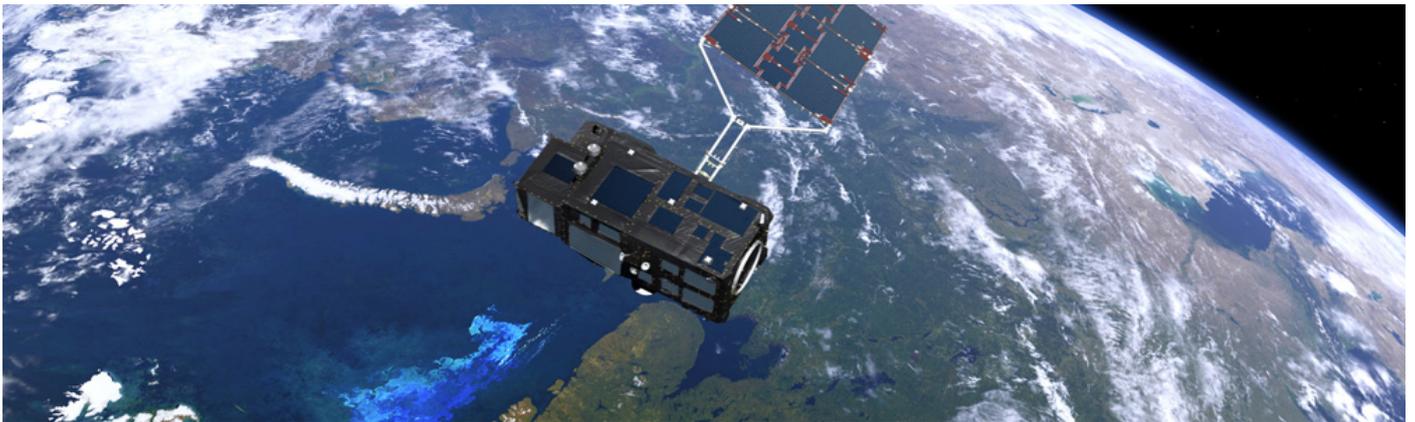# ACTIVITY 1

Coherence in assessment framework of chlorophyll a and nutrients

Part of the EU project JMP EUNOSAT

May 2019

Authors & contributors:

Anouk Blauw[1*], Marieke Eleveld[1], Theo Prins[1], Firmijn Zijl[1], Julien Groenenboom[1], Gundula Winter[1], Lilith Kramer[1], Tineke Troost[1], Alena Bartosova[2], Johannes Johansson[2], René Capell[2], Kari Eiola[2], Anders Höglund[2], Gavin Tilstone[3], Peter E. Land[3], Victor Martinez-Vicente[3], Silvia Pardo[3], Dimitry van der Zande[4]

[1] Deltares,   Boussinesqweg 1, 2629 HV Delft, The Netherlands.
* Corresponding author   Tel.:   +316 1583 8520
  E-mail address: Anouk.Blauw@deltares.nl

[2] SMHI
[3] PML
[4] RBINS

Key words:

North Sea, eutrophication, assessment, threshold values, assessment areas, satellite data, chlorophyll-a, nutrients

The JMP EUNOSAT project partners are:

RWS (NL), RBINS (BE), Deltares (NL), AU (DK), Ifremer (FR), PML (UK), MSS (UK), CEFAS (UK), SMHI (SE), IMR (NO), NIVA (NO), UBA (DE), BSH (DE), NLWKN (DE)

Summary

Excess nutrients stimulate algal growth, which can result in lack of oxygen and light under water for plants and animals. EU member states need to report every 6 years on the water quality of their seas for the Marine Strategy Framework Directive (MSFD). These reports so far show an representation of algae concentrations, expressed as chlorophyll a, due to the use of different threshold values for Good Environmental Status and monitoring methods between countries. Countries collaborating in OSPAR have defined threshold values as 50% above the natural background concentration. However, countries have used different approaches to estimate these background concentrations, leading to incoherent threshold values. The European Commission has asked EU member states to improve on this so future assessment reports will give a good overview of the level of eutrophication in the North Sea.

The JMP EUNOSAT project (Joint Monitoring Programme of the Eutrophication of the North Sea with Satellite data) has developed a new common method for monitoring and assessment of eutrophication in the North Sea. The project is organised in 3 activities with the following aims:

a) Derivation of threshold values for Good Environmental Status (GES) for nutrients and algae concentrations with a common a common method for all North Sea countries (activity 1);
b) Generation and validation of a coherent multi-algorithm satellite-based chlorophyll-a product for the North Sea and the suitability of these products for eutrophication assessments (activity 2)
c) Definition of coherent assessment areas with similar ecological and physical functioning (activities 1 and 2 together)
d) Development of a potential design of a future monitoring and assessment programme (activity 3)

This report describes the results of Activity 1 of the project comprising the definition of threshold values and assessment areas. Additionally, the potential of satellite data as a source of primary production data is explored.

For the definition of threshold values the project proposed to use the year 1900 as a reference year representing natural background concentrations. During this period human impact on the environment was relatively limited. Natural background concentrations are highest near mouths of large rivers, such as along the Dutch coast.

Nutrient inputs from rivers to the North Sea have been estimated by SMHI (Sweden) with their hydrologic model of Europe. Next, Deltares used these inputs in their models to estimate corresponding nutrient and algae concentrations throughout the North Sea. Field observations of nutrients and algae are only available for recent years. Therefore, we also ran the models for recent years, to check if the models can reproduce recent field observations. New threshold values have been derived as 50% above the model estimated concentrations for 1900.

The newly proposed threshold values are coherent across the whole Greater North Sea and reflects spatial patterns in fresh water influence: with higher threshold values near outflows of major rivers and near the coast. Offshore the newly proposed threshold values were generally lower than currently used threshold values and nearshore higher.

PML (UK) processed and analysed a dataset of 20 years of primary production estimates from satellite data for the Greater North Sea. These data corresponded well to observations in Atlantic waters. Unfortunately, observed data on primary production data in the North Sea were lacking, hampering validation in those waters.

Satellite data and Ferrybox data proved to be a valuable data source for model validation. The combination of model data and satellite data proved to provide a good basis for understanding ecosystem functioning, by showing and explaining spatial patterns of key ecosystem variables. These spatial patterns formed the basis of a new definition of assessment areas with similar levels and seasonal variability of chlorophyll, salinity and stratification.

Table of contents

1. Introduction

1.1. Background

Excess nutrients stimulate algal growth, which can result in lack of oxygen and light under water for plants and animals. Since 2012 EU member states need to report every 6 years on the water quality of their seas for the Marine Strategy Framework Directive (MSFD). Before 2012 countries around the North Sea have already reported regularly on the status of water quality the North Sea for OSPAR: the Oslo-Paris Convention. These reports so far show an incoherent representation of the eutrophication status, due to the use of different threshold values for chlorophyll a, the common indicator for algal biomass, and monitoring methods between countries. All countries united in OSPAR have agreed that concentrations of chlorophyll and nutrients should not be more than 50% above the natural background concentration. This is the threshold level (also called assessment level) to distinguish between a eutrophication problem area and a non-problem area, which is linked to Good Environmental Status of the MSFD. However, countries have used different approaches to estimate these background concentrations, leading to incoherent threshold values. This is most prominent in the North Sea, which is divided in national territories of eight countries. A joint assessment of the eutrophication status of the North Sea partially failed because of these inconsistencies (OSPAR, 2017a). The European Commission has asked EU member states to improve on this so future assessment reports will give a good overview of the level of eutrophication.

1.2. Objectives of activity 1

The JMP-EUNOSAT project (Joint Monitoring Programme of the Eutrophication of the North Sea with Satellite data) has developed a new common method for monitoring and assessment of eutrophication in the North Sea. Activity 1 of JMP-EUNOSAT includes the following objectives:

- Derive threshold values for nutrient and algae concentrations with the same method for all North Sea countries;
- Propose a new cross-border coherent definition of assessment areas with similar algae dynamics
- Develop a satellite product for primary production as a potential indicator for food webs: MSFD descriptor D4C4 – Productivity of the trophic guild is not adversely affected due to anthropogenic pressures.

The novel proposed threshold values for the North Sea should be coherent in different ways:

- Between countries around the North Sea
- Between the eutrophication indicators for dissolved inorganic nitrogen (DIN), dissolved inorganic phosphorus (DIP) and chlorophyll-a (chl-a)

1.3. Approach

The main causes of eutrophication in marine waters are nutrient inputs from land, through rivers and atmospheric deposition. The nutrient inputs from rivers are diluted with marine waters, resulting in elevated concentrations of nutrients and chlorophyll near the outflow of major rivers and near the coast.

Within OSPAR threshold values for eutrophication are traditionally determined as 50% above the estimated concentration of nutrients and chlorophyll under natural reference conditions. Countries have earlier used different reference years representing 'natural background conditions'. The project investigated the methods used for determination of background concentrations in all North Sea countries and agreed to use a common reference year representation natural background conditions: the period around 1900 (chapter 2). For a derivation of coherent

threshold values we used the same step-wise approach (as illustrated in Figure D.1) for all North Sea territorial waters:

1. We estimated nutrient loads to the North Sea from rivers under natural reference conditions, using the European model E-HYPE and observed data (by SMHI project: chapter 4);
2. We estimated nutrient concentrations in the North Sea under natural reference conditions, by combining the nutrient loads from E-HYPE with a transport model, assuming conservative transport (chapter 5);
3. We estimated chlorophyll concentrations, corresponding to the estimated nutrient concentrations under natural reference conditions, using different modelling approaches (chapter 6);
4. We estimated coherent threshold values based on the above modelling activities (chapter 7).
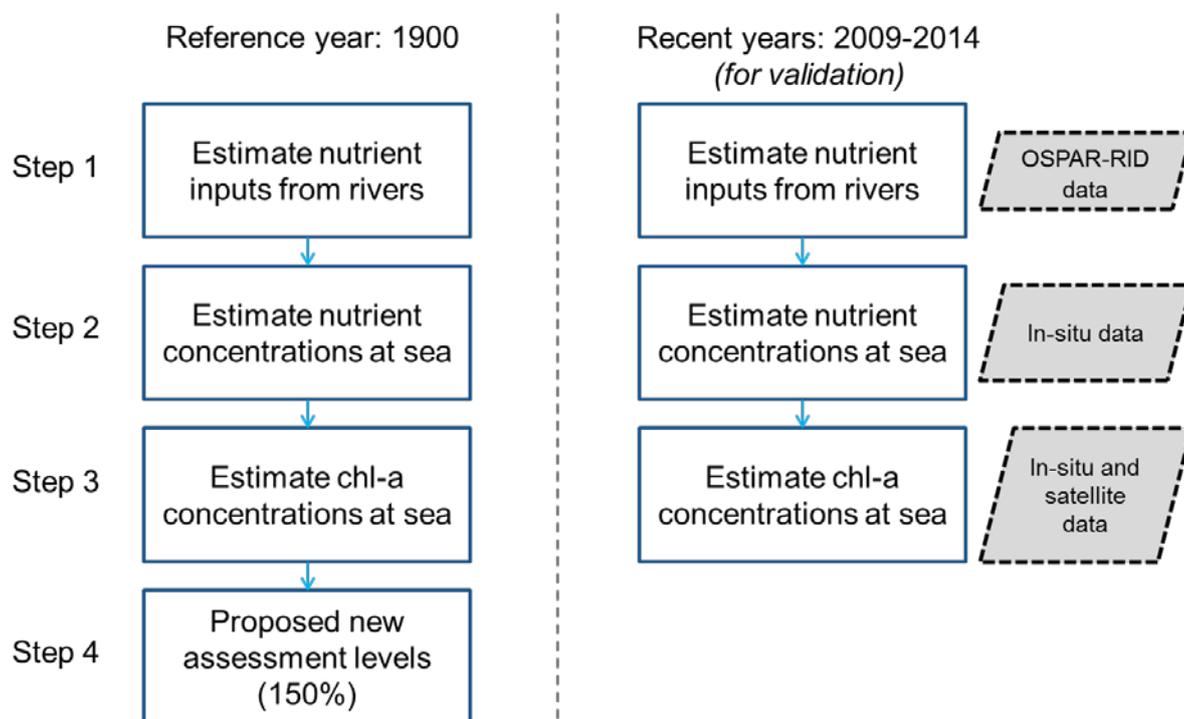
Figure 1.1: Schematized representation of the workflow used to estimate chlorophyll concentrations under reference conditions and validate with present conditions. Grey boxes indicate validation data used.

Since observed data for model validation were lacking for the period around 1900 we ran the same series of models for recent years. These were validated with recent observations on nutrient inputs from rivers (from OSPAR-RID database), marine in-situ data and satellite data. We collected observed data from all countries around the North Sea on nutrients, chlorophyll, salinity and light climate for validation of the models in this study (chapter 3). Satellite data on chlorophyll-a and suspended particulate matter were provided by RBINS as part of activity 2 of the project. PML provided satellite data on primary production (chapter 8). Finally we combined the ecosystem information from the models and the satellite data to propose new assessment areas with similar physical and ecological dynamics (chapter 9). This report concludes with conclusions and recommendations for next steps (chapter 10).

Most of the activities described in this report were performed by Deltares, with input from other project partners. The work on nutrient load estimation (described in chapter 4) was performed by SMHI and the work on satellite data on primary production (described in chapter 8) was performed by PML.

2. Inventory of approaches for definition of reference conditions

2.1. Method

The JMP-EUNOSAT project aims to determine coherent threshold values for the whole North Sea in similar ways as countries around the North Sea have used before. Within OSPAR, member states have agreed to follow a similar approach for determining threshold values. The threshold values are commonly defined as a justified area-specific percentage deviation from background conditions not exceeding 50%, i.e. ≤50% above chlorophyll and nutrient concentrations under "natural reference conditions". Therefore, an inventory was made at the start of the project to find out:

- how countries have defined "natural reference conditions"
- what methods and data were used to determine " natural reference conditions" for their coastal waters.

2.2. Definition of "natural reference conditions"

Natural reference conditions have been defined in different ways by countries. Table 2.1 gives an overview of the definitions in use by countries around the North Sea.

Table 2.1: Overview of definitions of "natural reference conditions" by countries around the North Sea

Country	Choice of reference
Norway	Expert judgement and data from 1950's
Sweden	1930
Denmark	1900 (in WFD)
Germany	1880
Netherlands	1930
Belgium	50% of loads of 1950
France	Geographical reference to relatively unimpacted sites
England and Scotland	Geographical reference: off-shore waters

Based on the above the year 1900 has been chosen as the reference year to estimate nutrient loads from rivers with E-HYPE, for the JMP-EUNOSAT project. This year is well within the range of reference years that have been used so far and before widespread use of inorganic fertilizers in agriculture. Also, in 1900 loads were not affected by the 1st and 2nd world war or the severe economic crisis from the 1930's.

2.3. Estimation methods in use

Table 2.2 gives an overview of methods that have been used by countries around the North Sea to estimate concentrations of nitrogen, phosphorus and chlorophyll in their North Sea waters, under natural reference conditions. The methods in the table are described in more detail below.

Table 2.2: Overview of methods used by countries around the North Sea to estimate natural reference concentrations of nutrients and chlorophyll-a.

Method \ country	NO	SE	DK	GE	NL	BE	FR	UK
Nutrient loads model				X		X		
Process-based oceanographic model				X	X	X		
Regression model		X		X		X		
Geographical reference site							X	X
WFD approach			X	X			X	X
Expert judgment	X							
Literature review	X		X					

2.3.1. *Nutrient loads model*

Eutrophication in coastal waters is mainly caused by nutrient loads from rivers. To estimate historic chlorophyll concentrations it is therefore useful to estimate historic nutrient loads from rivers. The nutrient loads have been estimated with models that take into account the effect of variables such as number of inhabitants, number of cattle, land use, soil type and land cover. The models have been validated with available historic data on nutrient loads. In Germany the MONERIS model has been validated with available data for 1870. In Belgium the RIVERSTRAHLER has been validated with observed data from 1950 – 1998. Then it was estimated that pristine conditions would have approximately 50% of the nutrient loads in the 1950's.

2.3.2. *Process-based oceanographic model*

Process based oceanographic models can simulate the transport of nutrient loads from rivers in the North Sea and estimate the associated chlorophyll concentrations, based on nutrient concentrations and ambient light and mixing conditions. Such models have been applied by Germany, the Netherlands and Belgium for the estimation of threshold values. In Germany and Belgium the nutrient loads have been estimated with nutrient load models.

2.3.3. *Regression model*

One commonly used type of model to estimate chlorophyll concentrations from ambient environmental conditions are linear regression models. In Sweden historical data for Secchi depth are available for the 1930's. These data have been used to estimate chlorophyll concentrations in the 1930's using a linear regression model, fitted with recent observed in-situ data.

In Belgium and Germany linear regression models between chlorophyll and nitrogen (winter DIN and totalN respectively) have been fitted and used to estimate chlorophyll concentrations under natural reference conditions.

2.3.4. *Geographical reference site*

In France and UK natural reference concentrations of chlorophyll have been estimated from present observed chlorophyll concentrations at monitoring sites that have a limited impact of eutrophication. In France bays in relatively pristine watersheds have been used as reference sites and in the UK, offshore site have been used as reference sites.

2.3.5. *WFD approach*

All countries have developed threshold values for the Water Framework Directive (WFD) for the first nautical mile in coastal waters. These threshold values have gone through an extensive inter-calibration process. The WFD approach uses 90-percentiles as chlorophyll metric rather than growing season means. Reference concentrations can be estimated through regression models or through geographical reference sites. The WFD uses 5 classes for ecological status: Bad, Poor, Moderate, Good and High. The MSFD threshold in nearshore coastal waters is comparable to the WFD boundary between Moderate and Good Ecological Status (GES).

France and the UK only have threshold values defined for WFD- water bodies. Germany has applied the WFD approach to define threshold values for WFD water bodies and MSFD assessment areas.

2.4.   Selection of estimation methods for the JMP-EUNOSAT project

The overview shows that there are 3 types of methods in use to estimate natural reference conditions:
- Regression models
- Process-based models
- Geographical reference sites

Both model types can fit a range of model applications to the same data. For regression models the choice of explanatory variables and data pre-processing can yield different results. The dependency on explanatory variables will differ between sites. In most areas nitrogen limitation is expected to play a dominant role in controlling chlorophyll concentrations. In other areas phosphate or light and mixing conditions may play a dominant role.

Process-based models can also yield very different results, depending on the approximation of input data, process-formulations and parameter settings and the observation data used for calibration. Most countries around the North Sea have their own process-based model. It is not feasible within the JMP-EUNOSAT project to run all these models with the coherent input data for nutrients and light conditions estimated within the project. We compared two process-based box model approaches with the regression model, using the same model inputs. We selected the BLOOM/ GEM model from the Netherlands since this is a model that we have available at Deltares and it has given good results in earlier ecological modelling studies (Blauw et al., 2009; Troost et al., 2013;2014; Los et al., 2014). We used the box model approach for individual monitoring locations (Los and Wijsman, 2007). Additionally we used an ERSEM model that has been validated for a monitoring station in English coastal waters as 1DV column model (Butenschön et al., 2016). ERSEM has been developed in international collaboration during a European research project and has often been used for ecological modelling in the North Sea (for example: Radach and Lenhart, 1995; Edwards, et al., 2012). There are different versions of ERSEM in use at different organisations. We used the version that can be readily downloaded as 1DV column model from a PML website (https://www.pml.ac.uk/Modelling_at_PML/Models/ERSEM).

It is complicated to define a coherent map of natural reference conditions based on geographical reference sites. Natural reference conditions are different everywhere depending on the level of freshwater influence and light and mixing conditions. Therefore we did not apply this method in our study.

3. Validation data

3.1. Data collection and pre-processing

Observed data have been received and processed from all countries around the North Sea. The observed data differed between countries in sampling depth, sampling frequency, the variables that were measured and the units that were used. We pre-processed the data to create a more coherent dataset:

We first:

- merged data from nearby latitude-longitude combinations, when each latitude-longitude combination had limited data, to create longer time series. This was especially relevant for Scottish and Norwegian off-shore data, since their sampling design does not imply fixed stations that are frequently visited. For the Norwegian data we used a grid with 0.02 degree mesh size to combine all observations within each grid cell into one 'monitoring location';
- converted all data to the same units;
- averaged all data of the upper 10 meters per day. We neglected data of deeper water layers since our analysis will focus on concentrations in the upper mixed layer that can be observed with satellite;
- We selected only the stations with sustained time series for several years with more than two observations during the periods 2003-2008 or 2009 - 2014. Three observations per season in six years is actually still a very low amount that is likely to result in a biased estimation of the season mean, particularly for chlorophyll-a that shows a relatively large variability within the growing season. We looked for a balance including as many as possible monitoring locations in the analysis, while accepting that this would result in a large uncertainty and possibly noisy spatial patterns in some areas.

With the resulting dataset of daily averages for the surface layer we calculated seasonal statistics per year:

- Winter means: December, January and February
- Number of DIN observations during winter months
- Growing season means: March through September
- Number of chlorophyll observations during the growing season

With the resulting dataset of season means per year we calculated long term season means. We used 6-year means because the reporting period for MSFD is every 6 years. Therefore, we split recent years into two periods of 6 years:

- 2003 – 2008
- 2009 – 2014

Although many countries use 90-percentiles as threshold values we use seasonal means. Since many of the monitoring locations are sampled only a few times per year the estimations of seasonal means are already quite uncertain. Estimation of 90-percentiles from such time series is even more uncertain, with a large impact of the random timing of sampling in or outside the spring bloom period (van der Zande et al., 2011). We used the number of observations per season as a filter to explore the effect of temporal under-sampling at some monitoring locations and to reduce noisy spatial patterns in in-situ data in our validation plots.

3.2. Resulting dataset

The resulting dataset of season means for 2009 - 2014 contains 198 monitoring stations with season means for at least chlorophyll-a and DIN. The number of monitoring locations and available variables per country are summarized in Table 3.1. Figure 3.1 shows the distribution of monitoring locations in the dataset on a map of the North Sea.

Table 3.1: Number of monitoring locations per variable per country in pre-processed data of season means for 2009 – 2014. Only locations with more than 2 chlorophyll-a observations during the growing season or 2 DIN observations during winter over 2009 – 2014 are included.

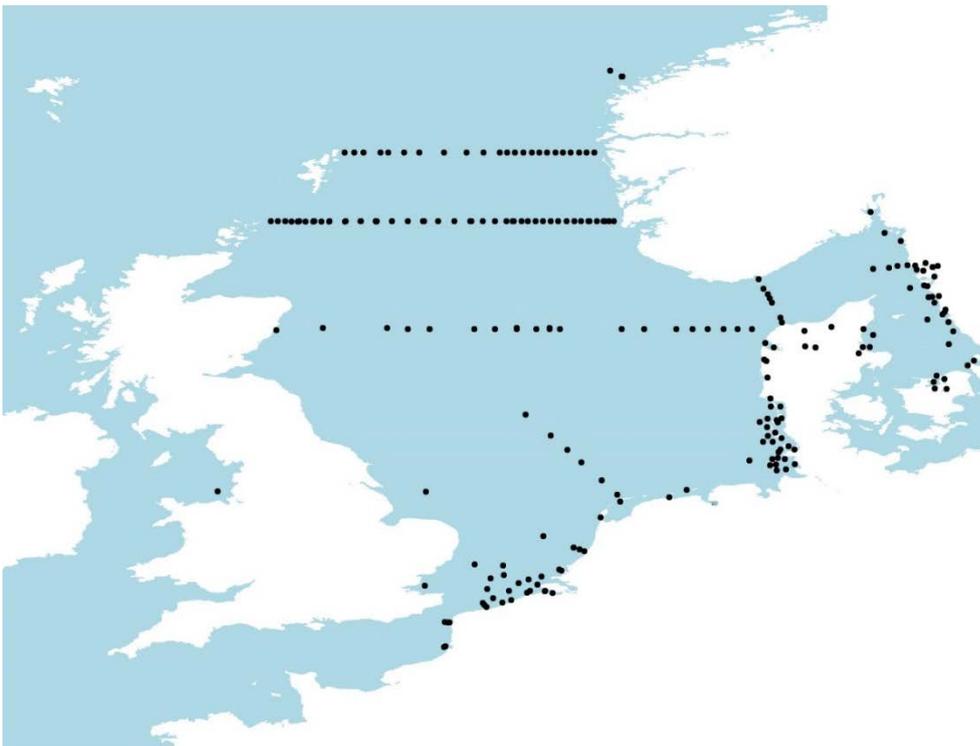|	chlfa	DIN	DIP	Si	sal	Kd	Secchi	spm
BE	14	12	12	12	12	-	23	12
DK	25	24	24	24	23	21	-	-
EN	4	4		4	4	4	-	-
FR	8	8	8	8	8	-	-	8
GE	18	22	22	22	22	-	20	
NL	24	20	20	20	20	13	4	20
NO	117	67	67	67	5	-	-	-
SC	13	13	13	13	13	-	1	-
SE	28	28	28	28	28	-	25	-
Total	251	198	194	198	135	38	73	40

Figure 3.1: Spatial distribution of monitoring locations in the validation dataset of season means of 2009 – 2014, with more than 2 observations of chlorophyll during the growing season or DIN in winter.

4. Nutrient loads from rivers

4.1.   Introduction

In order to establish reference nutrient and chlorophyll concentrations in the North Sea, reference nutrient concentrations and loads are needed for rivers and streams discharging to the study area (Figure 4.1). For the purpose of this project, reference conditions are assumed to reflect current climate, current status of water management (i.e., reservoirs and canal systems as in about 2010s), and population, industrial, and agriculture activities as in 1900s. We also calculated estimations of nutrient loads in recent years with the same approach as for 1900 to allow for model validation with recent observed data.

A pan-European model E-HYPE (Hundecha et al. 2016) developed and maintained by SMHI was used to evaluate the current riverine flows and nutrient loads from land to the study area at a fine spatial and temporal resolution (daily values for catchments with median size 215 km^2).
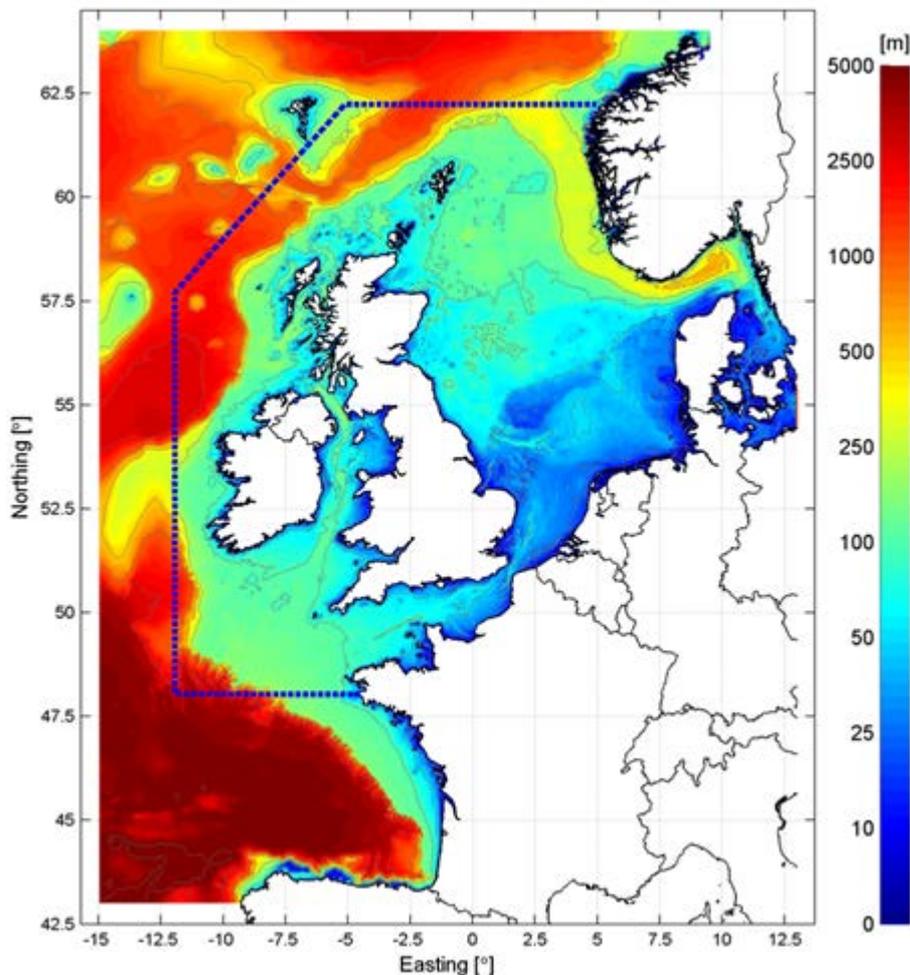

Figure 4.1. Study area

4.2. Methodology

4.2.1. Model Description

4.2.2. HYPE model

HYPE (Hydrological Predictions for the Environment) is an integrated rainfall-runoff and nutrient transport model developed by SMHI under a Creative Commons open source licence (Lindström et al., 2010). HYPE works on a sub-basin scale, as opposed to grid cells. Each sub-basin acts as an independent catchment reactor and sub-basins are hydrologically connected though a routing scheme. Model results are aggregated to and evaluated at the sub-basin scale.

Computations of fluxes and states are done in a time-stepping scheme using a hydrological response unit (HRU) approach within these sub-basins. HRUs are derived from landscape properties, e.g. mapped soil and land cover types. Conceptually, HRUs consist of up to three stacked subsurface storages (linear reservoirs) with a single groundwater table and a number of flow path conceptualizations for recharge and release (e.g. snow cover, evapotranspiration, saturation and infiltration excess, matrix flow, macropore flow). HYPE also includes a lake model, which is embedded in the river routing scheme. As an adaptation to nutrient modelling, HYPE explicitly accounts for soil porosity and field capacity/wilting point storage volumes. Nutrient pools are then tied to these volumes.

HRUs in HYPE are parameterized for the whole model domain, which allows for parameterization of ungauged areas. Data required for setting up the model include spatial data for delineation and HRU computation, forcing data (precipitation, air temperature), calibration data (stream discharge, nutrient concentrations), and land management data sources (e.g. point source releases, crop fractions, land management practices).

4.2.2.1. E-HYPE v 3.1.3

We used E-HYPE version 3.1.3 (released in August 2016) for the model simulations. This particular version uses Global Forcing Data v.1 (GFD, Berg et al 2018) instead of the WFDEI (WATCH Forcing Data methodology applied to ERA-Interim data) used in previous E-HYPE versions. This update resulted in better match with observed flows and concentrations. Additional updates from v. 3.1.2 included

- point sources updated to reflect 2012 discharges from the Urban Waste Water Treatment Directive database (UWWTD, 2014),
- lake and reservoirs descriptions corrected for several instances, and
- updated observed nutrient data were used for evaluation.

E-HYPE was calibrated using observed data from the full European domain in several phases since its inception (Donnelly et al., 2015; Hundecha et al. 2016, Bartosova et al., 2017). No additional calibration was done within this study. However, the performance of the model within the study area was verified using observed data.

4.2.2.2. Initial conditions for 1900

Daily nutrient loads were calculated for 25 years of current climate data (1989-2013). No change was made to initial conditions. Instead, the model was executed with 1979-2013 climate data and the model state was saved. Then, the model was restarted with the saved state again with 1979-2013 data and this model was run for 1979 - 1988. This

provided the model with 45 years of so-called warm-up period during which the processes and storages in the model could settle.

4.2.3. Nutrient Loads in 1900s

The purpose of this study was to determine background loads that can be used to derive ecological reference values and provide a benchmark to evaluate the level of impairment as well as the potential for restoration. This means that not all anthropogenic or other changes that occurred since 1900 were restored to the model. Namely, water management and climate were retained. The existing level of channelization, building and operation of reservoirs and dams, and the system of channels and dikes was kept in the model.

E-HYPE v.3.1.3 was used to simulate current loads from the study area. The model input files were then modified to approximate conditions in 1900s and the model was executed without any further adjustment in the model parameters to generate the loads corresponding to 1900. Model input files with land use, agriculture practices such as irrigation and fertilization, population-driven sources of pollution, and atmospheric deposition were modified based on available information.

We changed the inputs to the model step by step to see the effects of individual changes to the total change in nutrient loads between present days and 1900. In addition, we simulated a scenario with even more nutrient reduction than the year 1900: a synthetic "no human impact" scenario was developed following assumptions used in the Swedish assessment of background concentrations. Nutrient concentrations in discharges from point and rural sources were set to 0, atmospheric deposition was eliminated, and all agriculture land was converted to unmanaged grasslands (i.e., no fertilization).

4.2.3.1. Land Use

Land use data were acquired from History Database of the Global Environment (HYDE) developed under the authority of the Netherlands Environmental Assessment Agency (Klein Goldewijk et al., 2010, Klein Goldewijk et al., 2011). HYDE presents time series of land use and population developed on a 5-minute grid (about 85 km^2 grid cell around the equator).

HYDE contains geospatial layers that show a proportion of selected land uses for each grid. The land use includes grazing, irrigated crops excluding rice, irrigated rice fields, pastures, rangeland, rain-fed crops excluding rice, rain-fed rice fields, and partial summaries such as total irrigated cropland, total rain-fed cropland, total rice fields, and total cropland. The population-related geospatial layers contain population density, population counts (urban population, rural population, and total population), and urban area.

There are 15 different land uses included in E-HYPE v.3.1.3. These were matched with the land uses available from HYDE (Table 1). Four E-HYPE land uses were considered constant and were not modified for the 1900 simulation: lakes, glaciers, wetlands, and rivers. While wetlands have certainly changed since 1900s, HYDE does not contain any information on their extent nor their location. A series of scripts were developed to modify E-HYPE input data containing information on the land use proportion. E-HYPE's basic units (HRUs) are defined for unique land use, soil, and crop combination. That means that several HRUs can be assigned the same land use and have different soils and/or crops. However, HYDE includes only land use information, not the type of soil that supports the land use.

When modifying E-HYPE's HRUs from the current land use to the land use of 1900s, the increase or decrease in area for each particular land use was allocated to all HRUs with the corresponding land use to keep the proportion of soils

and/or crops the same. Any remaining change that was needed after modifying the land uses explicitly included in HYDE was allocated to the most prevalent land use among those in E-HYPE that are marked as Other land use in HYDE (Table 4.1). If none of these were present, all such remaining change was allocated to Mixed forest.

Table 4.1. Linking land uses from E-HYPEv.3.1.3. and HYDE

E-HYPE		HYDE
Land use code	Description	Land use
1	Lakes	Not modified
2	Glacier, permanent snow	Not modified
3	Sealed urban area	Urban area
4	Broad leaved forest	Other
5	Needle leaved forest	Other
6	Mixed forest	Other
7	Agricultural land - rainfed	Rain-fed crops
8	Agricultural land - irrigated	Irrigated crops
9	Agricultural land - Permanent crops	Rain-fed crops
10	Pastures	Pastures + rangeland
11	Irrigated pastures	Pastures + rangeland
12	Open with veg	Other
13	Open without veg	Other
14	Wetland	Not modified
15	River	Not modified

4.2.3.2. *Human waste*

HYDE data on urban and rural population in 1900 was used to estimate the number of people living in urban and rural settings in each catchment. At that time, waste water treatment technologies were in the early stages. Imhoff tank with sedimentation chamber and digestion chamber, was patented in 1906 (Lofrano and Brown, 2010). Activated sludge process used in the secondary treatment today was not yet discovered. Ardern and Lockett presented their research findings on activated sludge process in 1914 (Ardern and Lockett, 1914). However, four separate patent applications dealing with "Improvements in Apparatus for the Purification of Sewage or other Impure Waters" were filed in 1913 and 1914 in UK (Jones &Attwood, 1913ab; 1914; 1915), although they did not use this specific term, the basics of the process were clearly included (Alleman 2017).

Simple anaerobic septic tanks such as designed by Mouras in 1860s and patented by Cameron in 1895 or in some cities trickling filters might have been used (Lofrano and Brown, 2010). First trickling filters were installed at Salford near Manchester, UK in 1893 (Alleman, 2017). Septic tanks or cesspools were typically not well sealed and leaked significantly to soils. Some cities used "sewer farms" and put their sludge harvested from sewers on fields; this was not considered in the model as there are no records that would allow us to consistently determine which cities in the study area used this method, what proportion of the sewage was collected and disposed on fields, and what was the area and location of these fields.

Based on urban population density, the urban population was divided into two groups (Table 4.2): population connected to sewers that discharge to streams (i.e., contributing as point sources), and population with waste disposal to cesspools and septic tanks or no infrastructure (i.e., contributing as rural sources to soils). All rural population was assumed to contribute to soils. Population assumed to live in cities without a functioning sewer system was added to rural population for each respective catchment. These assumptions were applied across the whole study area regardless of any differences on regional or country level that might have existed for the lack of data on spatial distribution of this information.

Table 4.2: Percentage of urban population connected to a sewer and a percentage of the sewered load further available in the catchment.

Population density	Point sources		Rural sources	
	Percent population	**Percent load delivered**	**Percent population**	**Percent load delivered**
< 100 #/km^2	0%	Not applicable	100%	90%
100-500	50%	100%	50%	90%
>500	100%	90%	0%	Not applicable

The nutrient loads from the population were calculated using Population Equivalent (PE). The PE was estimated from diet and other factors relevant in 1900s (Schmid 2000, Smil 2000). Each person was assumed to produce 1 g P per day (0.37 kg/cap-year) and 5.5 g N per day (2 kg/cap-year). To preserve the water balance, the volume discharged by both point sources and rural sources was assumed to stay the same as in E-HYPE v.3.1.3. and the concentration was modified in a way that resulted in the desired total load.

In a few cases where the 1900s population analyses placed a point source in a catchment without a current point source, the load was allocated to the next available catchment downstream with a current point source. This reflects the fact that many sewers and treatment facilities are located downstream of urban centres. If there was no current rural population, a discharge of 100 m^3/day was added to that catchment with an appropriate concentration to deliver the expected load. For catchments where current conditions result in a rural source or a point source but there was no such source in 1900s, the concentrations were simply set to zero. No change was made to industrial sources as no reliable information was obtained.

4.2.3.3. _Fertilization and other agriculture practices_

The amount of land under agriculture was adjusted as a part of the land use adjustment process described above. The HYDE database information on irrigated and non-irrigated agriculture land was utilized during the land use adjustment and no further modification was necessary. No further adjustment was made for crop types other than secondary crops were removed where they existed.

The amount of fertilizers applied to crops was adjusted as follows. The maximum application rate was set to 100 kg N per hectare and 50 kg P per hectare (Smil 2000, van Grinsven et al 2015). The application rates were assumed to remain in the same relative proportion for different crops as in E-HYPE v.3.1.3., i.e., the crops with the highest current fertilizer application rate still used the highest application rate in 1900s.

The Haber-Bosch process that industrialized production of inorganic nitrogen fertilizers was first demonstrated in 1909 with a first industrial-level production starting in 1913 (Morris and Morris, 2001). All nitrogen applied to fields was thus assumed to be in organic form (manure) in 1900. The phosphorus application rate was assumed to be applied to crops as manure for 80% of the application rate and as inorganic fertilizers for 20% of the application rate. This gives application rates in accordance with Knudsen and Schnug (2016), Kyllingsbæk (2005) and Vinther (2012). Any phosphorus applied on other land uses (e.g. pastures) was assumed to be in the organic form only.

4.2.3.4. _Atmospheric Deposition_

Atmospheric deposition rates were based on Engardt et al (2017) who concluded that the current deposition rates of oxidized nitrogen are now 3-4 times higher than in 1900 and the annual deposition of reduced nitrogen is more than two times as high as in 1900. We assumed the spatial distribution of nitrogen deposition to be the same as in E-HYPE v.3.1.3 with deposition rates set to 33% of the current deposition rates. There is no atmospheric deposition of phosphorus in E-HYPE v.3.1.3.

4.3. Results and Discussion

4.3.1. _Model Performance Evaluation_

E-HYPE v3.1.3 is calibrated to river discharge and riverine nutrient concentrations using a large number of available observation sites distributed over the model domain. The model is calibrated for an optimised mean performance over the whole model domain using a representative gauged basin approach, resulting in considerable performance spread at individual sites.

For the purpose of this study we evaluated the performance of the E-HYPE model with respect to discharge and nutrient loads at gauging stations near river outlets to OSPAR sea basins. In addition, we compared the simulated discharges and loads with corresponding annual estimates from the OSPAR RID database at national aggregation level. The model calibration was conducted using soluble phosphorus (SP), total phosphorus (TP), inorganic nitrogen (IN), and total nitrogen (TN). However, due a significantly smaller number of calibrations sites with TN and SP available, only results for IN and TP are presented in this report.

To evaluate how the E-HYPE calibration can affect the oceanographic models, we limited the sampling sites used for the model performance evaluation to those within the vicinity of the sea. We defined the sampling site was "near-sea" when its drainage area represented at least 80% of the drainage area of its outlet to the sea. The average performance of the model calibration with respect to river discharge in the study area is good (Figure 4.2) with median NSE 0.58 (range from -2 to 0.86) and median relative bias -4 % (range -38 % to 67 %). The range of the relative bias for nutrient loads estimates is considerably higher: ranging between -91% and 408% for inorganic

nitrogen (IN, Figure 4.3) and -89% to 1099% for total phosphorus (TP, Figure 4.4). The median bias stays close to zero at -7 % for inorganic nitrogen and 3 % for total phosphorus concentrations.
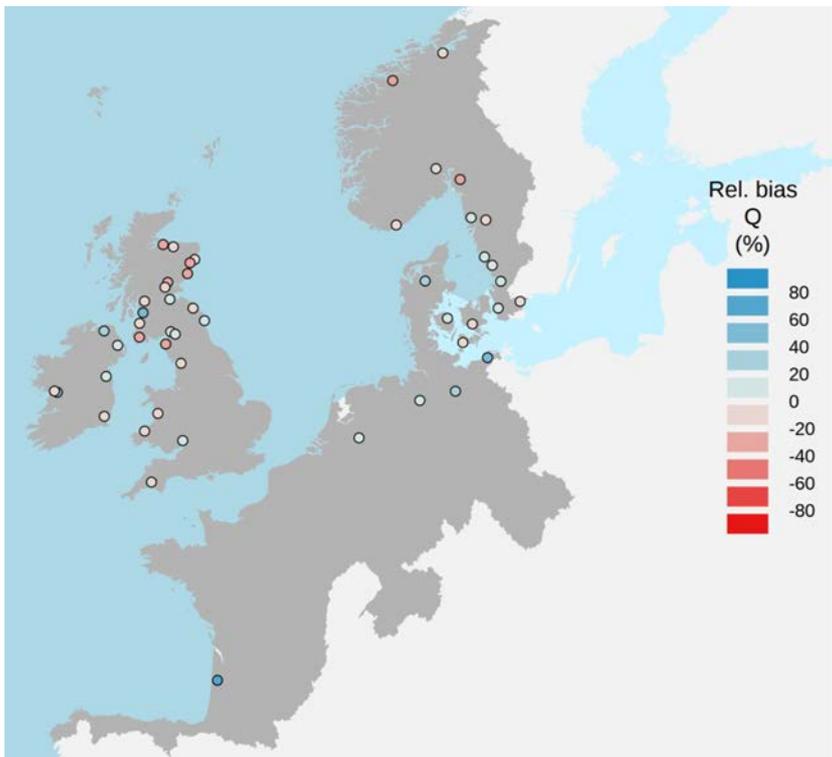

Figure 4.2. E-HYPE v3.1.3 model performance as relative bias (RE) for river discharge to OSPAR seas at near-sea gauging stations, 2000 to 2010 calibration period.
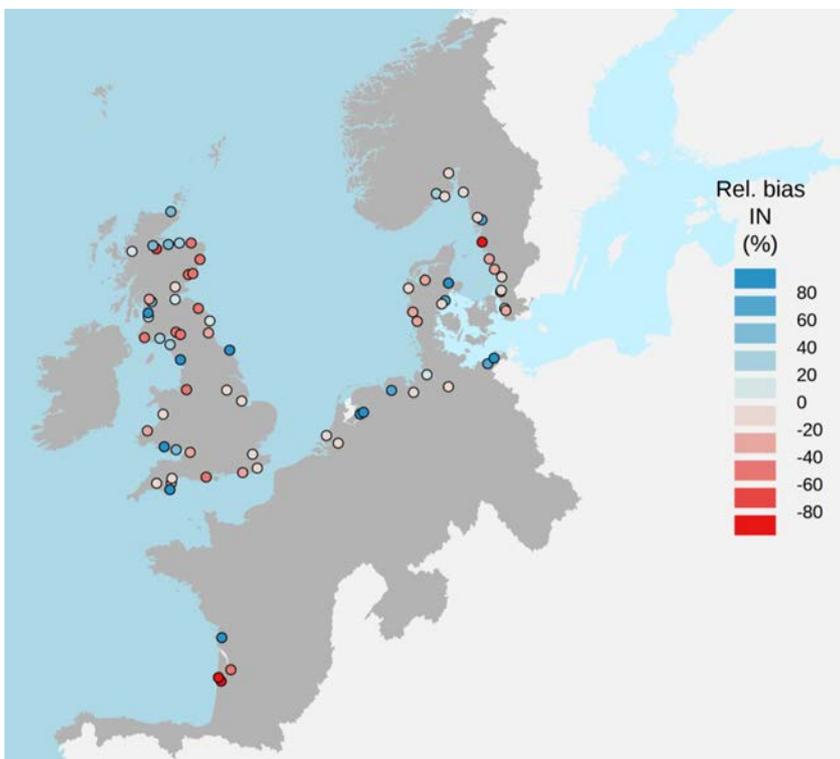

Figure 4.3. E-HYPE v3.1.3 model performance as relative bias (RE) for inorganic nitrogen concentration in river discharge to OSPAR seas at near-sea observation sites, 2000 to 2010 calibration period.
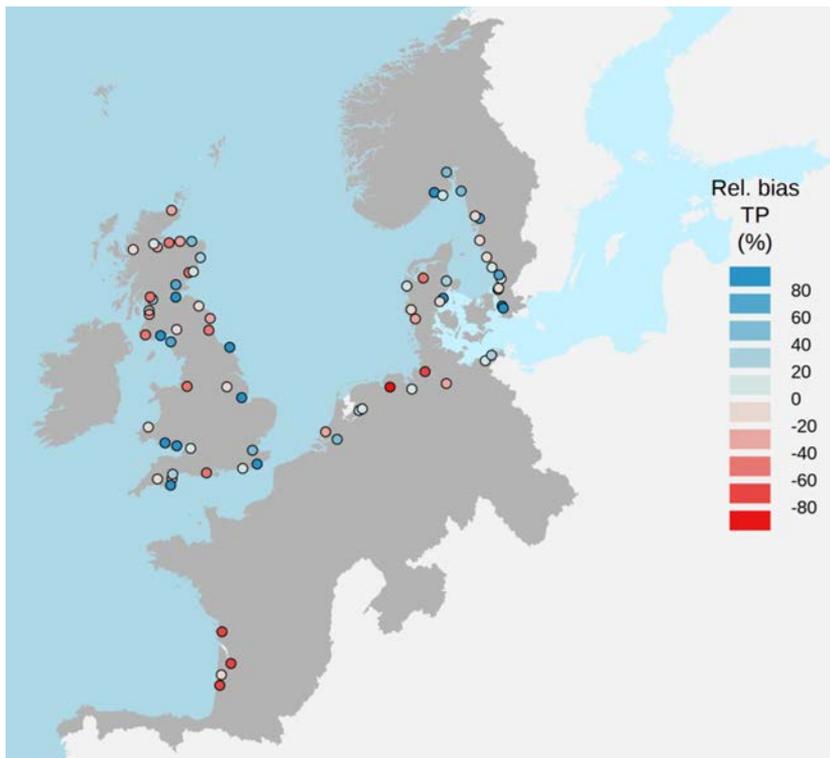
Figure 4.4. E-HYPE v3.1.3 model performance as relative bias (RE) for total phosphorus concentration in river discharge to OSPAR seas at near-sea observation sites, 2000 to 2010 calibration period.

Figure 4.5 illustrates the intra- and inter-annual variability of TN and TP concentrations for Elbe River at the most downstream station used in E-HYPE calibration. On average, observed TN concentrations increase gradually in late winter through early spring and start declining in March. Observed TP concentrations at this site are more stable with only a slight increase in August-September. The model simulation shows a larger variability in the concentrations during spring, especially for TP. However, it must be noted that observed concentrations are available for a small number of days every year with only 18-23 samples collected each month over the 11-year time period summarized in the figure, so the data may not be fully representative of the actual variability.
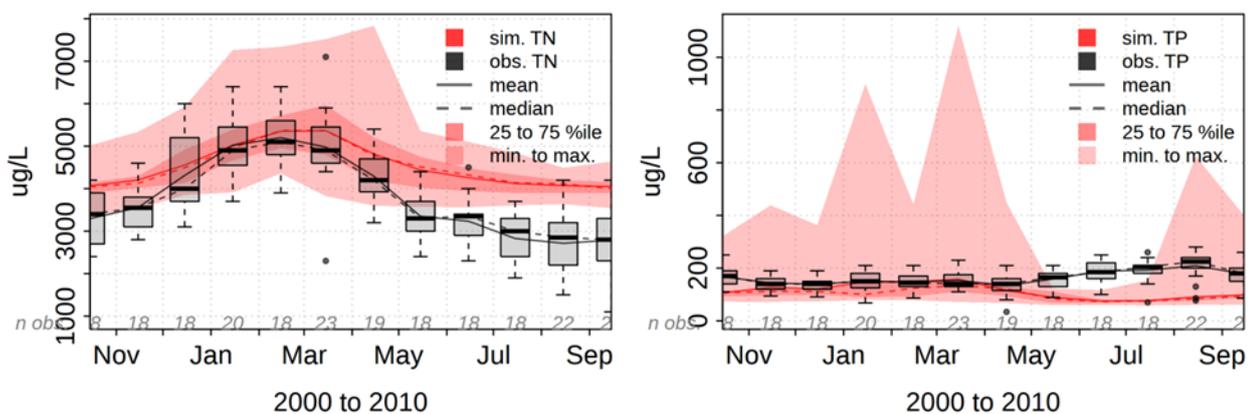

Figure 4.5. E-HYPE-modelled and observed riverine nutrient concentration regimes at near-sea observation site in river Elbe (DE), total nitrogen (TN, left panel) and total phosphorus (TP, right panel). Observed concentration ranges as monthly boxplots, daily model estimate ranges as red ribbons.

Total nutrient loads have also been compared with total nutrient loads per country as estimated by OSPAR countries in the OSPAR-RID dataset (Figure 4.6). The E-HYPE estimates are considerably higher than the OSPAR estimates in France, Germany, the Netherlands and the UK.
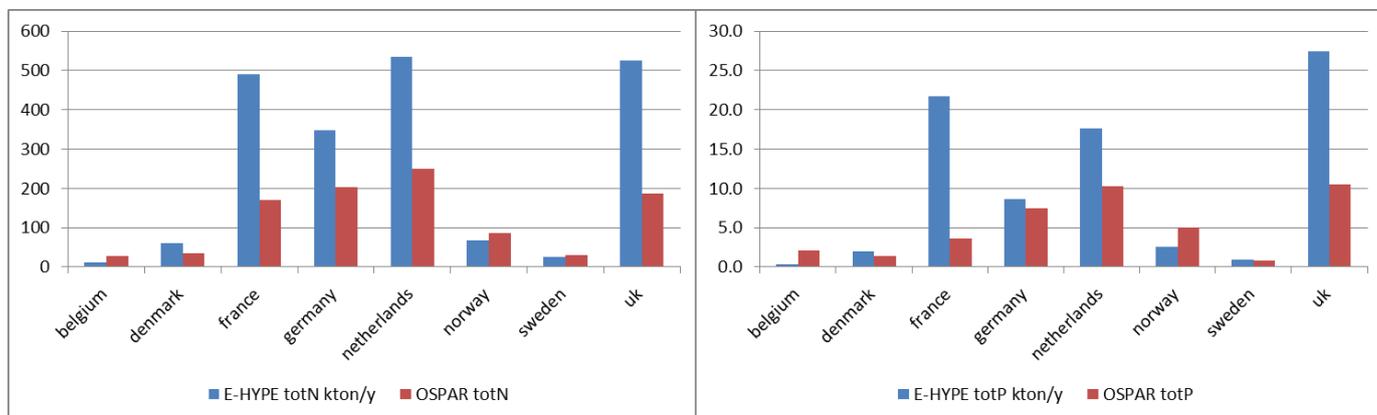

Figure 4.6: total nutrient loads per country in recent years as estimated by E-HYPE (blue) and OSPAR (red) for total N (left) and total P (right).

4.3.2. *Nutrient loads in 1900, compared to the present situation*

The changes in the model were implemented step by step with results from each step processed and analyzed in order to understand the impact of individual changes. The changes were implemented in this order:

a) Point source discharges
b) Rural source discharges
c) Land use changes
d) Fertilization changes
e) Atmospheric deposition changes

TN and TP loads to the study area for these 5 cumulative steps and the "no human impact" scenario were expressed in relation to the current loads (Figure 4.7). The apparent increase in the nutrient loads for rural sources is related to a larger population living in the rural areas and a limited number of treatment options for sanitary waste in the rural area in 1900s. Atmospheric deposition, fertilization practices, and point source reduction appear to have significant impact on the total N loads in 1900s when compared to the current loads. Total P loads were most affected by reduction of point sources.

The TN and TP loads estimated for 1900s is about 40% and 64% of the current loads, respectively (Table 4.3). The proportion varies by location; on country basis it varies between 23 % and 50% for TN load and between 25% and 85% for TP load.
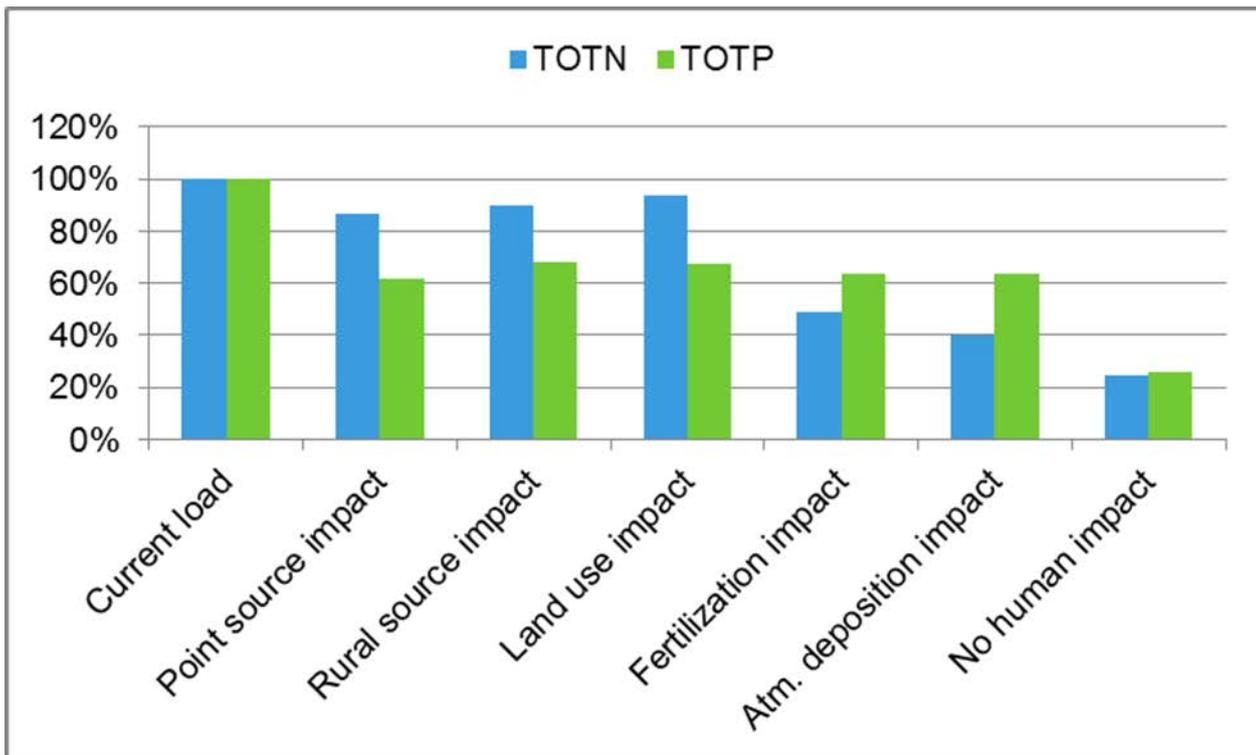
Figure 4.7. Proportion of the current loads associated with each step in changing the model from current state (2010s) to 1900s and "no human impact" scenario. Changes are cumulative, e.g. "rural source impact" simulation includes both point and rural sources adjusted to reflect 1900s.

Table 4.3. Nutrient loads in 1900s expressed as proportion of the current loads on a country-basis

Country	TN load (%)	TP load (%)
Belgium	23	56
Denmark	28	43
France	50	72
Germany	41	85
Ireland	35	50
Netherlands	36	59
Norway	38	49
Spain	27	25
Sweden	48	58
U.K.	42	49
Average	40	64

5. Nutrient transport modelling

5.1.   D-FLOW-FM transport model for the North Sea

For the nutrient transport modelling the new 3D unstructured grid model for the whole North Sea by Deltares has been used. The model grid is shown in Figure 5.1. It has a resolution of 1 nautical mile (1.8 km) in shelf waters of less than 100 meter water depth. A coarser grid is used in waters between 100 and 400 meters water depth, with grid cells of 2 nautical miles. For oceanic waters of more than 400 meter water depth a resolution of 4 nautical mile per grid cell has been used. Vertically the model has 20 equidistant layers.
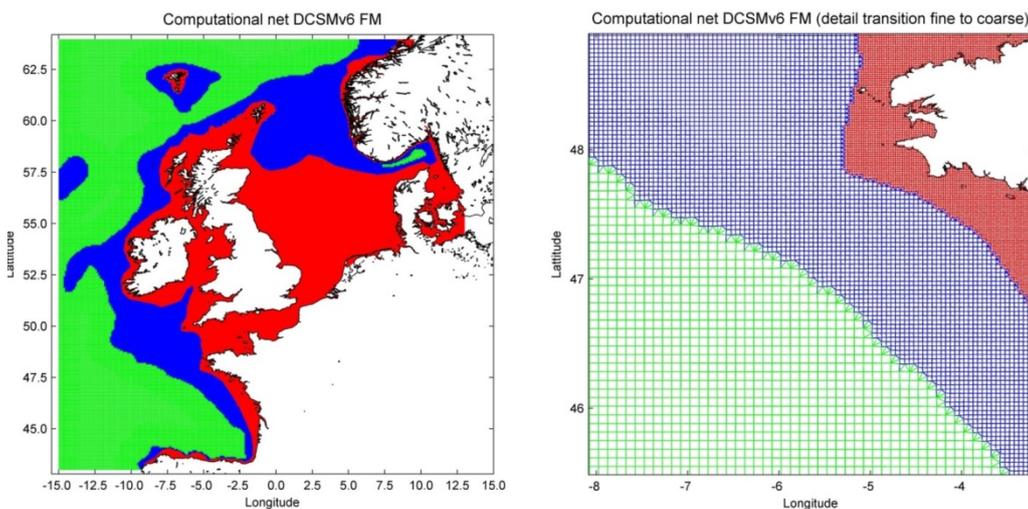

Figure 5.1: Model domain and grid size of the transport model used for simulating nutrient concentrations. left) model domain and water depth: green areas are more than 400 meter deep, blue areas between 100 and 400 meter deep and red areas less than 100 meters deep. right) detail of the model grid near Brittany, showing the grid size corresponding to the 3 depth classes.

Boundary conditions for temperature, salinity, DIN and DIP have been derived from the surface layer of the World Ocean Atlas as monthly averages per 0.25 degree latitude and longitude. Winter mean concentrations of DIN and DIP along the boundaries are approximately: 6 µM DIN and 0.5 µM DIP.

As meteorologic forcing the ECMWF ERA5 -interim dataset by the European Centre for Meteorologic Weather Forecasting (ECMWF) has been used for the actual years (2005 – 2013) that were run to simulate present conditions. The same meteorologic forcing (2005 – 2013) has been used to simulate the natural reference conditions around the year 1900.

5.2.   River inputs

River discharges and nutrient loads are derived as total bioavailable nitrogen and phosphorus from E-HYPE simulations. To mitigate the large overestimation of nutrient loads by E-HYPE for some countries, we decided to scale the nutrient loads so that they match exactly with the available data on nutrient loads per country per year. The E-HYPE model results brought further detail to these total numbers, by specifying the distribution of loads within the country and throughout the year. And most importantly E-HYPE provided information about differences between the reference year 1900 and present conditions. For the simulation of present conditions, E-HYPE results for 2005 – 2013 have been used. For the simulation of natural reference conditions E-HYPE results for the same

period (in terms of meteo and river discharges) have been used in combination with nutrient loads of around 1900 (see chapter 4).

Validation of E-HYPE estimations of nutrient loads to the North Sea showed that these deviated substantially from the nutrient loads as estimated at observation locations (see Figures 4.3 and 4.4) and estimated by OSPAR member states in the OSPAR RID dataset (see for example Figure 4.6). Apart from the official nutrient load estimates in the OSPAR-RID dataset there is an alternative dataset of nutrient loads available from the OSPAR ecological modelling working group (OSPAR ICG-EMO). The OSPAR-RID dataset provides only data for total loads per year, whereas the ICG-EMO dataset provides information per river and per day. We have compared these two datasets of nutrient loads to the North Sea to get an impression of the differences and possible reasons for differences. Table 5.1 shows the ratios between the discharges and loads in the OSPAR ICG-EMO dataset and the OSPAR RID dataset. The OSPAR ICG-EMO dataset does not provide data for Swedish discharges. For Danish rivers this dataset lacks data on DIN and DIP and for UK rivers it lacks data on total nitrogen. For the other countries and variables the ratio between discharges seems to be a rather good proxy for the ratio between nutrient loads in the two datasets. This suggests that the difference in total nutrient loads between the two datasets is explained by the number of rivers included in the dataset. For the Netherlands, where river discharge is concentrated in a few big river inputs, the estimated discharges and nutrient loads do not differ much between the two data sets. In France and Denmark the total discharge and nutrient loads in the OSPAR-RID dataset is approximately twice as high as in the OSPAR ICG-EMO dataset. In the UK, where the OSPAR ICG-EMO dataset was maintained, the discharges and nutrient loads are higher in the EMO dataset than in the RID dataset.

Based on the above we assume that datasets with a higher discharge have a more complete representation of all the freshwater sources and are therefore more reliable. This means that for the UK we use the OSPAR ICG-EMO dataset as most reliable dataset and for the other countries we use the OSPAR-RID dataset.

Table 5.1 Ratios between mean discharges and nutrient loads over 2003 – 2010 in the OSPAR ICG-EMO dataset and the OSPAR-RID dataset per country

EMO/RID	discharge	DIN	DIP	totN	totP
UK	1.5	1.5	1.1		1.7
F	0.4	0.4	0.6	0.6	0.5
GE	0.8	0.8	0.8	0.7	0.7
NL	1.0	1.1	0.9	1.1	1.0
DK	0.5			0.5	0.4

Figure 5.2 shows the Dutch outflow locations of catchments in E-HYPE. The outflows into Lake IJssel have been directly moved to the outflow locations to the Wadden Sea and North Sea at the sluices of Kornwerderzand and Den Oever in the Afsluitdijk. Retention processes in Lake IJssel have therefore been ignored in our current approach. In the Netherlands the flow of river water through the delta is manipulated so that more Rhine water enters the North Sea through Lake IJssel and the Wadden Sea than would be expected from naturally flowing water. This infrastructure is not included in E-HYPE resulting in a large overestimation of Rhine discharge from the 'Nieuwe Waterweg' near Rotterdam and an underestimation of discharges from Lake IJssel. This has been adjusted by a redistribution of the discharges and nutrient loads within the Netherlands.

Table 5.2 shows the scaling factors applied to the original outputs from the E-HYPE model, to correct for deviations from the OSPAR-RID and OSPAR ICG-EMO datasets and for the wrong distribution of Rhine river water over different outlets to the sea. For Norway and Ireland uncorrected nutrient loads from E-HYPE have been used. For Norway only

part of the coastline is included in the model which made it hard to compare the OSPAR-RID loads to the part of the coastline included in the model. Ireland does not border the North Sea and therefore it was not included in the above analysis of nutrient load validation.
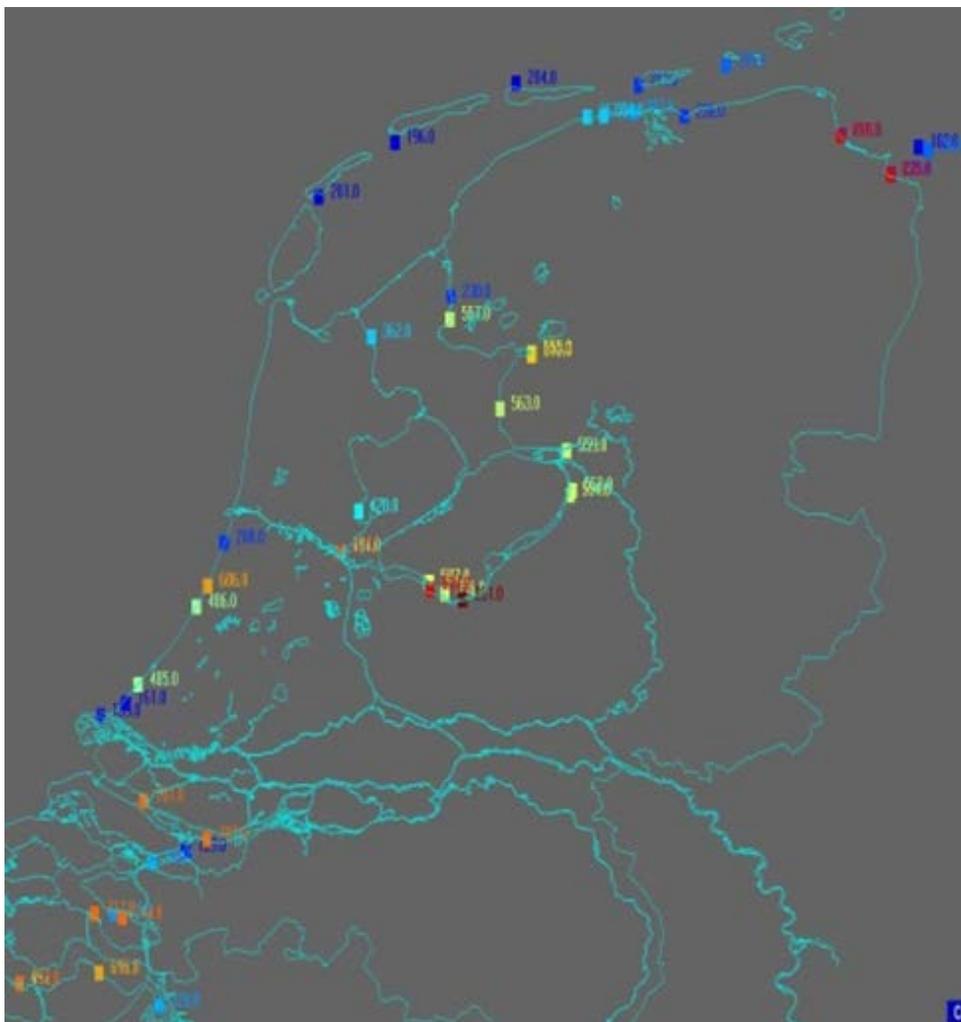

Figure 5.2: Dutch outflow locations of catchments in E-HYPE.

Table 5.2: Scaling factors applied to discharges and nutrient loads from the E-HYPE model, to calculate river inputs to the physical transport model.

Country	Source	Discharge	totN_conc	totP_conc
NL	Nwaterweg	0.58	0.56	0.90
	Haringvliet	1.57	0.56	0.90
	Ijsselmeer	2.24	0.56	0.90
UK	all	0.98	0.75	0.73
F	all	0.97	0.85	0.58
GE	all	0.77	0.73	1.21
DK	all	0.81	0.77	0.82
SE	all	1.15	1.10	1.23
NO	all	1.00	1.00	1.00

5.3.   Baltic inflow

The transport model domain does not include the Baltic Sea. Therefore, the inflow of water and nutrients from the Baltic Sea through the Kattegat needs to be included in the model as additional source. Monthly mean inflows of

water from the Baltic Sea have been estimated from an oceanographic model by SMHI (Meier et al., 2012; Eiola et al., 2009). The concentrations of DIN and DIP in the Baltic inflow have been estimated from measured concentrations near the surface at the Swedish monitoring location: K1. This approach assumes that the inflow of water from the Baltic Sea in predominantly in the form of brackish surface water. The outflow of North Sea water into the Baltic Sea along the bottom is ignored, since our study focuses on nutrients and chlorophyll in surface waters of the North Sea and Skagerrak. The input from the Baltic Sea through the 3 Danish straits is summed and the total is released at the model boundary, further east in the Baltic Sea. This is done to allow the model to simulate the exchange between brackish waters from the Baltic Sea and North Sea waters through the Danish straits and the resulting stratification pattern. Annex D describes in more detail the set-up of the oceanographic model by SMHI and how it has been used to estimate Baltic inflow fluxes to the nutrient transport model in JMP-EUNOSAT.

5.4. Validation results for recent years (2009- 2013)

For validation of transport patterns in the nutrient transport model and particularly the distribution of fresh river inputs in the sea we first validated the results for salinity. We've used the in-situ data, collected as part of the JMP-EUNOSAT project and the available Ferrybox data (from the COSYNA data portal: www.hzg.de/institutes_platforms/cosyna/data_management/). For most of the North Sea the salinity patterns in the model closely resembled the observations (Figure 5.3: green colours). There were two main exceptions: 1) in estuarine areas including the Wadden Sea the exchange between fresh water inputs and marine waters is underestimated leading to underestimation of salinities (Figure 5.3: blue colours); 2) the inflow of waters from the Baltic Sea is not well represented due to complex hydrodynamics not fully captured in the model. Figure 5.3 shows that the model had too low salinities south of the Danish straits (blue colours) and too high salinities north of the Danish straits (red colours).
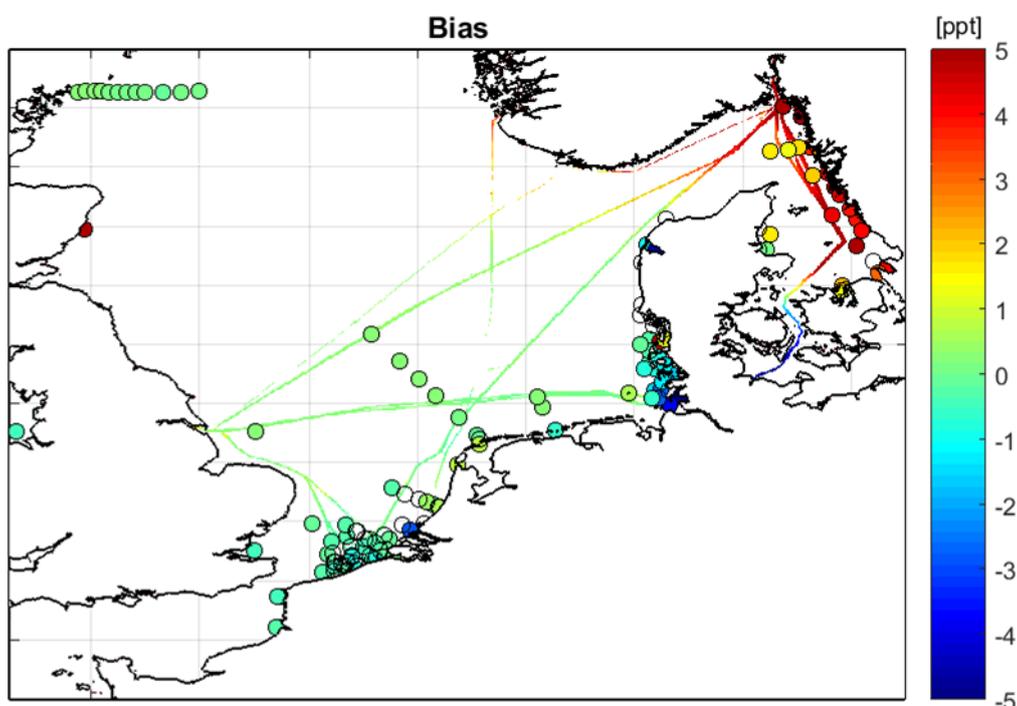

Figure 5.3: Transport model validation results for salinity. Colours represent the deviation of yearly mean salinity in the model (2009 – 2013) from the mean in available in-situ data: from blue (underestimation) to red (overestimation). Circles represent in-situ monitoring locations and lines represent Ferrybox transects.

A closer look at the validation results in the German Wadden Sea (Figure 5.4) shows that salinities are underestimated (blue) at water depths lower than approximately 10 m. The cross-shore gradient in the difference between observed

and simulated salinities is particularly clear in the Ferrybox data. Nearshore blue colours indicate underestimated salinities, whereas further offshore salinities are first slightly overestimated (yellow-green) and then correctly simulated (green). This pattern suggests that the exchange between fresh river inputs and marine waters is underestimated in shallow areas with deep gullies. We expect that a correct representation of the exchange through deep gullies in such estuarine systems requires a much higher resolution than the one nautical mile resolution that we've used in this model study.
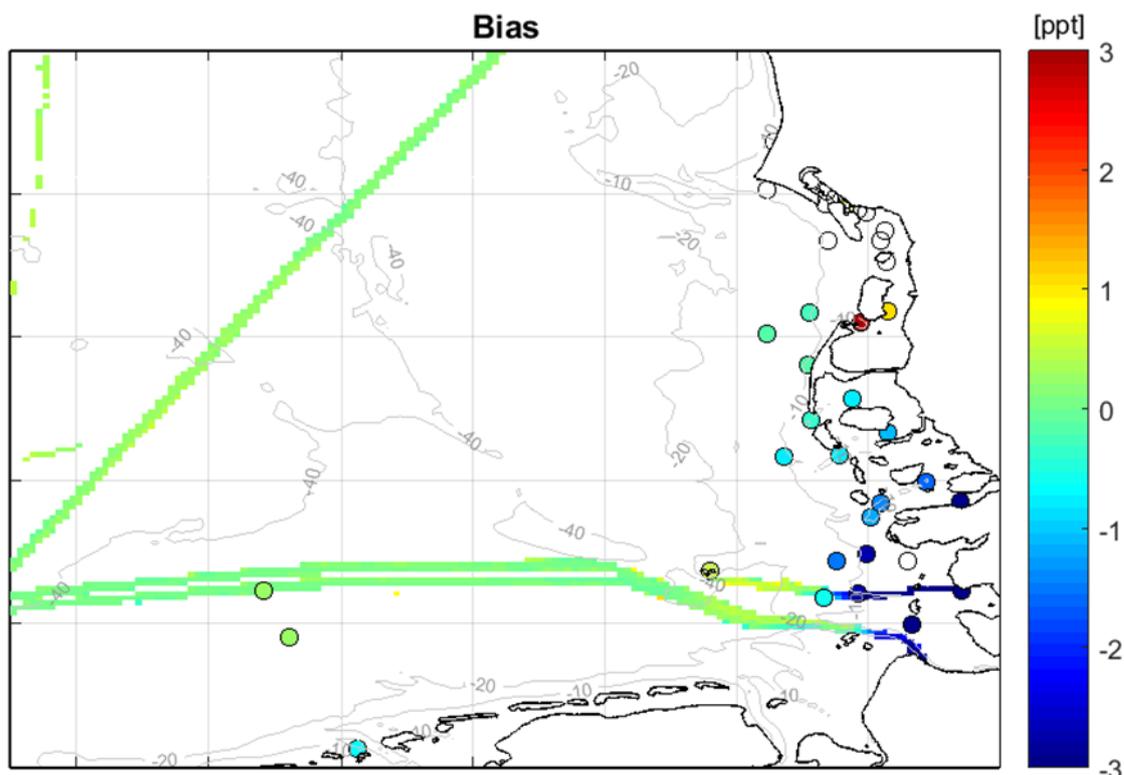

Figure 5.4: *Transport model validation results for salinity in the German Bight and Wadden Sea. Colours represent the deviation of yearly mean salinity in the model (2009 – 2013) from the mean in available in-situ data: from blue (underestimation) to red (overestimation). Circles represent in-situ monitoring locations and lines represent Ferrybox transects.*

Based on the above validation results we conclude that our model gives an appropriate representation of salinity, indicative of the distribution of river inputs, in parts of the model that meet the following criteria:

- Water depth is above 10 m
- Salinity is above 20 psu
- Longitude is below 9°.

Figure 5.5 summarizes the correlation between simulated and observed salinity in all available monitoring locations for JMP-EUNOSAT and in the selection of monitoring locations meeting the above criteria.
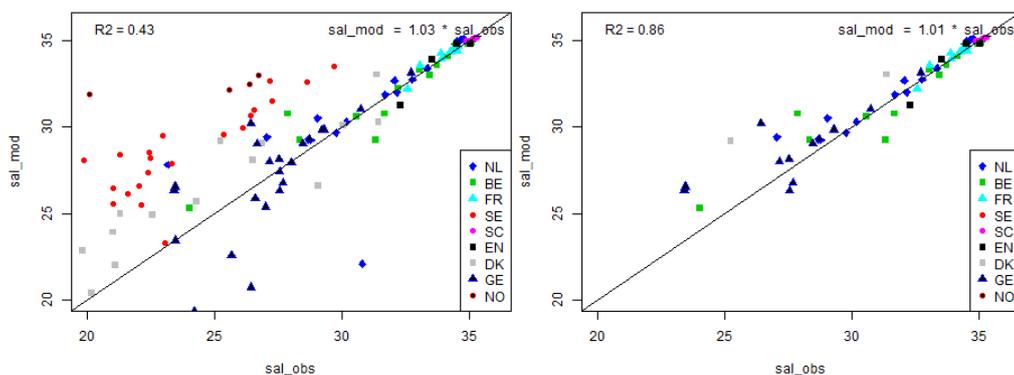

Figure 5.5: Correlation between observed salinity in situ (x-axis) and simulated salinity (y-axis) for all in-situ monitoring locations (left) and for a selection of monitoring locations (right) meeting our criteria for appropriate model fit.

Figure 5.6 shows the simulated salinities throughout the model, together with in-situ observations, after filtering the in-situ data with the above criteria. It shows that salinity in a large part of the model domain is considered as appropriately modelled, based on the above criteria.
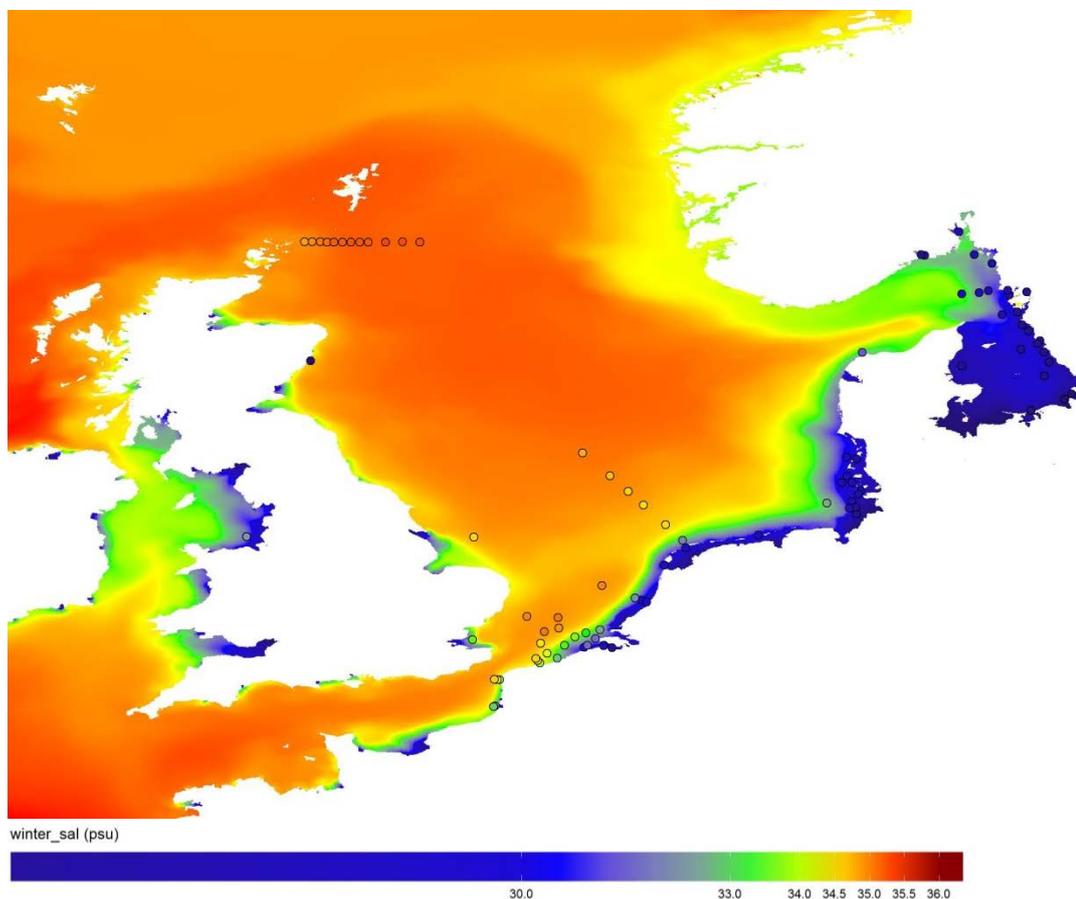

winter_sal (psu)

Figure 5.6: simulated winter mean salinities over 2009 – 2013 (background colour) and observed winter mean salinities at in-situ monitoring locations in the North Sea.

The next step in our model validation is the comparison of simulated nitrogen and phosphorus concentrations with available in-situ observations. For the input of nutrients from rivers we have used total nitrogen and total phosphorus loads from E-HYPE. OSPAR threshold values have been defined for winter mean dissolved inorganic nitrogen (DIN) and winter mean dissolved inorganic phosphorus (DIP). Therefore, we choose to define new threshold values also in terms of winter mean DIN and DIP.

Figure 5.7A shows the correlation between simulated winter mean total nitrogen and observed winter mean DIN, for those locations that meet the validation criteria for salinity. If the salinity is not simulated correctly, we don't expect the simulated total nutrient concentrations to be reliable either. The correlation is quite strong (R^2 = 0.84) and the linear regression model suggests that the in-situ winter mean DIN can be approximated by 0.81 * the simulated totalN. The correlation between simulated total phosphorus and observed winter mean DIP is less strong (Figure 5.7B; R^2 = 0.62). Still we used the linear regression models shown in Figure 5.7 to convert the model results of total N and P to maps of winter mean DIN and DIP.
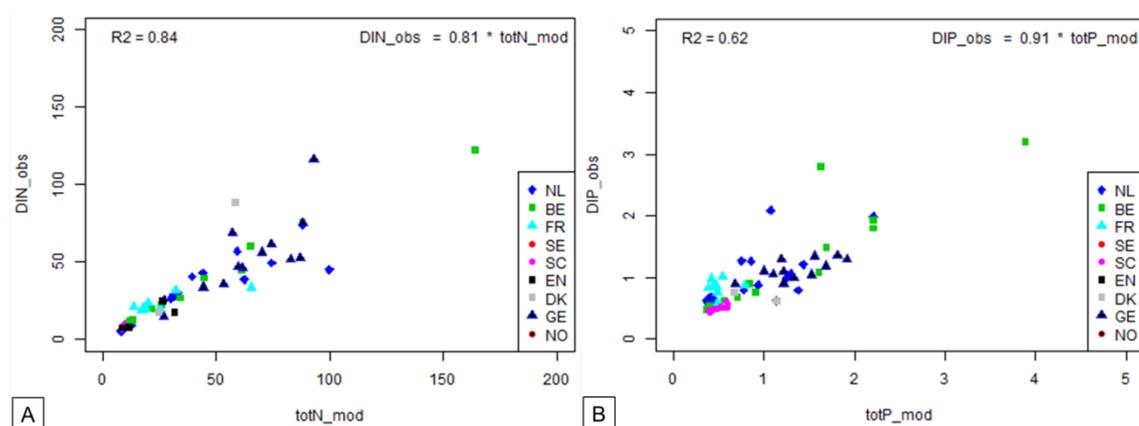

Figure 5.7: regression plots of: A) observed winter mean DIN with simulated winter mean total nitrogen and B) observed winter mean DIP with simulated winter mean total phosphorus

Figures 5.8 and 5.9 show the resulting maps for winter mean DIN and DIP, together with corresponding estimates based on in-situ monitoring data. Coastal gradients in winter DIN are generally well represented, except in the outflow of the Baltic Sea, all the way up the Norwegian coast. Concentrations in the southern North Sea are slightly underestimated. Winter DIP concentrations are well represented by the model in the coastal waters along the Belgian, Dutch and German coast and even in the outflow of the Baltic Sea and in the northern North Sea. However, in offshore waters of the southern and central North Sea and in the Channel DIP concentrations are underestimated by the model. The figures also show that spatial patterns in in-situ observations sometimes look noisy, which may be caused by a limited number of observations used to calculate the winter mean.
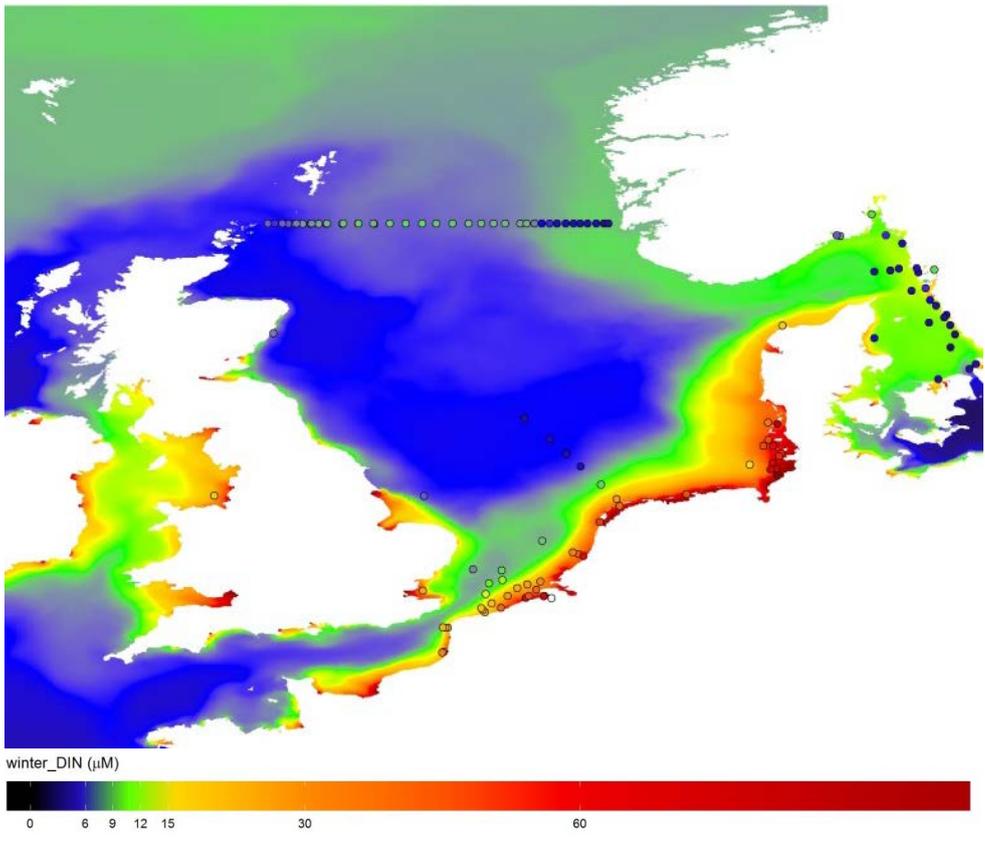
Figure 5.8: Spatial pattern of winter mean DIN in the model (background colour) and in observed in-situ data (circles)
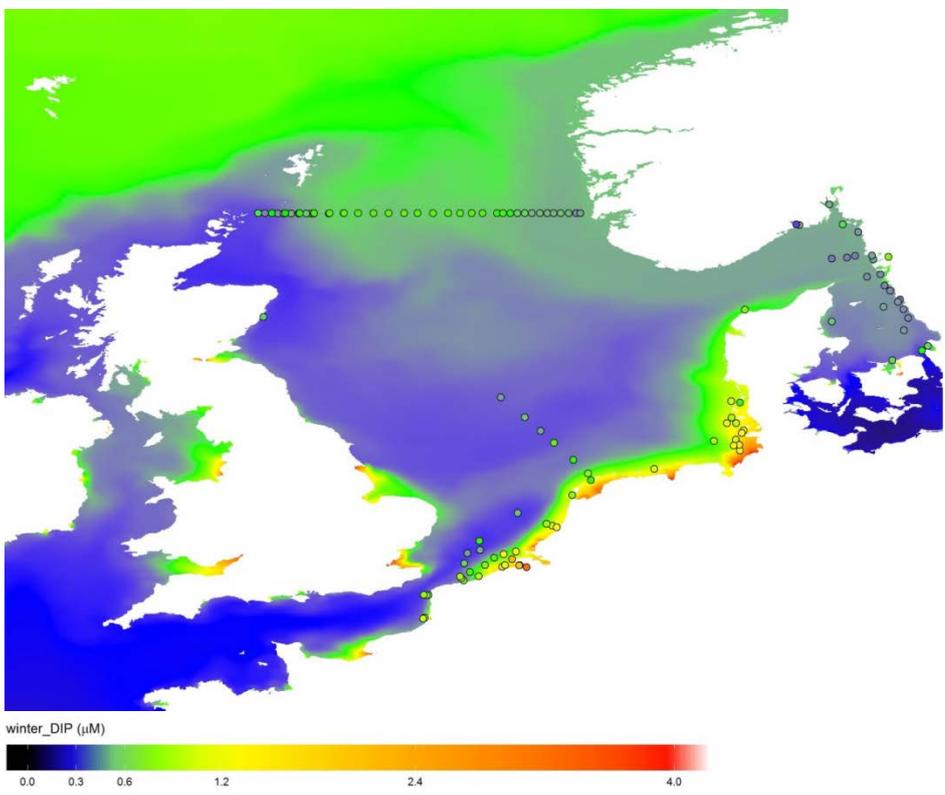

Figure 5.9: Spatial pattern of winter mean DIP in the model (background colour) and in observed in-situ data (circles)

6. Estimation of chlorophyll concentrations

6.1. Approaches

There are many modelling approaches and applications for the North Sea available for the estimation of chlorophyll-a concentrations. When model results are to be used for defining threshold values it is important to understand to what extent these ethreshold values are influenced by the choice of the model. Therefore, we have applied 3 modelling approaches in this study to get an indication of the sensitivity of the results to the choice of the model:

1. A linear regression model, relating winter mean nutrient concentrations to growing season mean chlorophyll-a concentrations;
2. The BLOOM model, which has been used by Deltares for decades to estimate chlorophyll concentrations in the North Sea and other marine systems worldwide.
3. The ERSEM model, which is another commonly used model for the North Sea. It is for example used in CMEMS (Copernicus Marine Environment Monitoring Service).

All modelling approaches were run with the exact same input data, to make the results comparable. Earlier model comparisons for the North Sea by OSPAR ICG-EMO have shown large differences between model outcomes. But it was hard to understand the reasons for the differences because in this model comparison not only the relations between environmental variables and chlorophyll-a differed but also the estimation of the environmental variables (nutrient concentrations, temperature, under water light intensity etc.).

The downside of this approach is that, in order to make the model inputs of the deterministic models BLOOM and ERSEM the same as for the regression model, they needed to be simplified from fully coupled 3D models to 0D box models.

The alternative approach would have been to make different North Sea models fully comparable by using the same inputs for meteorology, transport, boundaries and inputs of substances from rivers and the atmosphere. This approach did not fit in the time and budget available for this project. Therefore, we chose to first go through the entire approach (as shown in Figure 1.1) with relatively simple tools that were readily available. If this approach turns out to be feasible and acceptable, we can improve tools in follow-up work, where needed.

The deterministic approaches (BLOOM and ERSEM) gave unsatisfactory validation results. Therefore, we used only the linear regression model to estimate threshold values. The set-up and validation results of the two deterministic approaches are described in Annex A.

6.2. Linear regression model

Statistical models fit relations between variables in observed data. For this study we are mainly interested in explaining spatial patterns in chlorophyll-a by spatial patterns in the availability of nutrients and light. Therefore, we use the dataset of seasonal means over 2009 -2014 (described in chapter 3) for the model construction. The smoothing over seasons and 6-year periods reduces the noise in the data due to inter-annual variability and temporal under-sampling. The simplest statistical model to use is linear regression. This approach is also commonly used in the WFD and is easily understood, also by people that are not familiar with ecological modelling.

As explanatory variables we initially used winter mean DIN and winter mean DIP (December – February). We first fitted models with only one explanatory variable to compare the explained variance (R^2) between variables. Winter mean DIN could explain the largest part of the observed spatial variability in growing season mean chlorophyll-a

concentrations (Figure 6.1). Therefore, we decided to use the linear regression model based on winter mean DIN as our base model. From the linear regression model based on available in-situ data we concluded that a slope of 0.2 would be appropriate.
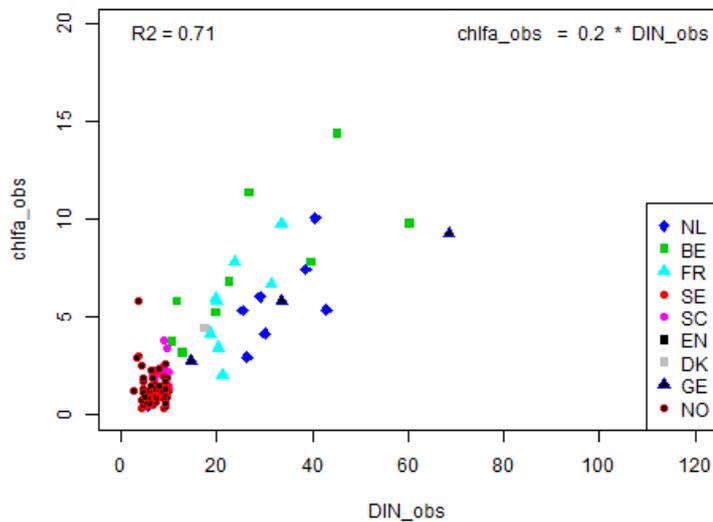

Figure 6.1: Linear correlation between observed growing season mean chlorophyll-a and winter mean DIN in data of 2009 – 2014.

Figure 6.2 shows the resulting spatial distribution of growing season mean chlorophyll-a concentrations calculated as 0.2 times the model estimated winter DIN concentrations (from Figure 5.8) in comparison with estimates based on in-situ data (circles) and satellite data (right panel). Overall spatial patterns and cross-shore gradients are fairly well represented by this simple regression model. There are a few exceptions where the model shows very different results than in-situ data and satellite data:

- In the Southern Bight of the North Sea the model seems to underestimate chlorophyll concentrations, both in comparison with satellite data and with local in-situ data.
- In the stratified central part of the North Sea, in the Irish Sea and along the Norwegian coast chlorophyll-a concentrations are overestimated. In reality organic matter and associated nutrients sink below the pycnocline leading to very low concentrations of nutrients and chlorophyll-a in surface waters. The satellite and our in-situ dataset represent only the top (approximately 1-10 m) layer of the water column.
- In the Dutch and German Wadden Sea chlorophyll-a concentrations in the model are higher than observed by the satellite and in-situ data. This is most likely caused by overestimation of nitrogen concentrations in these waters, due to too limited exchange of water between the Wadden Sea and offshore waters (see also Figure 5.4)
- Along the English east coast satellite data show higher chlorophyll concentrations than would be expected from simulated nitrogen availability alone.
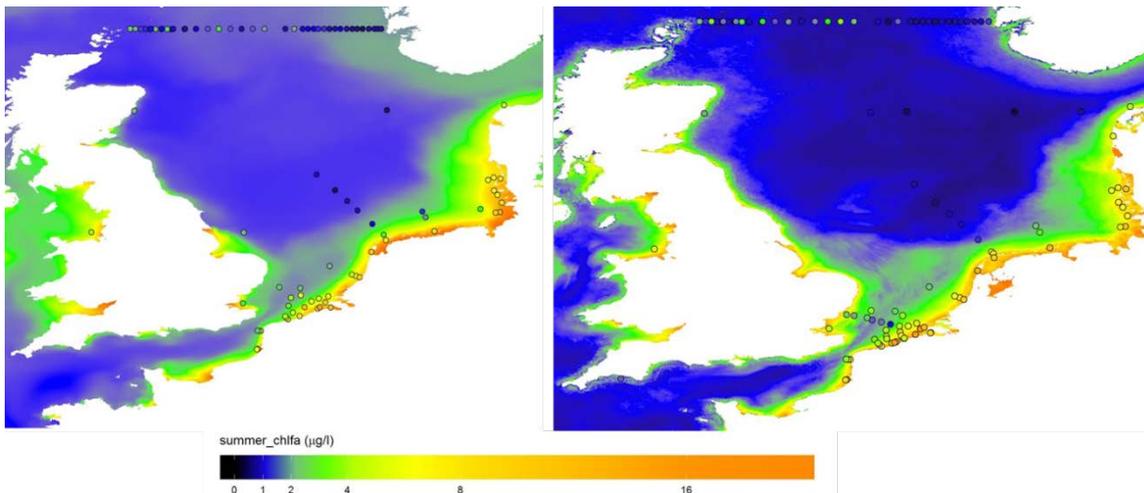
Figure 6.3: Regression model estimates of growing season mean chlorophyll-a concentrations for 2009 – 2014 (left panel, background colour) with JMP EUNOSAT satellite data for 2005 – 2010 (right panel, background colour). Estimates based on in-situ data for 2003 – 2008 are shown as circles.

We tried to improve the model performance by including other variables. Spatial variability in DIN, DIP and light climate (approximated by the extinction coefficient for visible light (Kd) or suspended matter concentrations (spm)) is strongly correlated: DIN, DIP, Kd and spm are all highest near river inputs and near the coast. Therefore, fitting a multiple linear regression model with all these variables would yield unrealistic models, due to collinearity effects (results not shown). As an alternative we checked if the residuals of the linear regression model with DIN were correlated with the availability of DIP or under water light climate. We fitted the (molar) ratio between winter mean DIN and DIP against the residual of the DIN model, based on the hypothesis that the model based on DIN alone may overestimate chlorophyll concentrations in areas with relatively low DIP concentrations. In those areas low DIP availability may limit phytoplankton growth, leading to lower chlorophyll-a concentrations. Figure 6.4 shows that indeed the highest model residuals occurred at NP ratios above circa 20. But these high residuals occurred predominantly in Dutch and German waters. In French and Belgian waters the residuals tend to be negative at high NP ratios. Overall the R^2 of the linear relation between the residuals and NP-ratio was very low (0.06) so we did not add this variable to our linear model based on DIN.
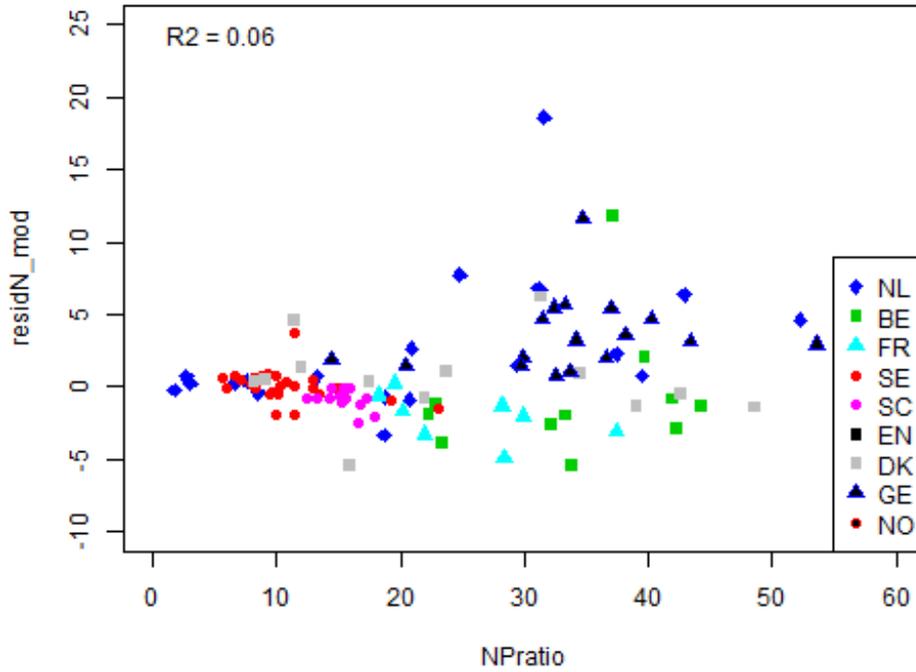
Figure 6.4: Relation between the residuals (model – observation of the DIN regression model) against NP ratio (winter mean DIN / winter mean DIP) in the JMP-EUNOSAT dataset.

To check for a possible improvement of the DIN-based regression model by including effects of light we tried different approaches. In theory the depth averaged availability of light in the water column (I_M) is determined by a combination of water depth (Z) (or mixed layer depth in stratified conditions) and water transparency (approximated by the extinction coefficient for photosynthetically active radiance Kd_{PAR}). In turbid waters the depth averaged availability of light can be approximated as:

$$I_M = Kd_{PAR} \times Z$$

The JMP-EUNOSAT included only limited data on Kd, Secchi or suspended matter concentrations. Therefore, we used satellite data of spm received from RBINS as part of JMP-EUNOSAT activity 2. These were estimated as long term monthly averages of total suspended matter concentrations in MERIS data of 2003 – 2011, processed with the MEG7.5 algorithm by RBINS. We estimated Kd using the same formula that is routinely used in the Deltares models for the North Sea (BLOOM / GEM):

$$Kd_{PAR} = 0.08 + 0.025 \times spm + 0.97 \times (1 - \frac{sal}{34.97})$$

We plotted both Kd x Z and Kd alone against the residual of the nitrogen model and the Kd alone showed a stronger correlation than Kd x Z (Figure 6.5). Although the R^2 of this relation was higher than the relation with NP-ratio we still did not consider it convincing enough to add it to the simple DIN based regression model.

Based on light limitation for growth one would expect lower algal biomass in turbid waters than in clear waters. Highest residuals of the DIN-based model at high Kd-values (indicating turbid waters) suggest that chlorophyll-a concentrations were higher in turbid waters than in clear waters. This may be explained by higher chlorophyll to carbon ratios in turbid waters. This is a well-known adaptation of phytoplankton to turbid conditions, when they need more pigments to harvest sufficient light.

Alternatively, chlorophyll-a concentrations can be higher in turbid waters because these indicate waters with strong vertical mixing and nearby sediment. This prevents algae from settling to the sediment or below the pycnocline and nutrients getting lost from the upper mixed layer of the water. Due to settling both suspended sediment concentrations and algae concentrations are low in stratified areas. This suggestion is further supported by the ratio of chlorophyll concentrations in satellite data with winter mean total nitrogen concentrations from our model (Figure 6.5). If algae would always produce the same amount of chlorophyll per unit of nitrogen (as our linear model assumes) this ratio would be constant throughout the North Sea. However, there are clear spatial patterns visible: with very low (blue) values in seasonally stratified parts of the North Sea (see also chapter 9 about definitions of assessment areas) and relatively high values in the southern North Sea and along the UK east coast. Differences in vertical mixing intensity are likely to explain part of these differences. The blue area in the German Bight may be (partly) explained by overestimation of total nitrogen concentrations in this area and / phosphorus limitation of algal growth.

Another possible explanation for spatial differences may be that our model only includes nitrogen inputs from rivers but not from atmospheric deposition. Figure 6.7 shows concentrations of NOx in the atmosphere as observed by satellites (OMI and GOME), with relatively high concentrations above the southern North Sea and Channel and along the English east coast.

Our linear model, fitted with in-situ data, assumes a constant ratio between chlorophyll and total nitrogen of 0.166 (0.81 * 0.2), but the satellite data in Figure 6.6 show large spatial variability around this mean value, which explains the areas where our model results do not correspond to satellite data and in-situ data (Figure 6.4). This illustrates the added value of the high spatial resolution of satellite data for gaining ecosystem understanding. Further analyses of spatial patterns in model results and satellite data are expected to lead to better system understanding and model improvement.
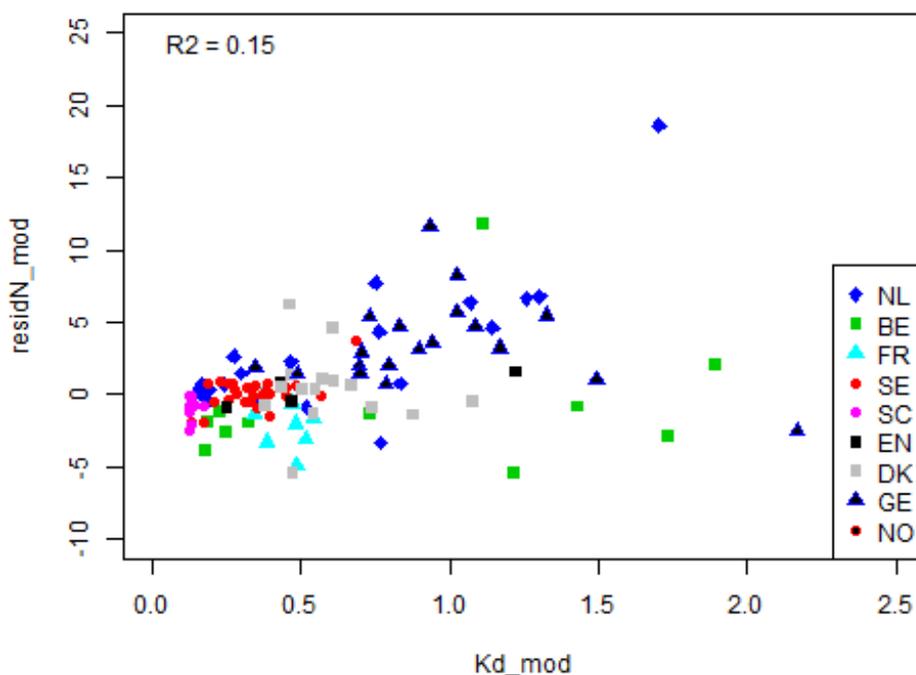

Figure 6.5: Relation between the residuals (model – observation of the DIN regression model) against the local extinction coefficient (Kd) estimated from satellite data and simulated salinity.
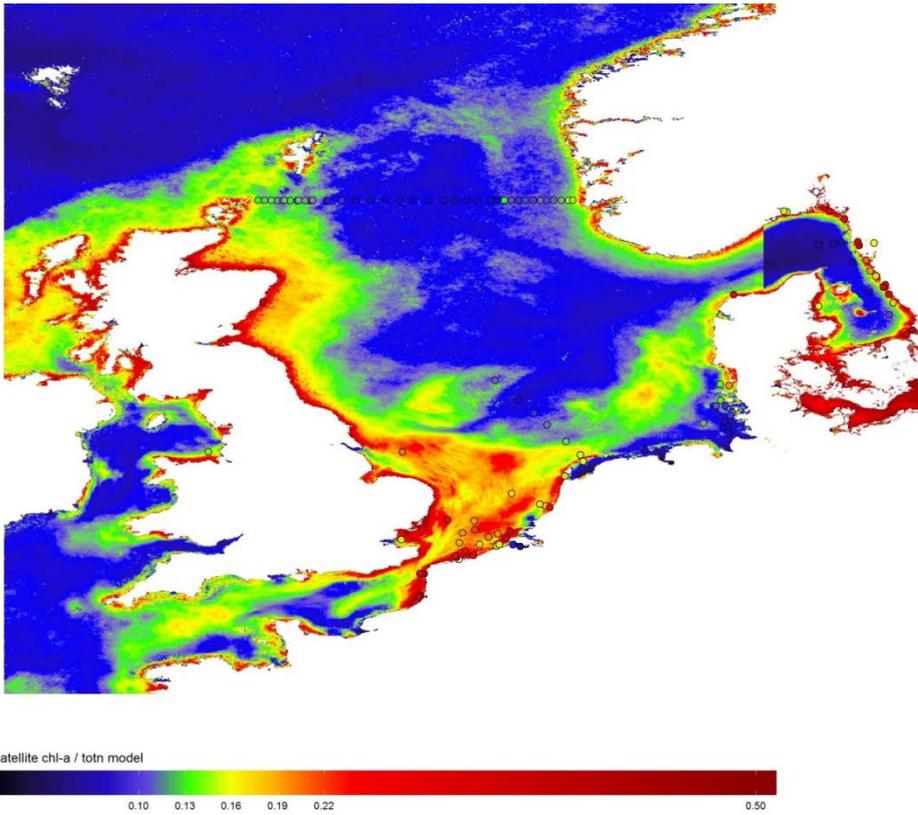
Figure 6.6: Ratio between satellite data of growing season mean chlorophyll-a concentrations and winter mean total nitrogen concentrations from the model.
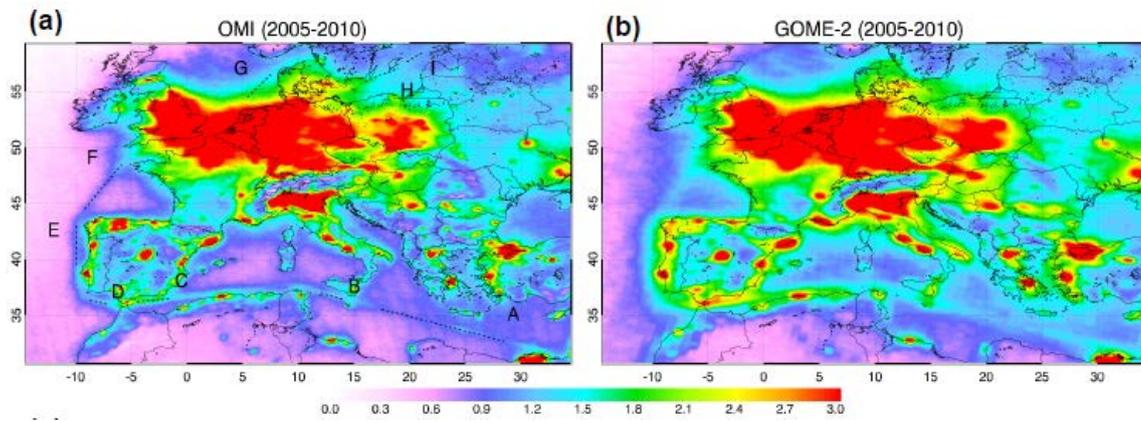

Figure 6.7: Concentrations NOx in the atmosphere from satellite data (from Yu et al., 2014)

7. Threshold values

7.1. Threshold values for nutrients

Threshold values are defined as maximally 50% above the natural background concentration, according to an earlier definition by OSPAR. We have estimated the spatial pattern of winter mean DIN and DIP concentrations with our model, with inputs representative for the period around 1900. The newly proposed threshold values for winter mean dissolved nutrients are calculated as the winter mean nutrient concentration in 1900 multiplied by 1.5.

The resulting threshold for winter mean DIN is compared to the current threshold for winter mean DIN, applied by OSPAR for their eutrophication assessments for the North Sea (Figure 7.1). The current threshold values are uniform per assessment area and show clear differences at country boundaries. The newly proposed threshold values are cross-boundary coherent and reflect mostly the gradients in fresh water influence. The newly proposed threshold values are defined at a high spatial resolution to allow for local comparisons with satellite data at the same 1 x 1 km spatial resolution. Generally, the newly proposed threshold values are higher than current threshold values near the coast and lower in offshore waters.
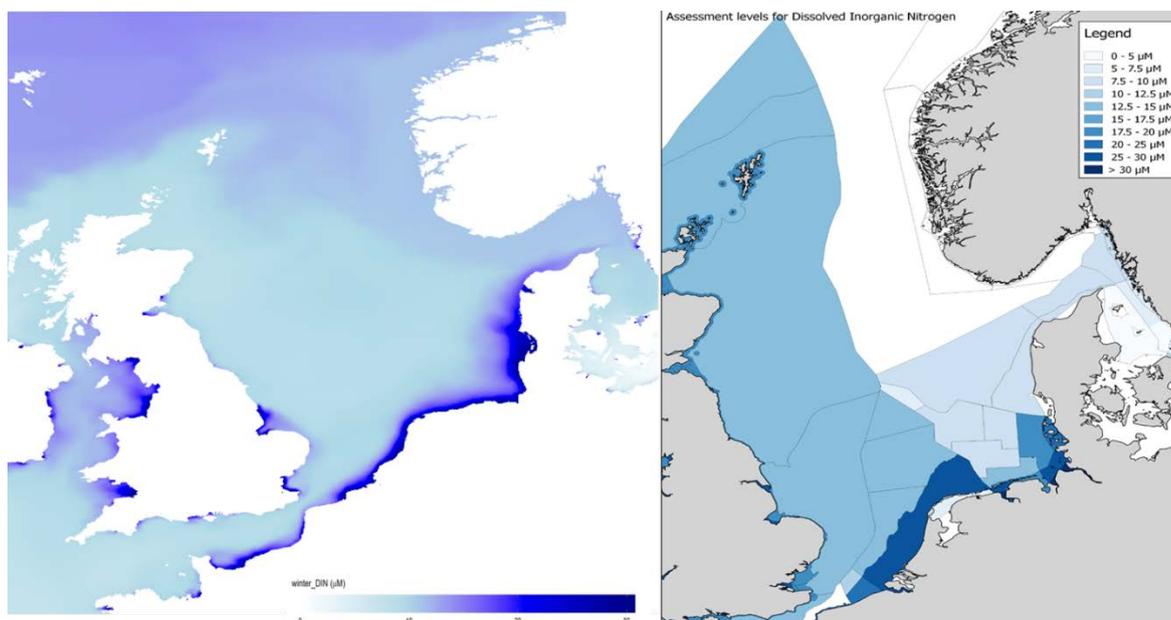

Figure 7.1: Comparison of newly proposed threshold values for winter mean DIN (left) and current OSPAR threshold values (right; OSPAR, 2017b). Lines in the right panel delimit territorial waters and subdivisions for eutrophication assessment.

Currently, threshold values for winter mean DIP have only been defined by the Netherlands, Germany and Denmark. For all 3 countries the novel proposed threshold values are generally higher than the present threshold values near the coast. Offshore the novel levels are lower than the current threshold values for the Netherlands and Denmark. For Germany the novel levels are higher than the present offshore threshold values.
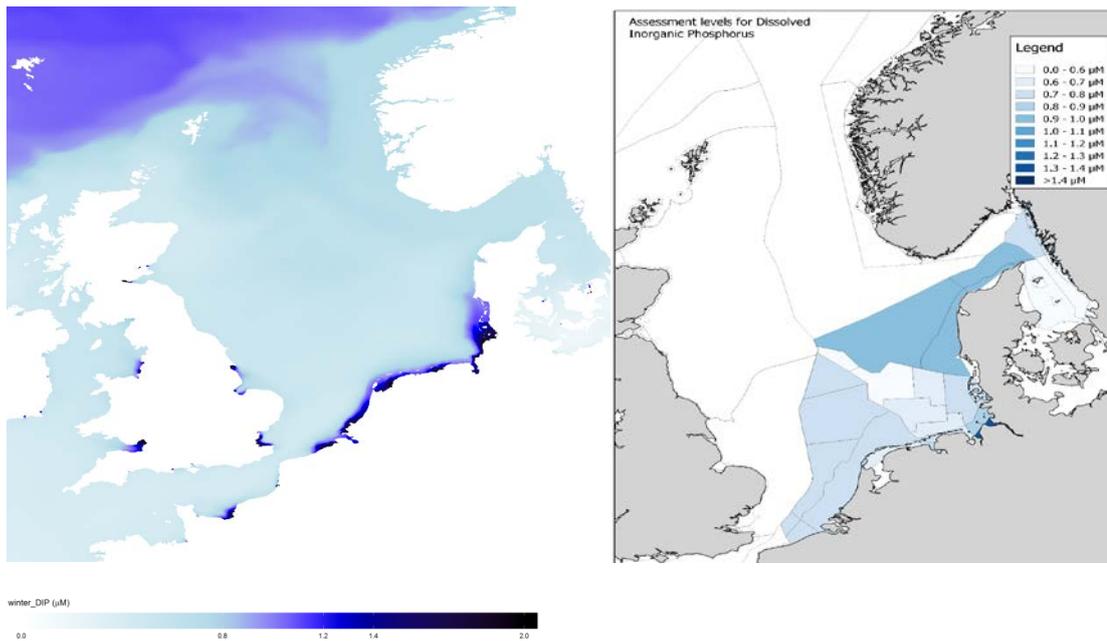
Figure 7.2: Comparison of newly proposed threshold values for winter mean DIP (left) and current OSPAR threshold values (right; OSPAR, 2017b). Lines in the right panel delimit territorial waters and subdivisions for eutrophication assessments.

7.2.   Threshold values for phytoplankton

The spatial pattern of the novel proposed threshold values for chlorophyll-a show the same pattern as the threshold values for DIN (Figure 7.3). In this study we have defined threshold values for eutrophication as seasonal means, since these can be estimated much more reliably from observed time series than 90-percentiles. In Figure 7.3 the novel proposed threshold values for growing season mean chlorophyll concentrations (left) are expressed as growing season 90-percentiles, to allow comparison with the colour scale of current OSPAR threshold values. The threshold for growing season 90-percentile is estimated as 2 times the growing season mean. The same approach was used by OSPAR to estimate Swedish and Danish 90-percentile threshold values from their growing season mean threshold values (Figure 7.3, right panel). Similar to the threshold values for nutrients, the novel threshold values for chlorophyll are higher than current levels nearshore and lower offshore. This is mainly due to the use of uniform threshold values per assessment area in the current approach.
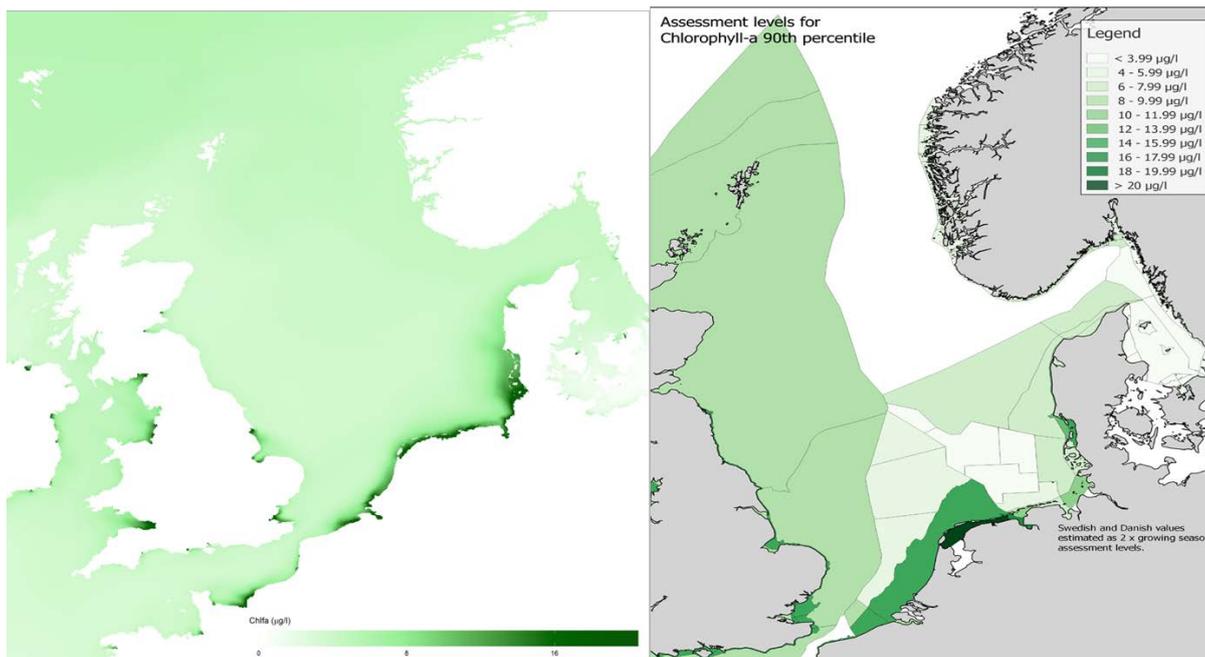
Figure 7.3: Comparison of newly proposed threshold values for growing season mean chlorophyll (left) and current OSPAR threshold values (right; OSPAR, 2017b). Lines in the right panel delimit territorial waters and subdivisions for eutrophication assessments.

7.3. Discussion

When the novel proposed threshold values for growing season mean chlorophyll are compared to the chlorophyll concentrations from the coherent JMP-EUNOSAT satellite product from activity 2 a large part of the southern North Sea is above the threshold (Figure 7.4). This area corresponds to the permanently mixed area with water depth below 35 m, that is identified in chapter 9 of this report (Figure 9.9). In this area our linear regression model for chlorophyll-a underestimates present chlorophyll-a concentrations (Figure 6.3). This results in an underestimation of chlorophyll-a concentrations in 1900 and hence of the threshold values in the southern North Sea. In this area the ratio between satellite estimates of chlorophyll and the model estimates of total nitrogen availability is relatively high (Figure 6.6). If we would apply the local ratio from Figure 6.6 instead of a constant ratio between growing season mean chlorophyll concentrations and total nitrogen the model would perfectly reproduce the present chlorophyll concentrations and the assessment result would be only affected by nitrogen availability from the model. In this case threshold values are only exceeded in near-shore waters (Figure 7.5).

The large difference between the two assessment results indicates the sensitivity of the assessment result to the modelling approach applied to estimate threshold values. Although the modelling approach with a spatially varying chlorophyll yield per unit of nitrogen gives a perfect fit with the satellite data it is not very informative on the reasons for spatial variability of this yield. In fact, this approach effectively compares model results of total N concentrations between the present and 1900, so the additional information about the ecosystem from the satellite data is not used. For example, the importance of atmospheric inputs of nitrogen in the southern North Sea would be masked from the assessment by locally adapting the chlorophyll to nitrogen ratio in the model.
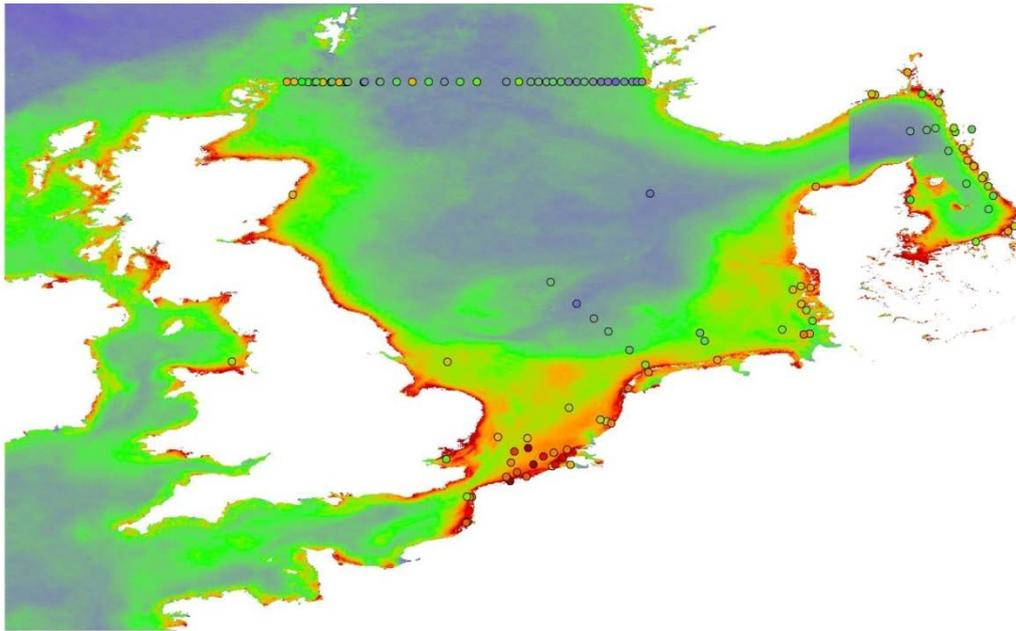
Figure 7.4: Ratio between the growing season mean chlorophyll concentration per pixel from the JMP-EUNOSAT satellite product and the novel proposed threshold. Orange and red colours indicate that present chlorophyll-a concentrations would exceed the novel proposed threshold.
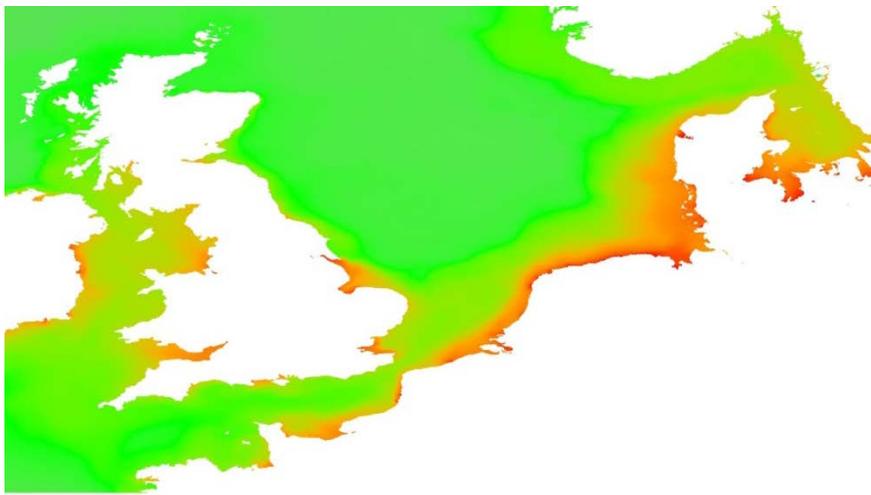

Figure 7.5: Assessment result (ratio between observed chlorophyll concentrations and threshold) if the threshold would be estimated with spatially variable ratios between chlorophyll and total nitrogen. Orange and red colours indicate that present chlorophyll-a concentrations would exceed the novel proposed threshold.

8. Estimation of primary production with satellites

Besides chlorophyll-a concentrations primary production rates would be a very interesting indicator for monitoring effects of eutrophication. This could also be an indicator for the MSFD (secondary) Foodwebs criterion D4C4: Productivity of the trophic guild is not adversely affected due to anthropogenic pressures. However, so far this indicator could not be used because there are hardly any monitoring data available for primary production. In the field this is a time consuming and expensive observation to make. Therefore, in JMP-EUNOSAT we have evaluated the applicability of primary production estimates from satellite data as additional indicator for eutrophication and foodweb assessments. More specifically the objectives were to:

1. Define regions of similarity in mean peak, timing and amplitude of primary production in the NE Atlantic using satellite data;
2. To define baseline values in primary production over the region using satellite;
3. To assess the variability in primary production over the region.

This chapter only summarizes the results for the above 3 objectives. More detailed results and information about the methods used to generate these results can be found in annex B. The primary production data from satellites were originally meant to be used as validation data for the process-based models, so that these models could estimate threshold values based on primary production in 19000, in the same way as the threshold values for chlorophyll in chapter 7. However, since the process-based model (in annex A) did not provide realistic values of chlorophyll in the North Sea, the primary production results were also considered not useful for the estimation of threshold values. Instead, baseline values representing current levels of primary production were derived from recent satellite data. These can be used in comparison with future primary production data to assess trends.

8.1. Regions with similar primary production patterns

To support the definition of assessment areas for OSPAR (*cf.* chapter 9), areas with similar patterns of primary production were identified using K-means clustering. The clustering was based on peak and timing of primary production in the satellite data. Eight regions were defined which broadly correspond to the principal hydrographic areas in the NE Atlantic (Figure 8.1). More details about the estimation of primary production from satellite data, its validation and the cluster analysis can be found in Annex B.
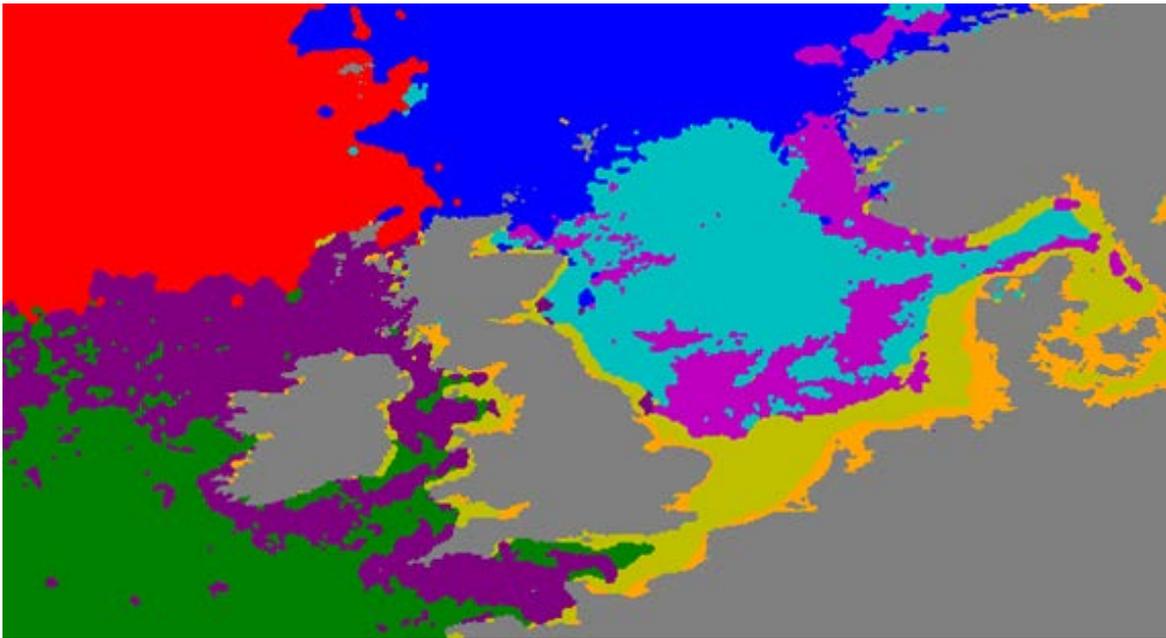
Figure 8.1. Regions of similar primary production patterns identified using k-means cluster analysis on peak and timing of primary production. Eight regions are identified: NE Atlantic (red), NW Atlantic (dark blue), Irish Sea and English Channel (purple), Central N Sea (light blue), Celtic Sea (green), NW European shelf seas (mauve) and NW European coast (orange).

8.2.   Baseline values of primary production

For the regions in Figure 8.1 baseline primary production rates were derived from satellite data from CMEMS for the period 1997 – 2017. Charactistic values for means and 90 percentiles for each region are shown in Figure 8.2. All areas show a somewhat similar seasonal pattern reflecting the intensity of solar irradiance over the season, with a peak in summer (around day 180). The peak values in summer are lower than would be expected if they would follow an identical pattern as solar irradiance. This reflects the influence of other drivers of primary production variability such as stratification. Primary production rates in the NW European coast (orange) are considerably higher than in the other areas. This is likely to reflect higher availability of nutrients in this area. Unfortunately, there were no monitoring data of primary production available from project partners to validate the satellite data of primary production in the North Sea. Estimates from literature could alleviate this issue in future work.

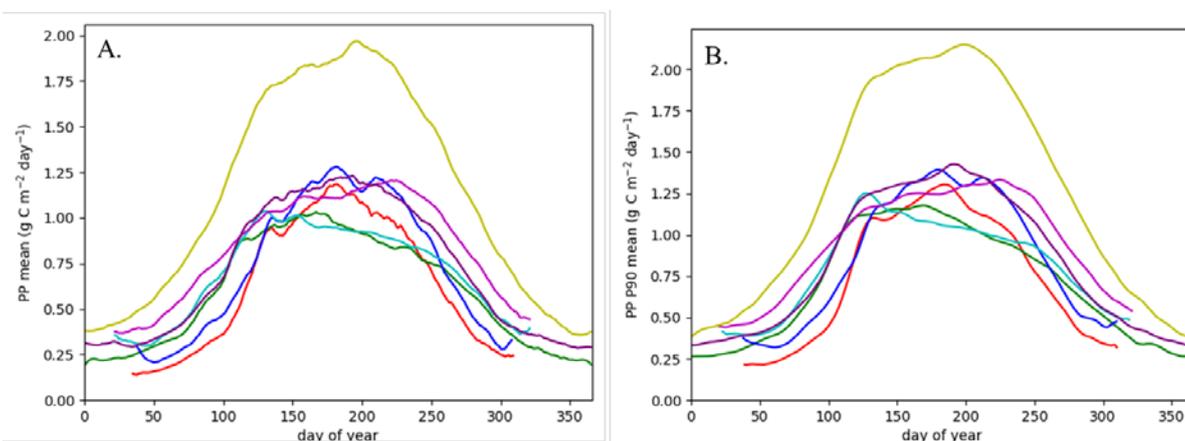
Figure 8.2. Climatological mean (A.) and mean 90th percentile (B.) of primary production (g C m^{-2} d^{-1}) for each of the Eight regions identified in Figure 8.1: NE Atlantic (red), NW Atlantic (dark blue), Irish Sea and English Channel (purple), Celtic Sea (green), NW European shelf seas (mauve) and NW European coast (orange).

8.3. Spatial and seasonal variability in primary production

Figure 8.3 shows the seasonal variability in satellite estimates of primary production at a high spatial resolution. These data form the basis for the K-means cluster analysis (Figure 8.1) and baseline values per region.
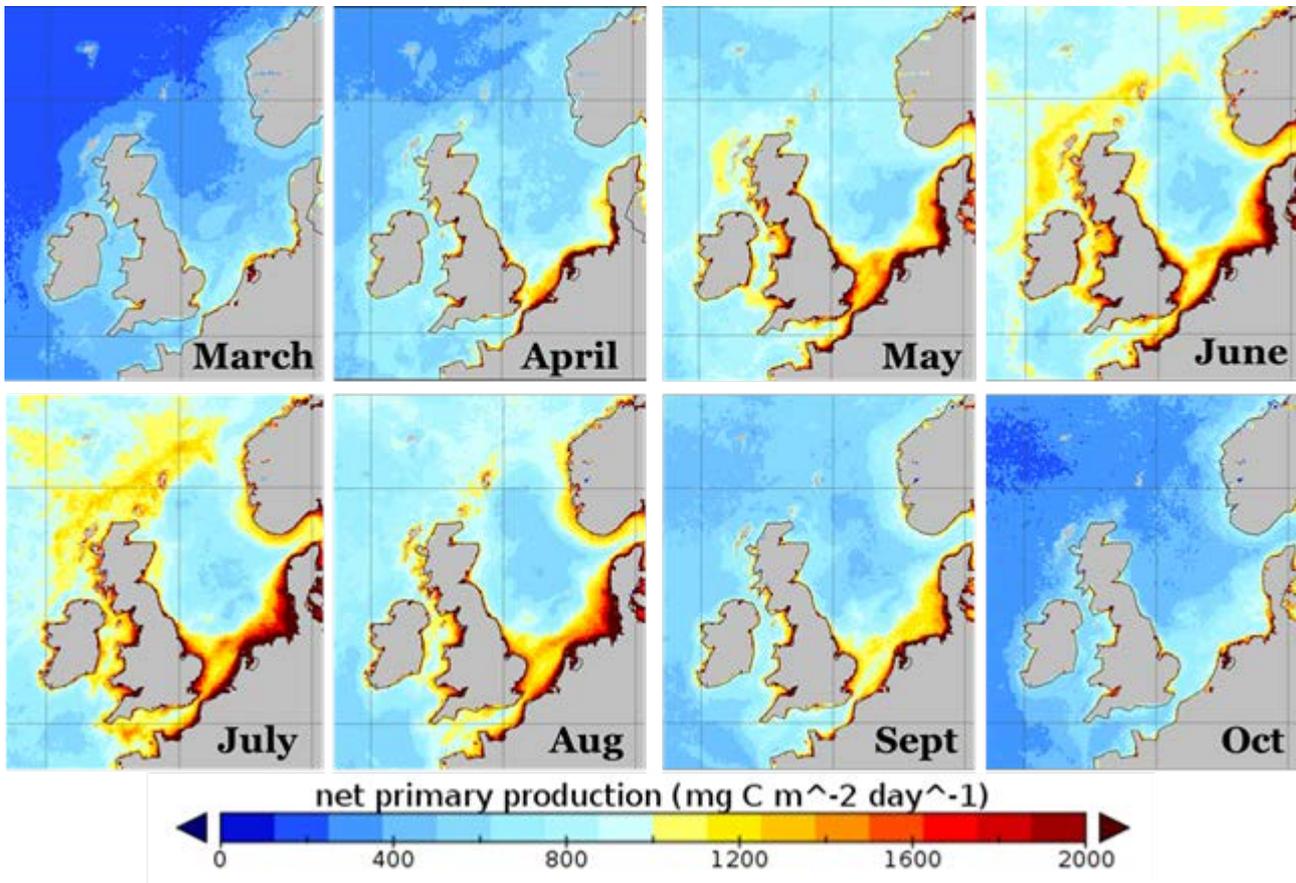

Figure 8.3. Mean monthly primary production from Copernicus Marine Environment Monitoring Service (CMEMS) data (1997-2017).

9. Assessment areas

9.1. Approach to define assessment areas

Current assessment areas used for OSPAR assessments (for example: OSPAR 2017) are defined by first splitting up the North Sea by national boundaries and then splitting up these in smaller areas, with similar ecological and physical functioning. Since countries have used different approaches and criteria for the definition of national sub-areas the resulting definition of assessment areas throughout the North Sea does not clearly reflect the ecological functioning of the North Sea (Figures 7.1 – 7.3). In the JMP-EUNOSAT project we have used a different approach:

1. We first defined areas with similar ecological and physical functioning throughout the North Sea;
2. Then split up these cross-border coherent areas into national sub-areas, so countries can take responsibility for their own part of the cross-border assessment areas.
3. These national sub-areas can be further subdivided into smaller areas, depending on preferences and practical considerations of countries. This would allow for example to assess changes in areas that are affected by specific river catchments.

The new assessment areas should reflect those characteristics of the North Sea ecosystem that are relevant for the assessment of eutrophication. The definition should result in assessment areas that share similar environmental conditions within one area, which can be distinguished from the conditions in other areas. Relevant environmental conditions include both physical, chemical and biological factors and anthropogenic pressures.

Physical factors determine phytoplankton growth conditions and are highly relevant to define assessment areas:

- Physical conditions:
    - Depth
    - Residence time
    - Vertical mixing (presence/absence of stratification)
    - Freshwater input

In addition, other abiotic and biological factors and anthropogenic pressures could also be considered, e.g.

- Physico-chemical and biological factors:
    - SPM levels
    - Zooplankton grazing
    - Benthic grazing

- Anthropogenic pressures:
    - Nutrient loading

9.1.1. Physical conditions; previous work by CEFAS

An analysis to identify distinct regions within the Greater North Sea based on hydrodynamical conditions was done by van Leeuwen et al. (2015). For this analysis the hydrodynamic model GETM (General Estuarine Transport Model; for more details see van Leeuwen et al., 2015) was used. This model has a spatial resolution (model grid size) of 6 nautical miles.

The model was used to distinguish regions with different stratification regimes. With this approach five distinct regimes were identified, i.e. permanently stratified, seasonally stratified, intermittently stratified, permanently mixed and the ROFI (Region of Freshwater Influence). Those regimes can be used to identify different areas based on hydrodynamic conditions only (Figure 9.1).

Results from the biogeochemical model ERSEM-BFM (for details see Lenhart et al. 2010) showed that the patterns of phytoplankton development and phytoplankton composition differ between those regions.

The areas presented in Figure 9.1 have been used as assessment areas by ICG-COBAM for the description of changes in phytoplankton and zooplankton communities as part of the assessment of biodiversity status (OSPAR, 2017a).
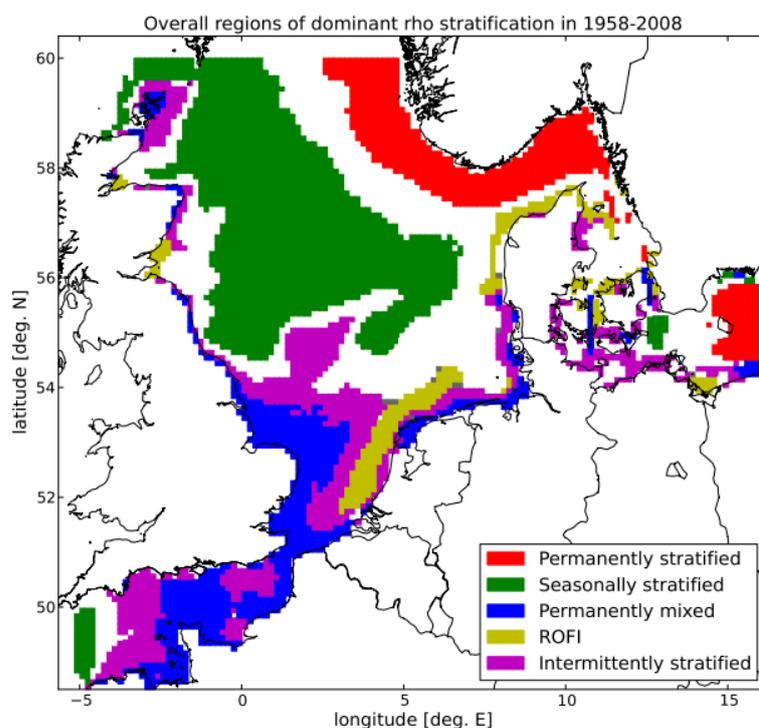
Figure 5. Time median results of the modelled, annual regions in the North Sea based on density stratification. Transparent areas indicate areas where the dominant regime occurs for less than 50% of the time (less visible due to minimal occurrence).

Figure 9.1. Modelled, annual regions in the North Sea based on density stratification (source: Figure 5 in van Leeuwen et al. 2015)

9.1.2. Physical conditions; JMP EUNOSAT results

As part of Activity 1, Deltares used the hydrodynamic model DCSMv6 FM (Dutch Continental Shelf model version 6) to model stratification and salinity, and those results were combined with data on bathymetry. Those three factors (stratification, salinity, depth) were then used to define assessment areas.

The DCSMv6 FM model has a spatial resolution (model grid size) of 1 nautical mile for all areas that are less than 100 m deep, so most of the North Sea (see Figure 5.1). It has been validated using data from the Dutch part of the North

Sea including data on stratification at the Oyster Grounds. The model application in the JMP-EUNOSAT project builds upon the assessment based on stratification by van Leeuwen et al. shown in Figure 9.1.

As a reference, the analysis of patterns in chlorophyll levels by RBINS (Activity 2 of JMP-EUNOSAT) was used. In this analysis by Desmit & van der Zande (Activity 2 report, Annex II), the chlorophyll signal from satellite data was decomposed into an inter-annual signal, a seasonal signal and a residual signal. The inter-annual signal can indicate long-term trends or regime shifts. The seasonal signal is an indication whether the blooms occur each year systematically in the same season or not. The residual signal gives an indication of the remaining variability and can give an indication of strongly varying conditions between years and seasons. With a statistical analysis using the various signals, areas with similar patterns can be identified and merged (Figure 9.2).

Largely similar areas appear in an analysis of patterns in primary production derived from satellite data, as developed by PML and described in chapter 8 of this report (Figure 9.3).
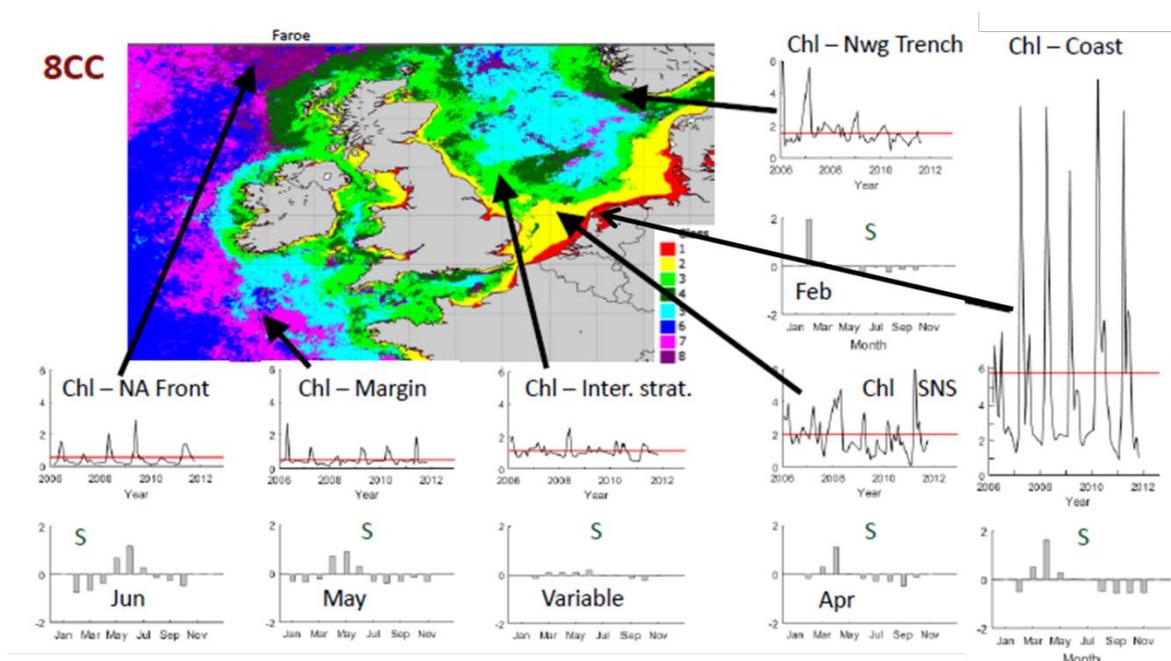

Figure 9.2: Eight clusters of CHL in the Greater North Sea (colours indicate the different clusters). Time series of CHL (µg L^{-1}) are shown with their grand mean (red line) and their seasonal signal for several pixels in typical areas (from left to right: the North Atlantic front, the margin between the ocean and the shelf, an area of intermittent stratification, an area in the Southern North Sea, the Norwegian Trench, and the Belgian Coastal Zone).
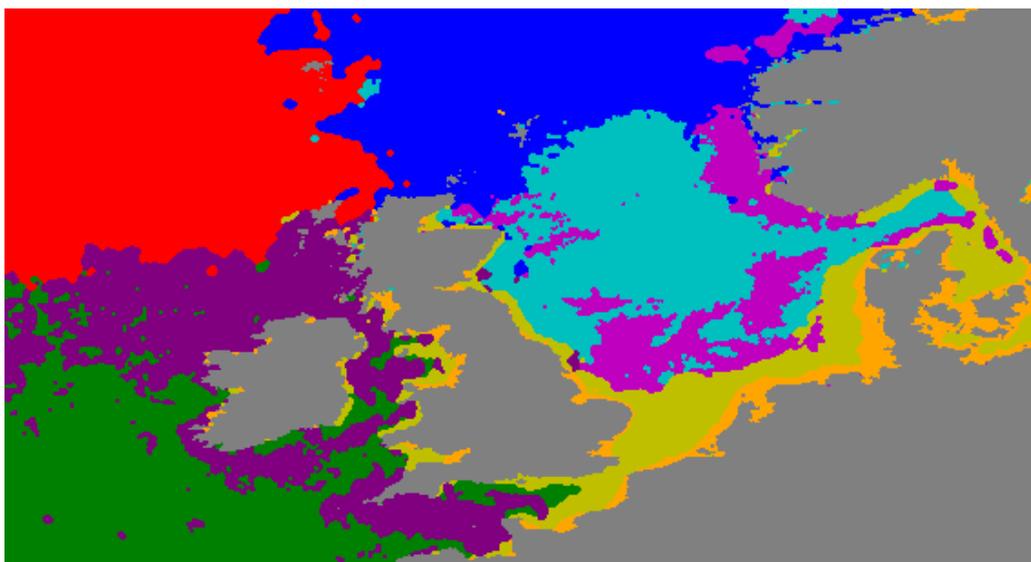
Figure 9.3. Regions of similar primary production patterns identified using k-means cluster analysis on peak and timing of primary production. Eight regions are identified: NE Atlantic (red), NW Atlantic (dark blue), Irish Sea and English Channel (purple), Central N Sea (light blue), Celtic Sea (green), NW European shelf seas (mauve) and NW European coast (orange). (Figure 8.1 from chapter 8)

9.1.3. Bathymetry

Some of the features in the spatial chlorophyll patterns are consistent with the bathymetry of the North Sea, namely the Dogger Bank, the Southern North Sea and the Norwegian Trench. Those features are best depictured by the 35 m (Dogger Bank and Southern North Sea) and the 250m depth contour (Norwegian Trench, Figure 9.4). The deep Atlantic is also separated by the 250 m depth contour.
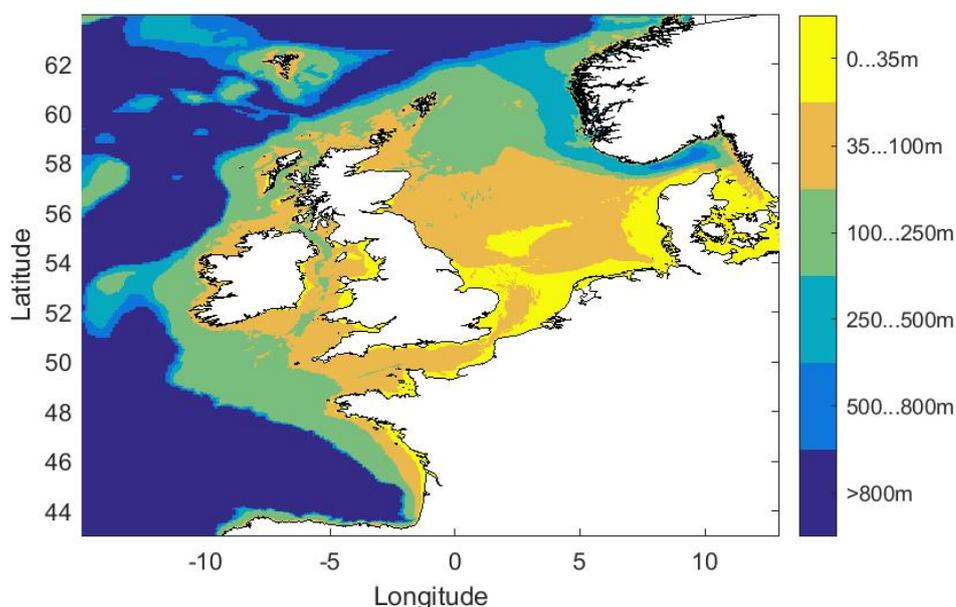

Figure 9.4. Depth contours

9.1.4. _Salinity_

The salinity in the top layer of the model was mostly higher than 34 psu across the model domain. A threshold of 32 psu was chosen to best approximate the coastal water type (red in Figure 9.2, orange in Figure 9.3 and light blue in Figure 9.5).
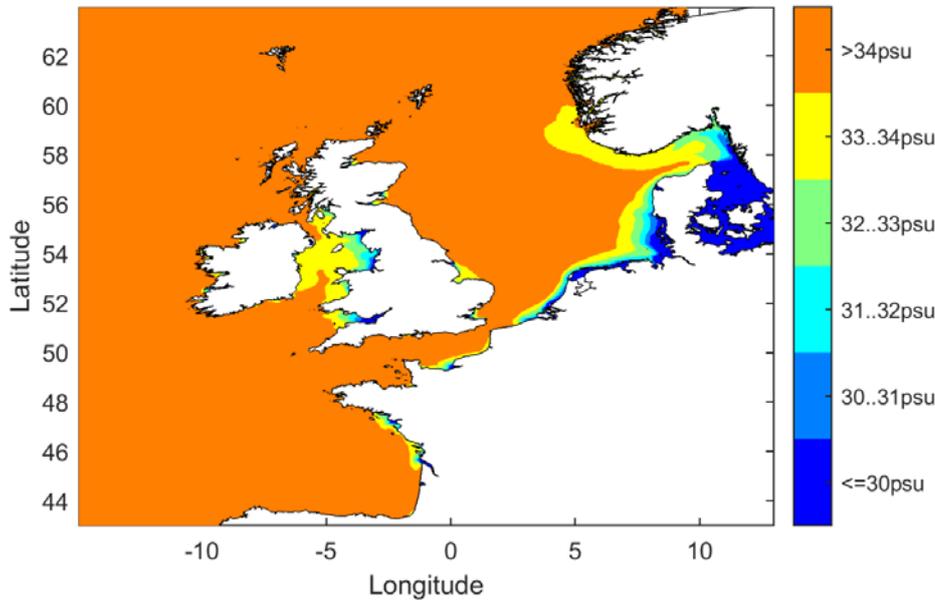

Figure 9.5. Salinity contours of the modelled salinity in the top layer.

9.1.5. _Stratification_

Stratification was determined based on the modelled monthly averaged density difference between the top and bottom layer in the model. A grid cell was classified as stratified when the density difference was larger than 0.75 kg/m^3 similar to van Leeuwen et al. (2015) (see red dotted line in Figures 1-4 in Annex C). Areas that are almost always stratified are the Norwegian trench and the waters off the French Atlantic coast. The Northern North Sea is only stratified in summer and mixed in winter. The shallow areas of the Dogger Bank and the Southern North Sea are always mixed. The Atlantic Ocean seems to be never stratified in our model. In reality the ocean is permanently stratified. The reason why our model cannot reproduce this stratification in these deep ocean waters is that it has 20 equidistant layers, which leads to a surface layer thickness over 100 meters along the oceanic boundaries. The pycnocline depth in the ocean is at less than 100 m from the surface so it is included within the upper layer of the model.

To differentiate the type of stratification (permanently, seasonally or intermittently) the number of consecutive months, in which grid cells are either mixed or stratified is calculated (Figure 9.6, Figure 9.7). The areas are then classified as shown in Table 9.1 and Figure 9.8.

Table 9.1. Stratification classes

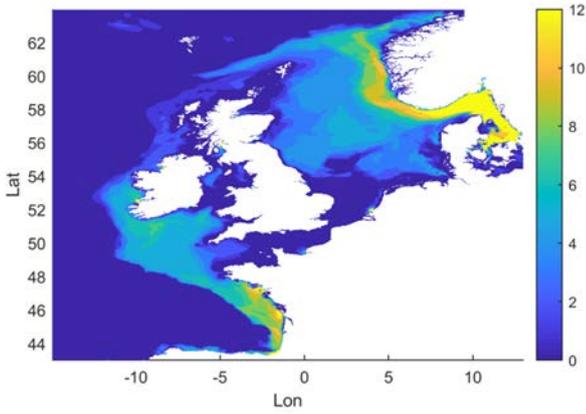| Stratification class	Number of consecutive months stratified	Number of consecutive months mixed
Permanently stratified	≥ 8	<8
Seasonally stratified	≥ 3 and <8	≥ 6
Intermittently stratified	≥ 1 and <3	≥ 6
Permanently mixed	= 0	≥ 10

Figure 9.6: Number of consecutive months, in which grid cells are stratified.
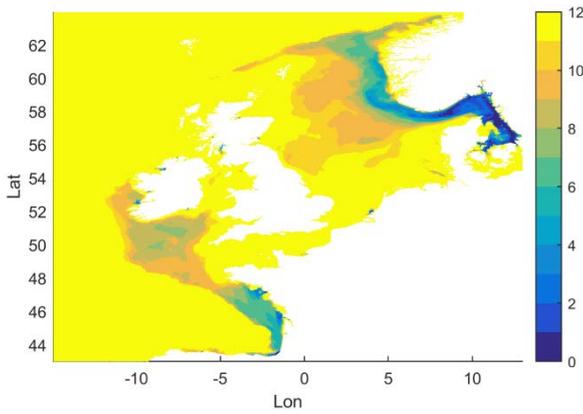

Figure 9.7: Number of consecutive months, in which grid cells are mixed.
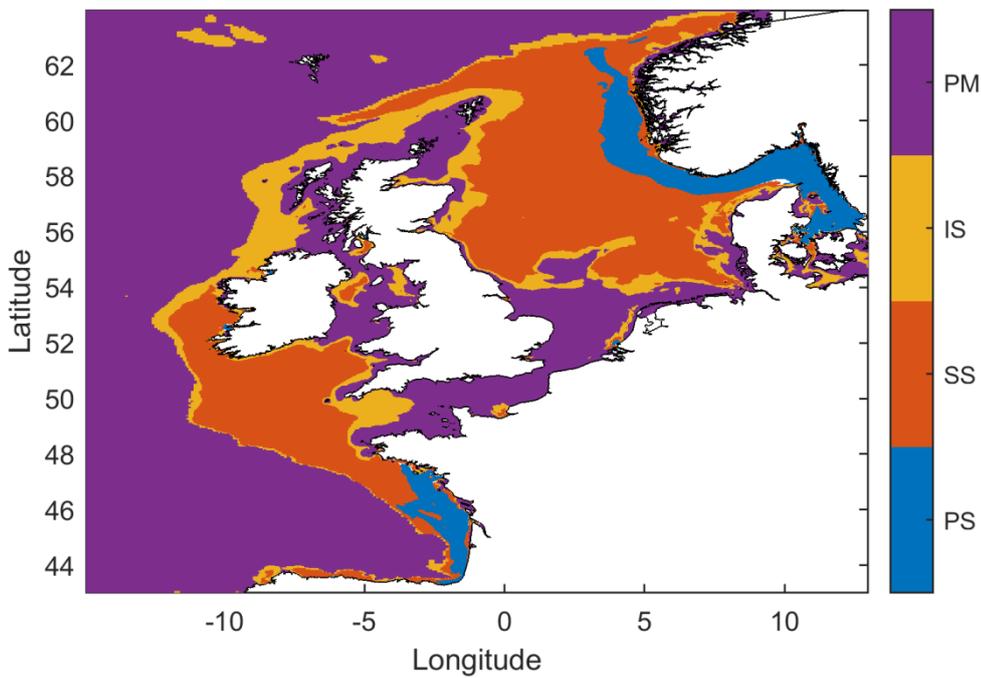

Figure 9.8. Stratification classes: Permanently stratified (PS), seasonally stratified (SS), and intermittently stratified (IS) or permanently mixed (PM).

9.2. Proposed new definition of assessment areas

Based on depth, salinity and stratification new assessment areas were defined (Figure 9.9). The criteria for these three factors (i.e. the values to subdivide areas) were set at specific values that ensured a good match with the areas defined by chlorophyll satellite observations as shown in Figure 9.2. For example, areas were subdivided at salinity of 32 and at depth of 35 m. Additionally, geographical areas were distinguished, such as the Channel, Irish Sea and Kattegat.

Figure 9.9 shows how the cross-border newly proposed assessment areas compare with national boundaries. Comparing the newly proposed assessment areas with the current assessment areas (Figure 9.10) the main difference is that different water types in the North Sea stand out clearly in the new approach and different water types (for example 'coastal waters' or 'Dogger Bank') are defined in the same way across national borders and form a coherent sub-area.
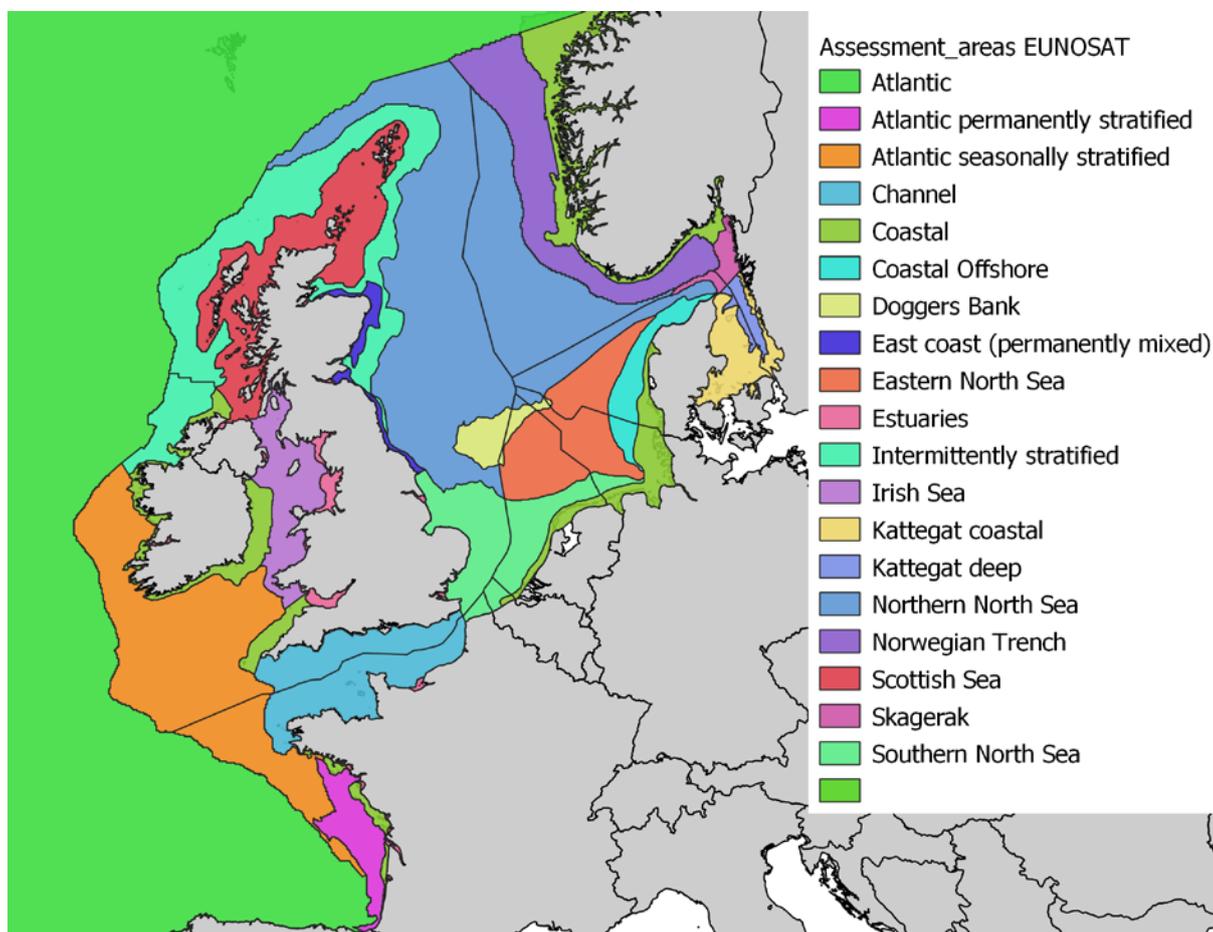

Figure 9.9: Proposal for ecologically relevant assessment areas based on duration of stratification, mean surface salinity and depth, with borders between EEZ's.
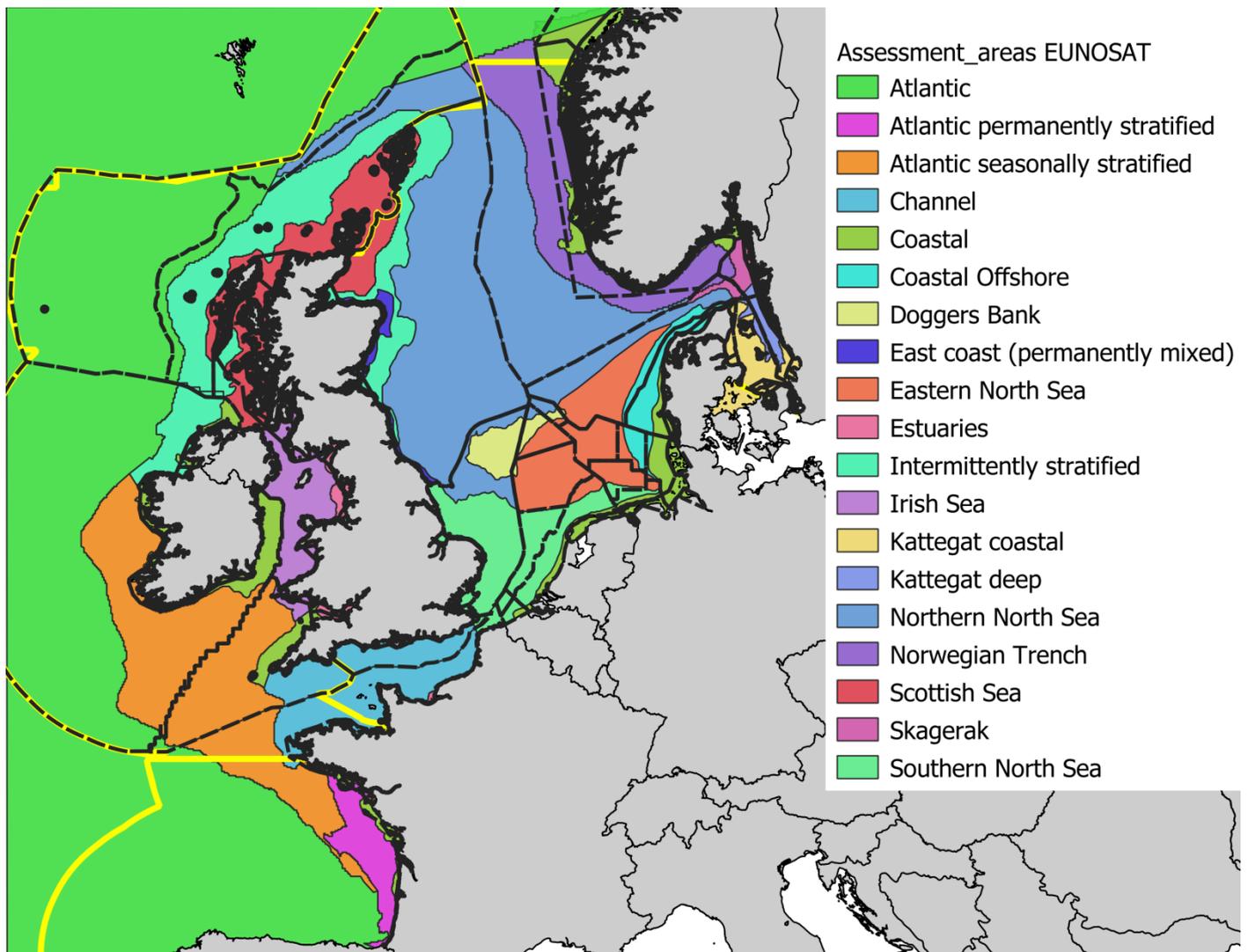
Figure 9.10: Comparison of 'new' assessment areas (as in Figure 9.9) with current COMP assessment areas (indicated with black broken lines). Borders between MSFD sub-regions are shown by yellow lines.

9.3. Discussion

The newly defined areas should be an improvement in the sense that the areas are ecologically relevant, meaning that within the areas there is some coherence in the environmental conditions determining phytoplankton growth, and consequently the areas can cross national borders. The assessment areas are defined based on physical characteristics that are relevant for phytoplankton growth. For the application in assessments, smaller units can be nested within the larger assessment areas, such as national boundaries and WFD assessment areas along the coast.

The results that are presented here are based on a modelling approach that describes a number of relevant physical conditions. The results can be considered a follow-up to work by Van Leeuwen et al. (2015), using a hydrodynamic model with a much higher spatial resolution and using information from satellite observations. Further validation and improvements of the underlying model(s) and ecosystem understanding may further improve the definition of assessment areas. Particularly for areas outside the North Sea (e.g. Celtic Seas, Bay of Biscay) further work might be needed to improve the model results. The same applies to the area of Skagerrak and particularly Kattegat, where due to the exchange with the Baltic Sea, hydrodynamic conditions are complex and have not been captured fully by the current model application.

The use of fixed borders to describe ecologically relevant assessment areas, in a dynamic system where physical conditions may change in space and time across those borders, may need some consideration. However, this is a question that also applies to the current assessment areas.

In the JMP-EUNOSAT project threshold values are defined at a 1 x 1 km grid (see chapter 7). So threshold values vary within the assessment areas. For some areas, particularly offshore, that do not have strong gradients in nutrient and algae concentrations within the area, uniform threshold values may be considered for practical reasons.

The new definition of assessment levels may have implications for monitoring efforts. Ideally each assessment area should have besides satellite data also in-situ monitoring data at a sufficient temporal resolution. Countries should check if this is still true for the newly proposed definition of assessment areas. The new assessment areas also offer opportunities for more efficient monitoring in cross-border assessment areas, such as the Dogger Bank. If concentrations and assessment levels are similar throughout the area, one shared monitoring location may be sufficient and it is not needed for every country to perform monitoring there.

In addition to the approach, presented here, defining assessment areas based on ecological characteristics, the assessment areas need to be evaluated based on:

- a risk-based approach: which areas have the highest sensitivity for eutrophication, which areas have the highest pressures (including transboundary effects)
- a management perspective: which choice of areas provides the right information on the progress towards good status. This includes finding a balance between areas that are not too large to mask the effects of elevated nutrient loads, and not too small to make it unfeasible for monitoring and reporting. This is particularly relevant for national subdivisions of the proposed cross-border assessment areas.

The assessment areas differ from the assessment areas that have been used by ICG COBAM for the description of pelagic habitats; the latter were based on work by CEFAS (Van Leeuwen et al. 2015). Coherence in assessments can be further enhanced if different OSPAR working groups / MSFD descriptors use the same definition of assessment areas. It is reasonable to expect that the physical and ecological properties that control spatial variability in algae dynamics also have a large impact on other parts of the ecosystem such as biodiversity and benthic life.

10. Conclusions and next steps

10.1. Conclusions

In the JMP-EUNOSAT project we developed a new method for coherent monitoring and assessment of eutrophication in the North Sea. We defined new threshold values and assessment areas and found new ways to use satellite data and Ferrybox data in the assessments and model validation. With this approach we could address all issues related to eutrophication assessments and we could demonstrate how a new coherent approach could work out in practice. The results showed that model results provide a useful basis for coherent threshold values and assessment areas. In combination with satellite data at the same high spatial resolution as the model more detailed assessment of chlorophyll concentrations can be done, that give more insight in underlying processes. We also identified Ferrybox monitoring as a promising additional resource of information at a higher spatial and temporal resolution as in-situ observations. The Netherlands and Norway collaborated on the implementation of an additional Ferrybox trajectory across the North Sea to better answer information needs on eutrophication in the North Sea (see Activity 3 report, Markager et al, 2019).

The two-year period to carry out the project and the budget put constraints on the amount of work that could be carried out with consequences for the level of detail that could be achieved. In order to fully address the need from OSPAR for a coherent eutrophication assessment it was decided to use the project to go through the complete workflow for the development of threshold values, accepting that inevitably not all steps in this workflow could be carried out at a high level of detail. It was recognized at the start of the project already, that there might be a need for follow-up at a more detailed level for some open issues.

By going through the whole monitoring and assessment process together we identified a range of knowledge gaps and sources of incoherence between countries. These have been discussed between project partners and invited guests at the final project meeting, resulting in an overview of remaining questions and next steps on a range of subjects.

10.2. Next steps

10.2.1. Definition of nutrient inputs to the sea

Riverine loads of nutrients were estimated using the E-HYPE model. With this model, current nutrient loads were estimated as well as nutrient loads in the reference year 1900. For validation of the model, the estimates of current N and P loads were compared to data on riverine loads from the OSPAR RID database and with data from a database of riverine nutrient loads maintained by OSPAR ICG-EMO. Those two validation data sets differ in temporal resolution (annual data versus daily estimates) and in the number of riverine sources that are included.

The validation of the E-HYPE data showed that there are, in some cases, large inconsistencies. A correction factor was applied when using the E-HYPE data in this project. However, for future applications improvements to the E-HYPE model should be made.

E-HYPE was used to estimate riverine nutrient loads for the reference year 1900, using a reconstruction of population density and land use. Water management was also very different in 1900 compared to the present situation. Many streams were not yet canalised and many dams have been built in rivers since 1900. These are likely to have a large impact on the retention of nutrients before reaching the sea (for example: Nordemann Jensen, 2017). These effects have not yet been taken into account in the present E-HYPE application for the JMP-EUNOSAT project. It is not clear yet if OSPAR countries would like this effect to be taken into account.

There have been several other studies providing estimates of historic nutrient loads (see e.g. Billen and Garnier 1997, Gadegast et al. 2014, Desmit et al. 2015), and Germany is currently carrying out an update of estimates using the MONERIS model. It will be worthwhile to make a comparison of E-HYPE model set-up and results with the other nutrient load models.

10.2.2. *Estimates of nutrient and chlorophyll concentrations at sea*

Nutrient concentrations at sea were estimated from riverine nutrient loads estimated with the E-HYPE model, and natural background levels in oceanic water. The concentrations in Atlantic Ocean water entering the North Sea were taken from a global database. Nutrient concentrations at sea were calculated with a 3D physical transport model, assuming conservative behaviour (no uptake or release processes). Other sources and sinks of nutrients (like atmospheric deposition, sedimentation of organic material) were not included in the model. Overall, the calculated salinity and nutrient concentrations showed a spatial distribution that fitted the in-situ validation data, indicating that the transport model gave a reasonable approximation. However, in specific areas (e.g. offshore waters of the Southern Bight; German coastal waters, Skagerrak/Kattegat) there were deviations from in situ data, indicating that improvements are needed. Deviations in modelled salinity point at a need to improve modelling of transport processes and exchange with the Baltic Sea, underestimation of nutrient concentrations in combination with underestimated salinities (Southern Bight) suggest that nutrient sources are underestimated, maybe due to the exclusion of atmospheric deposition.

Winter mean nutrient concentrations were used as the basis to estimate growing season chlorophyll concentrations. Winter concentrations of nutrients can be considered a proxy for nutrient availability during the growing season, as the concentrations reflect the spatial distribution of freshwater influence in the North Sea. However, the relation between winter nutrients and growing season nutrients may differ between different parts of the North Sea due to spatial differences in the contribution of natural and anthropogenic sources of nutrients and turnover rates.

Growing season mean chlorophyll concentrations were estimated from winter mean nitrogen concentrations, using a linear relation that was derived from in-situ data. It is likely that the nutrient-chlorophyll relation is not the same in all parts of the North Sea, due to differences in other growth factors. The results of an application of box models using the BLOOM and ERSEM models to calculate phytoplankton concentrations were not yet reliable enough to be used. However, an approach with more dynamical modelling of phytoplankton concentrations would most likely be an improvement to the currently used regression estimate.

From the above mentioned issues it becomes clear, that for a more accurate description of nutrient and chlorophyll concentrations at sea, the use of a 3D coupled physical-biochemical model and the addition of atmospheric inputs and the addition of exchange processes with the sediment, are recommended.

10.2.3. *Validation data*

For the validation of the model, in situ data were used that were supplied by the project partners. The model application would benefit from a higher availability of validation data, with a better spatial coverage. In the final workshop several suggestions for sources of additional data were made, in particular for areas where currently validation data were scarce. In addition, more Ferrybox data are expected to be available soon. The definition of assessment areas was based on a number of criteria, including model descriptions of stratification. To improve the modelling of the occurrence and duration of stratification, data from vertical profiles are needed. For the validation of satellite data of primary production in the North Sea primary production observations in the North Sea are needed.

10.2.4. _Monitoring_

If assessments change, with the definition of new assessment areas and the use of FerryBox data and satellite data in addition to in situ data, there may be a need to rethink the current monitoring programs. This could include the repositioning of monitoring stations to match the new assessment areas. Other requirements for in situ monitoring and the application of FerryBox may also need to be reconsidered. Now requirements for high temporal and spatial resolution monitoring of algae are covered by satellite data, the main purposes for in-situ monitoring are: 1) validation of satellite data, 2) time series continuation for trend detection and 3) observation of additional variables that cannot be observed with satellites.

10.2.5. _Definition of threshold values_

In this project, threshold values were defined following the approach common to OSPAR and WFD: threshold values are set as an area-specific deviation from a reference condition. This deviation should not exceed 50% according to the OSPAR Common Procedure (OSPAR 2013).

Threshold values for nutrients and chlorophyll were defined in this project as 50% above the values calculated for the reference year 1900. Although the use of a deviation of 50% is a more or less common approach, it might be necessary to reconsider this 'fixed' deviation from reference. In some cases, areas may be more sensitive to impacts from eutrophication (e.g. stratified areas) and a smaller deviation might be necessary to prevent negative impacts like oxygen deficiency. On the other hand, some areas may be more robust (for example, due to strong mixing) and could have a larger deviation provided that there are no other reasons, such as transboundary effects, to set a limit to the threshold values for nutrients and chlorophyll. Setting such region-specific deviation percentages requires sufficient ecological understanding of the local and transboundary effects of eutrophication and may therefore need further research.

It also needs to be considered how a percentage deviation from reference should be applied. Here, the deviation was applied for the whole North Sea, without considering if an area is strongly influenced by human pressures (e.g. coastal waters subject to freshwater discharges) or an area is hardly impacted by human pressures (offshore waters with no freshwater influence). An alternative approach could be to define an acceptable deviation from reference for the pressures, for example for riverine nutrient loads (see e.g. Kerimoglu et al., 2018). In this approach, not only riverine loads but also atmospheric deposition which is a significant source in offshore waters (Troost et al. 2013) needs to be taken into account.

For the water bodies under the WFD, threshold values have already been set for both nutrients and chlorophyll. For the rivers, threshold values for nutrient concentrations are available. At least in some countries (e.g. Germany, Netherlands) these threshold values have been set taking into account the downstream/transboundary effects of riverine nutrient loads on the ecological status of coastal waters. Although concentration threshold values cannot be easily converted to threshold values for loads, due to the fact that concentrations and discharges are correlated in many water systems, it might be useful to use these WFD threshold values in the model runs for a comparison with the threshold values derived in this project.

10.2.6. _Definition of assessment areas_

In this project, a proposal for new assessment areas was developed. This is a first approach that requires further fine-tuning. It focusses on the larger cross-border areas with similar physical and ecological dynamics. These need to be further subdivided along national boundaries and within countries, following a nested approach. For example, current assessment areas defined for the Water Framework Directive (WFD) can be used for further detail in assessments of near-shore areas.

References

Alleman, J.E., (accessed June 2017) The Genesis and Evolution of Activated Sludge Technology. https://www.elmhurst.org/DocumentCenter/View/301 and http://home.eng.iastate.edu/~jea/w3-class/456/article/article-aswisconsin.html

Ardern, E. and Lockett, W.T., "Experiments on the Oxidation of Sewage Without the Aid of Filters," Journal of the Society of Chemical Industry, Vol 33, No 10, pp 523 - 539, 1914 (presented at Manchester Section meeting on April 3, 1914).

Bartosova A., Donelly C., Strömqvist J., Capell R., and Tengdelous-Brunell J (2017). HYPE model for the Baltic Sea Basin. BONUS Soils2Sea Deliverable 5.1. Swedish Meteorological and Hydrological Institute, Norrköping, May 2017, www.Soils2Sea.eu

Berg P, Donnelly C, and Gustafsson D (2018). Near-real-time adjusted reanalysis forcing data for hydrology. Hydrology and Earth System Sciences 22(2):989-1000 DOI: 10.5194/hess-22-989-2018

Billen, G. and J. Garnier (1997). The Phison River plume: Coastal eutrophication in response to changes in land use and water management in the watershed. Aquatic Microbial Ecology 13: 3-17.

Blauw, A. N., Los, H. F., Bokhorst, M., & Erftemeijer, P. L. (2009). GEM: a generic ecological model for estuaries and coastal waters. Hydrobiologia, 618(1), 175.

Butenschön, M., Clark, J., Aldridge, J. N., Allen, J. I., Artioli, Y., Blackford, J., ... & Lessin, G. (2016). ERSEM 15.06: a generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels. Geoscientific Model Development, 9(4), 1293-1339.

Desmit, X., G. Lacroix, V. Dulière, C. Lancelot, N. Gypens, A. Ménesguen, B. Thouvenin, M. Dussauze, G. Billen, J. Garnier, V. Thieu, M. Silvestre, P. Passy, L. Lassaletta, G. Guittard, S. Théry, R. Neves, F. Campuzano, C. Garcia, L. Pinto, J. Sobrinho, M. Mateus and I.A. Kenov (2015). Ecosystem Models as Support to Eutrophication Management In the North Atlantic Ocean. RBINS, EMoSEM final report 31/8/2015, 174 pp.

Devlin, M. J., Barry, J., Mills, D. K., Gowen, R. J., Foden, J., Sivyer, D., & Tett, P. (2008). Relationships between suspended particulate material, light attenuation and Secchi depth in UK marine waters. Estuarine, Coastal and Shelf Science, 79(3), 429-439.

Donnelly, C., Andersson, J. C., & Arheimer, B. (2016). Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. Hydrological Sciences Journal, 61(2), 255-273.

Edwards, K. P., Barciela, R., & Butenschön, M. (2012). Validation of the NEMO-ERSEM operational ecosystem model for the North West European Continental Shelf. Ocean Science, 8(6), 983-1000.

Eilola, K., Meier, H. M., & Almroth, E. (2009). On the dynamics of oxygen, phosphorus and cyanobacteria in the Baltic Sea; A model study. Journal of Marine Systems, 75(1), 163-184.

Engardt M, Simpson D, Schwikowski M and Granat L(2017). Deposition of sulphur and nitrogen in Europe 1900–2050. Model calculations and comparison to historical observations. Tellus B: Chemical and Physical Meteorology Vol. 69 , Iss. 1,2017. DOI:10.1080/16000889.2017.1328945

Gadegast, M., U. Hirt and M. Venohr (2014). Changes in Waste Water Disposal for Central European River Catchments and Its Nutrient Impacts on Surface Waters for the Period 1878–1939. Water, Air, & Soil Pollution 225: 1914.

van Grinsven HJM, Bouwman L, Cassman KG, van Es HM, McCrackin ML, and Beusen AHW (2015). Losses of Ammonia and Nitrate from Agriculture and Their Effect on Nitrogen Recovery in the European Union and the United States between 1900 and 2050. J. Environ. Qual. doi:10.2134/jeq2014.03.0102

Hundecha Y, Arheimer B, Donnelly C , Pechlivanidis I (2016) A regional parameter estimation scheme for a pan-European multi-basin model. Journal of Hydrology: Regional Studies, 6, 90-111.
http://dx.doi.org/10.1016/j.ejrh.2016.04.002

Kerimoglu, O., F. Große, M. Kreus, H.-J. Lenhart and J.E.E. van Beusekom (2018). A model-based projection of historical state of a coastal ecosystem: Relevance of phytoplankton stoichiometry. Science of the Total Environment 639: 1311-1323.

Klein Goldewijk, K. , A. Beusen, M. de Vos and G. van Drecht (2011). The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years, Global Ecology and Biogeography20(1): 73-86. DOI: 10.1111/j.1466-8238.2010.00587.x.

Klein Goldewijk, K. , A. Beusen, and P. Janssen (2010). Long term dynamic modeling of global population and built-up area in a spatially explicit way, HYDE 3 .1. The Holocene20(4):565-573.
http://dx.doi.org/10.1177/0959683609356587 .

Knudsen L., Schnug E. (2016). Utilization of Phosphorus at Farm Level in Denmark. In: Schnug E., De Kok L. (eds) Phosphorus in Agriculture: 100 % Zero. Springer, Dordrecht. DOI: 10.1007/978-94-017-7612-7_11

Kyllingsbæk A (2005) Nutrient balances and nutrient surpluses in Danish agriculture 1979–2002: nitrogen, phosphorus, potassium DJF Report Nr. 116, Research Center Foulum, Danish Institute of Agricultural Sciences, p 100

van Leeuwen, S., P. Tett, D. Mills and J. van der Molen (2015). Stratified and nonstratified areas in the North Sea: Long-term variability and biological and policy implications. Journal of Geophysical Research: Oceans 120: 4670-4686.

Lenhart, H.-J., D.K. Mills, H. Baretta-Bekker, S.M. van Leeuwen, J.v. der Molen, J.W. Baretta, M. Blaas, X. Desmit, W. Kühn, G. Lacroix, H.J. Los, A. Ménesguen, R. Neves, R. Proctor, P. Ruardij, M.D. Skogen, A. Vanhoutte-Brunier, M.T. Villars and S.L. Wakelin (2010). Predicting the consequences of nutrient reduction on the eutrophication status of the North Sea. Journal of Marine Systems 81: 148-170.

Lindström, G., Pers, C.P., Rosberg, R., Strömqvist, J., Arheimer, B. (2010). Development and test of the HYPE (Hydrological Predictions for the Environment) model – A water quality model for different spatial scales. Hydrology Research, 41(3-4), 295-319.

Lofrano G and Brown J. (2010). Wastewater management through the ages: a history of mankind. Sci Total Environ. 2010 Oct 15;408(22):5254-64. doi: 10.1016/j.scitotenv.2010.07.062

Los, F. J., & Wijsman, J. W. M. (2007). Application of a validated primary production model (BLOOM) as a screening tool for marine, coastal and transitional waters. Journal of Marine Systems, 64(1), 201-215.

Los, F. J., Troost, T. A., & Van Beek, J. K. L. (2014). Finding the optimal reduction to meet all targets—applying Linear Programming with a nutrient tracer model of the North Sea. Journal of Marine Systems, 131, 91-101.

Meier, H. M., Andersson, H. C., Arheimer, B., Blenckner, T., Chubarenko, B., Donnelly, C., ... & Höglund, A. (2012). Comparing reconstructed past variations and future projections of the Baltic Sea ecosystem—first results from multi-model ensemble simulations. Environmental Research Letters, 7(3), 034005.

Morris P and Morris P, "From Fertile Minds" (review) American Scientist, 2001Nordemann Jensen, P. (ed) (2017). Estimation of nitrogen concentrations from root zone to marine areas in the year 1900. Scientific report from DCE: Danish Centre for Environment and Energy 214. Aarhus University and DCE

OSPAR (2017a). Intermediate Assessment 2017. https://oap.ospar.org/en/ospar-assessments/intermediate-assessment-2017/

OSPAR (2017b). Third Integrated Report on the Eutrophication Status of the OSPAR Maritime Area.   London, OSPAR Commission, Publication No 694/2017, 164 pp.

Radach, G., & Lenhart, H. J. (1995). Nutrient dynamics in the North Sea: fluxes and budgets in the water column derived from ERSEM. Netherlands Journal of Sea Research, 33(3-4), 301-335.

Smil V. (2000). PHOSPHORUS IN THE ENVIRONMENT: Natural Flows and Human Interferences. Annu. Rev. Energy Environ. 2000. 25:53–88

Schmid T. (2000). Food consumption and nitrogen flux. The impact of lifestyle in a long-term perspective Paper presented at the workshop "How do households use their resources?" Department of Technology and Social Change, Linköping University, April 2000.

Smith, A. F., Fryer, R. J., Webster, L., Berx, B., Taylor, A., Walsham, P., & Turrell, W. R. (2014). Setting background nutrient levels for coastal waters with oceanic influences. Estuarine, Coastal and Shelf Science, 145, 69-79.

UWWTD, 2014. Urban Waste Water Treatment Directive Database. Accesses Feburary 2016. http://www.eea.europa.eu/data-and-maps/data/waterbase-uwwtd-urban-waste-water-treatment-directive-4

Vinther (2012) Oversigt over Landsforsøgene (annually) Videncentret for Landbrug. Planter & Miljø, Agro Food Park 15, Dk-8500 Aarhus N, Denmark

Vries, I. de, Los, H., Jansen, R., Cramer, S., & Van der Tol, M. (1993). Risico-analyse eutrofiëring Noordzee (Vol. 93). DGW-report.

Van der Zande, D., Lavigne, H., Blauw, A., Prins, T., Desmit, X., Eleveld, M., Gohin, F., Pardo, S, Tilstone, G., Cardoso Dos Santos, J. (2019).   Coherence in assessment framework of chlorophyll a and nutrients as part of the EU project 'Joint monitoring programme of the eutrophication of the North Sea with satellite data' (Ref: DG ENV/MSFD Second Cycle/2016). Activity 2 Report.

Annex A: Results of process-based models

A.1 BLOOM model

The BLOOM model is a phytoplankton model that is incorporated in the Generic Ecological Model (GEM) by Deltares. This model has been used to estimate natural reference concentrations of nutrients and chlorophyll in Dutch coastal waters (de Vries et al., 1993). In a stand-alone box model version it can be used to estimate seasonal dynamics of chlorophyll-a, based on observed total bioavailable nitrogen, phosphorus and silicate and underwater light climate (Los and Wijsman, 2007). The stand-alone version of the model is used in the present study to estimate seasonal dynamics of chlorophyll-a.

The model has first been validated for the monitoring locations in the JMP-EUNOSAT dataset. We did not calibrate the model coefficients. We have approximated model inputs from available data in the following way:

- Total available nitrogen and phosphorus concentrations have been approximated by the observed winter mean DIN and DIP concentrations times a monthly scaling factor to represent that river inputs in coastal waters vary with the season. The scaling factor was calculated for each location and nutrient separately based on the results of the nutrient transport model. It was calculated for each month as the monthly mean total nutrient concentration divided by the winter mean (January and February).
- Total available silicate concentrations were estimated as constant yearly mean values based on observed winter mean salinity, assuming dilution of freshwater with silicate concentration of 10 gSi/m3 with saltwater with silicate concentration 0.013 gSi/m3. The freshwater concentration of 10 gSi/m3 is commonly used by Deltares as default for waters with missing data on silicate. The boundary concentration of 0.013 gSi/m3 is taken from the Deltares ecological model for the North Sea, which in turn was based on available observation data in offshore areas.
- Kd_{PAR} is approximated in the same way as described above for the linear regression model.
- Water depth has been taken from the nutrient transport model at the monitoring location and is applied as a constant value representing the mean water depth over the simulation time.
- Daylength and solar irradiance have been approximated from existing time series for Dutch coastal waters, in the absence of local data for the monitoring locations.

After the validation for monitoring locations the model has been run for the model segments from the nutrient transport model to create a map of chlorophyll-a concentrations. To limit the simulation time (to 3 weeks) we only simulated model segments with a water depth of less than 50 m. At larger water depths seasonal stratification plays an important role and that is not taken into account in the model. Therefore, we don't expect the model results in those waters to be reliable anyway.

The inputs for the model runs on the model segments were approximated in the same way as for the runs for the monitoring locations, except for total available nitrogen and phosphorus. Since observations of DIN and DIP were lacking for most model segments we used the monthly mean surface concentrations of total nitrogen and total phosphorus from the nutrient transport model. These were converted to total available nitrogen and phosphorus by multiplication with 0.57 and 0.8 respectively (from the regressions shown in Figure 5.8).

The growing season mean chlorophyll-a concentrations estimated by the BLOOM model correlated less well to observations than the linear regression model (R^2 = 0.24, Figure A.1). On the one hand this is surprising, since BLOOM takes into account several processes that can explain spatial variability in chlorophyll to nitrogen ratios as visualized in Figure 6.6. BLOOM adjusts the chlorophyll to carbon ratio of the algae under turbid conditions and it

takes into account possible effects of phosphorus limitation. On the other hand BLOOM (as other more complicated models) requires input on many more variables than the simple DIN-based linear regression model. All these inputs need to be estimated, which introduces a range of uncertainty to the model results. For some input variables, such as local solar irradiance and day length, spatially variable information was not available so we have used uniform values for the whole model area, based on Dutch data. In the 0D box model application applied in this study, stratification effects could not be taken into account.
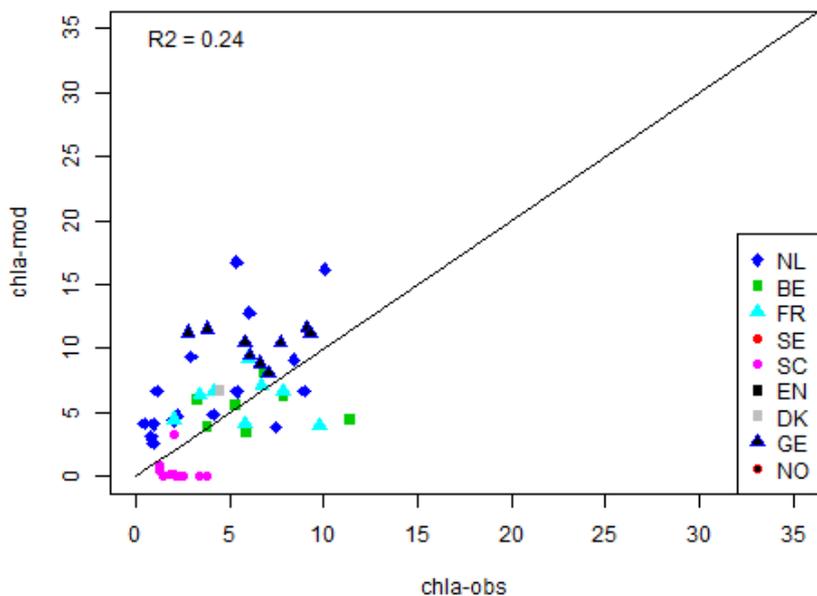
Figure A.1: validation results of the BLOOM model for monitoring locations in the JMP-EUNOSAT database (subset matching validation criteria in section 5.3).

Despite the low correlation between the BLOOM model results and observed growing season mean chlorophyll concentrations at the monitoring locations in the JMP-EUNOSAT dataset, we have applied the model to estimate a map of growing season mean chlorophyll concentrations. To this end a 0D box model was run for each grid cell from the nutrient transport model (see grid in Figure 5.1). Since this took a lot of simulation time and we don't expect the results to make sense in deep stratified areas we have only run the model for grid cells with a water column depth of less than 50 m. Figure A.2 shows the resulting map of estimated growing season mean chlorophyll concentrations from this model approach. The simulated chlorophyll concentrations are reasonable in nearshore waters, but offshore concentrations are dramatically overestimated. This may be due to the lack of vertical differences in our 0D modelling approach. In reality, nutrients and phytoplankton sink, resulting in relatively low concentrations near the surface. In stratified areas this effect is very clear, but it may also play a role in waters that do not stratify in summer, such as offshore waters in the southern Bight of the North Sea. Regular 3D applications give results that look more similar to the satellite data (Figure A.3).
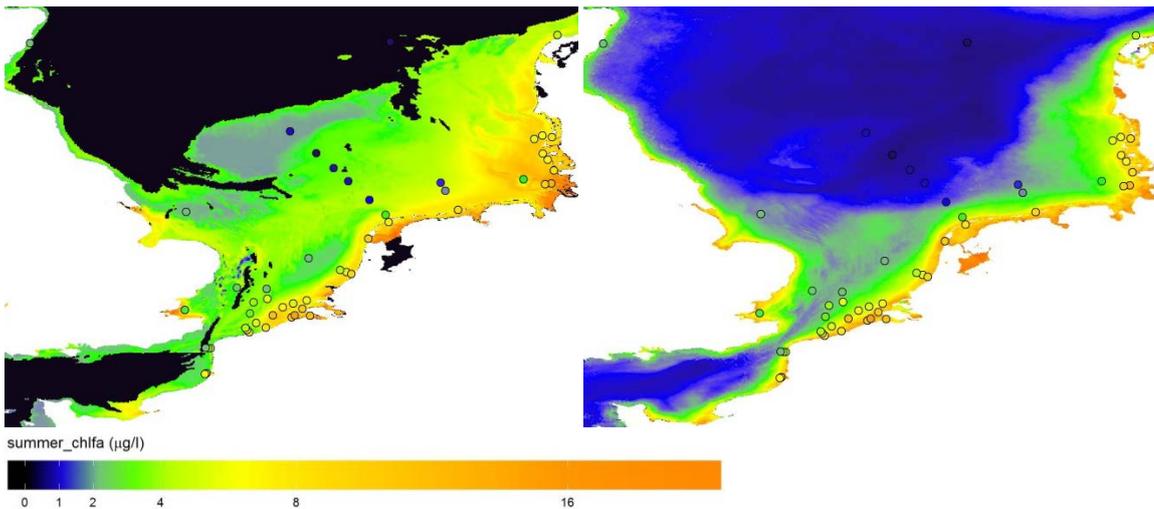
Figure A.2: BLOOM model estimates of growing season mean chlorophyll-a concentrations for 2009 – 2014 (left panel, background colour) with JMP EUNOSAT satellite data for 2005 – 2010 (right panel, background colour). Estimates based on in-situ data for 2003 – 2008 are shown as circles. Black areas have not been simulated since they are deeper than 50 m.
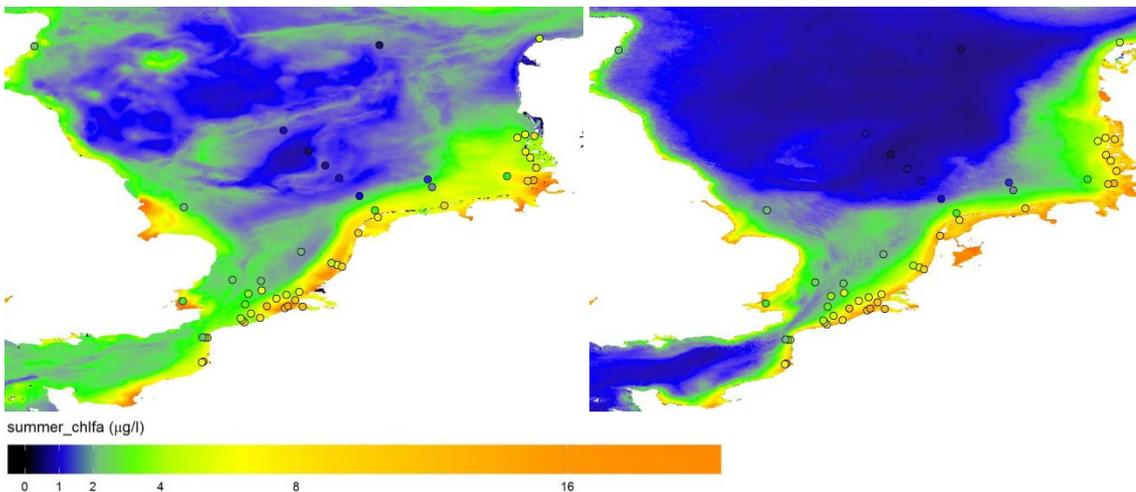

Figure A.3: Preliminary model results of the DFLOW-FM 3D BLOOM implementation: estimates of growing season mean chlorophyll-a concentrations for one year (left panel, background colour) with JMP EUNOSAT satellite data for 2005 – 2010 (right panel, background colour). Estimates based on in-situ data for 2003 – 2008 are shown as circles.

A.2   ERSEM model

We downloaded an ERSEM column model from: https://www.pml.ac.uk/Modelling_at_PML/Models/ERSEM. This model represents a 1DV column model with 50 layers that has been implemented and calibrated for the long term monitoring location L4, off the coast of Plymouth (Butenschön et al., 2016). The default model, as downloaded, produces chlorophyll-a and total nitrogen concentrations during the course of a year as shown in Figure A.4. During the onset of stratification a spring bloom develops in the upper mixed layer. Upon sinking of the organic matter after the spring bloom there is a draw-down of total nitrogen in the upper mixed layer and an accumulation of nitrogen below the pycnocline. After the break-down of stratification in fall water column averaged total nitrogen is lower than before the spring bloom.
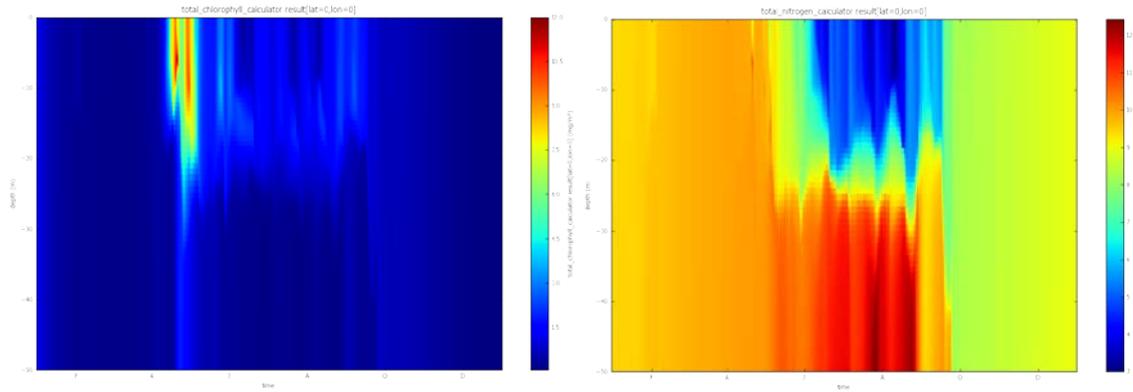
Figure A.4: Model results of the downloaded model for monitoring location L4 for chlorophyll-a (left) and total nitrogen (right). Colours represent concentrations, the y-axis represents depth and the x-axis time, from January to December.

We changed a few things to this basic set-up of the model to make the model representative for different monitoring locations, similar to the BLOOM box model. Also, we tried to make the model more similar to a 0D box model:

- The model implementation did not allow for reducing the number of layers in the model from 50 to 1 (personal communication J. Bruggeman, PML). Therefore we changed the salinity and temperature profiles in the model input to vertically uniform profiles, by using the surface temperature and salinity from the original simulation throughout the water column.
- We tried to create mass conservation and vertically uniform concentrations in the water column by reducing all settling velocities of organic matter to zero.
- We replaced the initial nitrate and phosphate concentrations of the simulation with the observed winter mean DIN and DIP concentration.
- The Si concentration was estimated from the winter mean salinity in the same way as described above for the BLOOM box model.
- We replaced the absorption coefficient of clear water by Kd_{PAR} approximated in the same way as described above for the linear regression model.
- Water depth has been taken from the nutrient transport model at the monitoring location and is applied as a constant value representing the mean water depth over the simulation time.

Figure A.5 shows the ERSEM results for monitoring location Noordwijk 20 km as an example. It shows a sharp bloom event in June, whereas observations (data not shown) show prolonged blooms from March through June. This suggests that light availability is limiting earlier blooms in the model. However the background extinction and water depth in the model correspond to local observations. Despite our efforts to create mass conservation in the water column, by reducing all exchanges with the sediment to zero, total nitrogen concentrations still show seasonal variability.

Figure A.6 shows the model fit for all monitoring locations in the JMP-EUNOSAT dataset. The estimated growing season mean chlorophyll concentrations are similar for all locations and are generally too low compared to field observations. There is no correlation between modelled and observed chlorophyll concentrations. Unfortunately, we don't understand the set-up of the model well enough to know why this is happening and how it can be changed.
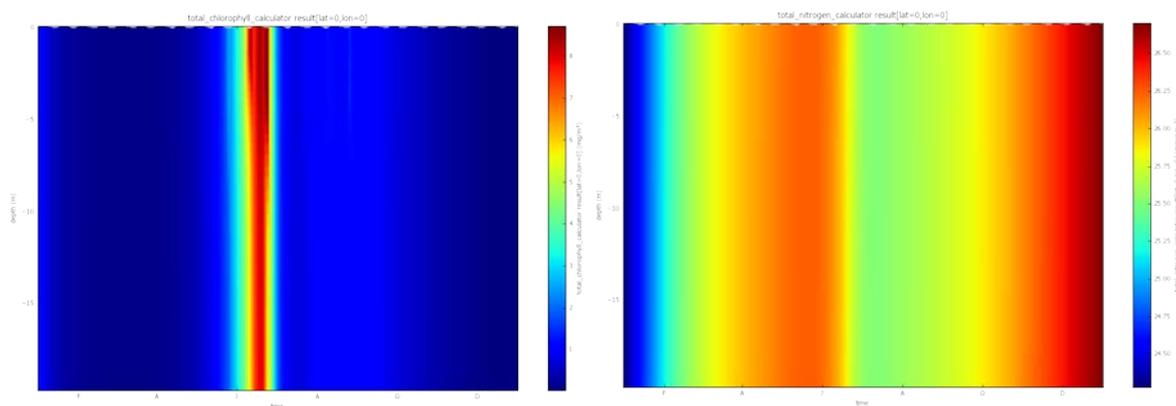
Figure A.5: Model results of the downloaded model for a new monitoring location (Noordwijk 20 km) for chlorophyll-a (left) and total nitrogen (right). Colours represent concentrations, the y-axis represents depth and the x-axis time, from January to December.
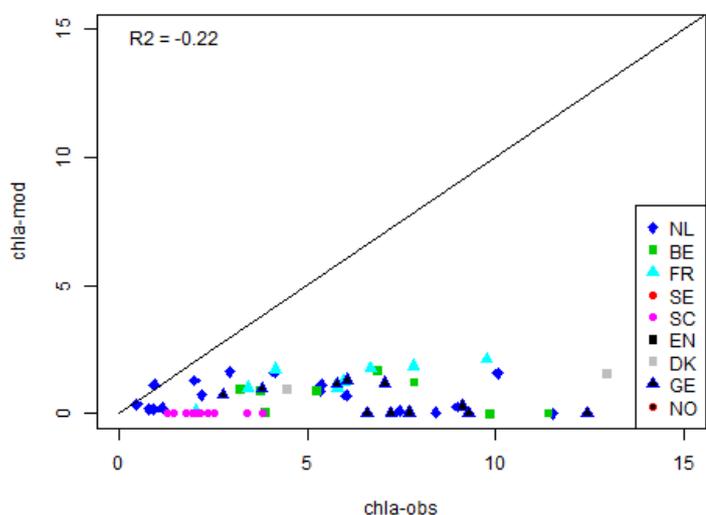

Figure A.6: validation results of the ERSEM model for monitoring locations in the JMP-EUNOSAT database (subset matching validation criteria in section 5.3).

A.3 Discussion on deterministic chlorophyll modelling

The linear regression model gives the best approximation of observed growing season mean chlorophyll concentrations. The process-based box model approaches should in theory give better results, but within this project is was not possible to implement them in a satisfactory way. We had expected that box models that had been used before would not need additional calibration and would give reasonable results by just replacing the inputs on environmental conditions by observed values. It turned out that there are many ways to estimate input values and finding the best approach is a labour-intensive task. Besides input on environmental conditions process based also need parameter estimations for a (often large) range of parameters. These parameter values have often been calibrated to a specific dataset for a specific region. Since there are so many parameters involved, the same model output and validation result can often be obtained by different combinations of parameter values. Therefore, calibrating a model for a specific area does not always yield a set of parameter values that is generically applicable for other areas as well. In this study we have taken two models that have been calibrated for a specific area and have applied them for a wide range of environmental conditions across the North Sea. This could also (partly) explain why the process-based models perform less good as the linear regression model, that has been fitted particularly for the validation data set.

Additionally, working with a new model without background knowledge on the technical model set-up, such as ERSEM in our case, bears risks of changing the model in the wrong way. We changed as little as possible, but the results suggest that we may still have violated the consistency of the model set-up, leading to erroneous results.

Annex B - Primary production in the North Sea, and Celtic Sea OSPAR regions: climatology, thresholds, time series.

B.1. Introduction

Phytoplankton are responsible for about half of the global primary production (PP) [Longhurst et al., 1995] and in the North Atlantic, represent a significant sink for carbon dioxide [Takahashi et al., 2009]. The magnitude of the spring bloom in the North Atlantic is one of the largest in the global ocean and is controlled by a combination of physical forcing and biological factors [Koertzinger et al., 2008] and its timing is strongly driven by physical forcing [Henson et al., 2009]. Primary production is driven by light, nutrients and temperature and therefore any anthropogenic perturbations in these parameters (such as climate change and eutrophication), are reflected in the magnitude and timing of daily production, which may ultimately significantly affect integrated annual rates. Phytoplankton production reflects several environmental pressures (e. g. hydrological changes, contaminants, nutrient inputs or climate changes), which cannot necessarily be detected through changes in Chl-a. This indicator is highly sensitive and can be used as an early warning indicator for direct pressure on food webs. It is an indicator of potential matter flow needed by higher trophic levels to produce biomass. PP can increase as a result of increases in nutrient loads resulting from eutrophication or due to an increase in temperature as a result of climate change. Though increases in PP are often associated with increases in fish stocks, this may not always be the case. It can also lead to an accumulation of organic matter [Nixon, 1995] thus promoting bacterial activity, which can then lead to undesirable changes in food webs, water quality, and aquatic chemistry [Cloern, 2001{Rabalais, 2004 #2139; Rabalais et al., 2004; Rabalais et al., 2009]}.

Marine eutrophication zone increased globally from the 1950's to the 1970's (McIntyre, 1995) in a number of coastal ecosystems [Lehmusluoto and Pesonen, 1973; Ryther and Dunstan, 1971{Kemp, 1997 #2141]}. Some northern European coastal regions recorded widespread increase in phytoplankton biomass and production, which led to a disruption in the balance between the production and the turnover of organic matter [Cloern, 2001]. In addition, The North Sea experienced abrupt increases in SST and pelagic primary production in the late 1980s (Reid and Edwards 2001, Reid et al. 2001, Beaugrand 2004), with cold winters occurring both before (in the late 1970s; Reid and Edwards 2001) and, to a lesser degree, after this period in the mid-1990s [Kroncke et al., 2013].Through a concerted effort to curb eutrophication in the 1980's, to reduce nutrient run-off into the coastal zone, some ecosystems experienced a reduction in pelagic primary production due to nitrogen (N) : phosphorus (P) limitation [Meyer et al., 2018].

 With the advent of ocean colour and the increase in the number, accuracy, resolution of Ocean Colour sensors, satellite data are being increasingly used to track eutrophication events and other anthropogenic disturbances in the marine environment which undoubtedly will aid the monitoring of coastal and shelf environments under the European Union Water Framework Directive (WFD) and Marine Strategy Framework Directive (MSFD) [Cristina et al., 2015; Novoa et al., 2012].The objective of the EU Marine Strategy Framework Directive (MSFD) is to enable the sustainable use of marine goods and services and to ensure that the marine environment is safeguarded for the use of future generations (European Commission, 2008). The MSFD has established comprehensive guidelines for Member States to develop and implement cost effective measures to protect and preserve the marine environment in order to achieve "good environmental status" (GES) by 2020. GES is evaluated based on 11 key Descriptor parameters (European Commission, 2008). Monitoring these key descriptors over large maritime areas poses a logistical and economic challenge for ship borne surveys. Some of these Descriptors are available from satellite data

which enhances the spatial and temporal coverage over which they can be reported. Using these data, a concerted effort on integrating them with in situ data to re-determine baselines and indicator threshold values is necessary (Garcia-Garcia et al. 2019). Baselines in this context refer to present characteristic levels and threshold values refer to 50% above characteristic levels for the period around 1900.

Under MSFD Descriptor 4 – Food webs there is a need to develop the secondary criterion D4C4 'Productivity of the trophic guild is not adversely affected due to anthropogenic pressures'. The use of primary production as an indicator has not yet been established in OSPAR but can contribute to the development of D4C4. In relation to the development of baselines for such an indicator long time series of primary production in the Western English Channel can be used to determine baselines for the NE Atlantic. Based on the work of [Boalch, 1987; Boalch et al., 1978], the annual PP should not exceed 300 gC m^{-2} yr^{-1} with daily values under 2-3 gC m^{-2} d^{-1} during phytoplankton blooms. More recently, [Cloern et al., 2014] compiled 1148 values of annual primary production values from 131 ecosystems, based on monthly carbon assimilation or oxygen production incubation assays. The median and mean values were 185 and 252 gC m^{-2} yr^{-1}, respectively over a very large range (from 105 to 1890 gC m^{-2} yr^{-1}) and 10-fold within ecosystems and a 5-fold from year to year.

The objectives of this report for the JMP EUNOSAT project were to:

1.   Define regions of similarity in mean peak, timing and amplitude of primary production in the NE Atlantic using satellite data;

2.   To define baseline values in PP over the region using satellite;

3.   To assess the variability in PP over the region.

B.2. Material and Methods.

B.2.1. Study area.

The study area is the North East Atlantic encompassing the Celtic, Irish and North Seas from 60°N, 10°E to 48°N, 15°W.

B.2.2 Remote Sensing data.

Primary production: A wavelength resolving PP model [Morel, 1991] was implemented following Smyth et al. [2005] using daily 4 km CMEMS OC5 Chl-a and Pathfinder v2009 AVHRR SST data to generate mean monthly satellite maps of PP from 1997 to 2017 (Figure B.1). The estimates of satellite PP are accurate to 20% in the Atlantic Ocean [G Tilstone et al., 2009]. To define the upper limit in PP over the region, the percentile 90 (P90) was determined to capture the maximum variability in PP (Figure B.2). K-means clustering based on peak and timing of satellite PP was then used to identify regions of similarity. Eight regions were defined which broadly correspond to the principal hydrographic areas in the NE Atlantic (Figure B.3).
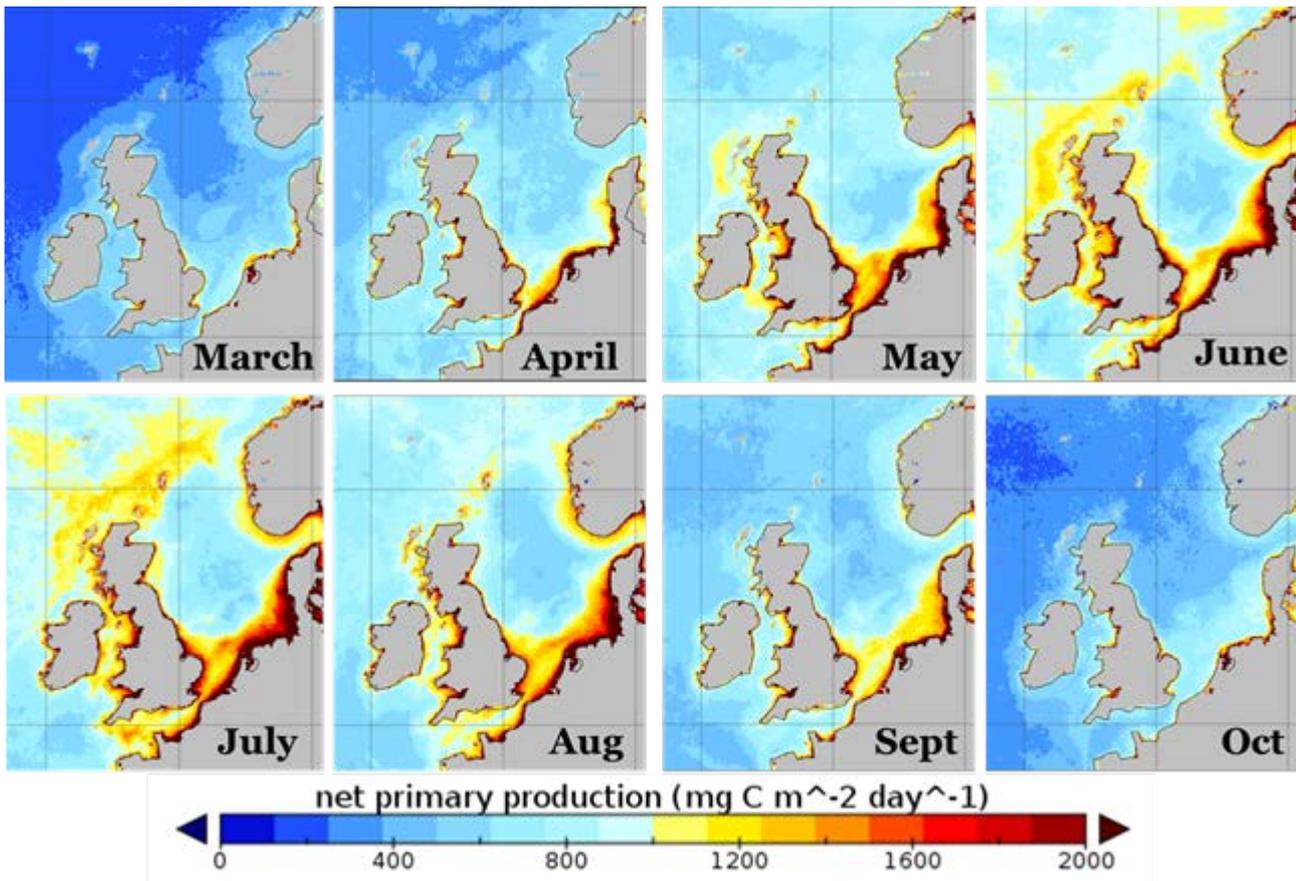
Figure B.1. Mean monthly primary production from Copernicus Marine Environment Monitoring Service (CMEMS) data (1997-2017).

B.3. Results.

Mean monthly PP from March to October illustrate the migration of the spring bloom from South to North over the region (Figure B.1). In the Celtic Sea and English Channel PP reached maximum values of ~900 mg C m^{-2} d^{-1} and ~1300 mg C m^{-2} d^{-1}, respectively. In the Central North Sea PP peaked at ~500 mg C m^{-2} d^{-1} during June and reached ~1200 mg C m^{-2} d^{-1} in NW European Shelf waters. In the North East Atlantic, PP reached a maximum in July at 1000-1200 mg C m^{-2} d^{-1}. PP was at its lowest over the entire area during March and October. To capture the upper limit in the variability in PP, maps of PP Percentile 90 (P90) were also determined (Figure B.2). These illustrated that Shelf areas have the highest PP from May-July (2 – 3 g C m^{-2} d^{-1}), whereas PP P90 for the Central N Sea is < 1 g C m^{-2} d^{-1}.
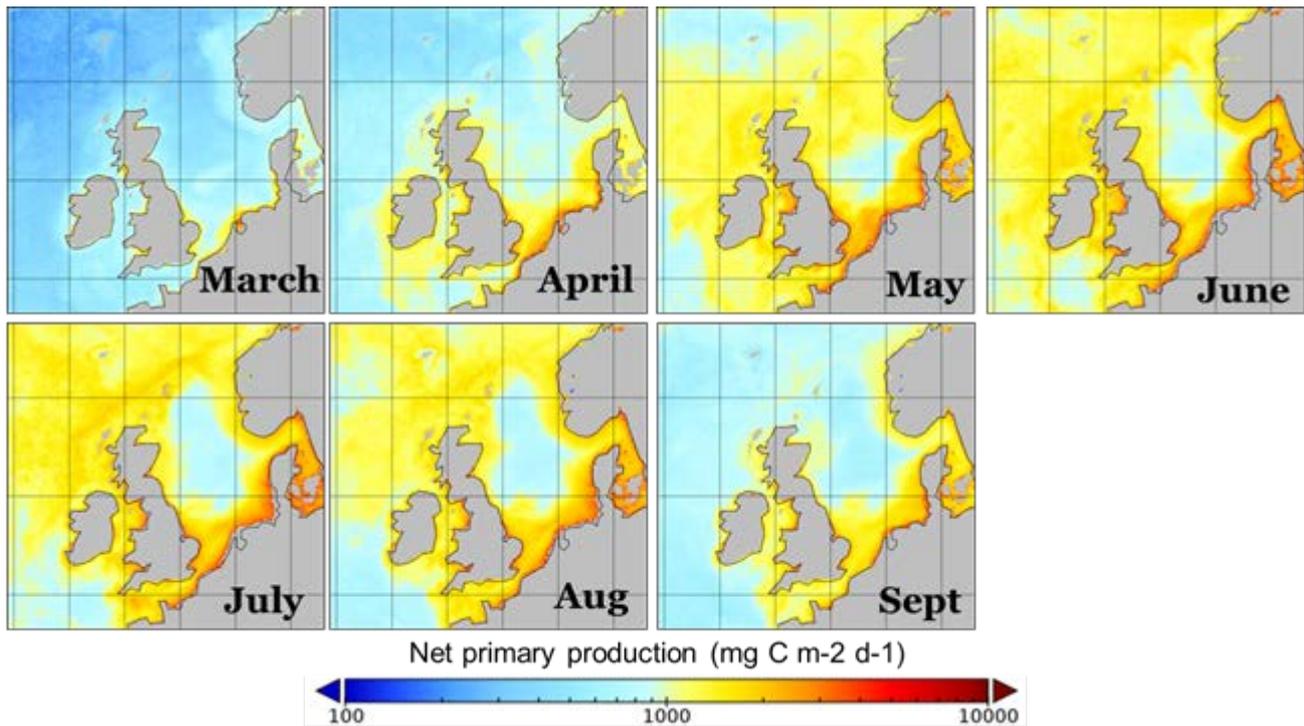
Figure B.2. Mean monthly primary production from Copernicus Marine Environment Monitoring Service (CMEMS) data (1997-2017).

Deploying k-means cluster analysis to characterise pixels with similar peak and timing of PP resulted in eight distinctive regions which we define as: the NE Atlantic (red), NW Atlantic (dark blue), Irish Sea and English Channel (purple), Central N Sea (light blue), Celtic Sea (green), NW European shelf seas (mauve) and NW European coast (orange; Figure B.3).
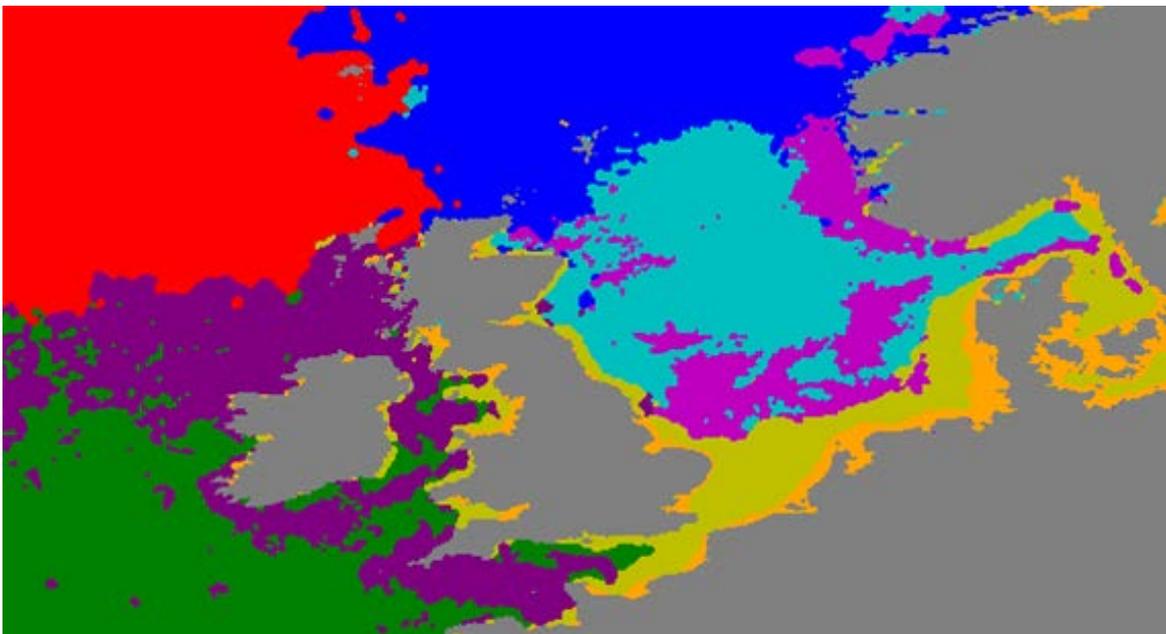

Figure B.3. Regions of similar primary production identified using k-means cluster analysis on peak and timing of PP. Eight regions are identified: NE Atlantic (red), NW Atlantic (dark blue), Irish Sea and English Channel (purple), Central N Sea (light blue), Celtic Sea (green), NW European shelf seas (mauve) and NW European coast (orange).

Mean PP and PP P90 for each region for the CMEMS time series (1997-2017) is given in Figure B.4. Each region showed a slightly different shape in PP climatology. For example, the Central N Sea exhibited the earliest maxima in April (~1000 mg C m^{-2} d^{-1}), followed by the Celtic Sea which peaked in May and then had the lowest PP through other months of the year. The NE Atlantic and Northern N Sea had a similar pattern, with an initial peak in April and PP maxima in June, though values in the Northern N Sea were always higher than the NE Atlantic (1250 and 1150 mg C m^{-2} d^{-1}, respectively) and there was a secondary peak in this region in August. Peak PP in the NW European shelf area occurred in July and was the highest of all regions (~2000 mg C m^{-2} d^{-1}). The North Sea frontal zones had the latest peak PP in August, which was ~1200 mg C m^{-2} d^{-1} (Figure B.1). The mean monthly PP P90 values were slightly higher compared to the mean (Figure B.4; Table B.1).
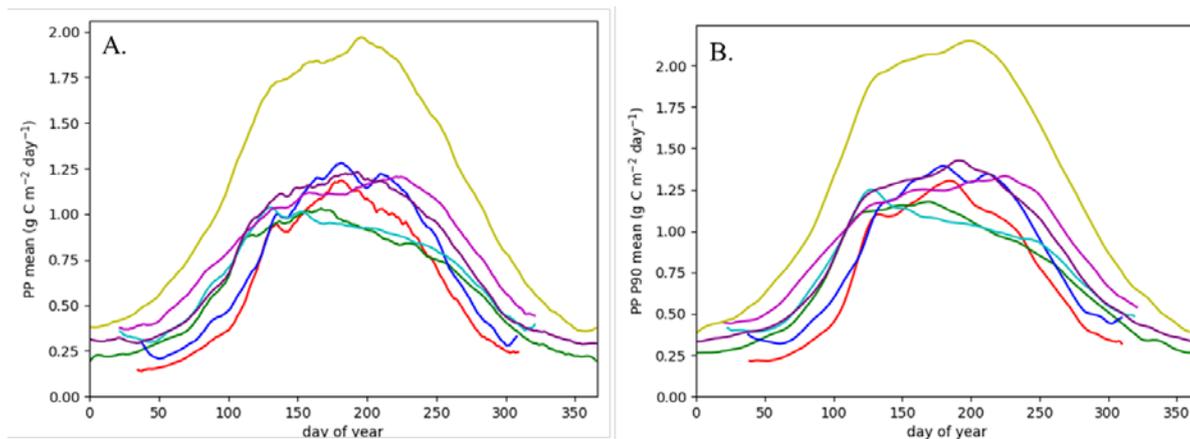

Figure B.4. Climatological mean (A.) and mean 90th percentile (B.) of primary production (g C m^{-2} d^{-1}) for each of the Eight regions identified in Figure B.3: NE Atlantic (red), NW Atlantic (dark blue), Irish Sea and English Channel (purple), Central N Sea (light blue), Celtic Sea (green), NW European shelf seas (mauve) and NW European coast (orange).

Table B.1. Baseline values in peak mean monthly and annual primary production in eight areas; NE Atlantic, Northern N Sea, English Channel and Irish Sea, Central N Sea, North Sea Transition zone, Celtic Sea, NW European shelf seas, NW European coast.

Area	Peak mean monthly PP (mg C m^{-2} d^{-1})	Peak mean monthly PP P90 (mg C m^{-2} d^{-1})	Peak mean annual PP (g C m^{-2} y^{-1})	Peak mean annual PP P90 (g C m^{-2} y^{-1})
NE Atlantic	1200	1250	190	215
Northern N Sea	1250	1300	235	255
Central N Sea	1100	1250	250	270
North Sea Transition zone	1250	1350	240	260
English Channel / Irish Sea	1250	1300	260	280
Celtic Sea	1100	1200	225	250
NW European shelf Seas	2000	2300	400	430

The 19-year time series of PP illustrated that highest PP occurred in the NW European Shelf region and the lowest in the Celtic Sea (Figure B.5). In the English Channel and Celtic Sea, the highest PP occurred in 2012, whereas in the NE Atlantic it occurred in 2013, in the Northern N Sea during 2002 & 2006, in the Central N Sea 2008 and in the NW European Shelf during 1999, 2007 and 2008. The lowest PP in the NE Atlantic, Northern and Central North Sea and the English Channel occurred in 2016 and was in 1999 in the Celtic Sea 2009 in the NW European Shelf Seas (Figure B.4). For each time series, the climatological mean PP P90 is plotted (solid black line). A repeated trend in exceeding the PP P90 indicates an environmental perturbation in the region or consistent errors in the satellite data used to compute the PP P90.
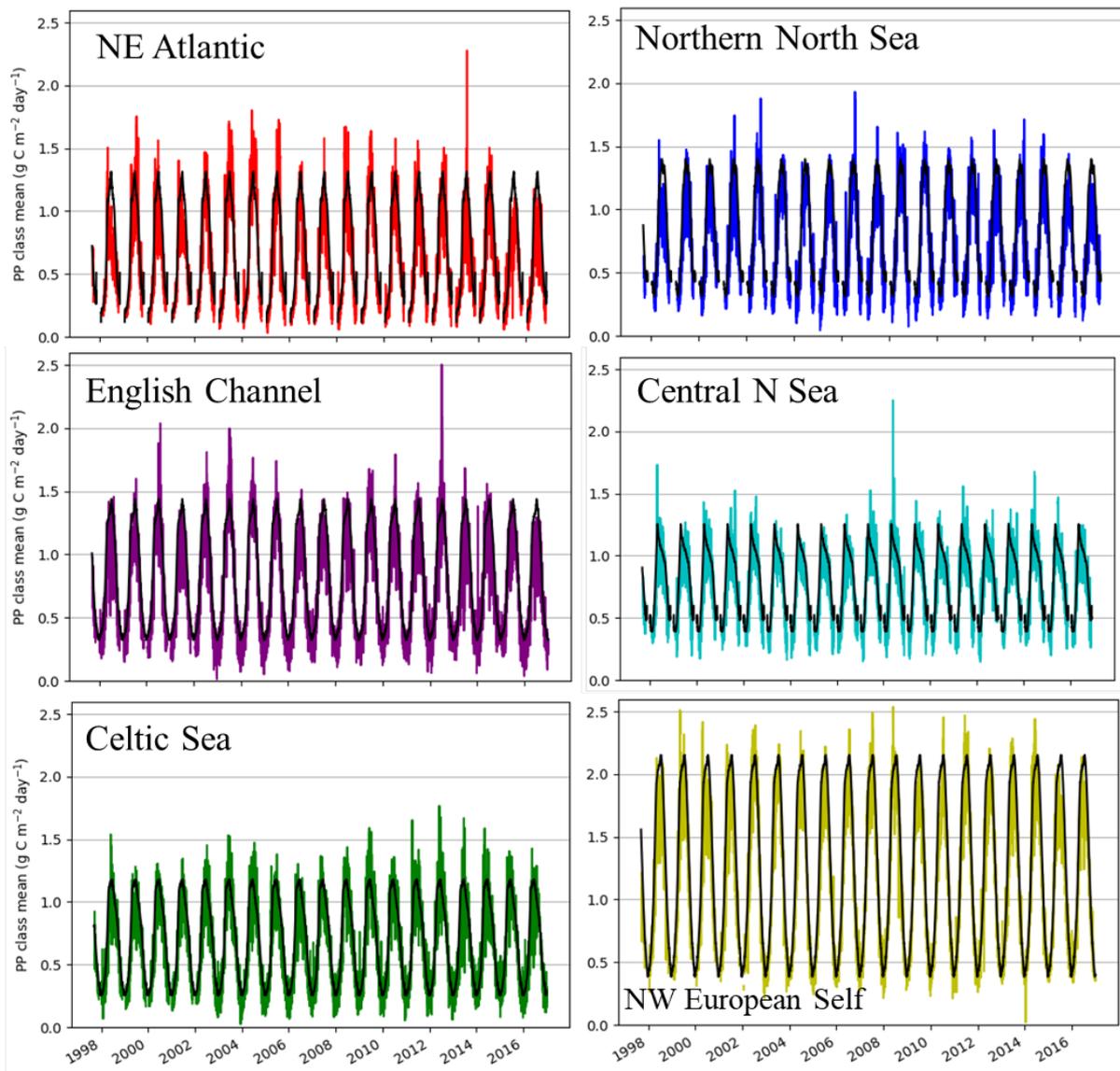
Figure B.5. Primary production (g C m^{-2} d^{-1}) time series in six of the eight regions identified in Figure B.3: NE Atlantic (red), NW Atlantic (dark blue), Irish Sea and English Channel (purple), Central N Sea (light blue), Celtic Sea (green), NW European shelf seas (mauve). Solid black line is the climatological mean PP P90 given in Figure B.4b.

Annual PP in all regions, except the NW European Shelf, were < 300 g C m^{-2} y^{-1}, with values in the English Channel consistently > 250 g C m^{-2} y^{-1}, values in the Northern, Central N Sea and Celtic Sea between 200-250 g C m^{-2} y^{-1} and values in the NE Atlantic between 150-200 g C m^{-2} y^{-1} (Figure B.5). Similar to mean monthly PP, highest annual mean PP in the Celtic Sea and English Channel was in 2012, in the Central N Sea during 2008 and in the NW European Shelf during 2008 (Figure B.5). By contrast, highest annual PP in the NE Atlantic occurred in 2003. Mean annual PP 90 were slightly higher and indicate the upper baseline for each region (Table B.1).
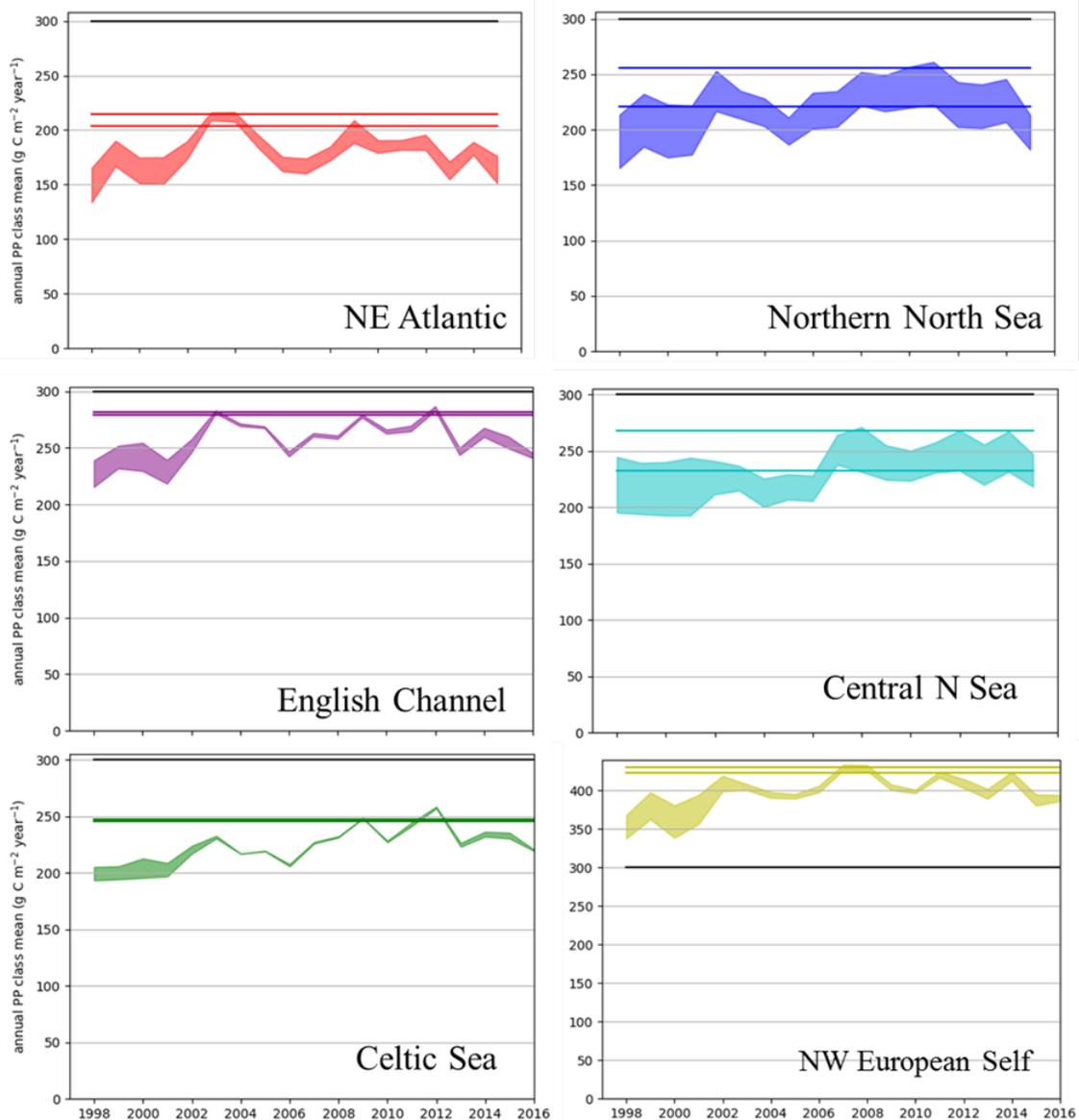
Figure B.6. Time series of annual Primary production (g C m^{-2} y^{-1}) in six of the eight regions identified in Figure B.3. The upper and lower limits of the time series are based on annual integration based on using the mean value for September and March for October to February (upper limit) and 0 PP over winter (lower limit). The coloured lines represent the new determined baseline values for each region based on these ways of integrating the data. The solid black line is the baseline derived by Boalch et al. (1968, 1978). NE Atlantic (red), NW Atlantic (dark blue), Irish Sea and English Channel (purple), Central N Sea (light blue), Celtic Sea (green), NW European shelf seas (mauve).
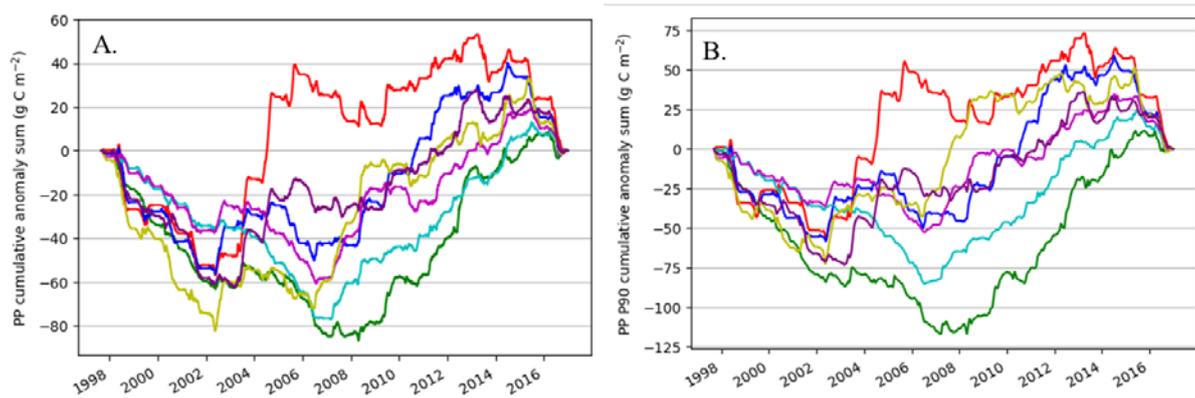
Figure B.7. Cumulative sum of mean monthly Primary production anomaly (g C m⁻² y⁻¹) in the eight regions identified in Figure B.3. NE Atlantic (red), NW Atlantic (dark blue), Irish Sea and English Channel (purple), Central N Sea (light blue), Celtic Sea (green), NW European shelf seas (mauve).

Cumulative sums of both the mean monthly anomaly in PP and the mean monthly PP P90 showed that in all areas except the NE Atlantic, that there was a decrease in PP from 1997 to 2002 followed by an increase from 2005, a further decrease to 2008, and an increase to 2017. In the NE Atlantic, there was a similar decrease in PP from 1997 to 2002 followed by a continual and higher increase to 2005, there was then a decrease to 2009 followed by another increase to 2013 and a decrease to 2016 (Figure B.7). Analyses are on-going to identify what are the principal factors that have causing these decadal oscillations in PP.

To quantify the uncertainty in the estimate of primary production, a data base of in situ 14C measurements (Figure B.8) was compiled for four of the eight PP regions identified in Figure B.3.
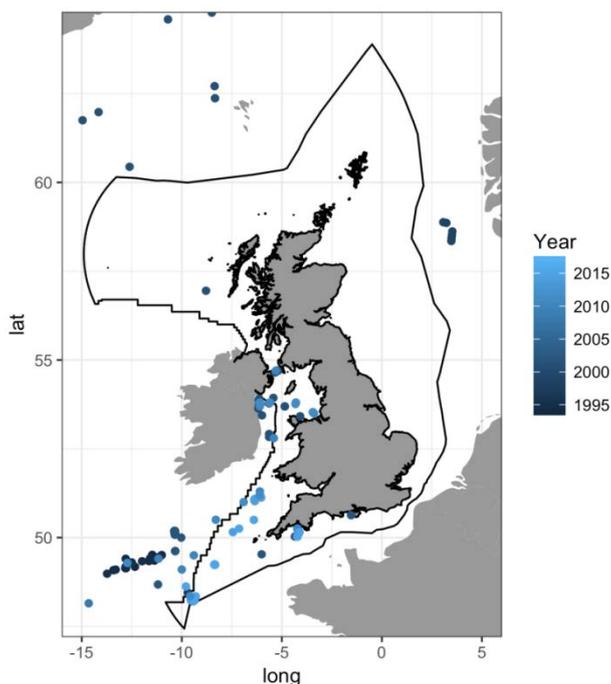

Figure B.8. Location of 14C based in situ measurements in the NE Atlantic.

For each region, the deviation between satellite (given in blue) and the in situ (given in orange) over the range of PP is computed (Figure B.9). When Y=X there is a perfect match between in situ and satellite data. A negative intercept indicates an overestimate in satellite PP. When slope >1 variability in in situ PP does not capture variability in

71

satellite data. For the Central North Sea, the NW European Shelf and the Celtic Sea there was good correspondence between the in situ and satellite data. At lower PP values in the Celtic Sea there was a greater deviation between these data. For the NE Atlantic, there was good correspondence between mid, high and low values, and greater deviation over the range between these, though fewer in situ data were available for validation in the NE Atlantic.
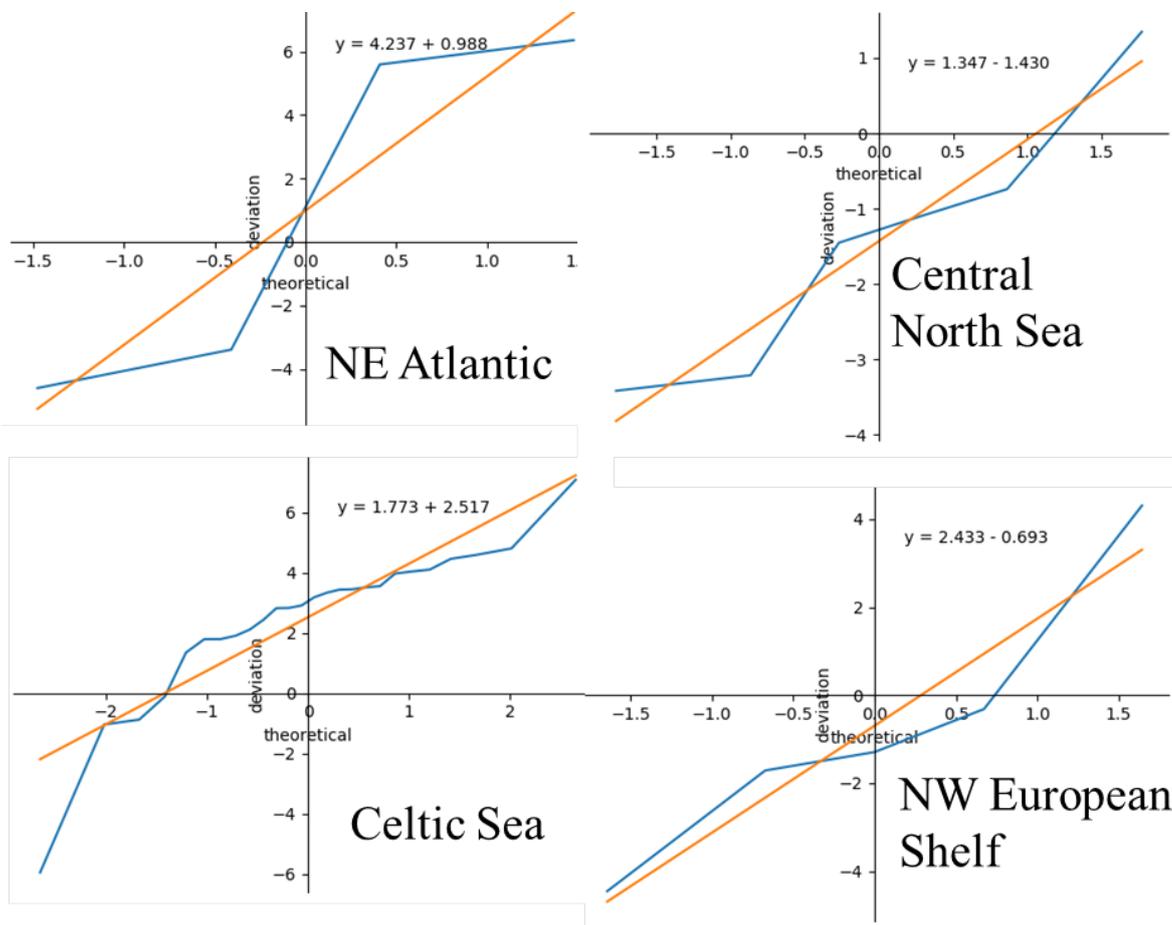

Figure B.9. Deviation between in situ 14C measurements and satellite estimates of primary production in the NE Atlantic, Central North Sea, NW European Shelf and Celtic Sea.

B.4. Recommendations for further research

For the coastal waters under the jurisdiction of the WFD (shown in orange in Figure B.3), a different parameterisation of the light field in the water column is required, under the influence of increasing Total Suspended Matter (TSM) and Coloured Dissolved Organic Matter (CDOM) which scatter (bb) and absorb (a) light thus modifying the irradiance field available to phytoplankton. Satellite Chl-a and Kd products can be used to evaluate the optical water type (OWT) where the established clear water, open ocean relationship between Chl-a and Kd breaks down, which can be subsequently used to identify which OWT are optically complex "case-2" waters influenced by TSM and CDOM. From this relationship the data can then be dealt with in one of two ways:

- To use the monthly OWT to mask out "case-2 " waters in the satellite imagery.
- To use the OWT on the satellite imagery to characterise the light field and to use this to compute PP in the coastal waters more accurately.

The elements in this process are outlined below:

CCI Optical water Type (OWT) definition of end member (as per [Jackson et al., 2017]):

- for Case 1 waters through satellite remote sensing reflectance (Rrs) data
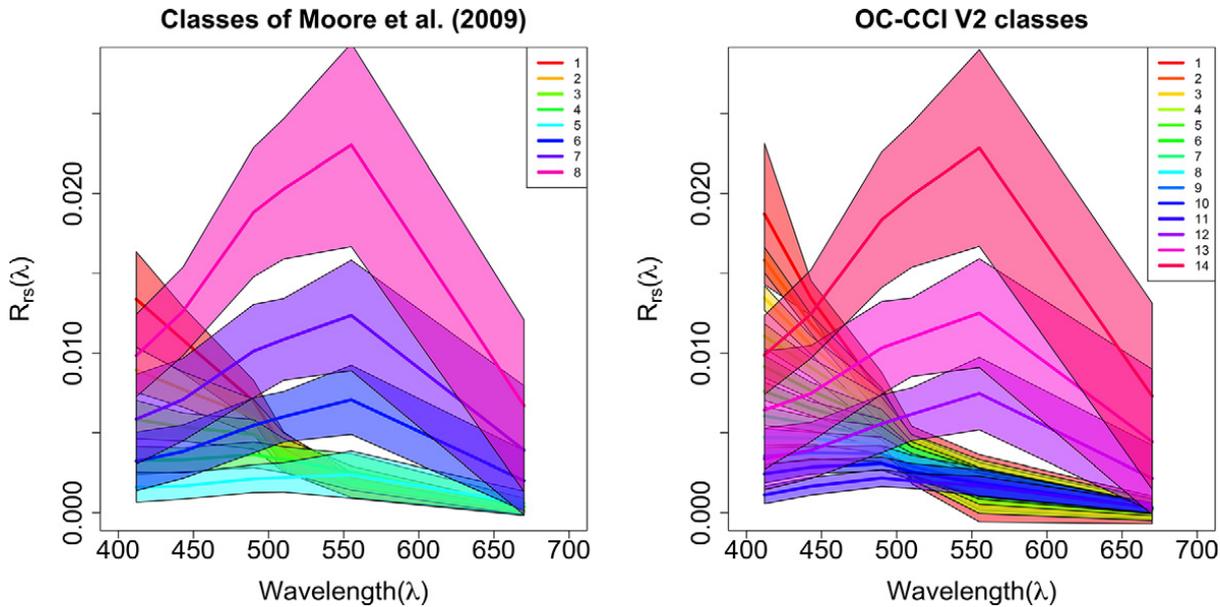- for coastal/coccolitophores based on in-situ

Figure B.10. Optical water types for the ESA Ocean Colour Climate Change Initiative (OC-CCI) products (figure from [Jackson et al., 2017]). Comparison between current OC-CCI OWT (right) and precedent OWT as defined by [Moore et al., 2009]. Solid thick lines are means of a given OWT and shaded areas are standard deviations.

Each pixel has a proportion of membership to the OWT. This proportion and membership to OWT is used to select which Chl-a (CHLOR_A OC-CCI product) algorithm is used (OC4, OC5 etc.) and to weight each algorithm by the proportion of membership. The computation of Kd (K_490 OC-CCI product) is the same in all water classes and uses [Lee et al., 2005]. If it is [Lee et al., 2005] then it is based on an empirical relationship between Kd and absorption (a) and backscattering coefficients (bb). Departures in the relationship between Kd and chlorophyll algorithm relationships also indicate where the Chla algorithm may potentially fail. Implementing this methodology will ultimately improve the accuracy of the computation of PP in the coastal zone.

In addition to identify coastal waters using the optical water type approach, the internal computation of light field in primary production models should be adapted to coastal waters. In particular, the most complex family of primary production models from ocean colour are those that consider explicitly the distribution of chlorophyll concentration (Chla) in water column (depth resolved) and the spectral composition of light (spectrally resolved) [Morel, 1991; Smyth et al., 2005]. In these algorithms, there are four input variables:

- Chlorophyll as a function of depth, Chla(z);
- Photosynthetically active radiation as a function of depth, wavelength and time, PAR(z,λ,t);
- Phytoplankton absorption cross section per unit of Chla as a function of wavelength, a*(λ);
- Net growth rate as a function of depth, wavelength and time, $\phi\mu$;

Of these four variables, only Chla(z) and PAR(z,λ,t) are obtainable by optical remote sensing. However, in comparison of 21 primary production models (Saba et al.,2011) found that the most important factor to reduce the skill of the models was accurate estimation of water column depth. Water column depth can also be used as a proxy for Case II waters, where inherent optical properties (IOP) vary independently of the Chla [H Gordon and Morel, 1983]. The authors found that surface Chla did not affect model skill at depths less than 250 m. The hypothesis is that the main cause of skill reduction in primary production models was the over-estimation of the euphotic zone in Case II waters and the underestimation of the euphotic zone in Case I waters, which point to the factors controlling PAR(z,λ,t) in the water column.

The attenuation of PAR in the water column in the [Smyth et al., 2005] spectrally and depth resolved model follows the Lambert-Beer law, is controlled by the diffuse attenuation coefficient, Kd (λ,Z). For their station-per-station based model, Kd (λ,Z) was computed using a simplified bio-optical model of absorption, scattering and backscattering for Chla, coloured dissolved organic matter absorption (aCDOM) and suspended particulate matter concentration (SPM). In situ results in [G H Tilstone et al., 2005], found that at concentrations of suspended particulate matter (SPM) up to 2.84 g/m^3 there were no significant improvements in the performance of a spectrally and depth resolved primary production model that included Chla and aCDOM and compared to one which included Chla, aCDOM and SPM.

[Smyth et al., 2005] suggest that one potential explanation for this lack of effect of SPM on the primary production model could be the neglect of the absorption effect for the SPM. This could be important, as Kd $\approx$a/μd (a for absorption coefficient and μd is the mean downwelling average cosine). Although difficult to quantify, mass-specific mineral absorption methods exist [Allali et al., 1995] and spectral measurements of mass specific absorption for minerals are available for single minerals [Stramski et al., 2007]as well as for in situ samples [Babin et al., 2003]. Additional simulations considering the effect of mineral absorption could therefore be used to improve the estimation of Kd (λ, Z) and therefore the accuracy of primary production models in coastal, case II waters.

To simplify the computation of primary production in coastal waters using satellite data, [Smyth et al., 2005] constructed look-up-tables using the radiative transfer model HYDROLIGHT at 0, 3, 5, 10, and 20m depth, for 27 values of Chla and 25 values of coloured dissolved organic matter absorption at 440 nm (aCDOM(440)) for one day at noon and 50N and a wind speed of 0 m/s. The use of a location/time bound value for Kd could be improved [Lee et al., 2015], since Kd is an apparent optical property, its value can differ significantly with solar zenith angle [H R Gordon, 1989]. It may be also necessary to match solar elevation, with the incident downwelling irradiance and Kd. In the satellite input (effectively PAR, which is then converted to downwelling irradiance) to the look-up-table though, latitude, date and time are variable, and not generally matched to the Kd computation. Although [Saba et al., 2011] relates the correlation between model-data misfit and latitude to the strong correlation between SST and latitude, the investigation, through a sensitivity analysis for instance, of the effects of differences between look-up-table Kd at a fixed and variable locations would help ascertain the magnitude of uncertainty relative to other potential sources of uncertainty in the model.

B.5. Conclusions

Analysis of 19 years of CMEMS Ocean colour satellite data were used to determine annual & daily primary production percentile 90 baselines for similar regions identified through k-means clustering. The baselines for each region are defined as:

- NE Atlantic: 1.25 g C m^{-2} d^{-1}; 215 g C m^{-2} y^{-1}.

- NW Atlantic: 1.30 g C m^{-2}d^{-1}; 255 g C m^{-2}y^{-1}.

- English Channel: 1.30 g C m^{-2} d^{-1}; 280 g C m^{-2} y^{-1}.

- Central N Sea: 1.25 g C m^{-2}d^{-1}; 270 g C m^{-2}y^{-1}.

- Celtic Sea: 1.2 g C m^{-2} d^{-1}; 250 g C m^{-2} y^{-1}.

- NW European shelf: 2.3 g C m^{-2} d^{-1}; 430 g C m^{-2} y^{-1}.

It should be noted that these baselines have been derived independent of the background concentrations for chlorophyll a, as described in the present JMP EUNOSAT Activity 1 report. For future use in OSPAR or MSFD assessments these should be aligned.

B.6. Acknowledgements

B.7.   References

Allali, K., A. Bricaud, M. Babin, A. Morel, and P. Chang (1995), A new method for measuring spectral absorption coefficients of marine particles, *Limnology and Oceanography*, *40*(8), 1526-1532.

Babin, M., A. Morel, V. Fournier-Sicre, F. Fell, and D. Stramski (2003), Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration, *Limnology and Oceanography*, *48*(2), 843-859.

Boalch, G. T. (1987), Changes in the phytoplankton of the western English Channel in recent years, *British Phycological Journal*, *22*(3), 225-235.

Boalch, G. T., D. S. Harbour, and E. I. Butler (1978), Seasonal phytoplankton production in the western English Channel 1964-1974, *Journal of the Marine Biological Association of the United Kingdom*, *58*(4), 943-953.

Cloern, J. E. (2001), Our evolving conceptual model of the coastal eutrophication problem, *Marine Ecology Progress Series*, *210*, 223-253, doi:10.3354/meps210223.

Cloern, J. E., S. Q. Foster, and A. E. Kleckner (2014), Phytoplankton primary production in the world's estuarine-coastal ecosystems, *Biogeosciences*, *11*(9), 2477-2501, doi:10.5194/bg-11-2477-2014.

Cristina, S., J. Icely, P. C. Goela, T. A. DelValls, and A. Newton (2015), Using remote sensing as a support to the implementation of the European Marine Strategy Framework Directive in SW Portugal, *Continental Shelf Research*, *108*, 169-177, doi:10.1016/j.csr.2015.03.011.

Gordon, H., and A. Morel (1983), *A review Lecture Notes on Coastal and Estuarine Studies*, 114 pp., Springer-Verlag NY.

Gordon, H. R. (1989), CAN THE LAMBERT-BEER LAW BE APPLIED TO THE DIFFUSE ATTENUATION COEFFICIENT OF OCEAN WATER, *Limnology and Oceanography*, *34*(8), 1389-1409, doi:10.4319/lo.1989.34.8.1389.

Henson, S. A., J. P. Dunne, and J. L. Sarmiento (2009), Decadal variability in North Atlantic phytoplankton blooms, *Journal of Geophysical Research-Oceans*, *114*, doi:C04013

10.1029/2008jc005139.

Jackson, T., S. Sathyendranath, and F. Melin (2017), An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications, *Remote Sensing of Environment*, *203*, 152-161, doi:10.1016/j.rse.2017.03.036.

Koertzinger, A., U. Send, R. S. Lampitt, S. Hartman, D. W. R. Wallace, J. Karstensen, M. G. Villagarcia, O. Llinas, and M. D. DeGrandpre (2008), The seasonal pCO_2 cycle at 49 degrees N/16.5 degrees W in the northeastern Atlantic Ocean and what it tells us about biological productivity, *Journal of Geophysical Research-Oceans*, *113*(C4), doi:10.1029/2007jc004347.

Kroncke, I., H. Reiss, and M. Turkay (2013), Macro- and megafauna communities in three deep basins of the South-East Atlantic, *Deep-Sea Research Part I-Oceanographic Research Papers*, *81*, 25-35, doi:10.1016/j.dsr.2013.07.005.

Lee, Z. P., K. P. Du, and R. Arnone (2005), A model for the diffuse attenuation coefficient of downwelling irradiance, *Journal of Geophysical Research-Oceans*, *110*(C2).

Lee, Z. P., J. Marra, M. J. Perry, and M. Kahru (2015), Estimating oceanic primary productivity from ocean color remote sensing: A strategic assessment, *Journal of Marine Systems*, *149*, 50-59, doi:10.1016/j.jmarsys.2014.11.015.

Lehmusluoto, P. O., and L. Pesonen (1973), EUTROPHICATION IN HELSINKI AND ESPOO SEA AREAS MEASURED AS PHYTOPLANKTON PRIMARY PRODUCTION, *Oikos*, 202-208.

Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill (1995), An Estimate of Global Primary Production in the Ocean from Satellite Radiometer Data, *Journal of Plankton Research*, *17*(6), 1245-1271.

Meyer, J., P. Nehmer, A. Moll, and I. Kroncke (2018), Shifting south-eastern North Sea macrofauna community structure since 1986: A response to de-eutrophication and regionally decreasing food supply?, *Estuarine Coastal and Shelf Science*, *213*, 115-127, doi:10.1016/j.ecss.2018.08.010.

Moore, T. S., J. W. Campbell, and M. D. Dowell (2009), A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product, *Remote Sensing of Environment*, *113*(11), 2424-2430, doi:10.1016/j.rse.2009.07.016.

Morel, A. (1991), Light and Marine Photosynthesis - a Spectral Model with Geochemical and Climatological Implications, *Progress in Oceanography*, *26*(3), 263-306.

Nixon, S. W. (1995), COASTAL MARINE EUTROPHICATION - A DEFINITION, SOCIAL CAUSES, AND FUTURE CONCERNS, *Ophelia*, *41*, 199-219, doi:10.1080/00785236.1995.10422044.

Novoa, S., G. Chust, Y. Sagarminaga, M. Revilla, A. Borja, and J. Franco (2012), Water quality assessment using satellite-derived chlorophyll-a within the European directives, in the southeastern Bay of Biscay, *Marine Pollution Bulletin*, *64*(4), 739-750, doi:10.1016/j.marpolbul.2012.01.020.

Rabalais, N. N., N. Atilla, C. Normandeau, and R. E. Turner (2004), Ecosystem history of Mississippi River-influenced continental shelf revealed through preserved phytoplankton pigments, *Marine Pollution Bulletin*, *49*(7-8), 537-547, doi:10.1016/j.marpolbul.2004.03.017.

Rabalais, N. N., R. E. Turner, R. J. Diaz, and D. Justic (2009), Global change and eutrophication of coastal waters, *Ices Journal of Marine Science*, *66*(7), 1528-1537, doi:10.1093/icesjms/fsp047.

Ryther, J. H., and W. M. Dunstan (1971), NITROGEN,PHOSPHORUS, AND EUTROPHICATION IN COASTAL MARINE ENVIRONMENT, *Science*, *171*(3975), 1008-&, doi:10.1126/science.171.3975.1008.

Saba, V. S., et al. (2011), An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe, *Biogeosciences*, *8*(2), 489-503, doi:10.5194/bg-8-489-2011.

Smyth, T. J., G. H. Tilstone, and S. B. Groom (2005), Integration of radiative transfer into satellite models of ocean primary production, *Journal of Geophysical Research-Oceans*, *110*(C10).

Stramski, D., M. Babin, and S. B. Wozniak (2007), Variations in the optical properties of terrigenous mineral-rich particulate matter suspended in seawater, *Limnology and Oceanography*, *52*(6), 2418-2433, doi:10.4319/lo.2007.52.6.2418.

Takahashi, T., et al. (2009), Climatological mean and decadal change in surface ocean pCO(2), and net sea-air CO(2) flux over the global oceans (vol 56, pg 554, 2009), *Deep-Sea Research Part I-Oceanographic Research Papers*, *56*(11), 2075-2076, doi:10.1016/j.dsr.2009.07.007.

Tilstone, G., T. Smyth, A. Poulton, and R. Hutson (2009), Measured and remotely sensed estimates of primary production in the Atlantic Ocean from 1998 to 2005, *Deep-Sea Research Part II-Topical Studies in Oceanography*, *56*(15), 918-930, doi:10.1016/j.dsr2.2008.10.034.

Tilstone, G. H., T. J. Smyth, R. J. Gowen, V. Martinez-Vicente, and S. B. Groom (2005), Inherent optical properties of the Irish Sea and their effect on satellite primary production algorithms, *Journal of Plankton Research*, *27*(11), 1127-1148.

Annex C – Seasonal differences in strength of stratification

Stratification was determined based on the modelled monthly averaged density difference between the top and bottom layer in the model. A grid cell was classified as stratified when the density difference was larger than 0.75 kg/m^3 similar to van Leeuwen et al. (2015) (see red dotted line in Figures 1-4 in Annex C).
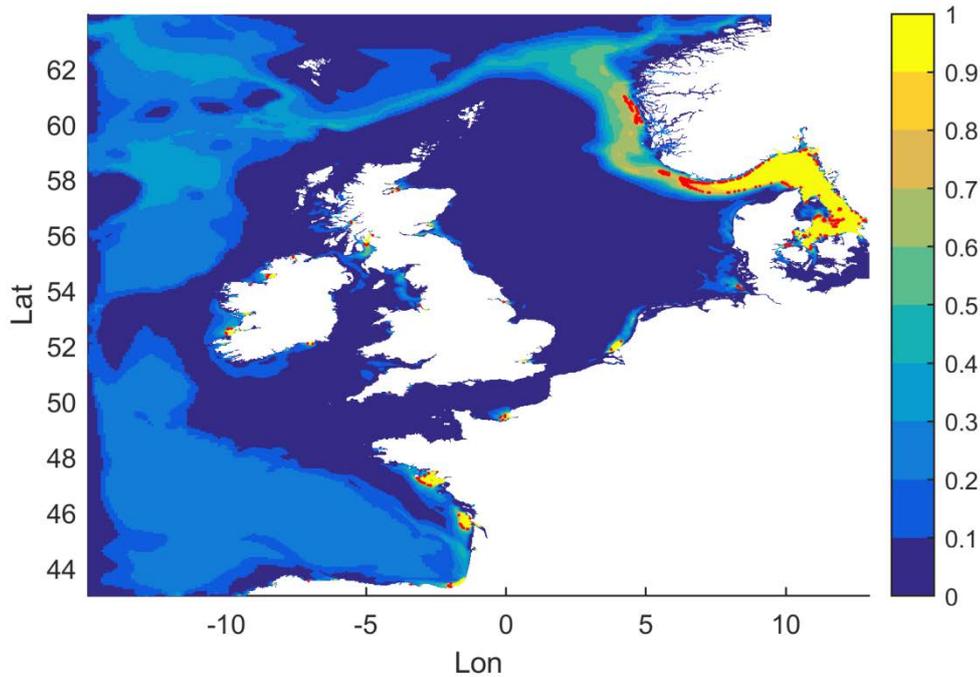

Figure C.1. Density difference between top and bottom layer in winter. The red dotted line indicates the 0.75 kg/m3 threshold.
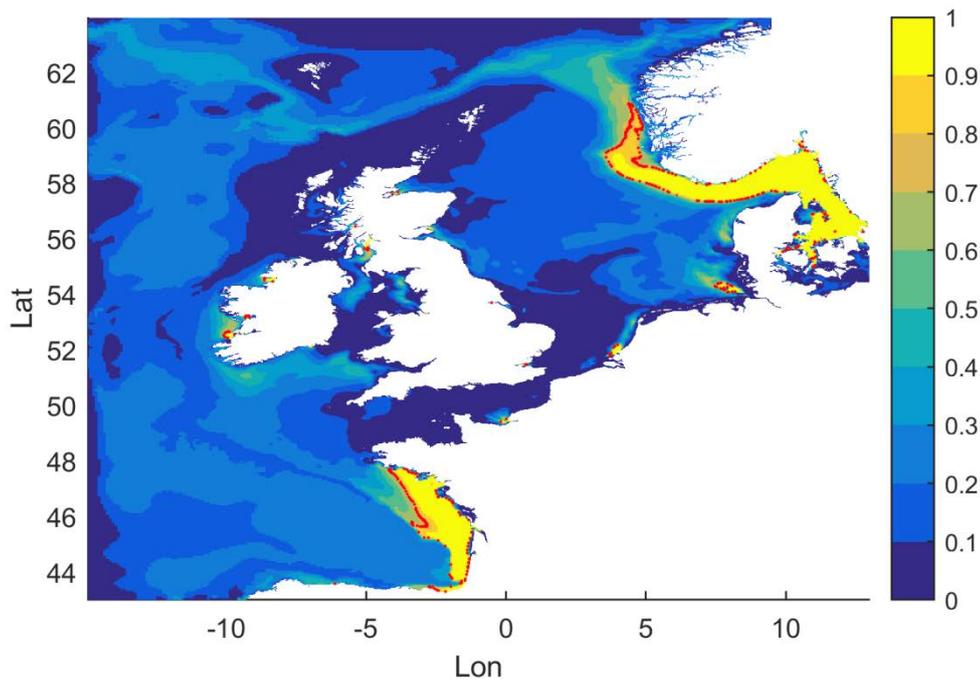

Figure C.2. Density difference between top and bottom layer in spring. The red dotted line indicates the 0.75 kg/m3 threshold.
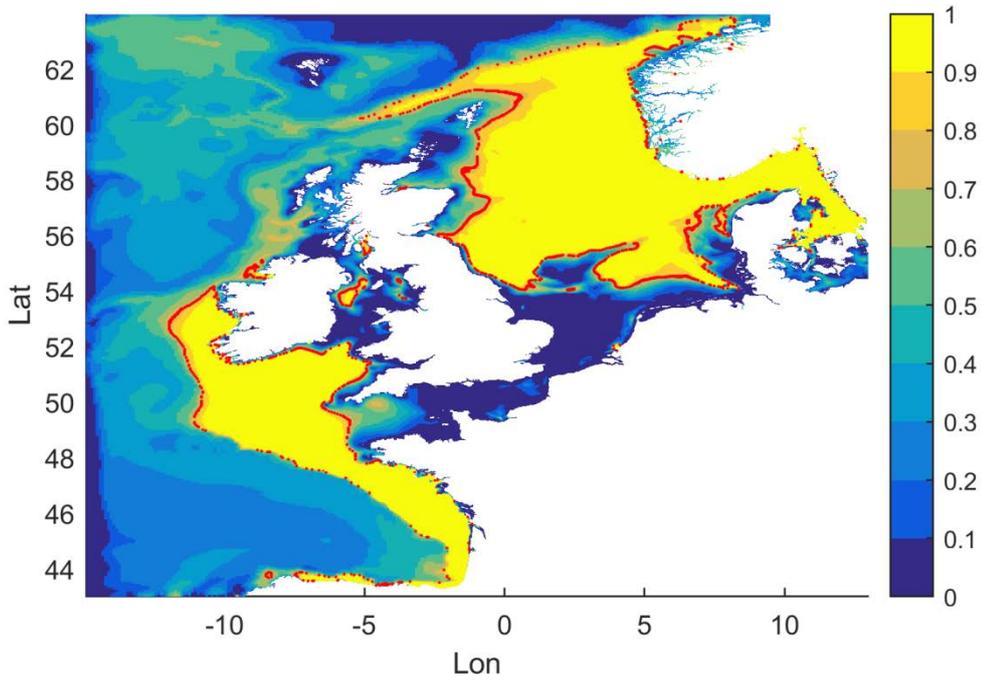
Figure C.3. Density difference between top and bottom layer in summer. The red dotted line indicates the 0.75 kg/m3 threshold.
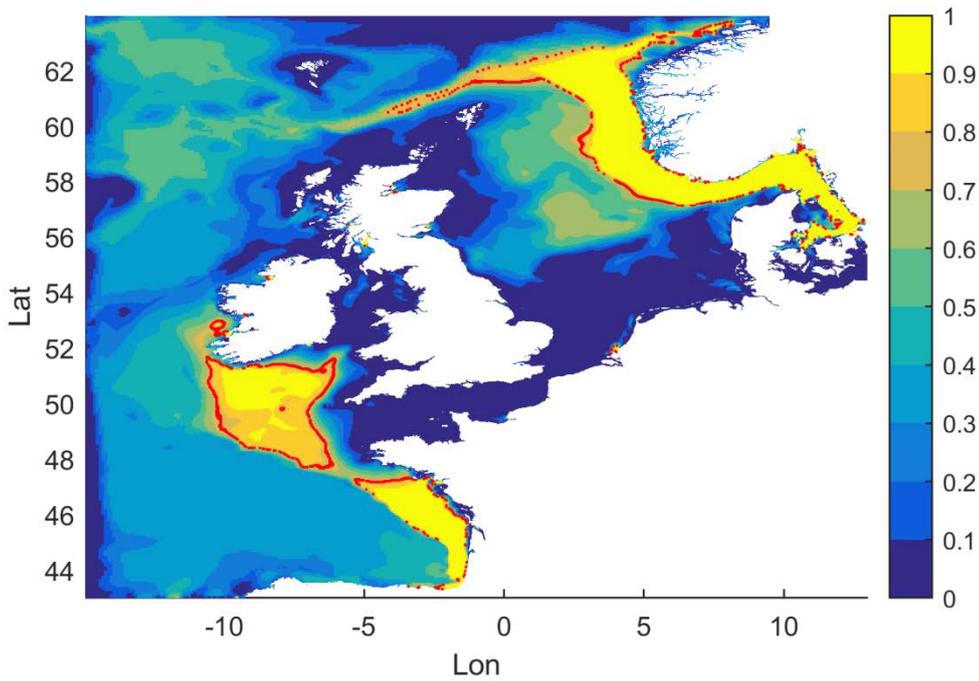

Figure C.4. Density difference between top and bottom layer in autumn. The red dotted line indicates the 0.75 kg/m3 threshold.

Annex D – Introduction to the RCO-SCOBI models and the experimental setup

The coupled physical RCO (Rossby Centre ice and Ocean model) (Meier et al. 2003; Meier 2005; 2007) and biogeochemical model SCOBI (Swedish Coastal and BIogeochemical Ocean model) (Eilola et al. 2009; Eilola et al. 2011; Almroth-Rosell et al. 2011) of the Baltic Sea (Fig. 1) used in the present work is described in detail by Meier et al. (2018). A brief introduction of RCO-SCOBI and the experimental setup is given below. For more detailed discussions and model references the reader is referred to the open access article by Meier et al. (2018). They evaluated model results to historical data and performed several sensitivity experiments with the model to disentangle the drivers of historical development of eutrophication and hypoxia in the Baltic Sea.
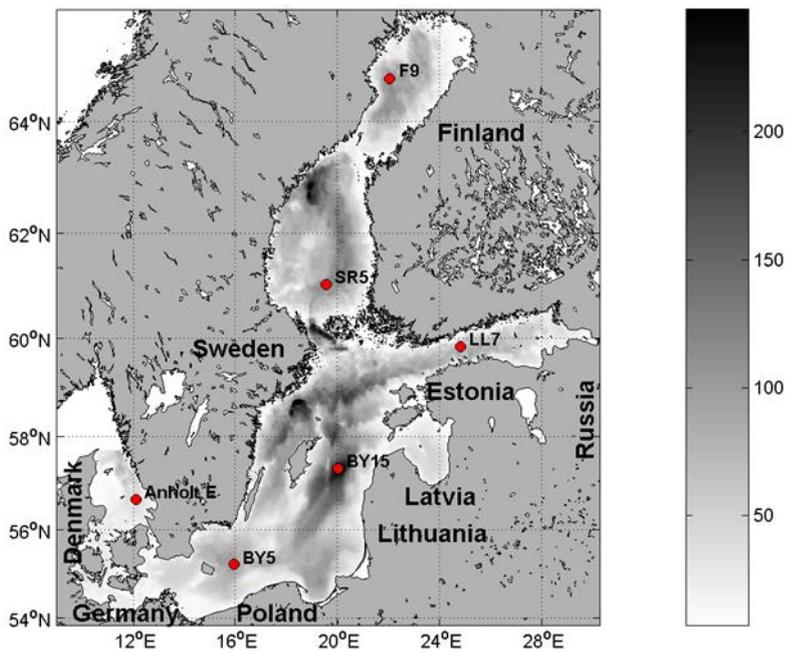

Figure D.1. The model domain of RCO-SCOBI. The model has an open boundary in the northern Kattegat. The color bar indicates the depth in meter.

D.1. Model description

RCO model

The ocean circulation model is coupled to a Hibler-type sea ice model and the subgrid-scale mixing in the ocean is parameterized using a k-ε turbulence closure scheme with flux boundary conditions and a flux-corrected, monotonicity preserving transport scheme is embedded without explicit horizontal diffusion. The model domain comprises the Baltic Sea and Kattegat with lateral open boundaries in the northern Kattegat (Fig. 1). In case of inflow, temperature, salinity, nutrients (phosphate, nitrate, ammonium) and detritus model results are nudged towards observed climatological profiles and, in case of outflow, a modified Orlanski radiation condition is used. Fluxes of heat, incoming long- and shortwave radiation, momentum, and matter between atmosphere, ocean and sea ice are parameterized using bulk formulae adopted to the Baltic Sea region. Inputs to the bulk formulae are state variables of the atmospheric planetary boundary layer like 2 m air temperature, 2 m specific humidity, 10 m wind, cloudiness and mean sea-level pressure, and ocean variables like sea surface temperature, sea surface salinity, sea

ice concentration, albedo, and water and sea ice velocities. Horizontal and vertical resolutions amount to 3.7 km and 3 m, respectively.

SCOBI model

SCOBI (Fig. 2) simulates the dynamics of nitrate, ammonium, phosphate, oxygen and hydrogen sulfide (as negative oxygen concentration), three phytoplankton groups (including nitrogen fixing cyanobacteria), zooplankton and detritus (one pool limited by the Redfield molar ratio C:N:P=106:16:1). The sediment contains nutrients in the form of benthic nitrogen and benthic phosphorus. Processes like assimilation, remineralization, nitrogen fixation, nitrification, denitrification, grazing, mortality, excretion, sedimentation, resuspension and burial are considered. Self-shading due to light attenuation by particulate organic matter is included. With the help of a simplified wave model resuspension of organic matter is calculated. SCOBI has a constant carbon (C) to chlorophyll (Chl) ratio C : Chl=50 (mg C (mg Chl)-1), and the production of phytoplankton assimilates carbon (C), nitrogen (N) and phosphorus (P) according to the Redfield molar ratio. The molar ratio of a complete oxidation of the remineralized nutrients is O2:C=138:106. An oxygen dependent fraction of the mineralized phosphorus in the sediments is adsorbed to sediment particles while the rest is released as a flux of phosphate to the overlying water. The phosphate adsorption process, simulating redox-dependent release and adsorption mechanisms of iron bound phosphorus, may reverse when the water turns anoxic and no adsorption occurs. Then some of the previously adsorbed phosphorus may also be released and added to the flux of mineralized phosphorus to the overlying water.
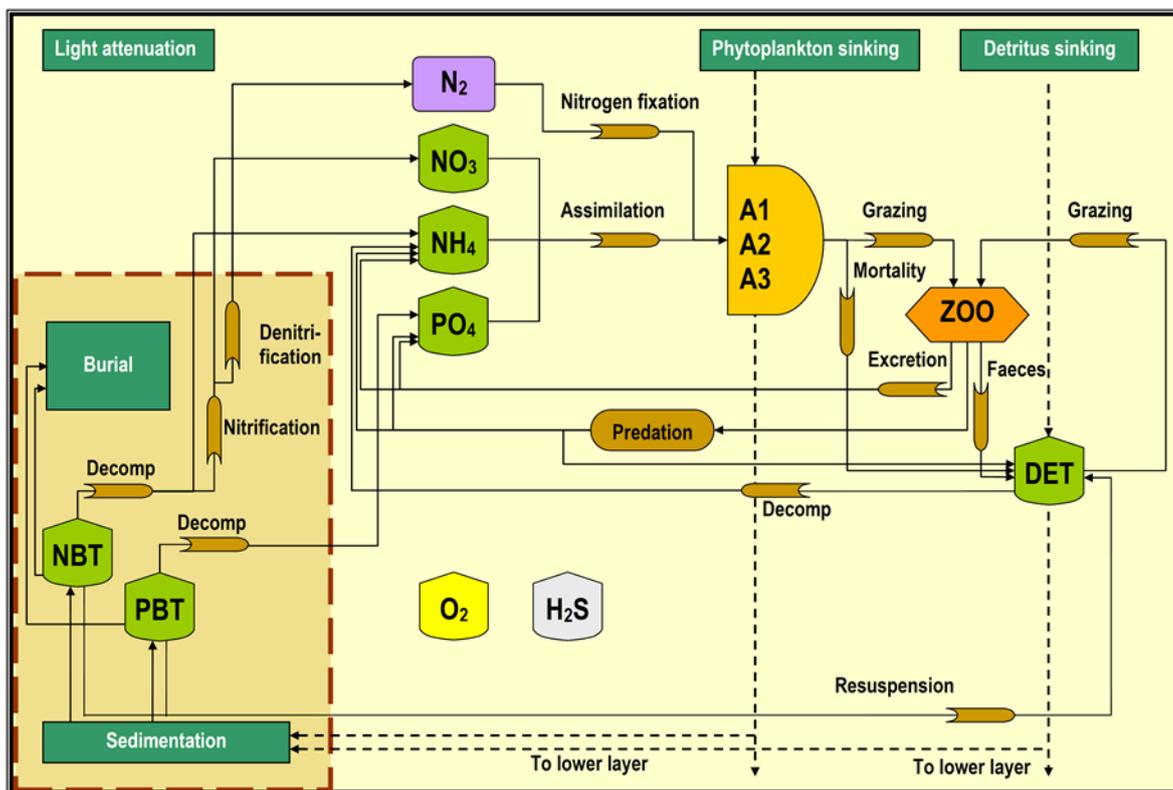

Figure D.2. The SCOBI model. Sediment variables and processes are shown in the lower left frame. Note that in the figure the process descriptions of light attenuation, oxygen and hydrogen sulfide are simplified for clarity.

D.1. Model set-up

Atmospheric forcing

The multivariate three hourly High RESolution Atmospheric Forcing Fields (HiReSAFF) for the period 1850–2008 were constructed based upon the analog-method as described by Schenk and Zorita (2012). The development of analogs for the HiReAFF forcing used regionalized (the Rossby Centre Atmosphere Ocean model) reanalysis data for 1958–2007 together with historical station data of daily sea-level pressure and monthly air temperature observations. The long-term reconstructed seasonal atmospheric forcing over the central Baltic Sea is shown e.g. in Fig. 1 in Gustafsson et al. (2012). More details about the methods and reconstructions are described by Meier et al. (2018), Gustafsson et al. (2012) and Schenk and Zorita (2012).

River runoff

For the calculation of monthly mean river flows several data sets were merged as described in detail by Meier et al. (2018). For 1850–1900, 1901–1920, 1921–1949, 1950–2004, 2005–2008 reconstructions by Hansson et al. (2011), Cyberski and Wroblewski (2000) and Mikulski (1986), observations from the BALTEX Hydrological Data Center (BHDC) (Bergström and Carlsson 1994), and hydrological model results (Graham 1999) were used, respectively. As the spatial resolution of the various data sets differ (total catchment, basins, individual rivers), the variability of the reconstructed runoff changes with time (see Fig. 3 in Meier et al. 2018).

Nutrient loads and atmospheric deposition

The bioavailable loads to the RCO-SCOBI model are shown by Meier et al. (2018) in their Fig. 4. The reconstruction of historical nutrient loads of nitrogen and phosphorus followed Gustafsson et al. (2012). For 1970–2006, nutrient loads from rivers and point sources were compiled from the Baltic Environmental and HELCOM databases (Savchuk et al. 2012). Estimates of pre-industrial loads for 1900 were based upon Savchuk et al. (2008). Between selected reference years of these two environmental states (1970–2006 and around 1900), nutrient loads were linearly interpolated taking intensified agriculture since the 1950s into account. Similarly, atmospheric loads were estimated (Ruoho-Airola et al. 2012). Nutrient loads contain both organic and inorganic phosphorus and nitrogen, respectively. For riverine organic phosphorus and nitrogen loads bioavailable fractions of 100 and 30% are assumed, respectively. Nutrient loads after 2006 were set to the values of the year 2006. As loads were calculated from runoff and annual mean nutrient concentrations (Eilola et al. 2011), the seasonal cycle is determined by the river discharge.

Lateral boundary data

Daily mean sea level elevations at the lateral boundary were calculated from the reconstructed, meridional sea level pressure gradient across the North Sea following Gustafsson and Andersson (2001). In case of inflow, temperature, salinity, nutrients (phosphate, nitrate, ammonium) and detritus values are nudged towards observed climatological seasonal (winter DJF, spring MAM, summer JJA, autumn SON) mean profiles for 1980–2005 at the monitoring station Å17 located at 58°N 16.5' and 10°E 30.8' in the southern Skagerrak. Nutrient concentrations before 1900 were assumed to be only 85% of present-day concentrations (Savchuk et al. 2008). A linear decrease of nutrient concentrations from 1950 and back in time to 1900 was assumed (Gustafsson et al. 2012). Between 1850 and 1900 nutrient concentrations are constant at the 1900 level. Meier et al. (2018) investigated the importance of the assumption of historical lateral boundary conditions for nutrients and concluded that the sensitivity of the Baltic Sea ecosystem to changes in the lateral boundary condition is relatively small in the model.

Initial conditions

Initial conditions were estimated based upon Savchuk et al. (2008). After a spin-up simulation for 1850–1902 utilizing the reconstructed forcing as described above, the calculated physical and biogeochemical variables at the end of the spin up simulation on 1 January 1903 were used as initial conditions for 1 January 1850.

References

Almroth-Rosell E, Eilola K, Hordoir R, Meier HEM, Hall P (2011) Transport of fresh and resuspended particulate organic material in the Baltic Sea—a model study. J Mar Syst 87:1–12. https://doi.org/10.1016/j.jmarsys.2011.02.005.

Bergström S, Carlsson B (1994) River runoff to the Baltic Sea: 1950–1990. AMBIO 23:280–287.

Cyberski J, Wroblewski A (2000) Riverine water inflows and the Baltic Sea water volume 1901–1990. Hydrol Earth Syst Sci 4:1–11.

Eilola K, Meier HEM, Almroth E (2009) On the dynamics of oxygen, phosphorus and cyanobacteria in the Baltic Sea; a model study. J Mar Sys 75:163–184. https://doi.org/10.1016/j.jmarsys.2008.08.009.

Eilola K, Gustafson BG, Kuznetsov I, Meier HEM, Neumann T, Savchuk OP (2011) Evaluation of biogeochemical cycles in an ensemble of three state-of-the-art numerical models of the Baltic Sea. J Mar Syst 88:267–284. https://doi.org/10.1016/j.jmarsys.2011.05.004.

Graham LP (1999) Modeling runoff to the Baltic Sea. AMBIO 27:328–334.

Gustafsson BG, Schenk F, Blenckner T, Eilola K, Meier HEM, Müller-Karulis B, Neumann T, Ruoho-Airola T, Savchuk OP, Zorita E (2012) Reconstructing the development of Baltic Sea eutrophication 1850–2006. AMBIO 41:534–548. https://doi.org/10.1007/s13280-012-0318-x.

Hansson DC, Eriksson C, Omstedt A, Chen D (2011) Reconstruction of river runoff to the Baltic Sea, AD 1500–1995. Int J Clim 31:696–703. https://doi.org/10.1002/joc.2097.

Meier HEM (2005) Modeling the age of Baltic Sea water masses: quantification and steady state sensitivity experiments. J Geophys Res 110:C02006. https://doi.org/10.1029/2004JC002607.

Meier HEM (2007) Modeling the pathways and ages of inflowing salt- and freshwater in the Baltic Sea. Estuar Coast Shelf Sci 74:717–734.

Meier HEM, Döscher R, Faxén T (2003) A multiprocessor coupled ice-ocean model for the Baltic Sea: application to salt inflow. J Geophys Res 108:3273. https://doi.org/10.1029/2000J C000521.

Meier HEM., K. Eilola, E. Almroth-Rosell, S. Schimanke, M. Kniebusch, A. Höglund, P. Pemberton, Y. Liu, G. Väli, and S. Saraiva, (2018). Disentangling the impact of nutrient load and climate changes on Baltic Sea hypoxia and eutrophication since 1850. Climate dynamics. https://doi.org/10.1007/s00382-018-4296-y.

Mikulski Z (1986) Inflow from drainage basin, in Water Balance of the Baltic Sea-Baltic Sea Environment Proceedings 16, pp 24– 34, Baltic Mar Environ Prot Comm, Helsinki, Finland.Ruoho-Airola T, Eilola K, Savchuk OP, Parviainen M, Tarvainen V (2012) Atmospheric nutrient input to the Baltic Sea for 1850– 2006: a reconstruction from modeling results and historical data. AMBIO 41:549–557. https://doi.org/10.1007/s13280-012-0319-9.

Savchuk OP, Wulff F, Hille S, Humborg C, Pollehne F (2008) The Baltic Sea a century ago—a reconstruction from model simulations, verified by observations. J Mar Syst 74:485–494.

Savchuk OP, Gustafsson BG, Rodriguez-Medina M, Sokolov AV, Wulff F (2012) Nutrient loads to the Baltic Sea 1970– 2006, Baltic Nest Institute Technical Report No.5, Stockholm, Sweden.

Schenk F, Zorita E (2012) Reconstruction of high resolution atmospheric fields for Northern Europe using analog-upscaling. Clim Past 8:1681–1703. https://doi.org/10.5194/cp-8-1681-2012.