JC017686>

Received 16 JUN 2021
 Accepted 23 NOV 2021

Author Contributions:

Conceptualization: Marc H. Taylor
Data curation: Marc H. Taylor, Pierre Hélaouët
Formal analysis: Marc H. Taylor, Anna Akimova, Bernhard Kühn
Investigation: Marc H. Taylor, Anna Akimova, Bernhard Kühn
Methodology: Marc H. Taylor, Pierre Hélaouët
Validation: Anna Akimova
Visualization: Marc H. Taylor
Writing – original draft: Marc H. Taylor, Anna Akimova, Astrid Bracher,

Marc H. Taylor¹ , Anna Akimova¹ , Astrid Bracher^{2,3} , Alexander Kempf¹ , Bernhard Kühn¹ , and Pierre Hélaouët⁴ 

¹Thuenen Institute of Sea Fisheries, Herwigstraße, Germany, ²Phytooptics Group, Physical Oceanography of Polar Seas, Climate Sciences, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany,

³Institute of Environmental Physics, Faculty of Physics and Electrical Engineering, University Bremen, Bremen, Germany,

⁴Marine Biological Association, The Laboratory, Citadel Hill, Plymouth, Devon, UK

Abstract Ocean biogeochemical provinces have historically been used to characterize regions of similar biogeography, allowing for large-scale comparisons and the definition of possible management units. Recent studies have shown that ocean color (OC) may be used as a proxy for a range of physical and biological parameters, allowing for a more direct approach in defining and monitoring provinces. The current work investigates whether OC-derived provinces (area of a given OC type) can be identified in the North Sea and if they can provide insight into the system's dynamics. The analysis workflow was organized into three main steps: 1. OC type classification based on clustering of normalized reflectance spectra, 2. OC type characterization following match-ups to other biological and physical covariates, and 3. Exploration of drivers of OC province dynamics. Based on data from 1997–2019, four main OC types were identified, which are shown to be related to bio-optical, physical, and zooplankton community covariates. Atlantic-associated OC types varied in their coverage of the North Sea on seasonal and interannual scales, with increased penetration related to periods of a weakened circulation. Interannual variability of OC province extent was also found to correlate with an index of the Subpolar Gyre. These results indicate that the dynamics of OC provinces in the North Sea are likely driven by the interaction between large-scale circulation and local wind-driven and baroclinic currents. The perspective offered by OC provinces helps elucidate spatiotemporal patterns of plankton and is proposed as a novel operational index for additional exploration of ecosystem functioning.

Plain Language Summary A water's color, from muddy brown to green or clear blue, reflects a range of information regarding its origin and characteristics. For example, we may be able to infer the amount of sediments or the concentration and composition of its algal community (phytoplankton), which form the base of the food web for many marine ecosystems. Using color measurements obtained from historical satellite imagery (1997–2019), North Sea water types were defined, which related to a range of parameters; including salinity, phytoplankton concentration, and associated zooplankton community. Of particular interest are the clearer, bluish water types associated with Atlantic waters that enter the North Sea around the area of the Shetland Islands, whose inflow variability is thought to be an important driver of the North Sea ecosystem over time. The results show that the coverage of Atlantic waters changes over seasonal and long-term time scales, which are related to the strength of currents and changes in the source waters arriving to the European shelf. In addition to providing insight to historical changes, the method's use of freely available satellite imagery makes it an attractive approach for future monitoring of the ecosystem.

1. Introduction

The concept of an ecological province has historically been used to identify and characterize regions of similar biogeography, which may serve as units for comparison, monitoring, and management (Sonnewald et al., 2020). Remote sensing of ocean color has proved to be an important tool in helping to delineate provinces in terms of differing phytoplankton biomass (i.e. chlorophyll *a* concentration, Chl*a*) and production. One of the earliest applications was that of Platt and Sathyendranath (1988), who used surface Chl*a* to estimate primary production throughout the water column. Their algorithm relied on the definition of ocean biogeochemical provinces (BG-CPs) that were similar in terms of seasonal physical characteristics and resulting vertical Chl*a* profiles. Longhurst's (1998, 2007) seminal work expanded on the approach to define BGCPs on a global scale using a mixture of biological and physical variables, such as circulation patterns, Chl*a* distribution, mixing rates, stratification,

© 2021 The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Alexander Kempf, Bernhard Kühn, Pierre Hélaouët

Writing – review & editing: Marc H. Taylor, Anna Akimova, Astrid Bracher, Alexander Kempf, Bernhard Kühn, Pierre Hélaouët

nutrient concentrations, and irradiance. Statistical, analytical, and biogeographic tests were used to verify the objectivity of province delineations. Biogeographic tests included links to fish distributions and zooplankton community similarity, which provided a possible link between BGCPs and higher trophic level patterns.

In both of these early studies, BGCPs were based on static boundaries for practical reasons, related to the frequency of observations, and a degree of observed stability in their spatial extent over time; however, it was acknowledged that the use of dynamic boundaries would be preferred as they would allow for a better understanding of spatial changes through time (Platt, 1999). With the increased availability and coverage of modern remote-sensing and modelling data, the detection and monitoring of dynamic BGCPs is now increasingly possible. For example, Reygondeau et al. (2013) developed a statistical approach to dynamically identify 56 BGCPs based on Longhurst's original definitions, plus four main parameters (bathymetry, Chl_a, surface temperature, and salinity) that allowed for the description of their spatiotemporal variability. Kavanaugh et al. (2014), using the related concept of a "seascape", found that dynamically-defined boundaries resulted in a nearly 2-fold increase of biochemical classification efficiency over statically-defined ones, demonstrating the importance of temporal changes in coverage.

With the recent availability of high spectral resolution (hyperspectral) radiometers, the remote detection of phytoplankton types that contain accessory pigments to Chl_a, with differing light absorption frequencies, may provide a more complete picture of the plankton community. The spatial distribution of different pigment compositions has been shown to largely conform to the static boundaries of Longhurst's (2007) provinces, lending support to the possibility of identifying BGCPs from remotely-sensed bio-optical properties alone (Bracher et al., 2020; Taylor et al., 2011). For example, Vantrepotte et al. (2012) applied water type classification to hyperspectral reflectance data in eastern English Channel and French Guiana and found good agreement between bio-optical and biogeochemical water properties, including phytoplankton, sediment, and mineral concentrations. Despite these advances using hyperspectral data, large-scale definitions of bio-optical provinces will likely continue to rely on a smaller number of more commonly measured reflectance bands given their measurement on global scales and spanning decades (Sathyendranath et al., 2019).

The observed link between OC types and underlying biological and physical characteristics opens up the prospect of monitoring province dynamics on various spatiotemporal scales. Furthermore, the concept of provinces (or seascapes) may provide a valuable framework for both identifying drivers of spatiotemporal variability of aquatic systems and the role of this variability in ecosystem functioning (Kavanaugh et al., 2014). The North Sea provides an interesting case study given that it is a well-studied system with complex oceanography, including exchange with the Atlantic via the English Channel in the south, the Norwegian Trench and around the Shetlands in the north, and exchange with the more brackish waters of the Baltic Sea to the east (e.g. Stündermann & Pohlmann, 2011). The system is of great economic importance to the surrounding countries in terms of fisheries resources and other economic activities (e.g. offshore wind energy, aquaculture), and real-time monitoring could be useful for a range of applications. Of particular interest is the measurement of Atlantic water inflow and penetration, whose long-term variability is thought to be a major factor in forcing regime shifts in the North Sea ecosystem (Alvarez-Fernandez et al., 2012; Beaugrand, 2004; Longhurst, 2007; Reid & Edwards, 2001). Despite advances in remote sensing, there has been a reliance on oceanographic models for insights into currents and flows in the North Sea, since observational data are often sparse in space (e.g. moorings) or time (occasional cruises). Our observation-based understanding of the multiple drivers of the North Sea currents and their variability therefore becomes hard to discern and the causes for resulting trends are difficult to diagnose (Huthnance et al., 2016). Thus, the aims of the study are to investigate whether OC-derived provinces (area of a given OC type) can be identified for the North Sea and to determine if they can serve as proxies for physical and biological parameters that are more difficult to monitor.

2. Methods

The analysis workflow was organized into three main steps (Figure 1): 1. OC type classification (Section 2.2), 2. OC type characterization (Section 2.3), and 3. Drivers of OC province dynamics (Section 2.4). The first two steps focused on the North Sea and adjacent areas (referred to as *Northeast Atlantic Shelves area*, 15°W – 15°E, 45°N – 65°N), while the third step is focused solely on the North Sea (Figure 2). The North Sea was delineated by the boundaries of Subarea 4, as defined by the International Council for the Exploration of the Sea (ICES) (Figure 2).

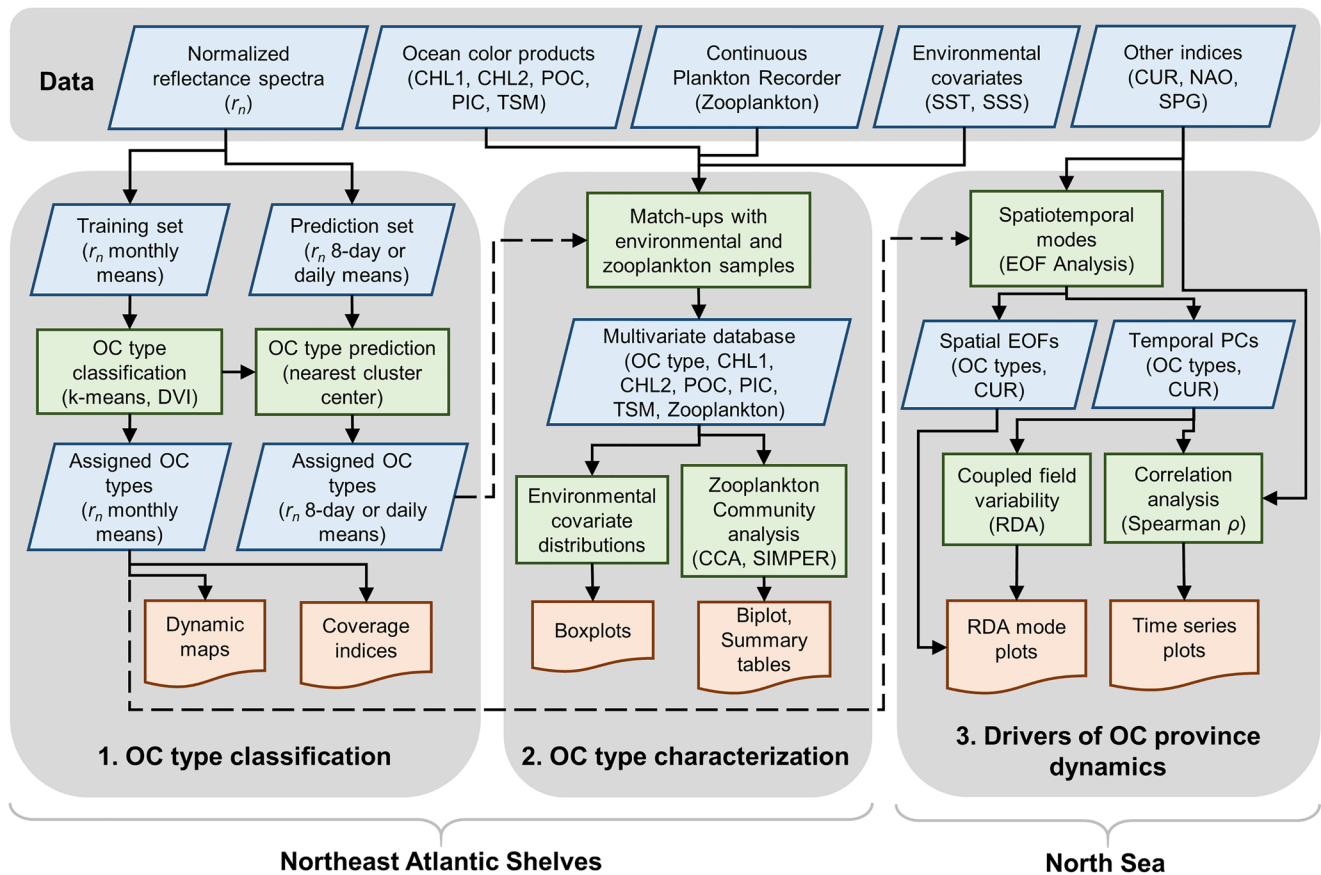


Figure 1. Ocean color analysis workflow. Colored shapes represent data (blue rhomboids), processes/analyses (green rectangles), and resulting outputs and plots (red rectangle with wavy base). Specific data sets or methods are shown in parentheses. Grey areas separate the initial data (top) and the three main methodological steps (below, numbered). OC type classification (step 1) and characterization (step 2) are conducted on the larger Northeast Atlantic Shelves area, while explorations of drivers to OC province dynamics (step 3) focuses on the smaller area of the North Sea proper. The assigned OC type classifications (step 1) feed into subsequent analyses (dashed lines).

In the first step, monthly OC reflectance data were used to train the classification of OC types in the Northeast Atlantic Shelves area, whose waters have the potential to influence North Sea conditions through exchange processes. The trained classification model was used to predict OC types of a second, finer temporal resolution (8-day and daily mean) data set, which was used for match-ups with other physical and biological variables. In the second step, the multivariate database of match-ups was used to characterize the OC types in terms of associated covariates and zooplankton community. In the third step, possible drivers of OC province dynamics in the North Sea were explored, including North Sea currents as well as large-scale atmospheric and oceanographic indices of the northeast Atlantic Ocean. Further descriptions of the data and analysis steps are described in the following sections. All analyses were conducted using the R Statistical software, version 3.6.2 (R Core Team, 2020).

2.1. Data

2.1.1. Zooplankton Data

Continuous Plankton Recorder (CPR, <https://www.mba.ac.uk/fellows/cpr-survey>) zooplankton data were used, spanning from 1997–2017. CPR Standard Areas (Figure 2, yellow boundary) were chosen that largely overlapped with the Northeast Atlantic Shelves area (Hélaouët, 2020). CPR samples contain abundance values for 100 zooplankton taxa, as estimated from a ca. 10 nautical mile transect segment. Associated metadata includes date, time, and geographic location of the transect center position. CPR sampling is done at 5–10 m depth, and zooplankton abundances correspond to a filtered water volume of about 3 m³ per 10 nautical mile transect (Batten et al., 2003).

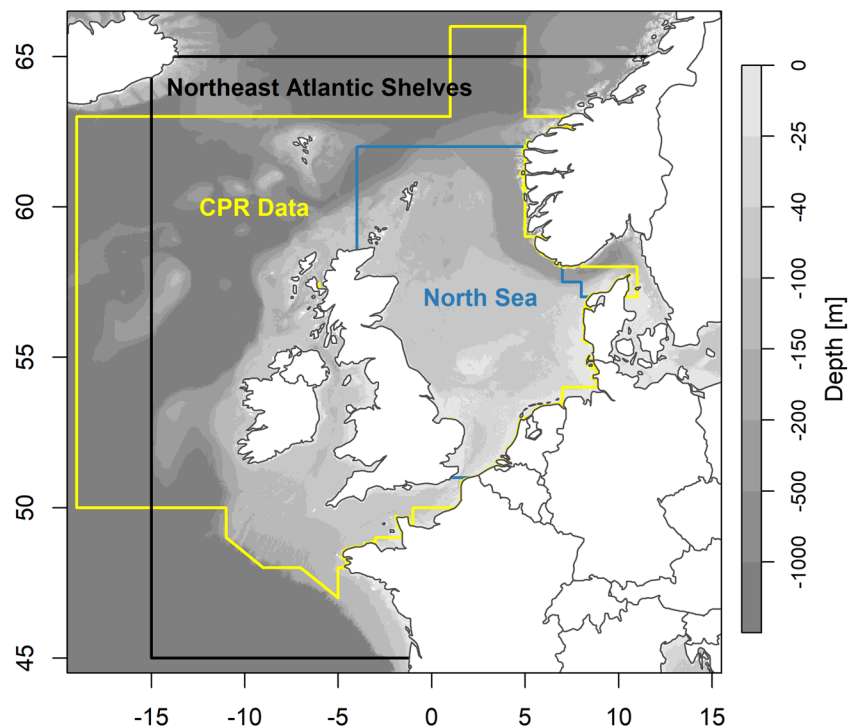


Figure 2. Map of the study area. Ocean color (OC) data from the Northeast Atlantic Shelves area (black boundary) was used in the training of OC types. Sample match-ups for zooplankton and other variables were restricted to the area of overlap with selected Continuous Plankton Recorder (CPR) Standard Areas (yellow boundary). The blue boundary shows the area used for a focused evaluation of historical OC type dynamics in the North Sea.

2.1.2. Ocean Color Products

OC data were acquired from Globcolour (<http://globcolour.info>, version R2017.0), which has been developed, validated, and distributed by ACRI-ST, France (Fanton d'Andon et al., 2009; Maritorea et al., 2010). The data set includes various products, many of which merge estimates from different satellite sensors (SeaWiFS, MERIS, MODIS AQUA, and VIIRS) to improve spatiotemporal coverage. Globcolour data products from the period of September 1997 - December 2019 were used in this study, with weighted mean values used in cases where estimates from more than one sensor were available (see Figure S1 in Supporting Information S1 for additional information on each sensor's temporal coverage and associated ocean color products).

The present work used normalized remote sensing reflectance (r_n , Sr^{-1}) from five bands (412, 443, 490, 555, and 670 nm) in order to classify OC types. These five bands have been measured by all sensors, thus allowing for the analysis to use data over a longer time period. Several other products were used to characterize OC types, including Chla from "case 1" waters (CHL1, mg m^{-3}), Chla from "case 2" (CHL2, mg m^{-3}), particulate organic carbon (POC, mg m^{-3}), particulate inorganic carbon (PIC, mol m^{-3}), and total suspended material (TSM, g m^{-3}). Although CHL1 is widely used for large scale estimates of Chla concentrations, CHL2 was also included in the analysis as it is considered a more appropriate estimate for waters where inorganic particles dominate over phytoplankton (typically in coastal waters) (Doerffer & Schiller, 2007). Spatial locations were not specifically designated as either CHL1 or CHL2 categories given the subjective nature of the assignment; rather, both values were used in characterization of OC types. 25 km horizontal resolution data products were used as an appropriate level for matching to the 10 nm (~ 19 km) CPR transect samples. Monthly mean values were used for OC type clustering, while daily and 8-day mean r_n values were used for later prediction and match-ups of OC types with other covariates. The lower temporal resolution monthly mean data were preferred for training over daily and 8-day mean data sets given the improved coverage and reduced sample size.

2.1.3. Physical Data

Sea surface salinity (SSS, g kg⁻¹), sea surface temperature (SST, °C), and depth-averaged currents (CUR, horizontal velocity, m s⁻¹) were obtained from the North-West European Shelf reanalysis (accessed via the E.U. Copernicus Marine Service website, <https://marine.copernicus.eu>). The data are produced with an ocean assimilation model, at 7 km horizontal resolution. The ocean model is NEMO (Nucleus for European Modelling of the Ocean), using the 3DVar NEMOVAR system to assimilate observations of surface temperature and vertical profiles of temperature and salinity. The data extend from 1993–2019. SSS was taken from the subsurface layer (10 m) in order to avoid the noisy values associated in the upper layer due to evaporation and precipitation. Data fields were interpolated on the 25 km resolution grid used by the OC data via bilinear interpolation (R package raster, Hijmans, 2020). Daily and 8-day mean SST and SSS were used for match-ups to OC samples (see Section 2.2), while monthly mean CUR was used for identifying coupled spatiotemporal patterns between currents and OC province dynamics (see Section 2.4).

Additional large-scale indices used in the analysis were the atmospheric index of the North Atlantic Oscillation (NAO) (Hurrell, 1995) (source: <https://crudata.uea.ac.uk/cru/data/nao/>) and the oceanographic index of the Subpolar gyre (SPG) (Häkkinen et al., 2011; Hátún & Chafik, 2018). The NAO describes the normalized pressure difference between a station on the Azores and one on Iceland. The SPG index describes variability of the sea surface height (SSH) field over the North Atlantic subpolar and subtropical gyres (30°N–65°N, 80°W–0°). For both NAO and SPG, yearly indices were calculated as the mean of the monthly values.

2.2. Ocean Color Type Classification

Monthly mean reflectance values were used for defining OC types due to their improved spatiotemporal coverage and more manageable number of samples. All data available for the Northeast Atlantic Shelves area were used in the analysis ($n = 1,244,874$). OC types were defined via clustering, similar to the approach of Mélin and Vantropotte (2015). Reflectance spectra were first normalized by dividing reflectance values by the area under the spectrum, resulting in an area of 1.0. Area under the spectrum was calculated with trapezoidal integration (R package caTools, Tuszynski, 2020). Clustering was conducted with *k-means*, using the default Hartigan-Wong algorithm (Hartigan & Wong, 1979) and four restarts to increase the stability of the cluster centers. The total number of clusters was selected according to the lowest dynamic validity index (DVI) proposed by Shen et al. (2005), which provides an objective metric for defining the optimal number of clusters. The DVI combines both a measure of compactness (minimize intra-cluster distances) with a measure of separateness (maximize between or inter-cluster distances), which allows for the identification of homogenous, well-separated clusters.

The cluster centers (mean r_n spectra) were then used to classify the OC type of two higher temporal resolution OC data sets (daily and 8-day mean reflectance). These data were used for match-ups with associated physical and biological covariates in order to further characterize the OC types (see next section). The higher temporal resolution reflectance spectra were normalized, as done with the monthly mean spectra, and prediction of OC type was based on the nearest neighbor Euclidean distance to the cluster centers (R package FNN, Beygelzimer et al., 2019).

2.3. Ocean Color Type Characterization

OC types were characterized in terms of associated covariates, both physical and biological in nature. In order to address whether differing levels of temporal aggregation affected the relationships, two temporal resolutions were used for all variables (daily and 8-day means). Since the measurement of OC data is hindered by low light conditions during winter months, only the months of March–October were used in the analyses. The predicted OC type data sets were matched to all remotely-sensed data of similar temporal aggregation, which included OC product parameters (CHL1, CHL2, POC, PIC, and TSM), and the abiotic reanalysis parameters (SST and SSS). The differences in the associated covariates among OC types were evaluated visually via boxplots displaying quantile distributions.

Canonical Correspondence Analysis (CCA) (ter Braak, 1986, 1987) was used to elucidate relationships between the zooplankton assemblages (i.e. constrained data) and environmental variables (i.e. constraining data), including OC type. CCA assumes a unimodal (e.g. Gaussian) relationship between environmental covariates and

zooplankton taxa, which implies an optimal set of environmental conditions associated with the maximum abundance (or biomass) of a given taxa. CPR samples were matched to environmental covariates based on the overlap of CPR sample coordinates to cells of the 25 km resolution grid used by other data. Rare zooplankton taxa, found in < 1% of samples, were removed from the community analyses in order to focus on more prevalent groups. CPR sample abundances were $\log(x+1)$ -transformed in order to down-weight the most abundant taxonomic groups. Environmental covariates included the continuous variables $\log(\text{CHL1})$, SST, SSS, and the categorical variables of month and OC type. The categorical factor for the month of the sample was also included to account for seasonal changes in abundance related to life history strategies, such as overwintering. Although CHL2 was used as a covariate in the characterization of OC types, it was not included as a constraining variable due to its more restricted (but subjective) association with coastal waters. Significance tests of the overall model fit, CCA axes, and marginal term effects were performed via an ANOVA-like permutation test (Legendre et al., 2011; Legendre & Legendre, 2012). The test statistic is “pseudo-F”, which is the ratio of constrained and unconstrained total inertia (Chi-square), divided by the degrees of freedom. Inertia is a measure of weighted variance among species, focusing on gradients in the overall community composition. The ratio of inertia explained by the CCA has been shown to be a poor indicator of the explained compositional variation and it is thus recommended to focus on the relative amounts of inertia explained by CCA axes and constraining variables when interpreting the model results (Økland, 1999). In the summary tables for each permutation test, relative importance is inferred from the reported Chi-square and F-values.

Similarity percentage analysis (SIMPER) (Clarke, 1993) was used to identify zooplankton taxa that differed between OC types. SIMPER performs pairwise comparisons of groups (here OC types) and quantifies the contribution of each taxa to the average between-group Bray-Curtis dissimilarity. Although abundance data were \log -transformed as with the CCA, highly abundant and variable species can have high contributions to dissimilarity, and thus permutation tests are performed to determine whether these differences are significant. The R package *vegan* (Oksanen et al., 2019) was used for all community analyses.

2.4. Drivers of Ocean Color Province Variability

Additional focus was placed on understanding the historical dynamics of OC types in the North Sea. The fraction coverage of total North Sea area (in km^2) by OC type was calculated from the monthly mean OC data. In order to summarize long-term and seasonal temporal patterns of the resulting coverage time series, a generalized additive model (GAM) was fit using the R package *mgcv* (Wood, 2011, 2020), which allows for the incorporation of smooth functions to describe nonlinear effects. The model assumed a quasi-binomial distribution with logit-link function, considered appropriate for proportional data, and contained two explanatory terms; 1) Year – Thin-plate regression spline applied to numeric date values for describing interannual patterns; 2) Month – Cyclic cubic regression spline applied to numeric month values for describing seasonal patterns.

For each OC type, an Empirical Orthogonal Function (EOF) analysis (also known as Principal Component Analysis) was conducted to extract the dominant spatiotemporal modes of variability; specifically, vectors describing temporal (principal components, PCs) and corresponding spatial (EOFs) modes of variability. The presence/absence (i.e. binomial) data of each OC type were \logit -transformed to approximate a normal distribution. Although missing values were few in non-winter monthly-averaged fields (~1%), gaps were filled via *Data Interpolating Empirical Orthogonal Functions* (DINEOF) (Beckers & Rixen, 2003) using the R package *sinker* (Taylor, 2020). Following interpolation, monthly anomalies were derived by subtracting the long-term monthly mean of each spatial grid, which removes the seasonal signal for analyses focused on interannual variability. An EOF was conducted on both the original and anomaly data fields; the original data were used to describe the dominant patterns in annual spatiotemporal variability, while the anomaly field was decomposed as a precursor to the Redundancy Analysis described in the next paragraph.

Redundancy analysis (RDA) (von Storch & Zwiers, 1999; Tyler, 1982; Rao, 1964) was used to elucidate physical drivers of OC type spatiotemporal dynamics on interannual scales. In particular, a field of depth-averaged current (CUR) anomalies (i.e. constraining data) was used to explain the field of each OC type's occurrence anomalies (i.e. unconstrained data). In contrast to the unimodal relationships described by CCA modes, RDA describes linear relationships between the constraining and unconstrained data sets. This assumption was deemed appropriate for evaluating the degree to which the spatiotemporal dynamics of OC provinces were related to water transport.

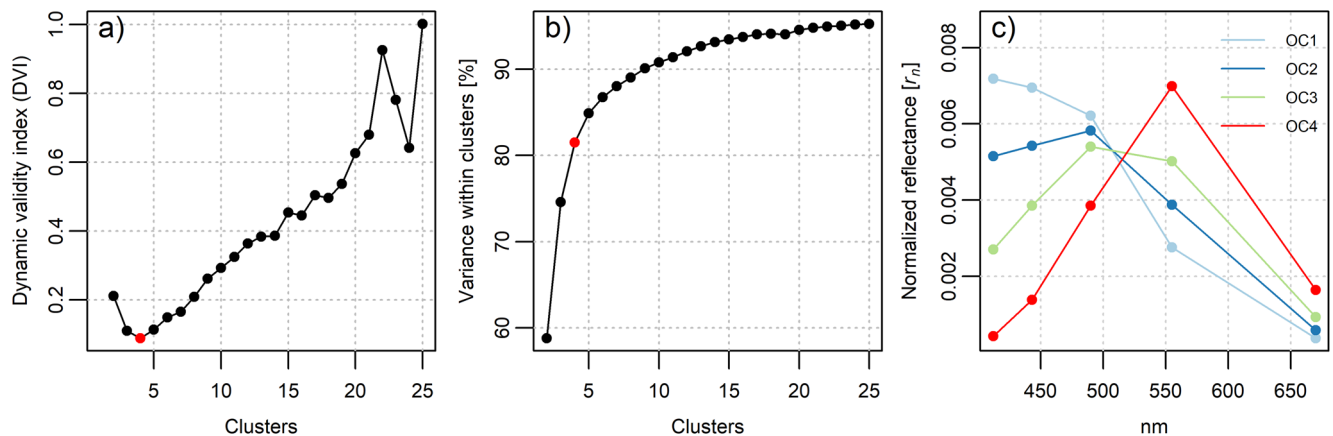


Figure 3. Ocean color reflectance clustering of monthly mean normalized reflectance (r_n) from the 5 merged bands (Northeast Atlantic Shelves area, September 1997 – April 2020). Minimization of the dynamic validity index (DVI) (a) was used to determine the optimal number of clusters ($n = 4$, red point). This optimal number corresponds to 81% of variance explained within the clusters (b). The resulting spectra associated with the cluster centers (c) were used to predict the ocean color type of subsequent, higher temporal resolution data (daily and 8-day averaged spectra).

Monthly anomalies were derived by subtracting the long-term monthly mean of each grid for all fields. Similar to Kauker et al. (2008), we first conducted an EOF on the CUR and OC type fields in order to identify a truncated number of temporal PCs that cumulatively explain > 90% of field variance. These PCs were then used as input data to the RDA rather than the original full data fields, which has several advantages over directly using the matrices of the data fields; principally, reduced computational effort, increased stability of solution, and avoidance of problems of collinearity in both the predictor and response data sets. Unscaled PCs (i.e. with units included) were used so that their importance in terms of explained variance of original field was maintained. Temporal patterns are a direct output of the RDA, while the associated spatial components are derived by projecting the unitless EOF patterns onto the RDA loadings associated with each PC. Significance of RDA patterns were tested via an ANOVA-like permutation test (Legendre et al., 2011; Legendre & Legendre, 2012). The R package *vegan* (Oksanen et al., 2019) was used for RDA and permutation tests. The test statistic is “pseudo-F”, which is the ratio of constrained and unconstrained total variance, divided by the degrees of freedom. Contrary to CCA, RDA uses linear relationships to best describe variance of the unconstrained data, and thus the fraction of total explained variance is an appropriate measure of the model overall performance.

Links between other environmental drivers (NAO and SPG) and the dominant modes of OC type spatiotemporal dynamics were explored using correlation analysis of annual mean values. Mean annual PC1 loadings derived from the EOF of monthly anomalies of occurrence were used for the OC types. Yearly indices of OC types were correlated to the NAO and SPG at lag 0 and lag 1 to address possible advective delays between conditions of the North Atlantic and the North Sea. Spearman's rank correlation coefficient (ρ) was used, which is a nonparametric metric capable of describing nonlinear, but monotonic, relationships.

3. Results

3.1. Ocean Color Type Classification

Clustering of the monthly mean spectra for the Northeast Atlantic Shelves area showed lowest DVI for 4 OC types (Figure 3). The variance contained within the clusters was > 80% of total variance, after which additional clusters showed only marginal improvements. Cluster mean spectra (values provided in Table S1 in Supporting Information S1) have been ordered according to decreasing normalized reflectance at 412 nm, which is a wavelength particularly associated with *Chl a* absorbance that aided later interpretation of OC types. For the remainder of the paper, these numerical identifiers (OC1 – OC4) will be used for the OC types. An example of the resulting OC type assignment to spectra from a single year (2013) is presented in Figure 4. The spatial pattern clearly shows a more Atlantic distribution for OC1 and OC2, while OC3 and OC4 are more coastally-distributed. OC4 is particularly associated with a littoral distribution.

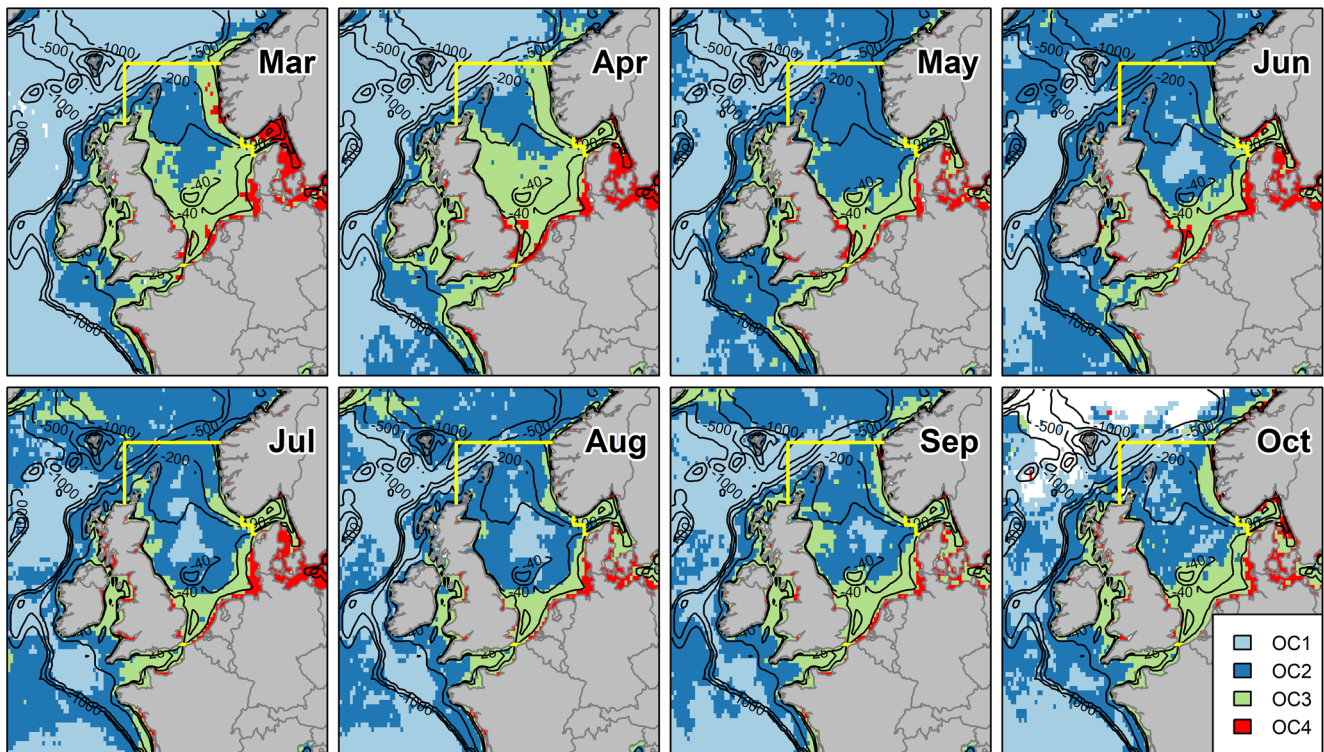


Figure 4. Example of the spatial distribution of ocean color (OC) types assigned to normalized monthly mean reflectance spectra for a single year (2013) in the Northeast Atlantic Shelves area. Bathymetry isolines (in black) are shown for the following depths: 25, 40, 100, 200, 500, and 1000 m. The boundary of the North Sea area is designated by the yellow border.

In general, the predicted classification of OC type was consistent between 8-day and daily match-ups (Table 1). Exact overlapping classification occurred in 76% of cases (Table 1 diagonal), while overlap with neighboring classifications (in terms of OC type order) occurred in 23% of cases (Table 1 sub- and superdiagonals). Overlap to more distant classifications occurred in 1% of cases. The high degree of overlap indicates that OC types are relatively stable over both time scales for a given grid. One exception was OC4, which was classified more frequently in daily than in 8-day mean reflectance data. The predictions of OC4 using daily mean data ($n = 47,647$) were split nearly equally between predictions of OC3 ($n = 22,780$) and OC4 ($n = 24,867$) using 8-day mean data, possibly indicating less persistence of this type over time.

3.2. Ocean Color Type Characterization

3.2.1. Bio-Optical and Physical Variables

23,571 of 42,563 CPR samples (56%) could be matched to the 8-day mean ocean color data (Figure 5), while daily mean data were matched in only 9,255 cases (22%) (Figure S2 in Supporting Information S1). Unmatched samples are due to missing remote OC estimates resulting from cloud coverage or low-light winter conditions. The number of matched samples was largely consistent across the considered months (March-October, Figure 5, bottom-right), although counts by year show lower values at the beginning of the series (Figure 5, top-right) due to the existence of a single ocean color sensor prior to 2003 (i.e. SeaWiFS, see Figure S1 in Supporting Information S1).

The OC types of the CPR match-up samples are associated with trends in a number of other OC products. Only 8-day mean match-ups are shown here (Figure 6), although very similar patterns were observed in the daily

Table 1
Contingency Table Comparing OC Type Classification Predictions in the Northeast Atlantic Shelves Area Based on Either 8-Day (Rows) or Daily (Columns) Mean Ocean Color Reflectance Bands ($n = 1,122,852$ Match-Ups)

8-day/ daily	OC1	OC2	OC3	OC4
OC1	313383 (28%)	39382 (4%)	1541 (<1%)	264 (<1%)
OC2	78770 (7%)	327131 (29%)	56751 (5%)	3041 (<1%)
OC3	5716 (1%)	51470 (5%)	186340 (17%)	22780 (2%)
OC4	321 (<1%)	1272 (<1%)	9823 (1%)	24867 (2%)

Note. Overlap in classification assignment are displayed as frequencies and as a percent of the total match-ups.

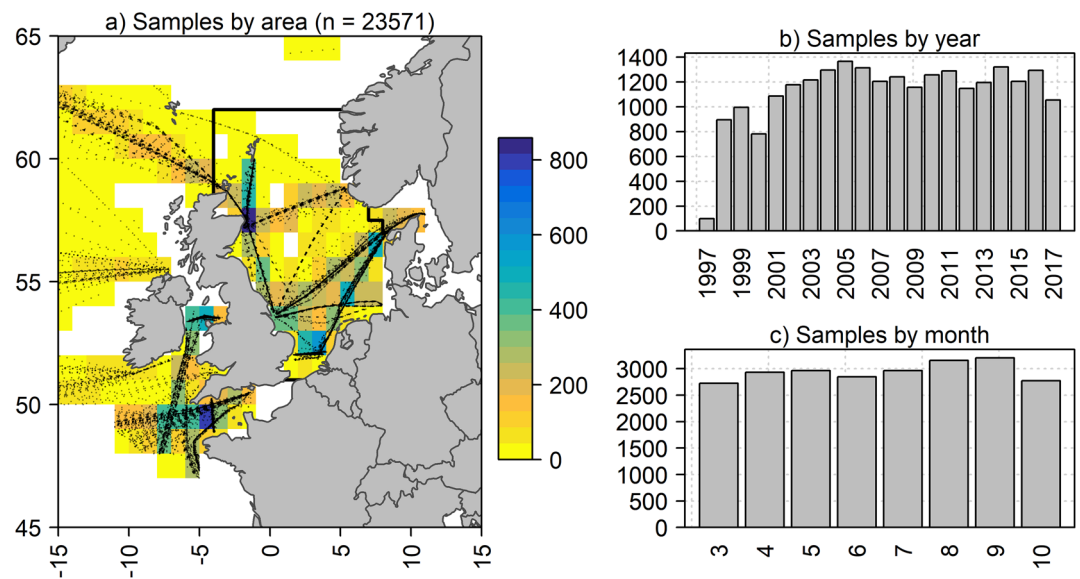


Figure 5. Continuous Plankton Recorder (CPR) samples matched to 8-day mean ocean color data in the Northeast Atlantic Shelves area. Panels show sample aggregations by a) area (individual samples are small points, with colors showing density per 1° grid), b) year (top-right), and c) month (bottom-right). The thick black boundary on the left panel indicates the focus area for North Sea OC province dynamics.

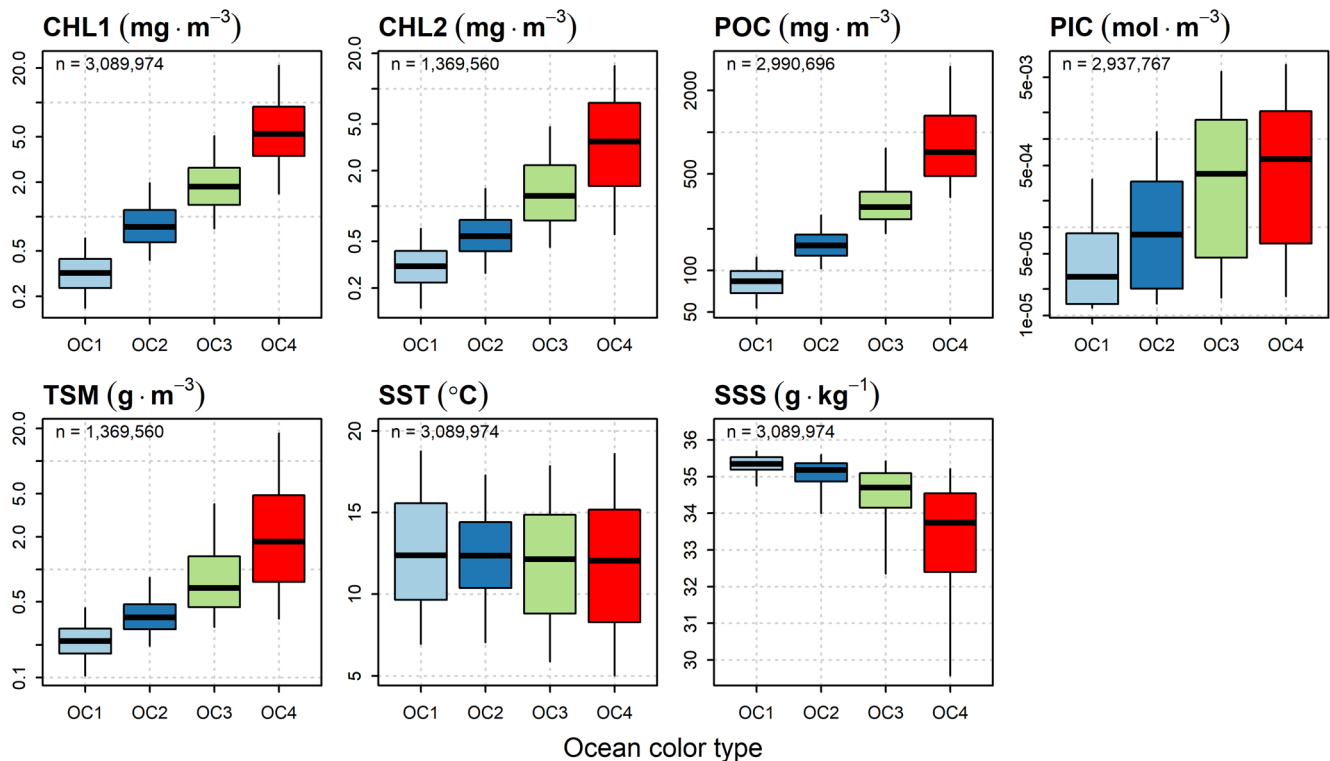


Figure 6. Boxplots of 8-day averaged parameters by ocean color (OC) type as matched to CPR samples. Optically-derived variables include Chla case 1 (CHL1), Chla case 2 (CHL2), Particulate Organic Carbon (POC), Particulate Inorganic Carbon (PIC), and Total Suspended Matter (TSM). Abiotic parameters include sea surface temperature (SST) and sea surface salinity (SSS). Y-axes of optically-derived variables have been log-scaled. Box limits represent the 25% and 75% quantiles, bold horizontal lines are the median, and whiskers are the 5% and 95% quantiles.

Table 2
Permutation Test of CCA Term Marginal Effects ($n = 999$) Relating Environmental Variables (8-Day Mean Match-Ups) to Changes in Zooplankton Community

	Df	Chi-square	Pseudo-F	p-value*
OC type	3	0.009	16.651	0.001
Month	7	0.074	60.262	0.001
SST	1	0.044	251.891	0.001
SSS	1	0.018	102.948	0.001
log(CHL1)	1	0.012	69.137	0.001
Residual	19870	3.481		

Note. Chi-square is a measure of the community inertia explained by each environmental variable. Explanatory variables OC type and Month are factors and others are continuous. *All variables showed significant effects at the $p \leq 0.001$ level.

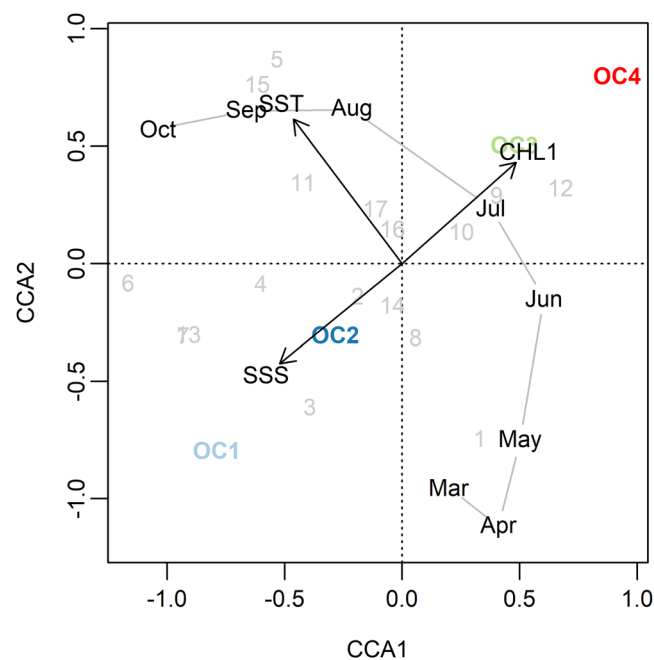


Figure 7. Biplot of leading Canonical Correspondence Analysis (CCA) axes relating explaining variables with zooplankton community. Species identification numbers are in grey, and explanatory variables in black (Month, log (CHL1), SST, and SSS) and color (OC type). SST, SSS, CHL1, and OC type are based on 8-day mean match-ups with CPR data. The numeric identifiers for the zooplankton groups are as follows: (Large-sized groups): 1. *Calanus finmarchicus*, 2. *Calanus helgolandicus*, 3. *Metridia lucens*, 4. *Candacia armata*, 5. *Labidocera wollastoni*, 6. *Paraeuchaeta hebes*, 7. *Euchaetidae*; (Small sized groups): 8. *Calanus spp.*, 9. *Temora longicornis*, 10. *Acartia spp.*, 11. *Centropages typicus*, 12. *Centropages hamatus*, 13. *Clausocalanus spp.*, 14. *Oithona spp.*, 15. *Corycaeus spp.*, 16. *Harpacticoida*, and 17. *Centropages spp.*

mean match-ups (Figure S3 in Supporting Information S1). In particular, OC types are ordered (OC1-OC4) along a gradient of increasing Chl_a (CHL1 and CHL2), POC, TSM, and along a decreasing gradient of SSS. As seen in Figure 4, these associations reflect spatial differences between the coastal and offshore environments. PIC increased with OC type number but shows less distinguishable distributions. SST showed no clear differences or trends among OC types.

3.2.2. Zooplankton Community

Following the filtering of CPR data to include most relevant taxa (i.e. taxa found in >1% of samples) and with match-ups to environmental parameters, 19,884 (17 taxa) and 7,821 (16 taxa) samples were included in the community analysis for 8-day and daily mean match-up data sets, respectively. Both data sets resulted in very similar results, and thus only 8-day match-up analyses are shown here. Daily-mean match-up results are available in the Supplementary Information (Figure S4 and Tables S4–S7 in Supporting Information S1).

The CCA showed that the environmental covariates significantly explained patterns in the zooplankton community (Table S2 in Supporting Information S1). The high residual inertia left unexplained by the CCA is likely due to several factors relating to a low signal to noise ratio in the zooplankton data, which may include: the large spatiotemporal coverage, the patchy nature of zooplankton distributions, and the less than optimal surface transect sampling of the CPR data set (as opposed to sampling through the water column). Despite these less than optimal characteristics, the sheer size of the data set allowed for the identification of significant community associations with the covariates.

All environmental terms were determined to be significant following a test of marginal effects (Table 2), with SST associated with the highest F-value. Variables that are more clearly seasonal in nature across the entire sampled area, e.g. SST and Month, explained a higher degree of community variance than other parameters. OC type, SSS and CHL1 all explained similar degrees of variation, however the F-value of OC type is lower due to the higher associated degrees of freedom. The biplot of the leading CCA axes shows two main directions of variability related to environmental variables (Figure 7); one associated with seasonal patterns, including Month and SST, and an orthogonal one, associated with SSS, OC type, and CHL1. Nine CCA axes were deemed significant at the $p \leq 0.05$ level, but a large majority of the explained inertia is contained within in the first two (Table S3 in Supporting Information S1). The CCA coordinates of zooplankton taxa (i.e. location of labels) correspond to the environmental conditions where higher abundances occur. As the biplot only displays the first two axes, direct interpretation can be misleading; yet the results of the SIMPER analysis (see below) general support the affinity of taxa to the most closely located OC type in the CCA biplot.

A summary of the SIMPER analysis can be seen in Table 3, where only the top species contributing to 95% dissimilarity between groups, and for which permutation tests show significant differences ($p < 0.05$), are presented. Many taxa show clear abundance gradients correlating to the OC types (ID numbers referred to in Figure 7 and Table 3 are provided after taxa names for reference). *Metridia lucens* (3), *Calanus finmarchicus* (1), *Oithona spp* (14), and *Clausocalanus spp* (13) are more associated with Atlantic waters of OC1 and OC2, as reflected by the mean abundances by OC type in the SIMPER

Table 3

SIMPER Analysis Summary Showing Zooplankton Groups That Significantly Contributed to Community Dissimilarity Between Ocean Color Types in the Northeast Atlantic Shelves Area (8-Day Mean Match-Ups, n = 999 Permutations for Each Comparison)

ID	Species	Size class	OC type A	OC type B	Ave. Diss.	SD	Ave. Abu. A	Ave. Abu. B	p-value
11	<i>Centropages typicus</i>	Small	1	2	0.115	0.140	2.819	3.017	0.001***
8	<i>Calanus spp</i>	Small	1	2	0.112	0.130	2.788	3.273	0.001***
2	<i>Calanus helgolandicus</i>	Large	1	2	0.087	0.088	2.264	2.769	0.001***
14	<i>Oithona spp</i>	Small	1	2	0.082	0.122	1.906	1.920	0.001***
1	<i>Calanus finmarchicus</i>	Large	1	2	0.055	0.082	1.171	1.062	0.001***
3	<i>Metridia lucens</i>	Large	1	2	0.036	0.061	0.812	0.595	0.001***
13	<i>Clausocalanus spp</i>	Small	1	2	0.025	0.072	0.583	0.465	0.001***
4	<i>Candacia armata</i>	Large	1	2	0.012	0.029	0.262	0.217	0.001***
10	<i>Acartia spp</i>	Small	1	3	0.137	0.161	2.527	4.020	0.001***
11	<i>Centropages typicus</i>	Small	1	3	0.109	0.140	2.819	2.410	0.001***
9	<i>Temora longicornis</i>	Small	1	3	0.106	0.140	0.679	3.497	0.001***
2	<i>Calanus helgolandicus</i>	Large	1	3	0.083	0.088	2.264	2.028	0.001***
14	<i>Oithona spp</i>	Small	1	3	0.073	0.120	1.906	1.237	0.001***
1	<i>Calanus finmarchicus</i>	Large	1	3	0.052	0.084	1.171	0.691	0.001***
3	<i>Metridia lucens</i>	Large	1	3	0.032	0.061	0.812	0.279	0.001***
13	<i>Clausocalanus spp</i>	Small	1	3	0.019	0.067	0.583	0.177	0.001***
10	<i>Acartia spp</i>	Small	1	4	0.152	0.168	2.527	4.591	0.001***
9	<i>Temora longicornis</i>	Small	1	4	0.124	0.150	0.679	3.991	0.001***
1	<i>Calanus finmarchicus</i>	Large	1	4	0.051	0.085	1.171	0.568	0.001***
3	<i>Metridia lucens</i>	Large	1	4	0.031	0.061	0.812	0.196	0.001***
9	<i>Temora longicornis</i>	Small	2	3	0.109	0.134	1.927	3.497	0.001***
12	<i>Centropages hamatus</i>	Small	2	3	0.020	0.061	0.338	0.579	0.001***
10	<i>Acartia spp</i>	Small	2	4	0.143	0.156	3.488	4.591	0.001***
9	<i>Temora longicornis</i>	Small	2	4	0.122	0.142	1.927	3.991	0.001***
12	<i>Centropages hamatus</i>	Small	2	4	0.022	0.065	0.338	0.608	0.015*
15	<i>Corycaeus spp</i>	Small	2	4	0.016	0.058	0.237	0.422	0.016*
10	<i>Acartia spp</i>	Small	3	4	0.151	0.165	4.020	4.591	0.001***
9	<i>Temora longicornis</i>	Small	3	4	0.139	0.155	3.497	3.991	0.001***
12	<i>Centropages hamatus</i>	Small	3	4	0.028	0.074	0.579	0.608	0.001***
15	<i>Corycaeus spp</i>	Small	3	4	0.020	0.066	0.373	0.422	0.001***

Note. Only top contributing species to average dissimilarity (Ave. Diss.) between group comparisons are included in summary table (95% cumulative dissimilarity cut-off, SD is standard deviation of dissimilarity). Average abundances (Ave. Abu.) are log-transformed values. Group identifiers (A vs. B) provide the specific OC type numbers compared (e.g., 1 = OC1, 2 = OC2, etc.). P-value significance levels are denoted by star symbols (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

analysis, and their CCA scores are generally orientated toward lowest Chla concentrations. Strongest association with the most Chla-poor OC1 is seen for *Metridia lucens* (3) and *Calanus finmarchicus* (1), which show a gradient of decreasing abundance with increasing OC type number. *Candacia armata* (4) was also found to be significantly higher in OC1 than OC2. *Acartia spp* (10), *Temora longicornis* (9), and *Centropages hamatus* (12) are strongly associated with Chla-rich waters, showing a gradient of increasing abundance with increasing OC type number, and are likewise oriented towards the region of higher Chla concentrations in the CCA biplot (Figure 7). Some species show a strong association with the seasonal axis in the CCA plot, including e.g. *C. finmarchicus* (1), which is oriented towards the cool waters and spring months, coinciding with the season associated with their ascendance to surface waters following overwintering.

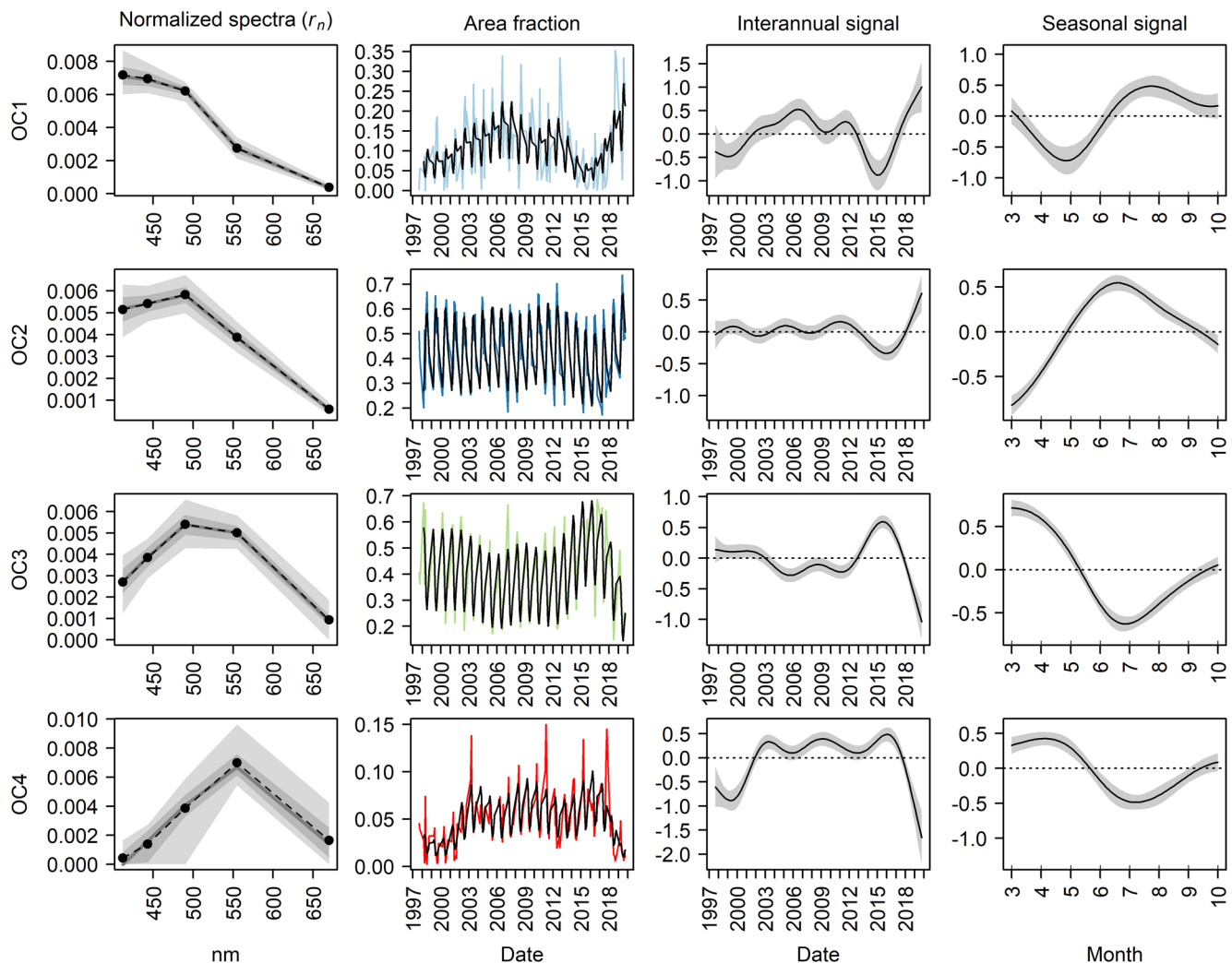


Figure 8. Summary of monthly average ocean color (OC) type contributions to the North Sea area. Far left panels show normalized spectra distributions for samples assigned to each OC type (shading: 5%, 25%, 75%, and 95% quantiles; grey line: median values; and dashed line with points: cluster centers). Center left panels show fraction coverage of total North Sea area by OC type (colored lines: observed coverage indices; and black lines: GAM prediction based on interannual and seasonal signals). Center right and far right panels show interannual and seasonal GAM terms (units are term contributions to the logit link-transformed response variable, area fraction) with confidence intervals shaded.

3.3. Drivers of Ocean Color Province Variability

The variations in OC province coverage in the North Sea can be seen in Figure 8, along with interannual and seasonal terms identified by the GAMs. Atlantic-associated OC1 and OC2 show increased North Sea coverage during summer and late summer, with OC2 peaking around June and OC1 peaking around August. Coastal-associated OC3 and OC4 show a bimodal seasonal signal with higher spatial coverage in spring (March) and later summer (August–September). Interannual variability is most notable during the period post-2014, where a strong decrease in the spatial coverage of Atlantic-associated OC types was observed from 2014–2016, followed by a strong increase. OC1 also shows lower areal coverage at the beginning of the time series. The leading EOF of each OC type shows the main seasonal spatiotemporal pattern in distribution (Figure S5 in Supporting Information S1).

For simplicity, only the RDA1 patterns for the more Atlantic-associated OC1 and OC2 types are presented here (Figures 9 and 10), but RDA results for all OC types are available in the Supplementary Information (Figures S6–10 and Tables S8–17 in Supporting Information S1). According to the permutation tests, CUR anomalies were significant predictors of the OC type anomalies, explaining between 15–20% of total variance, depending on OC

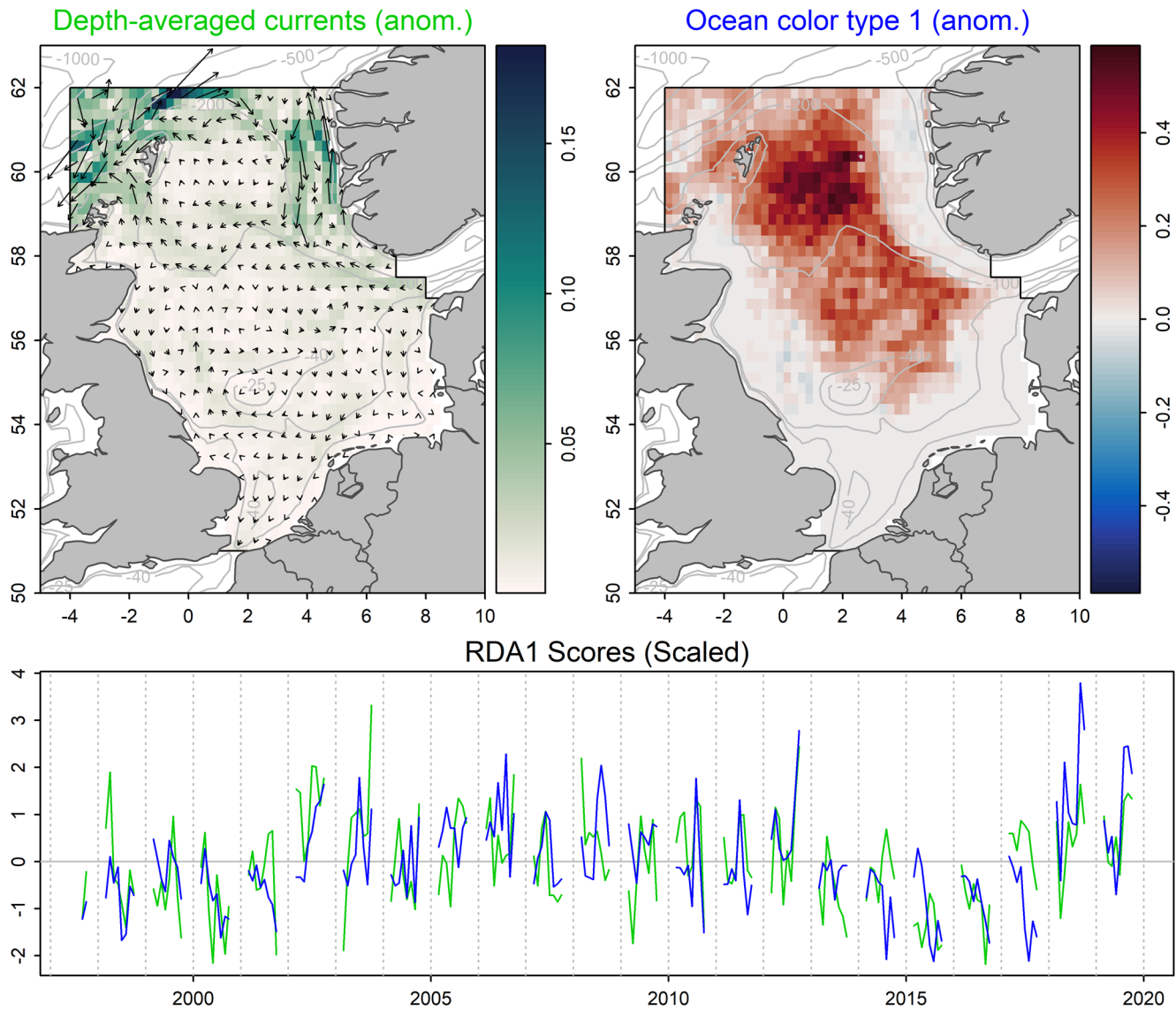


Figure 9. Leading Redundancy Analysis patterns (RDA1) between monthly anomalies of depth-averaged currents (CUR, left panel) (constraining field) and ocean color type 1 (OC1, right panel) occurrence (constrained field) in the North Sea (black boundary). Isobaths are shown in grey for reference. Spatial pattern values (upper panels) must be multiplied by the corresponding temporal scores (lower panel; green = currents and blue = ocean color type) to derive the strength and sign of the signal. For currents, darker colors and longer arrow vectors indicate higher velocity anomalies. Arrow direction indicates current orientation when RDA scores are positive, and are flipped 180° when they are negative. For OC type occurrence, darker colors indicate areas of stronger variation in anomalies associated with the temporal signal.

type. The leading RDA1 patterns of all OC types were significant at the $p \leq 0.05$, with RDA2 being only marginally significant for OC1. It should be noted that the cumulative explained variance of the significant RDA axes is lower than that of the full RDA model. Typically, this might indicate a need to remove insignificant explanatory variables from the constraining matrix, although this is not possible in the current application given that the PCs are component parts of a single field. In this special application, the total explained variation is an appropriate measure of the model's overall predictive skill, although only the leading RDA patterns are likely associated with larger-scale features in the OC field. The RDA1 patterns for the Atlantic-associated OC1 and OC2 types were largely opposite (out of phase) to those of OC3 given that periods of increased Atlantic inflow would naturally displace the more coastally-associated OC3.

Positive temporal RDA scores, which are associated with increased occurrence and extent of OC1 and OC2 in the North Sea, are particularly associated with a strengthened Norwegian Trench Inflow, located along the western

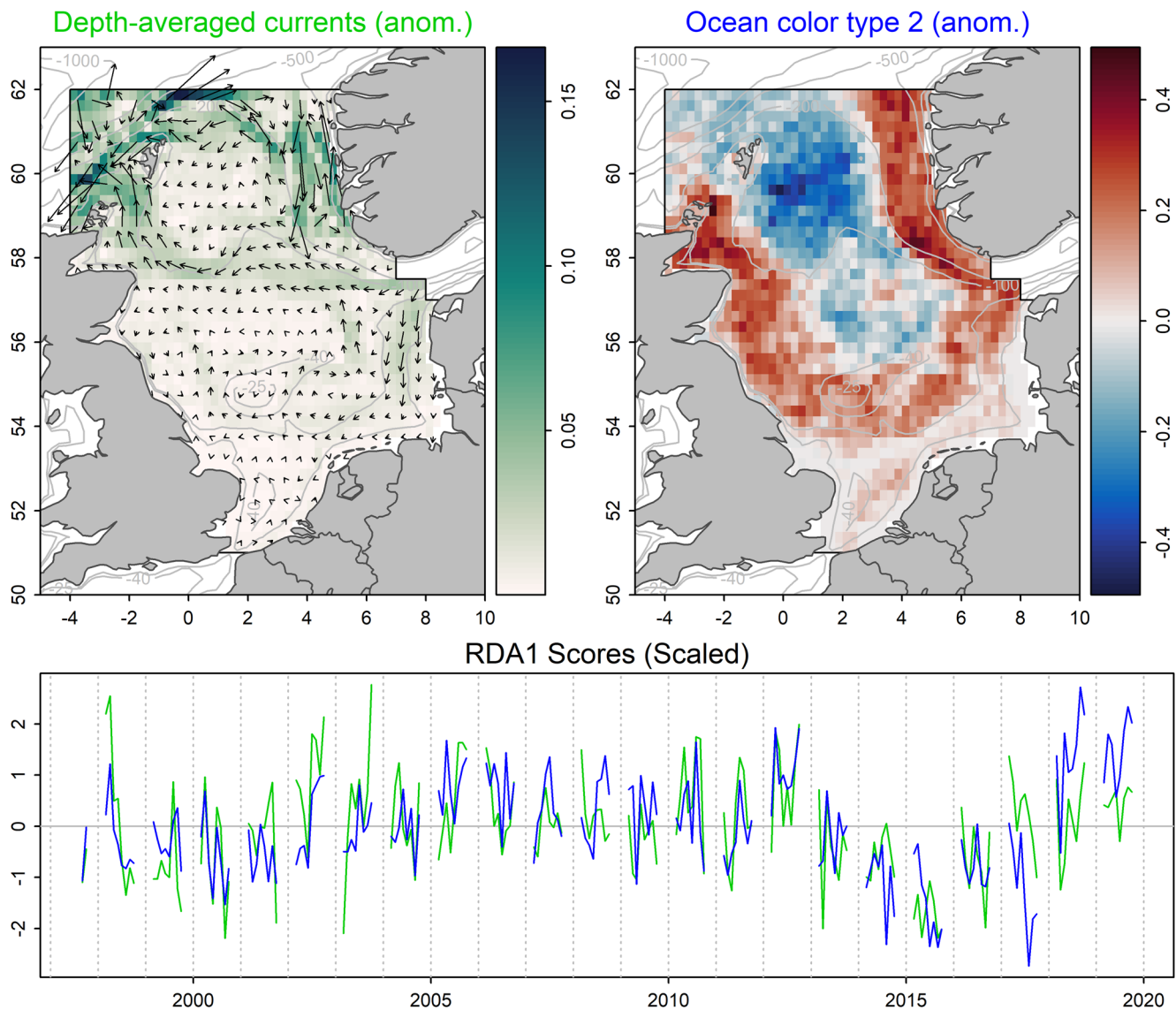


Figure 10. Leading Redundancy Analysis patterns (RDA1) between monthly anomalies of depth-averaged currents (CUR) (constraining field) and ocean color type 2 (OC2) occurrence (response field) in the North Sea (black boundary). See Figure 9 legend for additional details.

edge of the Norwegian Trench, and a weakened Fair Isle Current (FIC), flowing southward between Fair Island and mainland Shetland (Figures 9 and 10, top-left panels). A closer examination of the original current fields in both of these areas confirmed that negative anomalies reflect current weakening, with only rare cases of a reversal in direction. Current reversals of the dominant cyclonic circulation pattern in the North Sea are associated with occasional periods of easterly winds (Sündermann & Pohlmann, 2011). The increased occurrence of OC1 east of the Shetlands is associated with a lower occurrence of OC2 in that area as its distribution is shifted further to the south. The normally eastward flowing Dooley Current, found north of the 100 m depth isoline, also shows weakened flows during these positive RDA1 phases. This weakening is associated with a further southward penetration of the more coastally-associated OC2 waters along the Scottish coast and into the southern North Sea (Figure 10, upper right panel). The more Atlantic-associated OC1 (Figure 9, upper right panel) is mainly associated with an intensification of the Norwegian Trench Inflow, which shows southward oriented CUR anomalies (Figure 9, upper left panel). Decreased occurrence of OC1 and OC2 are linked with CUR anomalies in the opposite direction (i.e. flipped by 180°), which are especially associated with the period of 2014–2017 (Figures 9 and 10, lower panels). As stated above, the more coastally-associated OC3 patterns were largely opposite in their dynamics (Figure S8 in Supporting Information S1), especially when compared to a subsequent RDA using a

Table 4

Correlations Between Mean Annual North Sea Ocean Color (OC) Type Anomalies and Large-Scale Environmental Indices of the North Atlantic Oscillation (NAO) and Subpolar Gyre (SPG) Indices.

OC type ^a	NAO (lag 0)	NAO (lag 1)	SPG (lag 0)	SPG (lag 1)
OC1&2	-0.40	-0.14	0.48*	0.33
OC1	-0.38	-0.09	0.34	0.25
OC2	-0.43*	-0.13	0.41	0.26
OC3	0.38	0.12	-0.46*	-0.31
OC4	-0.07	-0.55*	0.32	0.20

Note. Correlations (Spearman ρ) are based on yearly averages from 1998–2019. Lags refer to the number of years shifted in the OC time series relative to the environmental indices.

^aOC temporal signals are the mean annual PC1 loading derived from the EOF of monthly anomalies of occurrence. *Correlations with an associated p-value at the ≤ 0.05 level.

field of combined OC1 and OC2 occurrence (OC1&2, Figure S10 in Supporting Information S1). This can be seen in the fact that the spatial patterns are opposite in sign (i.e. indicated by color). OC4, which is associated with the littoral zone, showed more unique dynamics (Figure S9 in Supporting Information S1), but with clear temporal similarities to OC3. The spatial RDA pattern for OC4 showed a greater influence of CUR anomalies associated with English Channel inflow and northward propagating continental coastal waters.

Spearman correlation coefficients (ρ) between the dominant yearly indices of OC type anomalies (PC1) and other environmental indices are presented in Table 4. Generally, correlations were higher at lag 0 than lag 1, and of similar magnitude for the NAO and SPG indices. The sign of the correlations is opposite between groups of Atlantic-associated OC types (OC1, OC2, OC1&2) and the more coastally-associated OC3. OC4 had the weakest correlations to the tested environmental indices overall, and showed the most ambiguous interannual pattern of the OC types. One exception was the correlation between OC4 at lag 1 and the NAO ($\rho = -0.55$).

Some of the strongest correlations were observed between the SPG index and both OC1&2 ($\rho = 0.48$) and OC3 ($\rho = -0.46$) indices, both at lag 0. The

NAO index was most strongly correlated with OC2 indices at lag 0 ($\rho = -0.43$). A plot of the time series from the strongest correlation between the SPG index and OC1&2 shows that the strongly negative OC1&2 anomalies of 2014–2017 are related to a stronger subpolar gyre (i.e. negative SPG index) (Figure 11). The largest outliers of the relationship are from the most recent years, where OC1&2 anomalies have returned to positive values while the SPG index has recovered but remains negative.

4. Discussion

4.1. Ocean Color Provinces and Their Characteristics

Longhurst (2007) included the North Sea with the much larger Northeast Atlantic Shelves Province ("NECS"), which encompassed the continental shelf of Western Europe, from Cape Finisterre in NW Spain to the Skagerrak north of Denmark, and the Baltic Sea. Nevertheless, he acknowledged that it could be further divided into smaller partitions, including one for the North Sea that extends from the Straits of Dover to the Shetland Islands. This definition more or less coincides with our delineation of the North Sea (Figure 2) and largely corresponds to the extent of the *Greater North Sea* ecoregion defined by ICES (ICES, 2004, 2020). This ecoregion has been delineated with ecosystem-based fisheries management in mind, including political, social, economic and management considerations. Several other works have defined even smaller divisions for the North Sea based on hydrographic regimes (Becker & Pauly, 1996; Laevastu, 1962; Lee, 1980; van Leeuwen et al., 2015) and fish assemblages (Ehrich et al., 2009; Frelat et al., 2017).

Our results also indicate that finer province delineations are likely more appropriate and that the use of dynamic boundaries, which could be monitored through the direct observation of OC types, could provide new insight into the hydrographic and biological changes of over time.

The four described OC types are consistently ordered along gradients of several abiotic and biotic parameters (Figure 6). These differences relate largely to whether the water is coastally- (OC3 and OC4) or offshore-associated (OC1 and OC2). OC4 corresponds to waters with the highest concentrations of Chl_a, POC and TSM, and the most nearshore distribution (Figure 4). Its spectral shape (Figure 3) is typical of very turbid water, which often contains high concentrations of mineral particles and dissolved organic matter (Babin & Stramski, 2004; Vantrepotte et al., 2012). Of the other abiotic parameters, SSS was the most clearly differentiated among OC types. A particularly narrow distribution of values is associated with OC1 (SSS above 35 g kg⁻¹),

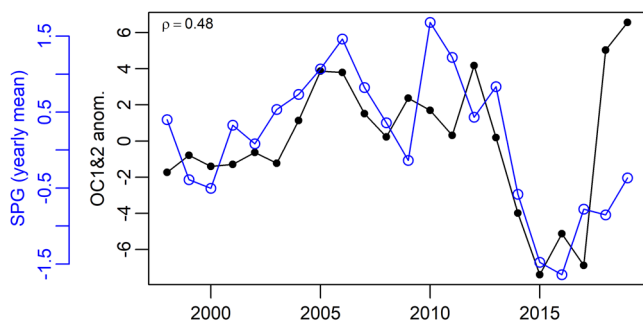


Figure 11. Mean annual OC1&2 anomalies versus Subpolar Gyre (SPG) index. The OC1&2 time series is the mean annual PC1 loading derived from the EOF analysis of monthly anomalies of occurrence. The Spearman correlation coefficient (ρ) of the two series is shown in the top left.

making it the most likely choice to use as an indicator of waters with Atlantic origin. SST was not distinguishable among OC types, although was significantly related to changes in the zooplankton community (Figure 7, Table 2). This discrepancy may be due to the complexity of SST patterns in 3D (e.g. dynamic of the stratification and heat exchange with the atmosphere, currents at different depth) and may more relate to seasonal patterns, such as the vertical migration of plankton (e.g. diel vertical migration).

OC3 and OC4 show a bimodal seasonal pattern with highest spatial coverage during the spring and late summer. These periods correlate with the typical annual pattern of phytoplankton blooms, which raises the question of whether the OC type classification is registering an increase in the coverage of low-Chl_a waters (OC1 and OC2) in summer simply due to the diminishment of blooms rather than variability in the inflow of Atlantic waters. Support for bloom depletion lies in the fact that OC1 and OC2 are associated with the area of seasonally-stratified waters (van Leeuwen et al., 2015), which may hinder nutrient resuspension to the photic zone while simultaneously facilitating losses via increased heterotrophic grazing (sensu the “Dilution–Recoupling Hypothesis” of Behrenfeld, 2010). On the other hand, support for Atlantic origin of these waters comes from the results of the coupled RDA patterns, which show a clear overlap between current variability and pathways of inflow for these Atlantic-associated OC types. Furthermore, OC1 and OC2 are clearly associated with salinities above 35 g kg⁻¹ (Figure 6), which is a signature of the Atlantic waters in the North Sea (e.g. Lee, 1980). Despite this evidence, the two mechanisms are not necessarily mutually exclusive, and further research is required to address the stability of OC types over time.

Various studies have focused on the dynamics of zooplankton in the Northeast Atlantic, and a large majority have relied on the CPR data set given its spatiotemporal coverage (e.g. Beaugrand, 2009; Hélaouët et al., 2013). The use of zooplankton as a possible indicator of Atlantic inflow into the North Sea has previously been a subject of study (e.g. Beare et al., 2002; Corten, 1999). In particular, the copepod species *Candacia armata* and *Metridia lucens* were identified as the most likely indicators of Atlantic waters given that their trends in the northwestern North Sea aligned with known inflow events (Corten, 1999). However, there was a lack of synchronicity with abundance trends in the area to the west of Scotland, which was assumed to be the source of inflow via FIC. Given the results of our work, this apparent disparity is explained by a stronger association of these indicator species to OC1 (Table 3), which is particularly related to inflows between the Shetland Islands and the Norwegian Trench (Figure 9), whereas OC2 is more related to inflow through the FIC (Figure 10), and may be more influenced by shelf waters. The extent of OC1 is also highest in late summer and early fall (Figure 8), which coincides with the months when both species were highly abundant in the northwestern North Sea area (Corten, 1999).

OC1 contains the highest number of significantly differing zooplankton taxa (Table 3). In addition to the two aforementioned taxa, *Calanus finmarchicus* is also strongly associated with OC1, and its abundance clearly declines with increasing OC type number. *C. finmarchicus* is known to be an important species of the Atlantic Subarctic and Arctic provinces (Hélaouët et al., 2011; Hélaouët & Beaugrand, 2007) but can enter into the North Sea with intrusions of Atlantic waters and can therefore be used as an indicator species. Combined with the clearly separated optical and abiotic characteristics, these findings would again suggest that OC1 shows promise for use as a proxy for waters of Atlantic origin. Taxa associated with the more coastally-associated OC3 and OC4 include *Centropages hamatus* and *Temora longicornis*, which have been previously described as neritic indicator species (Beare et al., 2002).

Although the OC types identified in this study show clear links to biotic and physical characteristics over the study period, extrapolation of these associations over longer time scales should be done with caution. Inferring links to other bio-optically derived characteristics (e.g. Chl_a) is likely to be robust given the common input data used for estimation (e.g. reflectance spectra); however, the association of OC types to specific plankton species may be less reliable given that they are not estimated remotely and may experience shifts in distribution over time that are not reflected by ocean color changes. For example, *C. finmarchicus* has shown a significant decrease in its abundance in the North Atlantic and North Sea (Beaugrand et al., 2003; Corten, 1999; Hélaouët et al., 2013; Pitois & Fox, 2006), which is also projected to continue in the coming decades (Hélaouët et al., 2011). This is linked to increased temperatures and a regime shift in the northeast Atlantic around the mid-1980s, which has resulted in a northward shift in its preferred habitat (Beaugrand, 2004; Hélaouët et al., 2013; Hélaouët & Beaugrand, 2007). However, planktonic aggregates, based on functional types or size, may show more consistency in relation to province dynamics on the regional scale of the North Sea. Ongoing work has shown that multispectral resolution data, such as were used in this study, are effective in identifying a suite of pigments and pigment groups

(Bracher et al., 2015) and have been effectively used to quantify major phytoplankton groups on a global scale (Xi et al., 2020).

4.2. Insights Into North Sea Hydrography

The ability to observe OC provinces remotely and in real-time makes their monitoring an attractive alternative over other more difficult to observe parameters. In terms of the oceanographic dynamics of the North Sea, spatiotemporal data on currents and salinity would also be obvious choices for the monitoring of water masses; however, their observation is expensive and therefore sparse in space and time. Hydrodynamic models offer valuable alternatives; yet, while their performance is steadily improving, some processes are still challenging to resolve (e.g. ocean-shelf exchange). Furthermore, the lag in new data releases may affect their utility for short-term forecasting. The links between currents and the dynamics of OC types in the North Sea potentially allow for a more straightforward and economical approach to monitoring inflow variability and the spatial distribution of water masses.

Our results indicate that the dynamics and extent of OC types in the North Sea are likely driven by a combination of large-scale circulation patterns of the eastern North Atlantic and local wind-driven processes. Interannual dynamics in OC provinces in the North Sea were shown to be correlated with the large-scale SPG index (Table 4). The strength of SPG modulates the warm, north-eastern limb of the Atlantic Meridional Overturning Circulation (AMOC), which in turn influences the flow and properties of waters onto the northwest European shelf and into the North Sea (Huthnance et al., 2016). Koul et al. (2019) found that weak phases of the SPG (i.e. positive indices) result in more saline water from the subtropical North-Atlantic dominating waters flowing into the North Sea, whereas less saline subpolar inflow water is typical during the periods of a strong SPG (i.e. negative indices). This is supported by the strongly positive correlation between the SPG index and interannual anomalies in OC1&2 (Table 4), which are associated with higher salinity (Figure 6). During the time period considered in our study, the SPG was relatively strong in the early 2000s (negative index), weakened gradually till 2010 (positive indices) and spun-up again afterwards (Figure 11). This strengthening (acceleration) of the SPG culminated during the period of 2014–2016, resulting in the lowest index measured since the existence of satellite altimetry (since 1993) (Chafik et al., 2019). The same period coincided with the lowest extent of Atlantic-associated OC types in the North Sea (Figures 8–10), whereas the peak sea level values of the SPG index in 2006 coincided with one of the highest extents of these OC types.

While variability in the SPG helps to explain variability in source waters arriving to the northwest European shelf, local-scale atmospheric processes influence the inflow and extent of these waters into the North Sea proper. In particular, the seasonal and interannual variability of OC1 and OC2 coverage (Figure 8) seem to be driven by wind- and density-driven currents in the northern and central North Sea. On seasonal scales, their coverage increases during the summer months, coinciding with periods of lowest wind (Sušelj et al., 2010) and water current velocities (Mathis et al., 2015). It has been suggested that the role of the baroclinic (i.e. density-driven) component of the Atlantic water inflow increases compared to the wind-driven component in summer (Dooley, 1983; Huthnance et al., 2016; Mathis et al., 2015).

Contrary to the positive correlation with the SPG index, a negative correlation was found between the extent of these OC types and the NAO index (Table 4). These inverse correlations are not surprising, since the NAO has been shown to drive to a large extent changes in the strength of the SPG and, thereby, the water properties in the eastern North Atlantic (Bersch et al., 2007). Curry and McCartney (2001) showed that the intensity of the SPG is a time-integrated oceanic response to persistent NAO anomalies. For the period considered in our study, the correlation between both indices was -0.47 ($p < 0.05$). Positive NAO indices are associated with enhanced westerly winds, increased inflow into the North Sea, and an intensified cyclonic (anti-clockwise) circulation (Emeis et al., 2015; Mathis et al., 2015). These patterns are reflected by the results of the RDA, which show that the southward penetration of Atlantic waters is associated with a weakening of the Dooley current and an intensification of the Norwegian Trench inflow (Figures 9 and 10). Using a modeling approach, Mathis et al. (2015) estimated that, on average, a large majority of the FIC and of the inflow east of the Shetland Islands (76%) turn eastward along the 100 m isobath of the Dooley Current, while 24% of inflows penetrate further southward to form the central recirculation cell north of the Dogger Bank. Of these waters, only 5% penetrate to the most southern reaches of the North Sea, where they merge with Atlantic waters entering via the English Channel. Our results suggest that the weakening of the Dooley current, in particular, allows for the penetration of OC2 into

these less-commonly reached southern areas of the North Sea, around the limits of the 40 m isobath and just south of the Dogger Bank (seen in Figure 10 for OC2 and Figure S10 in Supporting Information S1 for OC1&2). The fact that this penetration increases during periods of reduced wind-forcing is further support of the importance of baroclinic forces for Atlantic inflow on the interannual scale (Núñez-Riboni & Akimova, 2017).

Based on these findings, we propose that OC provinces have predictive skill in identifying water mass characteristics of the North Sea, which are largely driven by the interaction between large-scale circulation of the eastern North Atlantic and the wind-driven and baroclinic local currents. Although the two processes operate on differing spatial scales, a degree of synergy was observed during the study period, with both the strength of the SPG and the local wind-forced currents correlating with OC province dynamics. This is supported by the existence of strong external control on North Sea circulation (Holt & Proctor, 2008; Huthnance, 1997). Recent work of Pätsch et al. (2020) also shows that a similar interplay of regional and local processes helps to explain variability in salinity and nutrient concentrations in the North Sea. Specifically, while the SPG has been shown to affect the variability in salinity of waters entering into the northern North Sea following an advective delay of up to two years (see also Koul et al., 2019), nutrient concentrations are more immediately affected. The authors propose that atmospheric drivers are more likely driving nutrient dynamics. Given that OC province dynamics also show a higher correlation with unlagged indices of the SPG and NAO indices (Table 4), as well as their association to phytoplankton concentrations (Chla, Figures 6 and 7), it is likely that their spatial variability reflects similar changes in water mass characteristics that can influence primary production, such as nutrient concentrations.

4.3. Ecological Implications and Prospectus

The impact of OC province dynamics on ecological processes is likely to be a complex mixture of trophic and physical effects. The correlation between OC provinces and associated bio-physical properties has provided evidence of a link to lower trophic levels, which, in turn, are likely to have bottom-up effects to other components of the ecosystem. Furthermore, the spatially explicit perspective sheds new light on the dynamics of Atlantic waters inflow to the North Sea in terms of coverage rather than the more typical focus on (volumetric) fluxes. Both aspects are likely to be related, albeit inversely depending on where inflow is measured; however, the areal extent of OC provinces may be of particular relevance to ecological processes that are spatial in nature. For example, positive inflow anomalies in 1988/89 and 1998 were claimed to have been the cause of increased phytoplankton biomass in the central North Sea through enhanced nutrient availability (Reid et al., 2001). Our results suggest that these increases may not be a direct consequence of the increased influx of Atlantic waters, but rather due to a larger relative coverage of the more Chla-rich provinces, OC3 and OC4, which expand their distribution further northward as currents increase and Atlantic water penetration is reduced.

From this spatial perspective, monitoring OC provinces may further elucidate other ecologically-relevant processes in higher trophic levels, such as habitat changes or the transport of eggs and larvae from broadcast spawning species to important nursery areas. The area of inflow in the northwestern North Sea is considered an important spawning area for cod in spring (González-Irusta & Wright, 2016) and herring in fall (e.g. Dickey-Collas et al., 2010), and is generally associated with highest catches for many gadoid fish species (ICES, 2019). This habitat preference may be related to the observed differences in the zooplankton community associated with the Atlantic waters. For example, cod larvae and early juvenile stages are known to prefer larger calanoid copepods like *C. finmarchicus*, and decreases in cod recruitment in recent decades has been in part attributed to a possible mismatch in the spatial distribution of these zooplankton (Beaugrand et al., 2003).

Such mismatches may help to explain the large degree of recruitment variability in many fish stocks, which has implications for management and the ability to provide reliable short-term advice in the form of future catch quotas. The incorporation of environmental indices in models of historical stock-recruitment relationships has demonstrated significant improvements in fit (i.e. in addition to spawning stock biomass), and the use of more spatially-focused, smaller-scale environmental signals may be of particular importance to understanding recruitment dynamics (Akimova et al., 2016). Novel statistical approaches (e.g. Machine Learning) have also shown promise in the identification of environmental drivers for recruitment from large data sets of spatially explicit data (Kühn et al., 2021). The ever-growing suite of available environmental data sets, including ocean color, will likely improve our ability to explore environmental drivers to resource dynamics in the future. The ability to monitor OC provinces in real-time provides a unique operational index for use in other modelling and forecasting applications.

Data Availability Statement

GEBCO bathymetric data were used in figures (GEBCO Bathymetric Compilation Group 2020). CPR data are deposited at <https://www.mba.ac.uk/fellows/cpr-survey>, with the specific data extraction provided by Hélaouët (2020). Ocean color products are deposited at <http://globcolour.info>. North-West European Shelf reanalysis data are deposited at <https://marine.copernicus.eu>, with the Product ID: "NWSHELF_MULTYEAR_PHY_004_009". All processed data products, R-scripts, and yearly environmental indices (NAO, SPG) are deposited in Figshare (Taylor et al., 2021).

Acknowledgments

This study was in part conducted within the frame of the European Union's Horizon 2020 research and innovation program under the grant agreement No. 773713 (PANDORA). This study has been conducted using E.U. Copernicus Marine Service Information. Thanks to Léon Chafik and Hjálmar Hátún for providing the Subpolar Gyre Index. Thanks to Rabea Diekmann and Anne Sell for fruitful discussions that helped to shape the study. Thanks to David Zelený for helpful insights on RDA model interpretation and significance testing.

References

- Akimova, A., Núñez-Riboni, I., Kempf, A., & Taylor, M. H. (2016). Spatially-Resolved Influence of Temperature and Salinity on Stock and Recruitment Variability of Commercially Important Fishes in the North Sea. *PLOS ONE*, *11*(9), e0161917. <https://doi.org/10.1371/journal.pone.0161917>
- Alvarez-Fernandez, S., Lindeboom, H., & Meesters, E. (2012). Temporal changes in plankton of the North Sea: community shifts and environmental drivers. *Marine Ecology Progress Series*, *462*, 21–38. <https://doi.org/10.3354/meps09817>
- Babin, M., & Stramski, D. (2004). Variations in the mass-specific absorption coefficient of mineral particles suspended in water. *Limnology and Oceanography*, *49*(3), 756–767. <https://doi.org/10.4319/lo.2004.49.3.0756>
- Batten, S. D., Clark, R., Flinkman, J., Hays, G., John, E., John, A. W. G., et al. (2003). CPR sampling: the technical background, materials and methods, consistency and comparability. *Progress in Oceanography*, *58*(2–4), 193–215. <https://doi.org/10.1016/j.pocean.2003.08.004>
- Beare, D. J., Batten, S., Edwards, M., & Reid, D. G. (2002). Prevalence of boreal Atlantic, temperate Atlantic and neritic zooplankton in the North Sea between 1958 and 1998 in relation to temperature, salinity, stratification intensity and Atlantic inflow. *Journal of Sea Research*, *48*(1), 29–49. [https://doi.org/10.1016/S1385-1101\(02\)00131-4](https://doi.org/10.1016/S1385-1101(02)00131-4)
- Beaugrand, G. (2004). The North Sea regime shift: Evidence, causes, mechanisms and consequences. *Progress in Oceanography*, *60*(2–4), 245–262. <https://doi.org/10.1016/j.pocean.2004.02.018>
- Beaugrand, G. (2009). Decadal changes in climate and ecosystems in the North Atlantic Ocean and adjacent seas. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(8–10), 656–673. <https://doi.org/10.1016/j.dsr2.2008.12.022>
- Beaugrand, G., Brander, K. M., Lindley, J. A., Souissi, S., & Reid, P. C. (2003). Plankton effect on cod recruitment in the North Sea. *Nature*, *426*(6967), 661–664. <https://doi.org/10.1038/nature02164>
- Becker, G., & Pauly, M. (1996). Sea surface temperature changes in the North Sea and their causes. *ICES Journal of Marine Science*, *53*(6), 887–898. <https://doi.org/10.1006/jmsc.1996.0111>
- Beckers, J. M., & Rixen, M. (2003). EOF Calculations and Data Filling from Incomplete Oceanographic Datasets. *Journal of Atmospheric and Oceanic Technology*, *20*(12), 1839–1856. [https://doi.org/10.1175/1520-0426\(2003\)020<1839:ECADFF>2.0.CO;2](https://doi.org/10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2)
- Behrenfeld, M. J. (2010). Abandoning Sverdrup's Critical Depth Hypothesis on phytoplankton blooms. *Ecology*, *91*(4), 977–989. <https://doi.org/10.1890/09-1207.1>
- Bersch, M., Yashayaev, I., & Koltermann, K. P. (2007). Recent changes of the thermohaline circulation in the subpolar North Atlantic. *Ocean Dynamics*, *57*(3), 223–235. <https://doi.org/10.1007/s10236-007-0104-7>
- Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., & Li, S. (2019). *FNN: Fast Nearest Neighbor Search Algorithms and Applications (Version 1.1.3)*. Retrieved from <https://CRAN.R-project.org/package=FNN>
- Bracher, A., Taylor, M. H., Taylor, B. B., Dinter, T., Röttgers, R., & Steinmetz, F. (2015). Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. *Ocean Science*, *11*(1), 139–158. <https://doi.org/10.5194/os-11-139-2015>
- Bracher, A., Xi, H., Dinter, T., Mangin, A., Strass, V., von Appen, W.-J., & Wiegmann, S. (2020). High Resolution Water Column Phytoplankton Composition Across the Atlantic Ocean From Ship-Towed Vertical Undulating Radiometry. *Frontiers in Marine Science*, *7*, 235. <https://doi.org/10.3389/fmars.2020.00235>
- Chafik, L., Nilsen, J. E. Ø., Dangendorf, S., Reverdin, G., & Frederikse, T. (2019). North Atlantic Ocean Circulation and Decadal Sea Level Change During the Altimetry Era. *Scientific Reports*, *9*(1), 1041. <https://doi.org/10.1038/s41598-018-37603-6>
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral Ecology*, *18*(1), 117–143. <https://doi.org/10.1111/j.1442-9993.1993.tb00438.x>
- Corten, A. (1999). Evidence from plankton for multi-annual variations of Atlantic inflow in the northwestern North Sea. *Journal of Sea Research*, *42*(3), 191–205. [https://doi.org/10.1016/S1385-1101\(99\)00030-1](https://doi.org/10.1016/S1385-1101(99)00030-1)
- Curry, R. G., & McCartney, M. S. (2001). Ocean Gyre Circulation Changes Associated with the North Atlantic Oscillation. *Journal of Physical Oceanography*, *31*(12), 3374–3400. [https://doi.org/10.1175/1520-0485\(2001\)031<3374:OGCCAW>2.0.CO;2](https://doi.org/10.1175/1520-0485(2001)031<3374:OGCCAW>2.0.CO;2)
- Dickey-Collas, M., Nash, R. D. M., Brunel, T., van Damme, C. J. G., Marshall, C. T., Payne, M. R., et al. (2010). Lessons learned from stock collapse and recovery of North Sea herring: a review. *ICES Journal of Marine Science*, *67*(9), 1875–1886. <https://doi.org/10.1093/icesjms/fsq033>
- Doerffer, R., & Schiller, H. (2007). The MERIS Case 2 water algorithm. *International Journal of Remote Sensing*, *28*(3–4), 517–535. <https://doi.org/10.1080/01431160600821127>
- Dooley, H. (1983). Seasonal variability in the position and strength of the Fair Isle Current. In *North Sea Dynamics* (pp. 108–119). Springer. https://doi.org/10.1007/978-3-642-68838-6_7
- Ehrich, S., Stelzenmüller, V., & Adlerstein, S. (2009). Linking spatial pattern of bottom fish assemblages with water masses in the North Sea. *Fisheries Oceanography*, *18*(1), 36–50. <https://doi.org/10.1111/j.1365-2419.2008.00495.x>
- Emeis, K.-C., van Beusekom, J., Callies, U., Ebinghaus, R., Kannen, A., Kraus, G., et al. (2015). The North Sea — A shelf sea in the Anthropocene. *Journal of Marine Systems*, *141*, 18–33. <https://doi.org/10.1016/j.jmarsys.2014.03.012>
- Fanton d'Andon, O., Mangin, A., Lavender, S., Antoine, D., Maritorena, S., Morel, A., et al. (2009). GlobColour—the European Service for Ocean Colour. In *Proceedings from 2009 IEEE International Geoscience & Remote Sensing Symposium (IGARSS)*. IEEE International Geoscience & Remote Sensing Symposium (IGARSS).
- Frelat, R., Lindgren, M., Denker, T. S., Floeter, J., Fock, H. O., Sguotti, C., et al. (2017). Community ecology in 3D: Tensor decomposition reveals spatio-temporal dynamics of large ecological communities. *PLOS ONE*, *12*(11), e0188205. <https://doi.org/10.1371/journal.pone.0188205>

- GEBCO Bathymetric Compilation Group 2020. (2020). *The GEBCO_2020 Grid - a continuous terrain model of the global oceans and land. (Version 1) [Network Common Data Form]*. British Oceanographic Data Centre, National Oceanography Centre. <https://doi.org/10.5285/A29C5465-B138-234D-E053-6C86ABC040B9>
- González-Irusta, J. M., & Wright, P. J. (2016). Spawning grounds of Atlantic cod (*Gadus morhua*) in the North Sea. *ICES Journal of Marine Science: Journal Du Conseil*, 73(2), 304–315. <https://doi.org/10.1093/icesjms/fsv180>
- Häkkinen, S., Rhines, P. B., & Worthen, D. L. (2011). Warm and saline events embedded in the meridional circulation of the northern North Atlantic. *Journal of Geophysical Research: Oceans*, 116(C3). <https://doi.org/10.1029/2010JC006275>
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108. <https://doi.org/10.2307/2346830>
- Hátún, H., & Chafik, L. (2018). On the Recent Ambiguity of the North Atlantic Subpolar Gyre Index. *Journal of Geophysical Research: Oceans*, 123(8), 5072–5076. <https://doi.org/10.1029/2018JC014101>
- Hélaouët, P. (2020). *2020 Marine Biological Association of the UK (MBA) CPR Data request Marc Taylor Thunen-Institute March 2020 [Text/csv]*. The Archive for Marine Species and Habitats Data (DASSH). <https://doi.org/10.17031/1658>
- Hélaouët, P., & Beaugrand, G. (2007). Macroecology of *Calanus finmarchicus* and *C. helgolandicus* in the North Atlantic Ocean and adjacent seas. *Marine Ecology Progress Series*, 345, 147–165. <https://doi.org/10.3354/meps06775>
- Hélaouët, P., Beaugrand, G., & Edwards, M. (2013). Understanding Long-Term Changes in Species Abundance Using a Niche-Based Approach. *PLoS ONE*, 8(11), e79186. <https://doi.org/10.1371/journal.pone.0079186>
- Hélaouët, P., Beaugrand, G., & Reid, P. C. (2011). Macrophysiology of *Calanus finmarchicus* in the North Atlantic Ocean. *Progress in Oceanography*, 91(3), 217–228. <https://doi.org/10.1016/j.pocean.2010.11.003>
- Hijmans, R. J. (2020). *raster: Geographic Data Analysis and Modeling (Version 3.0-12)*. Retrieved from <https://CRAN.R-project.org/package=raster>
- Holt, J., & Proctor, R. (2008). The seasonal circulation and volume transport on the northwest European continental shelf: A fine-resolution model study. *Journal of Geophysical Research*, 113(C6), C06021. <https://doi.org/10.1029/2006JC004034>
- Hurrell, J. W. (1995). Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. *Science*, 269(5224), 676–679. <https://doi.org/10.1126/science.269.5224.676>
- Huthnance, J. (1997). North sea interaction with the north atlantic ocean. *Deutsche Hydrographische Zeitschrift*, 49(2–3), 153–162. <https://doi.org/10.1007/BF02764030>
- Huthnance, J., Weisse, R., Wahl, T., Thomas, H., Pietrzak, J., Souza, A. J., et al. (2016). Recent Change—North Sea. In M. Quante, & F. Colijn (Eds.), *North Sea Region Climate Change Assessment* (pp. 85–136). Springer International Publishing. https://doi.org/10.1007/978-3-319-39745-0_3
- ICES. (2004). Information and advice about appropriate eco-regions for the implementation of an ecosystem approach in European waters. *ICES Advice*, 1(2), 115–131.
- ICES. (2019). Report of the Working Group on Mixed Fisheries Advice Methodology (WGMIXFISH-METHODS). *ICES CM 2018/ACOM*, 68(96), 15–19. October 2018.) (p. 96).
- ICES. (2020). *Definition and rationale for ICES ecoregions*. <https://doi.org/10.17895/ICES.ADVICE.6014>
- Kauker, F., Köberle, C., Gerdes, R., & Karcher, M. (2008). Modeling the 20th century Arctic Ocean/Sea ice system: Reconstruction of surface forcing. *Journal of Geophysical Research*, 113(C9), C09027. <https://doi.org/10.1029/2006JC004023>
- Kavanaugh, M. T., Hales, B., Saraceno, M., Spitz, Y. H., White, A. E., & Letelier, R. M. (2014). Hierarchical and dynamic seascapes: A quantitative framework for scaling pelagic biogeochemistry and ecology. *Progress in Oceanography*, 120, 291–304. <https://doi.org/10.1016/j.pocean.2013.10.013>
- Koul, V., Schrum, C., Düsterhus, A., & Baehr, J. (2019). Atlantic Inflow to the North Sea Modulated by the Subpolar Gyre in a Historical Simulation With MPI-ESM. *Journal of Geophysical Research: Oceans*, 124(3), 1807–1826. <https://doi.org/10.1029/2018JC014738>
- Kühn, B., Taylor, M. H., & Kempf, A. (2021). Using machine learning to link spatiotemporal information to biological processes in the ocean – A case study for North Sea cod recruitment. *Marine Ecology Progress Series*, 664, 1–22. <https://doi.org/10.3354/meps13689>
- Laevastu, T. (1962). *Water types in the North Sea and their characteristics*. Meteorology & Oceanography Division, Hawaii Institute of Geophysics.
- Lee, A. J. (1980). North Sea: Physical Oceanography. In F. T. Banner, M. B. Collins, & K. S. Massie (Eds.), *Elsevier Oceanography Series* (Vol. 24, pp. 467–493). Elsevier. [https://doi.org/10.1016/S0422-9894\(08\)71359-X](https://doi.org/10.1016/S0422-9894(08)71359-X)
- Legendre, P., & Legendre, L. (2012). *Numerical ecology*, Developments in environmental modelling (3. Engl. ed). Elsevier. Amsterdam
- Legendre, P., Oksanen, J., & ter Braak, C. J. F. (2011). Testing the significance of canonical axes in redundancy analysis: Test of canonical axes in RDA. *Methods in Ecology and Evolution*, 2(3), 269–277. <https://doi.org/10.1111/j.2041-210X.2010.00078.x>
- Longhurst, A. R. (1998). *Ecological geography of the sea*. Academic Press.
- Longhurst, A. R. (2007). *Ecological geography of the sea* (2nd ed). Academic Press.
- Maritorena, S., d'Andon, O. H. F., Mangin, A., & Siegel, D. A. (2010). Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, 114(8), 1791–1804. <https://doi.org/10.1016/j.rse.2010.04.002>
- Mathis, M., Elizalde, A., Mikolajewicz, U., & Pohlmann, T. (2015). Variability patterns of the general circulation and sea water temperature in the North Sea. *Progress in Oceanography*, 135, 91–112. <https://doi.org/10.1016/j.pocean.2015.04.009>
- Mélin, F., & Vantrepotte, V. (2015). How optically diverse is the coastal ocean? *Remote Sensing of Environment*, 160, 235–251. <https://doi.org/10.1016/j.rse.2015.01.023>
- Núñez-Riboni, I., & Akimova, A. (2017). Quantifying the impact of the major driving mechanisms of inter-annual variability of salinity in the North Sea. *Progress in Oceanography*, 154, 25–37. <https://doi.org/10.1016/j.pocean.2017.04.004>
- Økland, R. H. (1999). On the variation explained by ordination and constrained ordination axes. *Journal of Vegetation Science*, 10(1), 131–136. <https://doi.org/10.2307/3237168>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2019). *vegan: Community Ecology Package (Version 2.5-6)*. Retrieved from <https://CRAN.R-project.org/package=vegan>
- Pätsch, J., Gouretski, V., Hinrichs, I., & Koul, V. (2020). Distinct Mechanisms Underlying Interannual to Decadal Variability of Observed Salinity and Nutrient Concentration in the Northern North Sea. *Journal of Geophysical Research: Oceans*, 125(5), e2019JC015825. <https://doi.org/10.1029/2019JC015825>
- Pitois, S. G., & Fox, C. J. (2006). Long-term changes in zooplankton biomass concentration and mean size over the Northwest European shelf inferred from Continuous Plankton Recorder data. *ICES Journal of Marine Science*, 63(5), 785–798. <https://doi.org/10.1016/j.icesjms.2006.03.009>
- Platt, T., & Sathyendranath, S. (1988). Oceanic Primary Production: Estimation by Remote Sensing at Local and Regional Scales. *Science*, 241(4873), 1613–1620. <https://doi.org/10.1126/science.241.4873.1613>

- Platt, T. & Sathyendranath, S. (1999). Spatial Structure of Pelagic Ecosystem Processes in the Global Ocean. *Ecosystems*, 2(5), 384–394. <https://doi.org/10.1007/s100219900088>
- Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, 329–358.
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>. Vienna, Austria
- Reid, P. C., & Edwards, M. (2001). Long-term changes in the pelagos, benthos and fisheries of the North Sea. *Senckenbergiana Maritima*, 31(2), 107–115. <https://doi.org/10.1007/BF03043021>
- Reid, P. C., Holliday, N. P., & Smyth, T. J. (2001). Pulses in the eastern margin current and warmer water off the north west European shelf linked to North Sea ecosystem changes. *Marine Ecology Progress Series*, 215, 283–287. <https://doi.org/10.3354/meps215283>
- Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., & Maury, O. (2013). Dynamic biogeochemical provinces in the global ocean. *Global Biogeochemical Cycles*, 27(4), 1046–1058. <https://doi.org/10.1002/gbc.20089>
- Sathyendranath, S., Brewin, R. J. W., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., et al. (2019). An Ocean-Colour Time Series for Use in Climate Studies: The Experience of the Ocean-Colour Climate Change Initiative (OC-CCI). *Sensors*, 19(19), 4285. <https://doi.org/10.3390/s19194285>
- Shen, J., Chang, S. I., Lee, E. S., Deng, Y., & Brown, S. J. (2005). Determination of cluster number in clustering microarray data. *Applied Mathematics and Computation*, 169(2), 1172–1185. <https://doi.org/10.1016/j.amc.2004.10.076>
- Sonnwald, M., Dutkiewicz, S., Hill, C., & Forget, G. (2020). Elucidating ecological complexity: Unsupervised learning determines global marine eco-provinces. *Science Advances*, 6(22), eaay4740. <https://doi.org/10.1126/sciadv.aay4740>
- von Storch, H., & Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press. Retrieved from <https://doi.org/10.1017/CBO9780511612336>
- Sündermann, J., & Pohlmann, T. (2011). A brief analysis of North Sea physics. *Oceanologia*, 53(3), 663–689. <https://doi.org/10.5697/oc.53-3.663>
- Sušelj, K., Sood, A., & Heinemann, D. (2010). North Sea near-surface wind climate and its relation to the large-scale circulation patterns. *Theoretical and Applied Climatology*, 99(3–4), 403–419. <https://doi.org/10.1007/s00704-009-0149-2>
- Taylor, B. B., Torrecilla, E., Bernhardt, A., Taylor, M. H., Peeken, I., Röttgers, R., et al. (2011). Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance. *Biogeosciences*, 8(12), 3609–3629. <https://doi.org/10.5194/bg-8-3609-2011>
- Taylor, M. H. (2020). *sinkr: Collection of functions with emphasis in multivariate data analysis (Version 0.6)*. Retrieved from <https://github.com/marctaylor/sinkr>
- Taylor, M. H., Akimova, A., Bracher, A., Kempf, A., Kühn, B., & Hélaouët, P. (2021). *ocprovinces: Ocean color provinces of the North Sea*. Retrieved from <https://doi.org/10.6084/m9.figshare.16691371>
- ter Braak, C. J. F. (1986). Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. *Ecology*, 67(5), 1167–1179. <https://doi.org/10.2307/1938672>
- ter Braak, C. J. F. (1987). The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, 69(1–3), 69–77. <https://doi.org/10.1007/BF00038688>
- Tuszynski, J. (2020). *caTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc (Version 1.18.0)*. Retrieved from <https://CRAN.R-project.org/package=caTools>
- Tyler, D. E. (1982). On the optimality of the simultaneous redundancy transformations. *Psychometrika*, 47(1), 77–86. <https://doi.org/10.1007/BF02293852>
- van Leeuwen, S., Tett, P., Mills, D., & van der Molen, J. (2015). Stratified and nonstratified areas in the North Sea: Long-term variability and biological and policy implications: North Sea stratification regimes. *Journal of Geophysical Research: Oceans*, 120(7), 4670–4686. <https://doi.org/10.1002/2014JC010485>
- Vantrepotte, V., Loisel, H., Dessailly, D., & Mériaux, X. (2012). Optical classification of contrasted coastal waters. *Remote Sensing of Environment*, 123, 306–323. <https://doi.org/10.1016/j.rse.2012.03.004>
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>
- Wood, S. N. (2020). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation (Version 1.8.31)*. Retrieved from <https://CRAN.R-project.org/package=mgcv>
- Xi, H., Losa, S. N., Mangin, A., Soppa, M. A., Garnesson, P., Demaria, J., et al. (2020). Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data. *Remote Sensing of Environment*, 240, 111704. <https://doi.org/10.1016/j.rse.2020.111704>