OSPAR COMMISSION

ICG-EMO report on model comparison for historical scenarios as basis to derive new threshold values

Hazardous Substances and Eutrophication Series

2022

ICG-EMO report on model comparison for historical scenarios as basis to derive new threshold values

This report was produced by the Intersessional Correspondence Group on Eutrophication Modelling, the authors are:

Hermann Lenhart[1], Anouk Blauw[2], Xavier Desmit[3], Liam Fernand[4], René Friedland[5], Birgit Heyden[6], Onur Kerimoglu[7], Geneviève Lacroix[3], Annelotte van der Linden[2], Johan van der Molen[8], Martin Plus[9], Theo Prins[2], Itzel Ruvalcaba Baroni[10], Tiago Silva[4], Christoph Stegert[11], Daniel Thewes[1], Tineke Troost[2], Lauriane Vilmin[2], Sonja van Leeuwen[8]

1. University Hamburg, *Germany*
2. Deltares - Delft, *The Netherlands*
3. RBINS-OD- Nature - Brussels, *Belgium*
4. Cefas - Lowestoft, *United Kingdom*
5. IOW – Warnemünde, Germany & JRC - Ispra*, EU*
6. AquaEcology, Oldenburg, *Germany*
7. University Oldenburg, *Germany*
8. NIOZ, Texel*, The Netherlands*
9. IFREMER - Brest, *France*
10. SMHI, *Sweden*
11. Helmholtz-Zentrum Hereon, *Germany*

OSPAR Convention

The Convention for the Protection of the Marine Environment of the North-East Atlantic (the "OSPAR Convention") was opened for signature at the Ministerial Meeting of the former Oslo and Paris Commissions in Paris on 22 September 1992. The Convention entered into force on 25 March 1998.

The Contracting Parties are Belgium, Denmark, the European Union, Finland, France, Germany, Iceland, Ireland, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

Convention OSPAR

La Convention pour la protection du milieu marin de l'Atlantique du Nord-Est, dite Convention OSPAR, a été ouverte à la signature à la réunion ministérielle des anciennes Commissions d'Oslo et de Paris, à Paris le 22 septembre 1992. La Convention est entrée en vigueur le 25 mars 1998.

Les Parties contractantes sont l'Allemagne, la Belgique, le Danemark, l'Espagne, la Finlande, la France, l'Irlande, l'Islande, le Luxembourg, la Norvège, les Pays-Bas, le Portugal, le Royaume-Uni de Grande Bretagne et d'Irlande du Nord, la Suède, la Suisse et l'Union européenne.

Table of Contents

Executive summary 6

Récapitulatif 7

1 Introduction 8

1.1 Background for this model study 8

1.2 Outline of the report 10

2 ICG-EMO model comparison study 11

2.1 Setup for model comparison study 11

2.2 Definition of pre-eutrophic scenarios for reference condition 14

3 Analysis of model results for historic scenario as basis for assessment 17

3.1 Horizontal distribution 17

3.2 Time series 21
3.2.1. Time series results from the different scenarios 21
3.2.2. Time series results compared to observations 22

3.3 Interpretation of model results 24

4 Derivation of new threshold values for assessment areas 25

4.1 Threshold values for all COMP4 assessment areas 25

4.2 Weighted model ensemble approach 26
4.2.1 Method for a given area A 26

4.3 Products from the weighted model ensemble approach 28

5 Derived threshold values 46

5.1 Threshold values for all COMP4 assessment areas 46

5.2 Horizontal distribution of new threshold values 48

6 Review of steps towards present assessment 54

6.1 Definition of steps towards present assessment 54

6.2 Evaluation of changes from previous COMP3 to newly proposed threshold values 54

7 Assessment based on newly derived threshold values 56

7.1 COMPEAT assessment based on model derived threshold values 56

7.2 Problems of matching model derived threshold values with values based on in-situ data 59

7.3 Alternative approach of matching modelling and in-situ data based on threshold values 61
7.3.1 Comparison of model outputs under CS and HS1 conditions 61
7.3.2 Estimation of thresholds and assessment boundaries 64
7.3.3 Realtive (Schernewski style) approach 65

7.4 Taylor diagram analysis 68

8 Reflection on ICG EMO approach; dealing with uncertainty **69**

8.1 Plausibility of model results *69*

8.2 Plausibility of pre-eutrophic scenarios *72*

8.3 Model response to nutrient reduction *73*
Memory Effect 73
Baltic boundary condition 74

8.4 Weighted ensemble application and threshold estimates *75*

8.5 Concluding remarks *76*

9 Discussion **77**

9.1 Definition of pre-eutrophic conditions *77*

9.2 ICG-EMO modelling approach *77*

9.3 Definition of threshold levels *78*

9.4 Results for specific areas *79*

9.5 DIN versus DIP *80*

9.6 Other assessment parameters *80*

9.7 Recommendations *80*

9.8 Final remarks *81*

9.9 Highlights *82*

10 Acknowledgements **83**

11 References **85**

ICG-EMO Report - Annex 1: Technical background information on modelling work **89**

Model description **89**

Definition of the scenarios **1**

2) Results of the E-HYPE historic model run **43**

3) Task for OSPAR Contracting Parties **45**

Fluxes of N and P in the Scheldt River proposed in the rationale *48*

Previous modelling studies *48*

Conclusion *48*

Conclusion **55**

Annexes (as contained in the associated zip file)

Annex 1: Technical background information on modelling work

Annex 2: Rationale and description of a pre-eutrophication nutrient scenario to derive harmonized OSPAR assessment levels for eutrophication parameters in Region II

Annex 3: Time series analysis at selected
in-situ observation sites

Annex 4: Model weight Information

Annex 5: Products for selected COMP4 assessment areas

Executive summary

During the third application of the Common Procedure (COMP3, 2017) the need to improve coherence in the eutrophication assessment was recognised. This improvement requires a change from nationally defined assessment areas to cross-boundary and ecologically relevant assessment areas and a harmonized approach to define more coherent threshold levels for nutrients and chlorophyll. Also, a methodology should be developed to harmonize threshold levels of nutrients and chlorophyll, making them more consistent with each other.

During the meeting of HASEC in March 2020, ICG-EMO was asked to set up a modelling approach and define new threshold levels to be used in the next eutrophication assessment (COMP). New, ecologically relevant, assessment areas have been defined, based on the initial work by van Leeuwen et al. (2015) and further work within the JMP-EUNOSAT project, followed by subsequent elaboration by TG COMP.

The model approach by ICG-EMO focuses on the establishment of new threshold levels in each assessment area for the main eutrophication assessment parameters: concentrations of dissolved inorganic nitrogen, dissolved inorganic phosphorus, and chlorophyll. The threshold values have been estimated with a stepwise weighted ensemble approach (Figure 1):

1. Estimate nutrient inputs to the NE Atlantic under pre-eutrophic reference conditions.
2. Estimate marine nutrient and chlorophyll-a concentrations under reference conditions, with an ensemble of models using nutrient inputs from step 1.
3. Estimate average concentrations per assessment area under reference conditions.
4. Calculate the cost function and normalised weights associated with each model and assessment area. This is done on the basis of the ICES *in-situ* observation data and the model results from the Current State simulation.
5. Multiply averaged pre-eutrophic concentration per model with their weights and aggregate the final weighted ensemble means per parameter and assessment area.
6. Derive threshold values as 150% of nutrient and chlorophyll-a concentrations under reference conditions from the weighted ensemble mean.

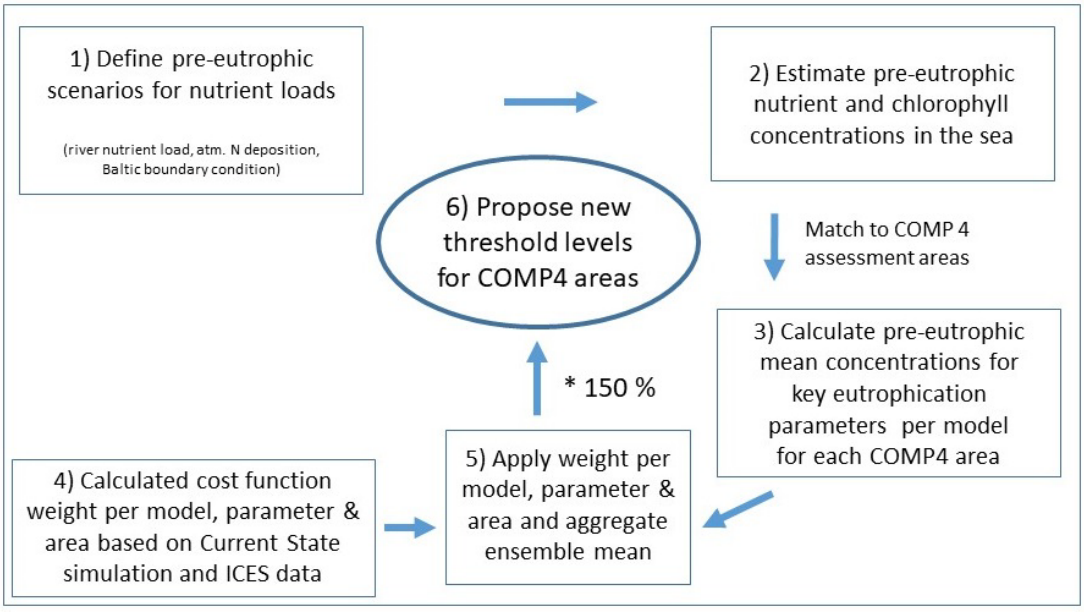


Figure 1: Conceptual diagram to achieve new threshold level for the COMP4 assessment areas.

Récapitulatif

Lors de la troisième application de la procédure commune (COMP3, 2017), la nécessité d'améliorer la cohérence de l'évaluation de l'eutrophisation a été reconnue. Cette amélioration nécessite de passer de zones d'évaluation définies au niveau national à des zones d'évaluation transfrontalières et écologiquement pertinentes, et d'adopter une approche harmonisée pour définir des niveaux seuils plus cohérents pour les nutriments et la chlorophylle. De plus, une méthodologie devrait être développée pour harmoniser les niveaux seuils des nutriments et de la chlorophylle, les rendant plus cohérents entre eux.

La réunion du comité HASEC en mars 2020 a demandé à l'ICG-EMO de mettre en place une approche de modélisation et de définir de nouveaux niveaux de seuil à utiliser dans la prochaine évaluation de l'eutrophisation (COMP). De nouveaux domaines d'évaluation, écologiquement pertinents, ont été définis, sur la base des travaux initiaux de van Leeuwen et al. (2015) et d'autres travaux dans le cadre du projet JMP-EUNOSAT, suivis d'une élaboration ultérieure par le TG COMP.

L'approche de modélisation établie par l'ICG-EMO est axée sur l'établissement de nouveaux seuils dans chaque zone d'évaluation pour les principaux paramètres d'évaluation de l'eutrophisation : les concentrations d'azote inorganique dissous, de phosphore inorganique dissous et de chlorophylle. Les valeurs seuils ont été estimées par une approche d'ensemble pondérée par étapes (figure 1) :

1. Estimer des apports de nutriments dans l'Atlantique nord-est dans des conditions de référence pré-eutrophes.

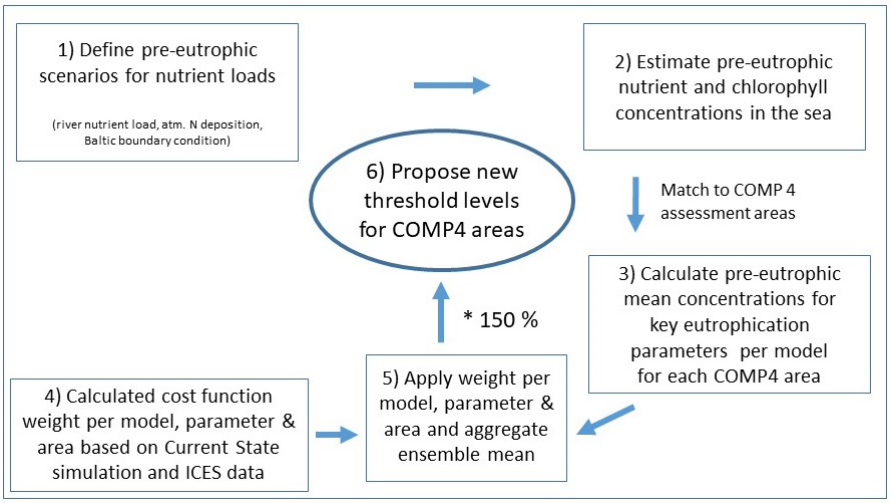
2. Estimer les concentrations de nutriments marins et de chlorophylle-a dans des conditions de référence, avec un ensemble de modèles utilisant les apports de nutriments de l'étape 1.

3. Estimer les concentrations moyennes par zone d'évaluation dans des conditions de référence.

4. Calculer la fonction de coût et les poids normalisés associés à chaque modèle et zone d'évaluation. Ceci est fait sur la base des données d'observation in-situ et des résultats du modèle de la simulation de l'état actuel.

5. Multiplier la concentration pré-eutrophique moyenne par modèle avec leurs poids et agréger les moyennes d'ensemble pondérées finales par paramètre et par zone d'évaluation.

6. Déterminer les valeurs seuils à 150 % des concentrations de nutriments et de chlorophylle-a dans les conditions de référence à partir de la moyenne d'ensemble pondérée



1 Introduction

The objective for the present study is to derive threshold values for the newly defined COMP4 assessment areas based on model simulation from a number of ICG-EMO partners. This report presents the aggregated results, as well as all relevant information.

1.1 Background for this model study

So far, the OSPAR COMP assessment reports have been based on national threshold levels for the eutrophication indicators, as presented in Figure 5.7 of OSPAR's 3rd application of the COMP (OSPAR 2017). A new attempt is needed to define coherent assessment areas in combination with consistent assessment levels. In a first step OSPAR (HASEC meeting in 2019) referred to a model study from the JMP-EUNOSAT project (Enserink et al., 2020; EUNOSAT report). The COMP3 report identified some issues in their approach to assess the eutrophication status in the OSPAR maritime area. Some of these have been addressed in this exercise (and already during the preliminary discussions with OSPAR delegates in Hamburg 2019). For instance, the definition and application of new assessment areas fitting better to natural gradients, freshwater plumes and marine fronts offers more geographical homogeneity in the assessment levels and reduces the risk of contrasting classifications, especially between neighbouring countries. Also, these new assessment areas allow more refined classifications, reducing the issue of large areas being classified as "problem areas" due to a small coastal fraction. This new geographical partition following natural features may also be an example for other assessment procedures with a view to reducing disparities between the Common Procedure and the WFD, or the MSFD, and are based on the ecohydrodynamic regions first described in van Leeuwen et al. (2015)

The new area definition was presented during the combined ICG-EMO and TG-COMP workshop, which resulted in a number of refinements for certain regions based on the description of Contracting Parties. The latest version (v7e) was accepted during the last HASEC meeting in March 2020 [HASEC 2020].

To reach the final goal of deriving ecologically relevant threshold values for the newly defined assessment areas, the ICG-EMO group was asked for support [HASEC 2020]. One key product that needed to be achieved was a model-based identification (as in Schernewski and Neumann, 2005; Kerimoglu et al. 2018; Stegert et al. 2021) of the so-called "reference condition", which provides a representation of the marine environment under historic or pre-eutrophic condition. The new threshold values could be obtained in a purely formal approach by adding an allowable deviation of max. 50% to these reference conditions, which reflects the common practice under OSPAR and the Water Framework Directive (WFD), as described by the WFD CIS Guidance Document No. 5 (2003). However, in the first application, which was presented by ICG-EMO to HASEC in March 2021 in a report, some problems became visible.

The preliminary results suggested lower threshold levels than the levels applied in COMP3 in many assessment areas, particularly for nitrogen and chlorophyll. This is due to the models' assessment based on the whole area, rather than on limited observations biased towards the coast. In contrast to these findings, nutrient threshold increased in some offshore assessment areas.

Therefore, we present in this report a new approach which takes into account a weighted ensemble approach by Almroth and Skogen (2010). In this approach measure the model performance in comparison to the *in-situ* measurements by a cost function calculation. With this approach we take advantage of the fact that it links the model results with the official ICES data that are used for the

assessment within the COMPEAT tool. In the present report an updated data coverage for Chl-a could be used by the inclusion of satellite Chl-a data.

The model approach by ICG EMO focuses on the establishment of new threshold levels for the concentrations of phosphate, dissolved inorganic nitrogen and chlorophyll, to be applied to the new assessment areas. The threshold values have been estimated with a stepwise approach (Figure 1.1):

1. Estimate nutrient inputs to the NE Atlantic under pre-eutrophic reference conditions.
2. Estimate nutrient and chlorophyll-a concentrations under reference conditions, with an ensemble of models using nutrient inputs from step 1.
3. Estimate average concentrations per assessment area under reference conditions.
4. Calculate cost function weight based on the Current State simulation in comparison with the ICES *in-situ* observation data.
5. Multiply averaged pre-eutrophic concentration per model with cost function normalised weights and aggregate the final ensemble means per parameter
6. Derive threshold values as 150% of nutrient and chlorophyll-a concentrations under reference conditions from weighted ensemble mean.

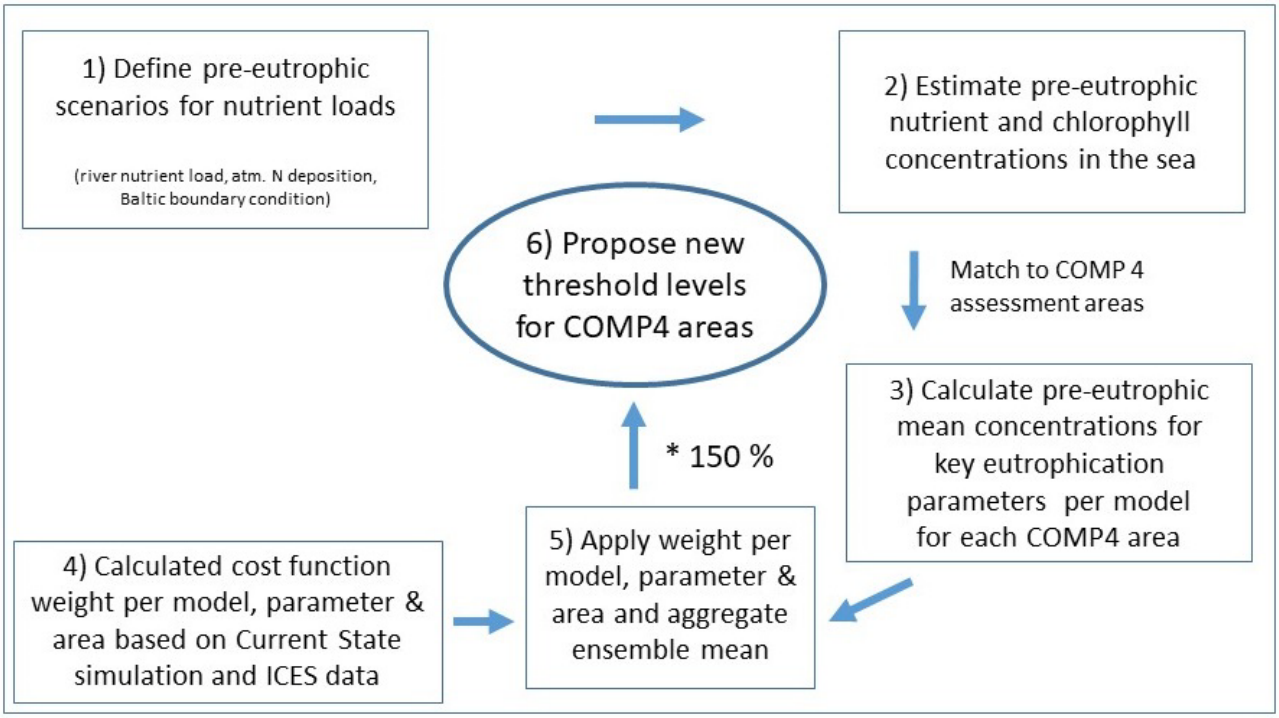


Figure 1.1: Conceptual diagram to achieve new threshold level for the COMP4 assessment areas.

1.2 Outline of the report

Chapter 2 provides an essential overview on the setup that was used for the model comparison study. Chapter 3 gives an overview of central findings from the model output and the analysis of these results. In chapter 4, first the theoretical background of the weighted ensemble approach by Almroth and Skogen (2010) is provided, as well as the practical aspects of the application for this ICG-EMO model study. Then the weighted aggregated ensemble results from the model study are presented for selected COMP4 assessment areas and the resulting threshold levels are indicated. In chapter 5 the resulting new threshold estimates for all COMP4 assessment areas based on the ICG-EMO modelling comparison are presented. In chapter 6 the steps from COMP3 over the JMP EUNOSAT approach towards the present threshold levels are illustrated, and these new threshold values compared against the previous results, like from the COMP3 assessment. Based on these new threshold values an assessment is carried out and the results from the COMPEAT tool is shown in chapter 7.

After the first presentation of the new threshold estimates by ICG-EMO in the HASEC report delivered in March 2021, a number of aspects have been highlighted in the discussion between ICG-EMO and TG-COMP to overcome the mismatch between the regional model averages vs. the biased coastal observation averaging. The steps that ICG-EMO worked on since the HASEC meeting took into account salinity correction, the Schernewski et al. (2015) tried for a Baltic Sea application, the analysis of scatter plots and the test of the weighting routine in relation to a Taylor diagram. At the end of chapter 7, all these approaches are described and evaluated for their possible use within the ICG-EMO model study. With the implementation of the weighted ensemble approach at the HASEC meeting in September 2021, a number of questions were raised mainly on the uncertainty of the model results and the related thresholds. ICG-EMO summarized these questions from these delegates and answered them in the overall context of the model study in chapter 8, like the pre-eutrophic scenario setup, the model quality and model responsiveness and of course the weighted ensemble approach. The report finishes with the discussion of the model results and their analysis in relation to provide new COMP4 assessment threshold estimates in chapter 9.

At the end of the report additional technical information is provided in two Annexes. Annex 1 contains the user guide for the model comparison followed by the model description of the models that take part in the model exercise. Annex 1 also provides an overview of the individual model results for the key eutrophication parameter and the aggregated representation for the COMP4 areas. In Annex 2 all information for the setup of the historic scenarios is aggregated, including the rationale from the Member States as discussed within the pre-eutrophic expert group. The Annexes 3 to 5 from the full report (HASEC HOD (2) 21/2/1 Add.3 to 5) are missing in this shorter version since these Annexes contain supplementary as graphs and tables.

2 ICG-EMO model comparison study

The overall goal of this study is to derive new threshold values for OSPAR assessment in newly defined assessment areas. The first objective is to derive the „reference“ values related to the new assessment areas, where „reference“ refers to the level in a pre-eutrophication period. Estimating pre-eutrophic conditions in marine areas is precisely a task for marine modellers, provided they receive the right inputs for their models. To define the scenario setup (i.e. model inputs) for these historic model simulations, a subgroup of TG-COMP was established, called the “pre-eutrophic group”. They agreed on two historic scenarios, each one with different estimates for the pre-eutrophic nutrient inputs to the sea, mainly from rivers, but also estimates for pre-eutrophic atmospheric N deposition and exchange of nutrients through the Baltic Sea outflow.

The goal should be achieved by a model comparison study from a number of ICG-EMO partners in a combined effort, using a weighted ensemble approach to take into account the strengths of both models and observations (Skogen et al. 2021). The basic idea is to run the different models in a setup which allows for maximum comparability by the use of common forcing (i.e. nutrient inputs), considering the constraints of available time and the financial support that is offered to the different modelling groups.

2.1 Setup for model comparison study

Based on the definition of the historic scenario setups by the pre-eutrophic group the ICG-EMO modelling group needed to plan the practical steps for the model application. The most important part was to achieve a common understanding between the modellers about the general model setup with the aim to achieve comparable model results. Followed by practical considerations like provision of the necessary model forcing, the definition of the parameters that need to be stored for post-processing as well as an EXCEL workbook, developed by Sonja van Leeuwen, to report on the results in a common format. All these details are documented in a user guide that can be found in Annex 1. The goal of this approach is to achieve aggregated estimates for the threshold values for the newly defined assessment areas based on these model applications.

The basic setup follows the modelling steps applied in the JMP EUNOSAT project, as explained in chapter 1.1. Firstly, a simulation covers the current state (CS) conditions for the assessment period 2009-2014. Secondly, the historic scenario runs aim to represent the so-called “historic” or “pre-eutrophic” environmental state of the marine environment. Both historic scenarios are run over 6 years (plus spin-up) with the exact same meteorological conditions as in the current state run for 2009-2014, and by applying estimates for the river nutrient loads, atmospheric N deposition and boundary conditions as delivered by the “pre-eutrophication group”. The results of these scenario simulations represent a “reference” situation.

The historic scenarios (HS1, HS2) consist of two different definitions for the historic conditions. For both scenarios the same assumptions on the atmospheric N deposition and changes for the Baltic boundary condition are applied, whereas the river load information will be different and for each scenario the complete data for the river loads will be provided individually.

HASEC decided that the default scenario HS1 should be the one that was developed in the JMP EUNOSAT project, which is based on estimates from the Swedish E-HYPE catchment model. Since the E-HYPE model has some drawbacks within the P load estimates, a second hybrid scenario with corrections only for P load has been set up and agreed upon by the pre-eutrophic expert group. While

the changes in the P loads from Germany and the Netherlands were based on the historic estimates from the German MONERIS model that are related to individual rivers, Denmark has provided additional national estimates for historic P loads for the coastal areas. An overview on the characteristics of the two scenarios is described in more detail in the following chapter 2.2.

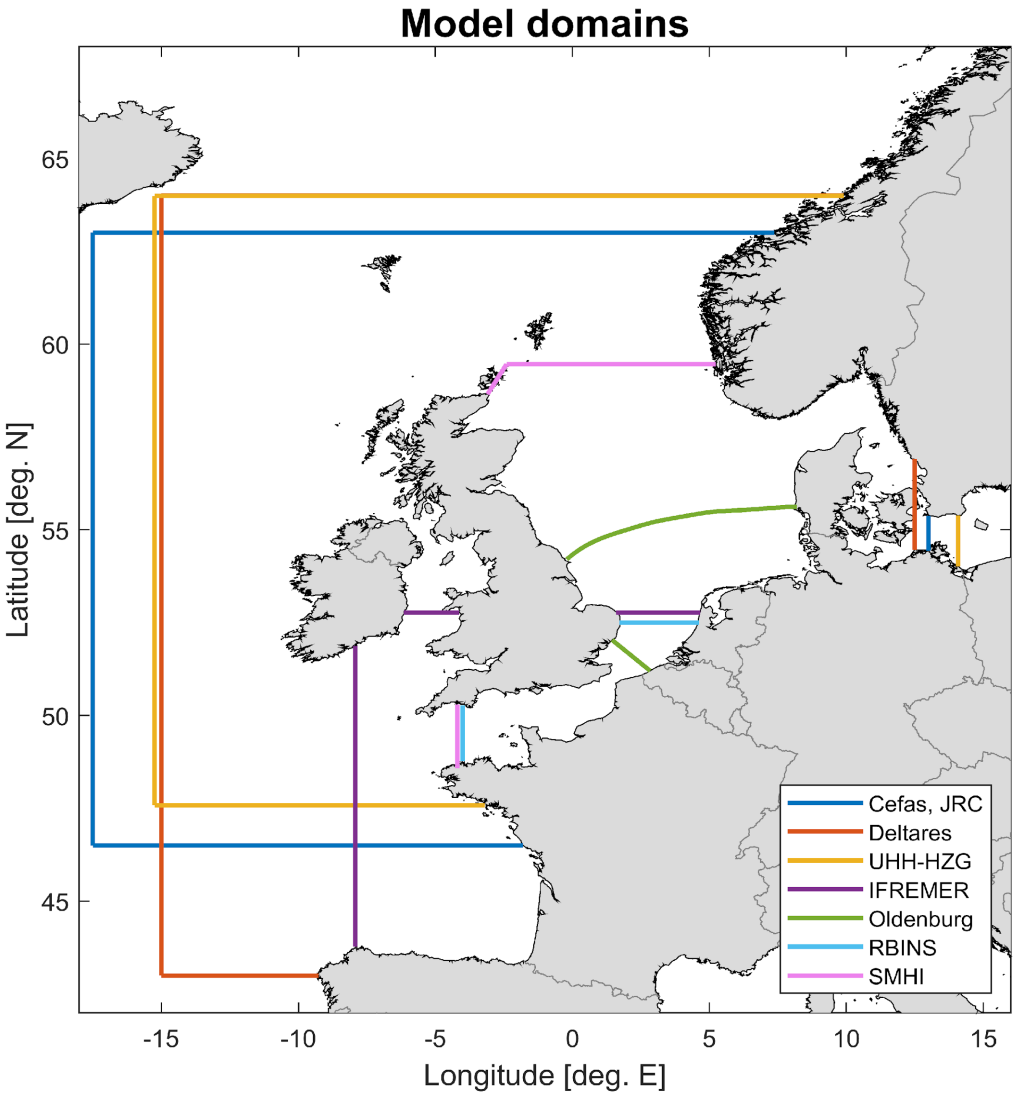


Figure 2.1: Overview of the eight model domains of the ecosystem models that run the historic scenarios.

In total eight models have participated in this model comparison. Sonja van Leeuwen has produced the map (Fig. 2.1) which illustrates the coverage of different model domains. Four models (Deltares, UHH-HZG, Cefas, JRC) cover the whole Northwest Continental Shelf region while three models have a regional focus: IFREMER (Channel and Bay of Biscay), RBINS (Channel) and Oldenburg (Southern North Sea). SMHI is the only institute to use a model domain that includes both the North Sea and the Baltic Sea, and therefore needs no boundary condition for the relevant outflow region to the east.

Model coverage is therefore highest in the Southern Bight region of the North Sea. At the model domain boundaries, assessment areas are not fully represented. All participants provided coverage percentages for all assessment areas. ICG-EMO imposed a limit of 80% coverage, below which model

results for assessment areas were not used in the ensemble approach to calculate the thresholds. The only exceptions were the Atlantic (ATL, now three models in total) area for UHH-HZG and the northern North Sea (NNS, now four models in total) area for SMHI, which have coverages of 67% and 78% respectively. These assessment areas were included despite insufficient coverage to allow for better model ensemble representation in those areas.

Table 2.1: Overview of the model contributions from each partner for the assessment. The list of requested variables was salinity (S), winter DIN (DIN), winter DIP (DIP), winter total N (TotalN), winter total P (TotalP), winter N:P ratio (N:P), summer Chlorophyll-a (Chla), summer Chlorophyll 90th percentile (Chla90th), near-bed O2 (O2), near-bed O2 saturation (O2sat), light extinction coefficient Kd (Kd), Secchi disc depth (Secchi) and net primary production (netPP). All models were asked to simulate 2006-2014 and submit results for 2009-2014, using the period 2006-2008 as spin up time.

Institute	years	Current state	Historic Scenario 1	Historic Scenario 2
Cefas (UK)[1]	2009-2014	DIN, DIP, Chla		
Deltares (NL) CS HS1, HS2	 2009-2014 2009-2013	S, DIN, DIP, Chla, Chla90th, Kd, Secchi, netPP	S, DIN, DIP, Chla, Chla90th, Kd, Secchi, netPP	S, DIN, DIP, Chla, Chla90th, Kd, Secchi, netPP
IFREMER (FR)	2009-2014	S, DIN, DIP, TotalN, TotalP, N:P, Chla, Chla90th, O2, O2sat, Kd, netPP	S, DIN, DIP, TotalN, TotalP, N:P, Chla, Chla90th, O2, O2sat, Kd, netPP	S, DIN, DIP, TotalN, TotalP, N:P, Chla, Chla90th, O2, O2sat, Kd, netPP
JRC (EU)	2009-2014	All variables	All variables	All variables
Oldenburg (DE)	2009-2014	All variables	All variables	All variables
RBINS (BE)	2009-2014	S, DIN, DIP, TotalN, TotalP, N:P, Chla, Chla90th, Kd, Secchi, netPP	S, DIN, DIP, TotalN, TotalP, N:P, Chla, Chla90th, Kd, Secchi, netPP	S, DIN, DIP, TotalN, TotalP, N:P, Chla, Chla90th, Kd, Secchi, netPP
SMHI (SE)	2009-2014	All variables	All variables	All variables
UHH-HZG (DE)	2009-2014	All variables	All variables	All variables

Table 2.1 shows the variables submitted by each institute in time for the presented analysis. Cefas could not submit scenario results due to technical problems, while Deltares provided results for both scenario in a slightly reduced time interval.

[1] Cefas results for the current state are from a previous simulation. As such, the used riverine loads were not identical to those of the presented exercise, though they are very close.

2.2 Definition of pre-eutrophic scenarios for reference condition

The "pre-eutrophic group" (a subgroup of TG-COMP) has prepared and set up the definition of the historic scenarios. They agreed on two scenarios with different estimates for the pre-eutrophic nutrient inputs to the sea, including river loads, atmospheric N deposition and boundary conditions for the Baltic Sea outflow. Under the lead of the ICG-EMO Convenor, Hermann Lenhart, a report was finalized, which describes the basis for the estimates of nutrient inputs into freshwater systems used in the catchment models to estimate the pre-eutrophic riverine nutrient loads to the sea. The report also takes on board the national statements from the Contracting Parties to the overall narrative for the scenarios. In order to provide a clear link to the scenario definition the main report from the pre-eutrophic group is placed within Annex 2.

The first historic scenario is related to the previous work from the JMP EUNOSAT project based on the Swedish E-HYPE model with historic loads estimates for TN and TP. A second scenario was defined as a hybrid approach. In this second scenario, the basic distribution is taken from the first E-HYPE scenario, but since the E-Hype model has some drawbacks, especially within the P load estimates, the second hybrid scenario includes corrections in the P load only. While the changes in the P loads from Germany and the Netherlands were based on the historic load estimates from the German MONERIS model, which were related to individual rivers, Denmark has provided additional national estimates for historic P loads for their coastal areas. This setup was also supported by the ICG-EMO model community since the focus of only changing the P load estimates offers better comparability between the model studies. Validation results of the E-HYPE model for N, P concentrations are available only for the Baltic so far, see Capell et al (2021).

Table 2.2: Estimates of pre-eutrophic condition for a selection of individual rivers for scenario HS1 (TN and TP) and the 2nd scenario (TP only). When the 2nd scenario has different TP loads, these are highlighted in bold numbers (taken from Tab. 4 in Annex 2). TN and TP loads show the historic river loads expressed as % of the current (2009-2014) loads.

Contracting Party	**River**	**TN load (%)** Scenario HS1	**TP load (%)** Scenario HS1	**TP load (%) Scenario HS2**
Belgium	IJzer	23	61	61
Belgium	Gent-Oostende Canal	17	76	76
Belgium	Schipdonk Canal	25	49	49
Belgium	Leopold Canal	25	49	49
Denmark	Omme	30	38	**36**
Denmark	Skjern	30	38	**36**
Denmark	Stora	32	44	**36**
Denmark	Vida	30	30	**36**
France	Seine	45	71	71
France	Loire	50	92	92
France	Garonne	70	74	74
France	Dordogne	57	82	82
Germany	Elbe	51	95	**26**

Contracting Party	**River**	**TN load (%)** Scenario HS1	**TP load (%)** Scenario HS1	**TP load (%)** **Scenario HS2**
Germany	Ems	26	60	**17**
Germany	Weser	37	74	**24**
Germany	Eider	23	73	**8**
Ireland	Blackwater	35	55	55
Ireland	Suir	34	57	57
Ireland	Barrow	34	57	57
Ireland	Boyne	31	50	50
The Netherlands	Meuse	38	44	**32**
The Netherlands	Rhine	43	72	**32**
The Netherlands	Lake IJssel East	22	34	**33**
The Netherlands	Lake IJssel West	21	21	**33**
The Netherlands	North Sea Canal	30	27	27
The Netherlands	Schelde	46	81	81
Norway	Glomma	44	50	50
Norway	Skien	47	76	76
Norway	Otra	48	91	91
Norway	Kvina	37	80	80
Spain	Deba	44	34	34
Spain	Oiartzun	31	21	21
Spain	Urola	44	34	34
Spain	Urumea	31	21	21
Sweden	Gota alv	56	62	62
Sweden	Lagan	48	57	57
Sweden	Nissan	48	45	45
Sweden	Atran	48	66	66
United Kingdom	Tweed	56	83	83
United Kingdom	Humber	34	33	33
United Kingdom	Thames	35	38	38
United Kingdom	Tay	63	100	100

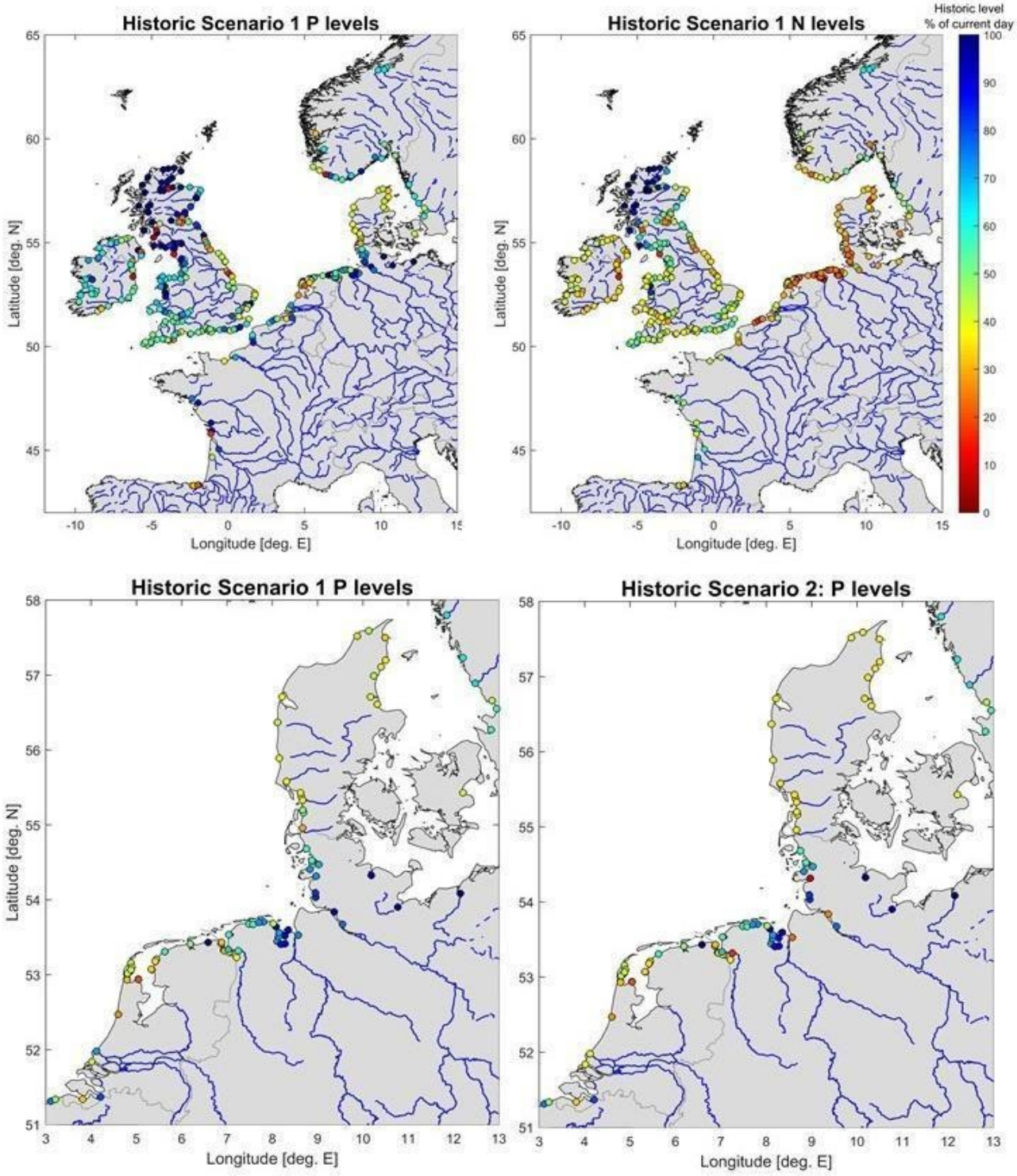


Figure 2.2: Overview of the two historic scenarios. The upper row shows loads for P (top left) and N (top right) in the 1st scenario (as a % of current loads). Since only the P levels are reduced for the 2nd hybrid scenario, the lower row provides an overview of the P loads in both scenarios (as a % of current loads) with focus on the Continental Coastal area.

Sonja van Leeuwen has illustrated the loads of the individual rivers related to these two scenarios (Fig. 2.2). The upper part illustrated the level for P and N loads for the 1st scenario. Since the 2nd scenario only has lower P loads at the coast of the Netherlands (e.g. Rhine from 72% to 32% of current loads), Germany (e.g. Elbe 96% to 26% of current loads) and Denmark (e.g. Stora 44% to 36% of current loads) this area is highlighted within Fig.2 (lower row). More detailed information on the change in the P loads from scenario HS1 to scenario HS2 is provided in Table 2.2.

3 Analysis of model results for historic scenario as basis for assessment

To illustrate the model results, we provide an overview of the model results in this chapter. We show results for the current state (years 2009-2014; abbreviated as CS) that was used for validation and the two historical scenarios HS1 and HS2. We also show comparisons between the results for CS, HS1 and HS2.

3.1 Horizontal distribution

In the JMP-EUNOSAT project (Enserink et al., 2019, Blauw et al., 2019), Deltares estimated threshold values for eutrophication in the North Sea with a similar approach as the current ICG-EMO initiative, but with a simplified modelling approach. The underlying hydrodynamic transport model was the same as in the current ICG-EMO modelling work, but only simulated the transport of nutrients and did not include any biochemical processes (like primary production, mineralization, denitrification, etc.). Chlorophyll concentrations were estimated from a linear relation between winter nutrient concentrations and summer chlorophyll derived from statistical relations based on *in situ* data. The use of the more advanced Deltares 3D biogeochemical model in the current model study (that includes the biochemical processes and 3D concentrations profiles along the oceanic model boundaries) enabled us to better reproduce winter DIN and DIP concentrations. DIP concentrations in coastal waters, particularly in the German Bight, are overestimated. The spatial pattern of chlorophyll concentrations now shows the effect of lower concentrations in the stratified areas of the central North Sea (Figure 3.1), but overall chlorophyll concentrations are overestimated in these areas. Chlorophyll concentrations in the southern North Sea and the Wadden Sea are better reproduced with the more advanced model.

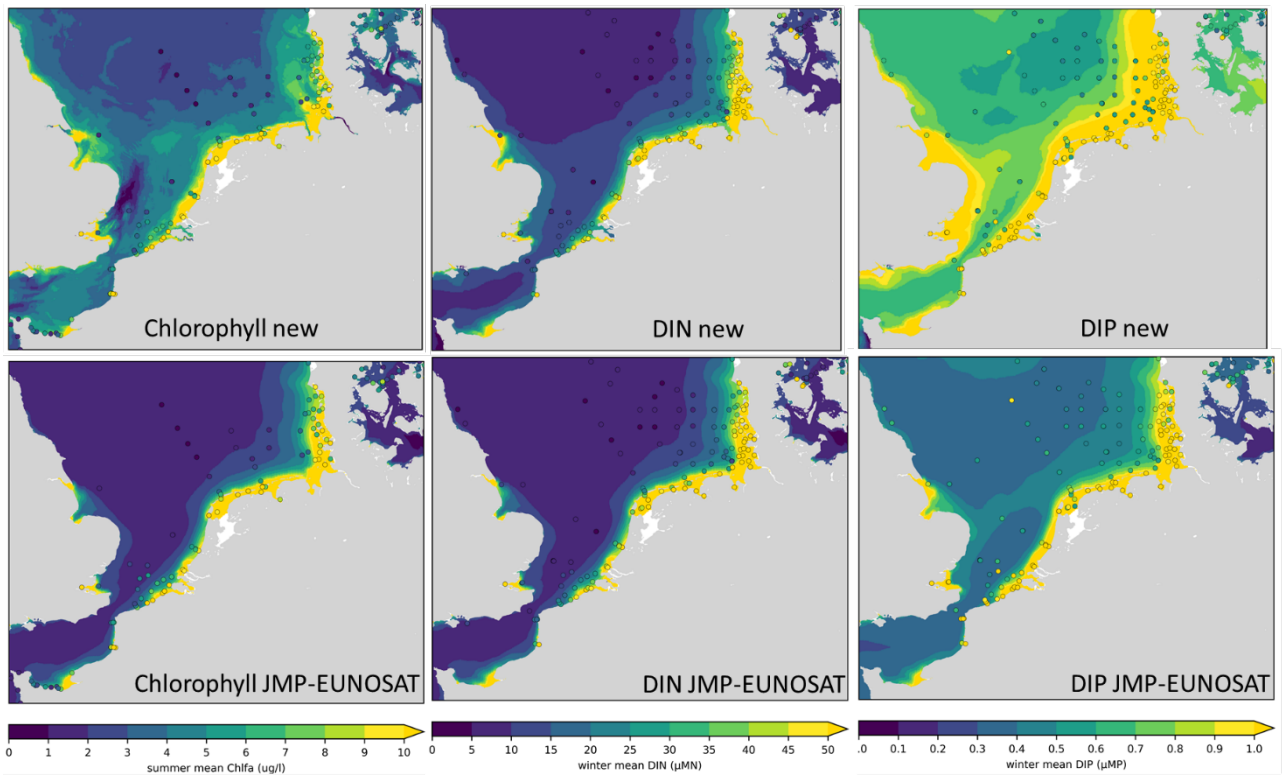


Figure 3.1: Comparison of Deltares model simulation from the present model application (upper row) in comparison to the JMP-EUNOSAT application, based on tracer model results (lower row). The horizontal plots show the mean spatial distribution for Chlorophyll-a, DIN and DIP in comparison to validation data (coloured dots) for the years 2009-2012.

Model results for the reference scenario HS1 show large reductions in winter DIN concentrations compared to recent years: over 50% in the regions of freshwater influence stretching along the Belgian up to Danish coasts (Figure 3.2, upper panels). Winter DIP concentrations under the reference scenario HS1 are around 25% lower than current concentrations, while the Elbe plume shows a lower reduction for DIP than surrounding coastal waters. The reduction in chlorophyll concentrations under scenario HS1 is considerably less than the reduction in DIN concentrations: around 25% compared to the current state. In the JMP-EUNOSAT project it was assumed that chlorophyll concentrations scaled linear with winter DIN concentrations, leading to reductions in chlorophyll proportional to reductions in DIN. Therefore, in the current results the chlorophyll concentrations in scenario HS1 are higher than those estimated by JMP EUNOSAT. The cause for a less than proportional decrease in chlorophyll compared to DIN is that phytoplankton growth is not only controlled by nitrogen but also by the availability of phosphorus and light. Also, phytoplankton can adapt to nitrogen limiting conditions by reducing their nitrogen content.

Comparing the results between the two reference scenarios (Fig 3.2, lower panels) shows strongest differences in the Elbe plume. In scenario HS2, winter DIP concentrations are 50 to 75% lower, compared to scenario HS1. Chlorophyll concentrations are 25 to 50% lower in the Elbe plume. Also, in the Rhine plume and Ems plume, DIP and chlorophyll concentrations are further reduced in the scenario HS2, but less than in the Elbe plume.

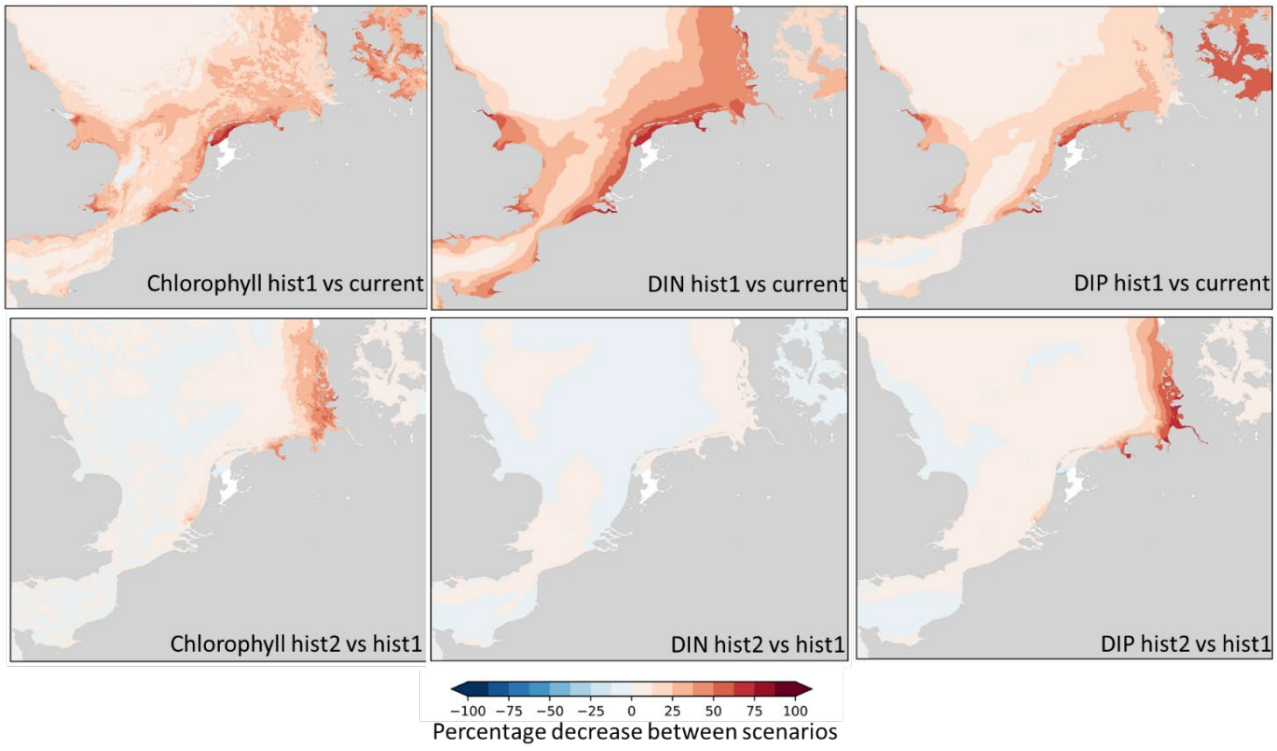


Figure 3.2: Results from the Deltares model for the year 2009. The upper row displays the percent decrease in the scenario HS1 compared to the current state run (red: concentrations HS1<CS; blue: concentrations HS1>CS), while the lower row shows the difference between scenarios HS1 and HS2 (red: concentrations HS2<HS1; blue: concentrations HS2>HS1).

The MIRO&CO model from RBINS (BE; Dulière et al. 2019) focuses on the Channel region and the Southern Bight of the North Sea. In this model the ICES observations for winter DIN are reasonably well reproduced in the simulations of the current state (Figure 3.3, upper panels). Concentrations of winter DIP and summer chlorophyll are underestimated in Belgian and Dutch coastal waters.

Model results for the reference scenario HS1, compared to recent years, show reductions in winter DIN and DIP concentrations similar to the Deltares model results along the Belgian and Dutch coasts (Figure 3.3, middle panels). The reduction in summer mean chlorophyll concentrations seems to be proportional to the reduction in nutrient concentrations in the southern North Sea. Although not directly visible on these graphs, it has been shown that winter DIP tends to limit the size of the spring bloom of chlorophyll in the coastal waters of the southern North Sea (Billen et al. 2011, Desmit et al. 2015), while N loads tend to control the species succession (Lancelot et al. 2007, Desmit et al. 2018) and are more correlated with annual mean chlorophyll (Loebl et al. 2009). These aspects are somewhat merged into the figures 3.2 and 3.3. The model results for the reference scenario HS2 (Figure 3.3 lower panels) differ only from the first reference scenario in the Rhine plume area, which is the only river in the model domain with changed P loads. This difference in P loads induces a seemingly proportional difference in chlorophyll within the Rhine plume, which underlines the role of P loads on the mean chlorophyll accumulation.

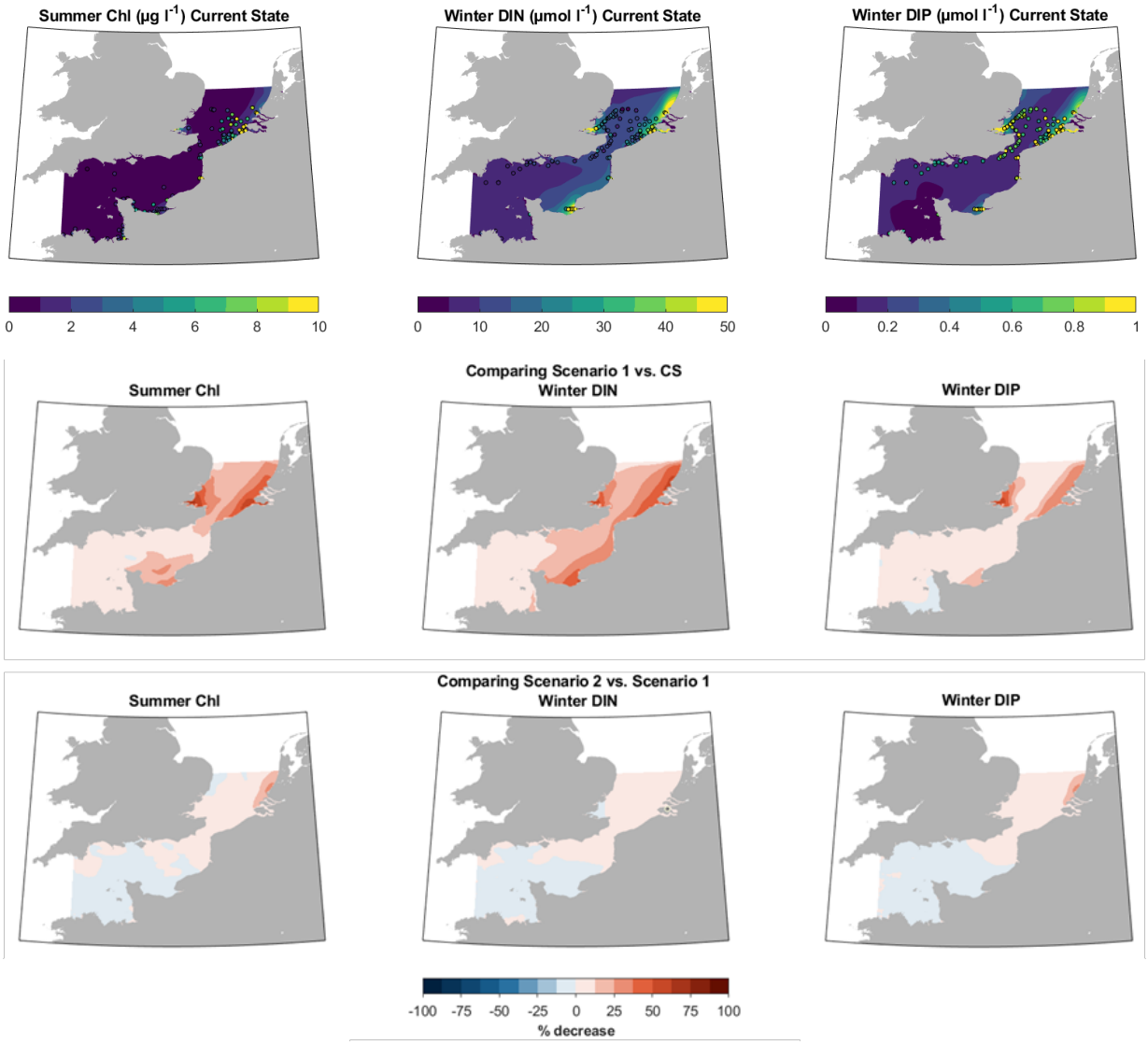


Figure 3.3: Results from the MIRO&CO model (RBINS, BE) for the period 2009-2014. The upper row presents the horizontal distribution for Chlorophyll-a, DIN and DIP concentrations in the CS run, in comparison to validation data for the years 2009-2014. The middle row shows the relative difference between scenario HS11 and CS, and the lower row the comparison between scenario HS2 and CS (red: concentrations HS1<CS; blue: concentrations HS1>CS).

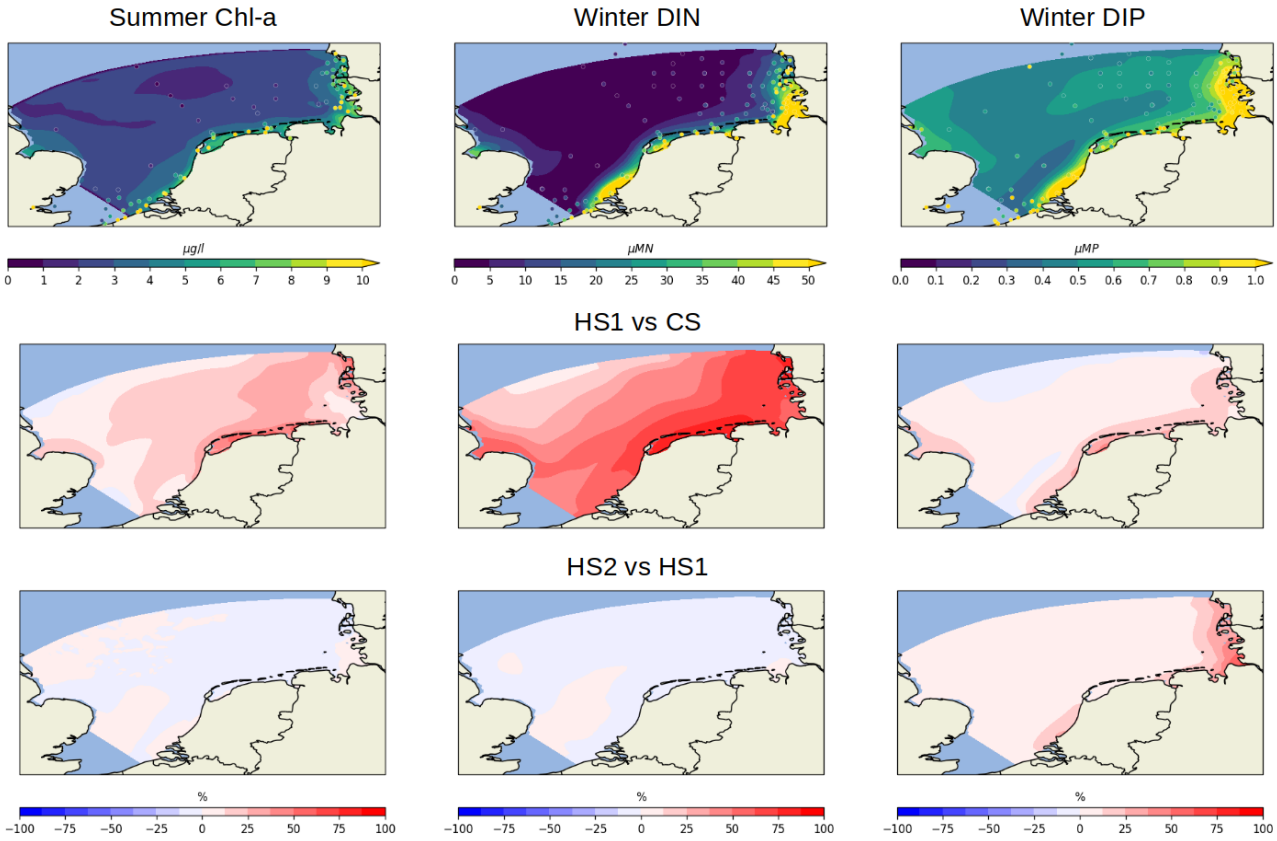


Figure. 3.4: Results from the GP model. The upper row presents the horizontal distribution for DIN, DIP and Chlorophyll-a concentrations obtained with the CS run, in comparison to validation data for the years 2009-2014. The middle row displays the percent difference between the scenario HS1 vs. the current state run, while the lower row shows the difference between the scenarios HS1 and HS2 (red: concentrations HS1<CS; blue: concentrations HS1>CS).

The GPM model (Kerimoglu et al., 2020) has its focus on the southern North Sea (Fig. 3.4). Spatial patterns of winter DIN in coastal waters are fairly similar to the results of the RBINS and Deltares models. Concentrations of winter DIP are higher than the other two models and therefore correspond better to observations. Summer chlorophyll concentrations show less steep gradients from nearshore to offshore than the Deltares model. The historic scenarioHS1 shows similar reductions in DIN and chlorophyll as the Deltares model: with relatively low reference concentrations of chlorophyll in the Elbe plume. Reference scenario HS2 also leads to strongest reductions of DIP in the Elbe plume. But contrary to the Deltares model this does not lead to strong reductions in chlorophyll concentrations.

An overview of the individual model results for the key eutrophication parameter for both historic scenario and the aggregated presentation for the COMP4 areas is displayed at the end of Annex1.

3.2 Time series

3.2.1. Time series results from the different scenarios

Time series illustrate the different results of the two reference scenarios in the Rhine plume (Noordwijk) and Elbe plume (Süderpiep) (Fig. 3.7). In the Rhine plume the blue and green line do not differ much. In the Elbe plume, on the contrary, reference scenario HS1 (blue line) is fairly similar to the current situation (red line) whereas reference scenario HS2 (green line) is much lower for chlorophyll. The time series for DIN at Süderpiep show that DIN concentrations are less depleted in spring under reference scenario HS2 than in scenario HS1, presumably because further phytoplankton growth is limited by phosphorus limitation in HS2. In the simulation of the current state (red line) DIN concentrations are insufficiently depleted in spring, compared to observations.

Time series of chlorophyll at Süderpiep from the ECOHAM model show different patterns than the Deltares model (Figure 3.8). In this model both reference scenarios show similar seasonal patterns as the current state simulation. Time series of DIN also show stronger depletion of DIN concentrations in spring in scenario HS1 than in scenario HS2, similar to the Deltares results.

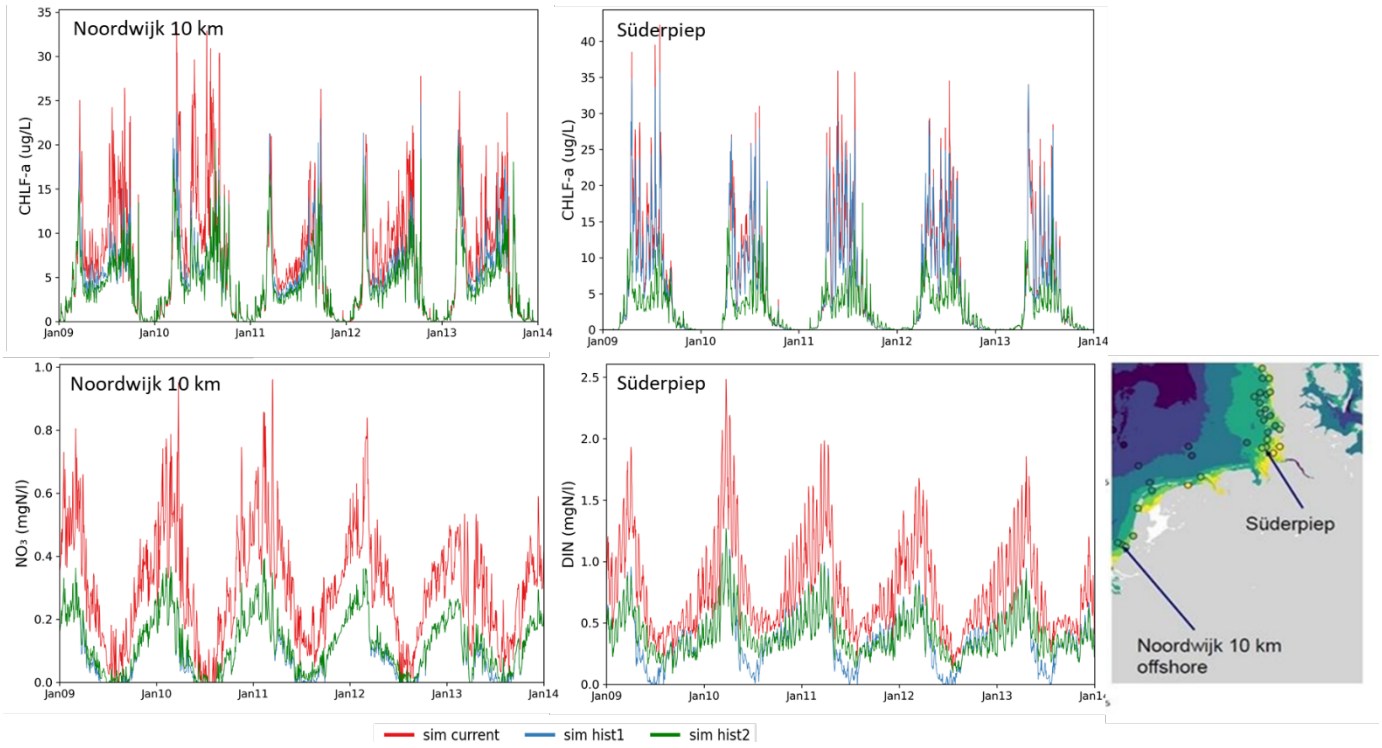


Figure 3.7: Concentration time series for Chlorophyll-a (upper row) from the Deltares model for 2009 and 2010, for the two stations Noordwijk 10 and Süderpiep for the CS run and the two scenarios.

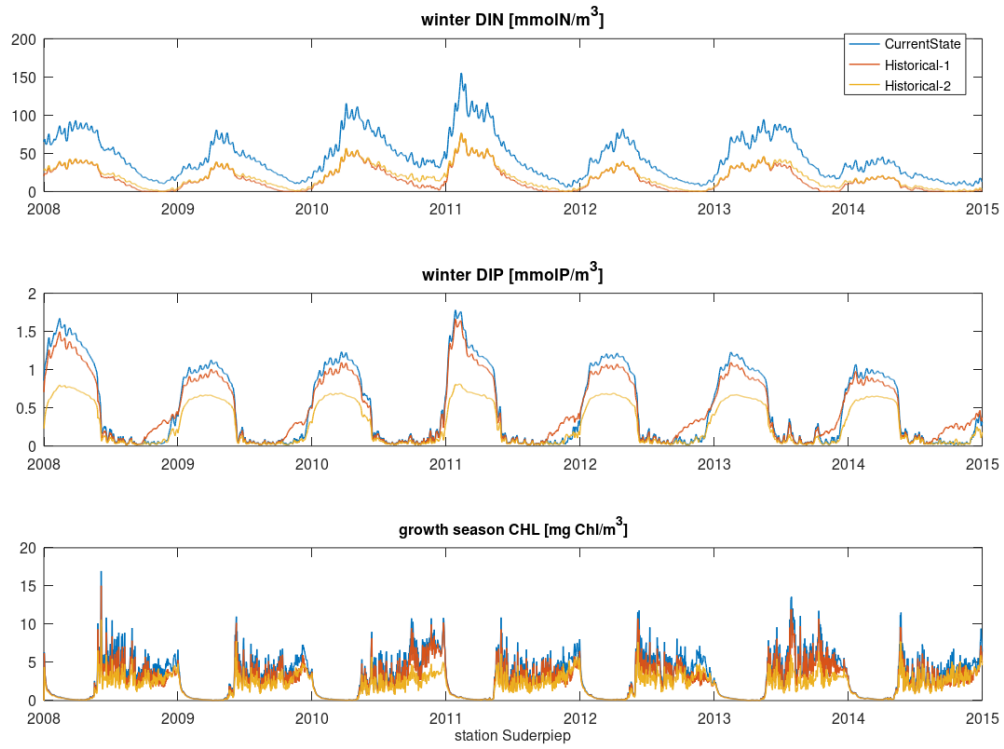


Figure 3.8: Concentration time series for DIN, DIP and Chlorophyll-a from the ECOHAM model for station Süderpiep for the CS run and the two scenarios.

3.2.2. Time series results compared to observations

Based on observation data available at the COMPEAT tool, we selected a comprehensive set of stations. Stations were chosen considering the number of observations and to cover as much of the OSPAR area as possible.

For the comparison of model results with data it is problematic that many stations are located quite near-shore, while for most parts of open sea areas insufficient observations are available. ICG-EMO asked the TG-COMP delegates at several meeting for a selection of COMP4 assessment areas and related in-situ observation sites, for which time series could be derived for validation. The selection by the TG-COMP delegates, including the rationale behind the selection, was described in the document TG-COMP(4) 21/1/info.1.

For the selected stations, a comparison of the time series from the observations with the different model outputs with respect to DIN, DIP and chlorophyll-a was done and basic model quality statistics were computed for each station. Further, the multi-annual monthly mean values were computed, allowing to compare the annual cycles between models and observations. As an example, Figure 3.9 shows the results for station TERSLG50 for DIN.

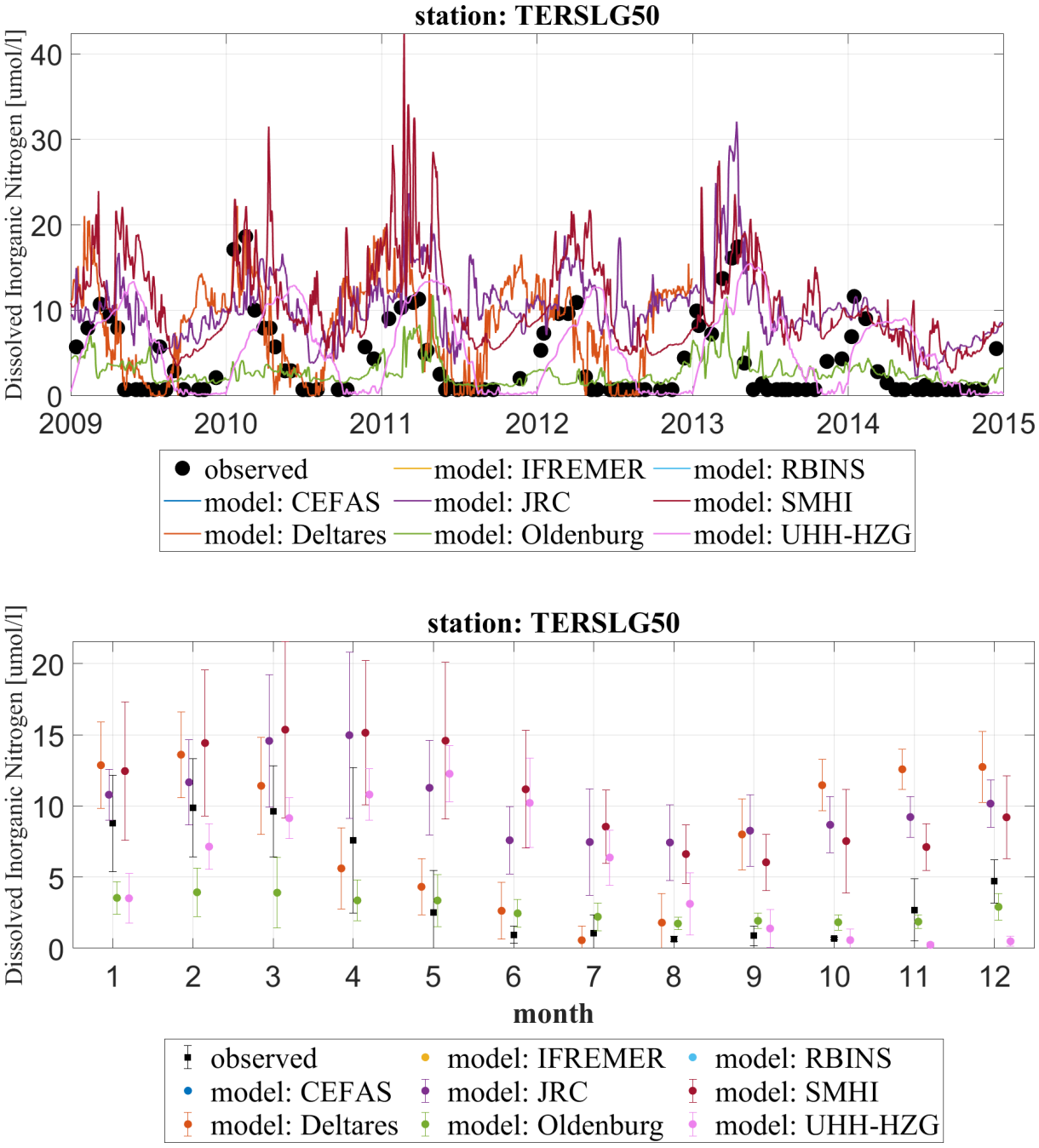


Figure 3.9: Comparison of observed DIN concentrations with the different model results (note that not all models are covering this station) at station TERSLG50 showing the full time series (top panel) and the multi-annual monthly means (lower panel)

Finally, scatter plots were produced, showing the quality of each model at each station for each of the three variables DIN (Fig. 8.1), DIP Fig. 8.2 and Chl (Fig. 8.3). The approach was following Edman & Omstedt (2013), by using a combination of the correlation factors and the RMSD (scaled by the standard deviation) between model results and observations. Thereby, values near to (0, 0) indicate best model results. More explanation is provided at the end of chapter 8 on the Taylor approach.

3.3 Interpretation of model results

The model results described above represent only a first preliminary analysis of a selection of the models involved in the ensemble of model simulations involved in this study. This preliminary analysis indicates that spatial patterns of observed winter DIN concentrations are generally well reproduced: with highest concentrations in river plumes and low concentrations offshore. Spatial patterns of observed winter DIP concentrations are reproduced less well: with underestimations particularly in offshore waters. The validation results for chlorophyll differ between regions and models and require more elaborate analysis. The response of summer chlorophyll to reductions in nutrient loads under reference conditions differs between models and areas. In the model results described above the response to reduction in nitrogen loads leads to a proportional decrease in chlorophyll concentrations in one model and in another to roughly 50% lower reductions. Also, the response to reductions in phosphorus loads (comparing scenarios 1 and 2) differs between models: in one model this does not result in reductions of chlorophyll concentrations whereas in other models the same reduction in phosphorus loads lead to considerable reductions in summer chlorophyll. These results indicate that there is considerable uncertainty about the nutrient and chlorophyll concentrations under reference conditions.

Reference scenario HS1 shows relatively low reductions in DIP concentrations in the Elbe plume compared to the current situation and surrounding coastal waters. Reference scenario HS1 is based on results of the catchment model E-HYPE for the year 1900. Reference scenario HS2, based on the catchment model MONERIS for the year 1880, shows much stronger reductions in DIP concentrations, particularly in the Elbe plume as this scenario has much lower riverine P loads in the German rivers. The model results show that the differences between the two reference scenarios are strongest in the Elbe plume. Therefore, we recommend to further compare the model assumptions in the Elbe catchment between the two model applications to assess which approach best approximates the 'pre-eutrophic' conditions in a coherent way with other areas. Possibly the E-HYPE simulation assumes that the Elbe river was already heavily affected by eutrophication around 1900.

4 Derivation of new threshold values for assessment areas

In this chapter the results from a detailed analysis for the CS run, based on yearly values compared to in-situ data are presented. We also show results for the two historic scenarios from the different models using a common format. Here we only provide a short overview for a few areas to illustrate how the weighted ensemble approach works. Before we present these products, we provide the theoretical background of the weighted ensemble approach by Almroth and Skogen (2010), as well as the practical aspects of the application for this ICG-EMO model study which finally led to the threshold estimates. As these newly derived thresholds are related to the new COMP4 assessment area these will be introduced first. The analysis was carried out by Sonja van Leeuwen, based on the return information on the model statistic as introduced into the EXCEL workbook (for more details see Annex 1).

4.1 Threshold values for all COMP4 assessment areas

Below we provide an overview of the COMP4 assessment areas (Fig 4.1) for better orientation.

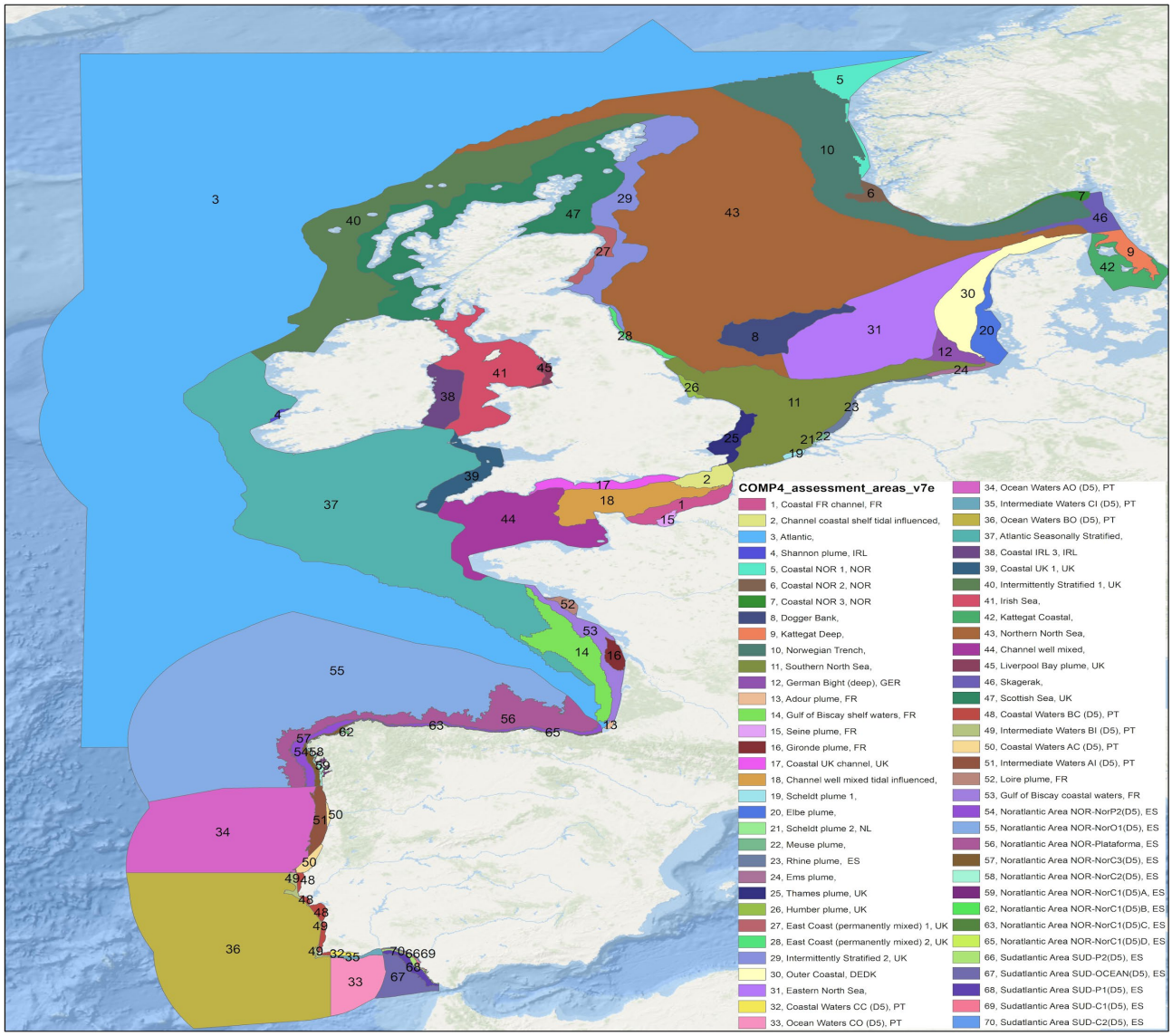


Figure 4.1: Overview on COMP4 assessment areas (internal numbering from ICG-EMO)

4.2 Weighted model ensemble approach

We want to determine new thresholds per parameter and per area on the basis of model results for the Historic Scenario (HS1), as illustrated in Fig 1.1. However, models perform differently in the areas. To use both models and observations, benefitting from the strength of both (Skogen et al. 2021), we use the Weighted Model Average (WMA) approach derived from Almroth and Skogen (2010). This gives a weight to each model in each area, with the weight calculated as the difference between the mean concentration derived from the model and the mean concentration from the measurements, divided by the standard deviation from the measurements. The weight is based on the information of the Current State simulation, and the weighted ensemble approach provides a more stable representation of this Current State than the one based on the unweighted one (direct calculation of the threshold levels from mean values from all the ensemble members).

The problem when applying this weighted approach for the ICG-EMO study is that we do not have observations corresponding to the two pre-eutrophic scenarios. The solution is to compute cost functions and weights on the basis of the Current State run (CS), and then apply them on pre-eutrophic model results to compute the WMA.

4.2.1 Method for a given area A

The basic assumption is that there is "no best model" as all models have their own strengths and weaknesses. The model quality will depend on the area, the time and the actual state variables under considerations. Almroth and Skogen (2010) referred to an ASMO workshop in 1996 as an example for a previous application. In addition they pointed out that weighted ensembles are generally an improvement from unweighted ones and are more stable than the individual ensemble members. However, a big challenge is to find enough observations to enable a meaningful computation of the weights. Here the ICES data used in the COMPEAT tool seems to be a valid platform to be used for this exercise. It should be pointed out that for this report the database of chlorophyll-a observations was extended by the inclusion of satellite data. To consider the quality of existing chlorophyll-a measures in comparison to satellite EO data, the confidence rating which is implemented in the COMPEAT tool was applied. The combination of *in-situ* data and EO data therefore was dependent on confidence classes: If the combined temporal and spatial confidence of *in-situ* Chl-a is high, 50:50 (*in-situ*/EO) is used. If confidence is moderate, then 30:70 (in-situ/EO), if confidence is low, then 10:90 (*in situ*/EO) (Leujak & Heyden, personal communication).

Consider a parameter *P* (e.g. annual Chl) in area *A*. We have observations for *P* in that area (annual mean and annual standard deviation) and also model values for *P* from *N* different models for the CS simulation.

1. We agreed to work on the values averaged over the period 2009-2014 and NOT on the yearly mean values. This is because we have not enough observations for the latter option.

2. We agreed to consider model values only when the model covers at least 80% of the surface of the Area *A*. When it is not the case, the model value is not used.

3. Compute the cost function *C* for parameter *P* and model *i* :

$$C_i^P = \left| \frac{mean\ P_i^{model\ in\ CS} - mean\ P^{obs}}{std\ P^{obs}} \right|$$

4. Observation data is paramount in this exercise. When there is data for the mean and not for the standard deviation (std), one can recalculate the std at the cost of an assumption. The assumption is: the distribution of a parameter value in one area is close to the mean distribution of that parameter values across all areas. Through this, one may thus assume that the ratio std/mean can be averaged across all areas and then used with the mean value in a specific area to re-assess the corresponding std. Consider area A in an ensemble of R areas:

$$std\ P_A^{obs} = \frac{mean\ P_A^{obs}}{R} * \sum_{i=1}^{R} \frac{std\ P_i^{obs}}{mean\ P_i^{obs}}$$

5. The weight (*W*) associated to parameter *P* and model *i* is:

$$W_i^P = \frac{1}{C_i^P + B}$$

Where *B* is an arbitrary constant taken as 0.1 by Almroth and Skogen (2010) to avoid *W* to go to infinity when *C* is very small (i.e., when the model performs really well). If there is no observation in area *A*, then W_i^P = 1 in that area.

6. We can calculate such a weight W_i^P for each model on the basis of the CS run. The weight of an individual model is normalized with the number of models involved in area *A*:

$$Wnorm_i^P = \frac{N}{\sum_{i=1}^{N} W_i^P} * W_i^P$$

7. Then, we can use these normalized weights to calculate the weighted mean in area *A* (WMA) on the basis of the HS1 run:

$$WMA_{HS1}^P = \frac{1}{\sum_{i=1}^{N} Wnorm_i^P} * \sum_{i=1}^{N} \left(Wnorm_i^P * mean\ P_i^{model\ in\ HS1} \right)$$

Without observation, *WMA* is simply the average of model values.

8. The WMA_{HS1}^P is the best estimate of parameter *P* in area *A* under the historic conditions (i.e., pre-industrial) taking into account the performances of all models in that area. We use that value as a basis to calculate the new threshold:

$$New\ threshold\ for\ P\ in\ area\ A = WMA_{HS1}^P * 1.5$$

9. Future discussions should address the validity – or legitimacy – of applying a factor 1.5 (i.e., +50%) upon the pre-industrial values given the accuracy of the new method determining these pre-industrial values. It may be sensible to consider instead a factor based on the spread of the models around the multi-model mean or around the observation value in each area.

This approach implies the approximation that we can neglect the differences in the model responsiveness to the two historic scenarios in this study, since we only rely on the cost-function calculated for the Current State (CS) run.

This approach implies the approximation that we can neglect the differences in the model responsiveness to the two historic scenarios in this study, since we only rely on the cost-function calculated for the Current State (CS) run.

This approach implies the approximation that we can neglect the differences in the model responsiveness to the two historic scenarios in this study, since we only rely on the cost-function calculated for the Current State (CS) run.

4.3 Products from the weighted model ensemble approach

The products Fig. 4.2 to 4.4 are displayed in an identical fashion, with different parameters involved, like Fig.4.2A, Fig 4.3A and Fig 4.4A display the model analysis and the derived thresholds for the parameter Chl mean for three different areas. In a similar way Fig 4.2B-4.4B display the same graphs for the parameter winter DIN and Fig 4.2C-4.4C for winter DIP, respectively.

In each of these figures the order is the same, starting with the upper graph (A) which represents the mean concentration for the individual years displayed against the ICES data which are used within the COMPEAT tool for the assessment. First this provides an overview of the simulated model concentrations in comparison to the *in-situ* data per year. This display also provides an overview on the variability between the years, both for the observation and the model results from the CS run. Second it also highlights the availability of in-situ data entering the COMPEAT tool for the time interval from 2006 – 2014, the period of the assessment, including the number of observations which form the basis for this assessment.

The middle graph (B) displays the mean concentration over the simulation period of the parameters based on the individual model results, as well as the mean of all models, including the standard deviation, in relation to the mean of the observation for this period. In addition, the weighted mean is displayed with STD in order to display the impact of the weighting approach on the resulting model concentration. One has to note that this is an illustration of the effect of the weighting on the resulting concentration of the CS run, whereas the threshold calculation are based on the weighted historical concentrations from the scenario runs HS1 and HS2. For the parameter Chl mean (Fig. 4.2A – 4.4A), the additional information on the impact of the inclusion of the satellite Chl data is provided. Here the mean of the observation only related to in-situ data and only with EO data is displayed in addition to the overall observational mean.

The lower graph (C) represents the mean concentration from the models integrated over the simulation period 2009 to 2014. In addition, the historic concentration for the scenario HS1 and HS2 is displayed for each model that provides information to this area. Finally, the thresholds based on these historic concentrations are displayed as a result of the weighted ensemble approach.

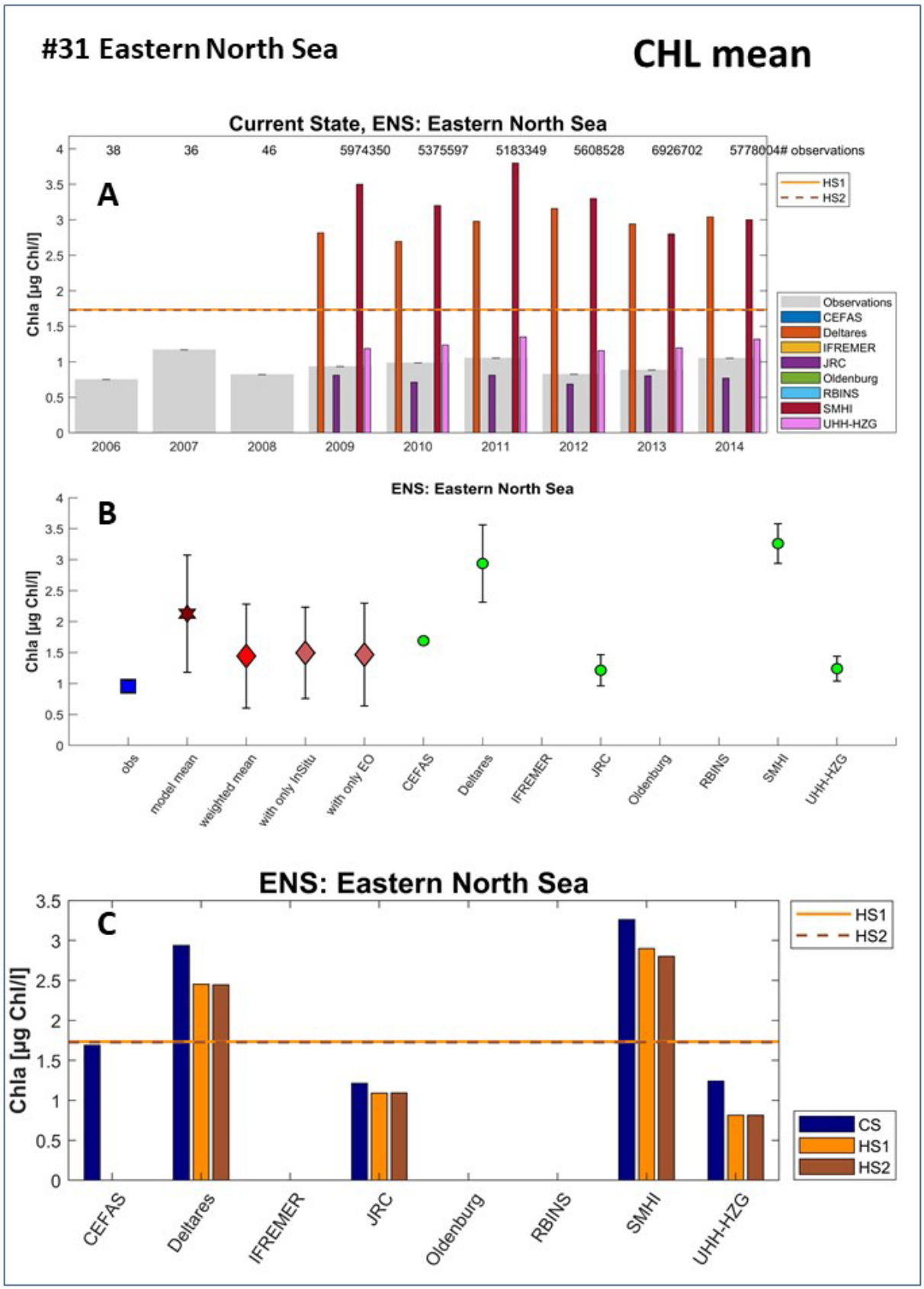


Figure 4.2A: Model analysis and derived threshold values for Chl mean for *area #31 Eastern North Sea.*

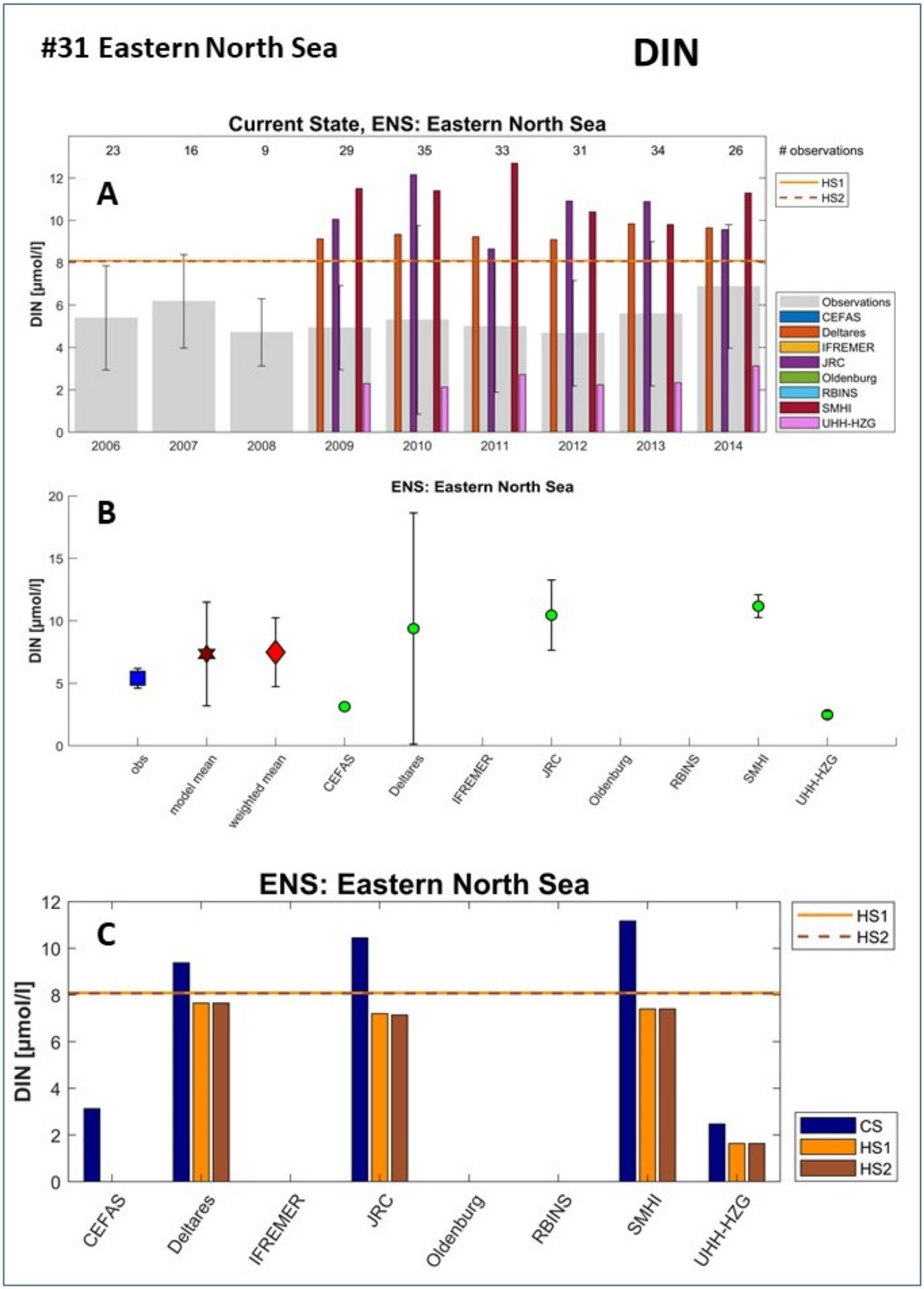


Figure 4.2B: Model analysis and derived threshold values for DIN for *area #31 Eastern North Sea.*

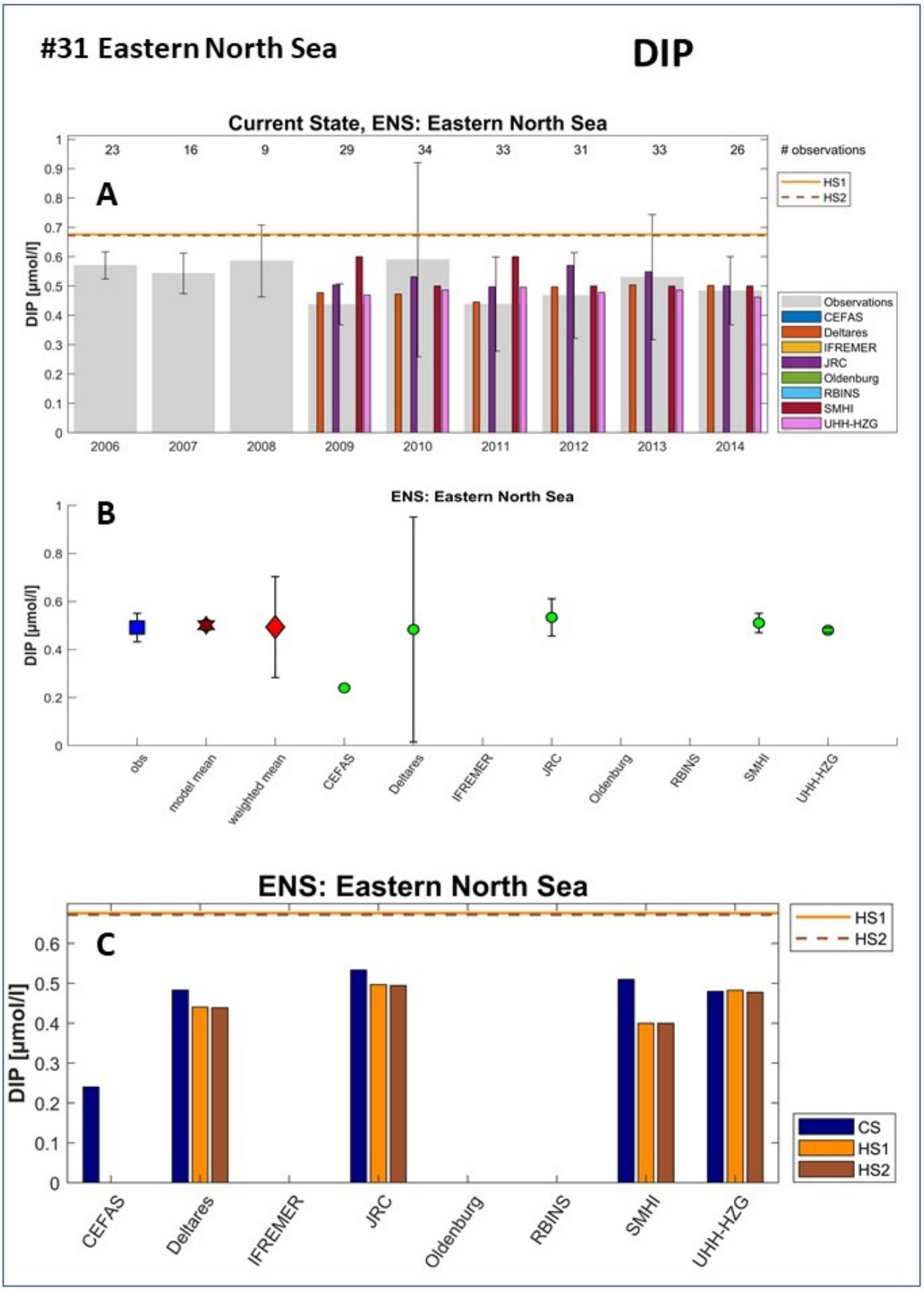


Figure 4.2C: Model analysis and derived threshold values for DIP for *area #31 Eastern North Sea.*

Fig 4.2A-C shows the results from the Eastern North Sea (area #31). In Fig. 4.2A, in the upper graph (A) one can notice that the Chl mean data for the years 2009 – 2014 are drastically increased by the combination with the EO data, as is illustrated by the amount of observation above the grey bar plots in the graph. In the middle graph (B) one can see that the mean of the observation is generally lower than the model values. While the model mean overestimates this Chl observational mean, however the weighted mean value reduced the concentration toward the observational value. This is true also for the weighted mean Chl values based on only the in-situ data or the EO data as well. In the lower graph (C) the mean concentration from the CS run as well as for the two historic scenario HS1 and HS2 are shown, including the resulting threshold for both scenario. As one can see there is hardly any difference between the Chl thresholds for this area. In the Current State (CS) simulation the Chl values from CEFAS, and UHH-HZG clearly fall below the threshold, while the other exceed the threshold values.

Fot the nutrient concentration the data coverage is rather good for both DIN (Fig. 4.2.B) and DIP (Fig 4.2.C). The simulated winter concentrations for DIN (Fig 4.2B) overestimate the observation for most models, what can be seen in the middle graph B. Here the weighted mean value slightly increases the mean model value. In the lower graph (C), as well as in the upper graph (A) representing the simulated yearly concentration values, one can see that three models show CS values above the DIN thresholds. For DIP (Fig. 4.2C) the mean winter concentrations from the models match very well with the observations. There is hardly any difference between the observation mean value in comparison to the model mean or the weighted model mean in the middle graph (B). With a rather low reduction between the CS mean DIP values and the ones from the two scenarios, all models concentration from the CS run fall below the threshold values.

Fig 4.3A-C shows the result for the Elbe Plume (area #20), which has a good observational coverage, both for the nutrients and, improved by the EO data, for Chl as well, as displayed in the upper graph (A) in Fig. 4.3A. Most models underestimate the Chl observation mean value (see middle graph B). The weighted mean raises the concentration value up, nearly independent of the basis of in-situ or EO data. The lower graph (C) compares the threshold values from both scenario, which differ considerably because of the applied P reduction in the Elbe area, with the integrated concentration over the simulation period 2009 – 2014. Nearly all models display CS values less than the two thresholds, except for Deltares, where the values lies just below the HS1 threshold, and Cefas which shows a much higher value compared to both thresholds.

The mean model value for DIN in the Elbe Plume area (Fig. 4.3.B) as well as the weighted ensemble value are at the same level as the mean of the observations (middle graph B). Only two models, Uni Oldenburg and UHH-HZG display CS values blow the thresholds. Both thresholds are nearly identical, since there is no difference in N reduction applied for the two scenarios. For DIP (Fig. 4.3c) the model mean slightly underestimates the mean of the observation (middle graph B). The weighted ensemble mean lifts the value nearly to the same level as the mean observation. There is a strong difference between the threshold values for HS1 and HS2 because of the applied further P reduction in the HS2 scenario. The CS concentrations from most models fall below even the lower HS2 threshold. Only the value for SMHI exceeds both thresholds, while the values Deltares and Oldenburg reach just above the HS2 value.

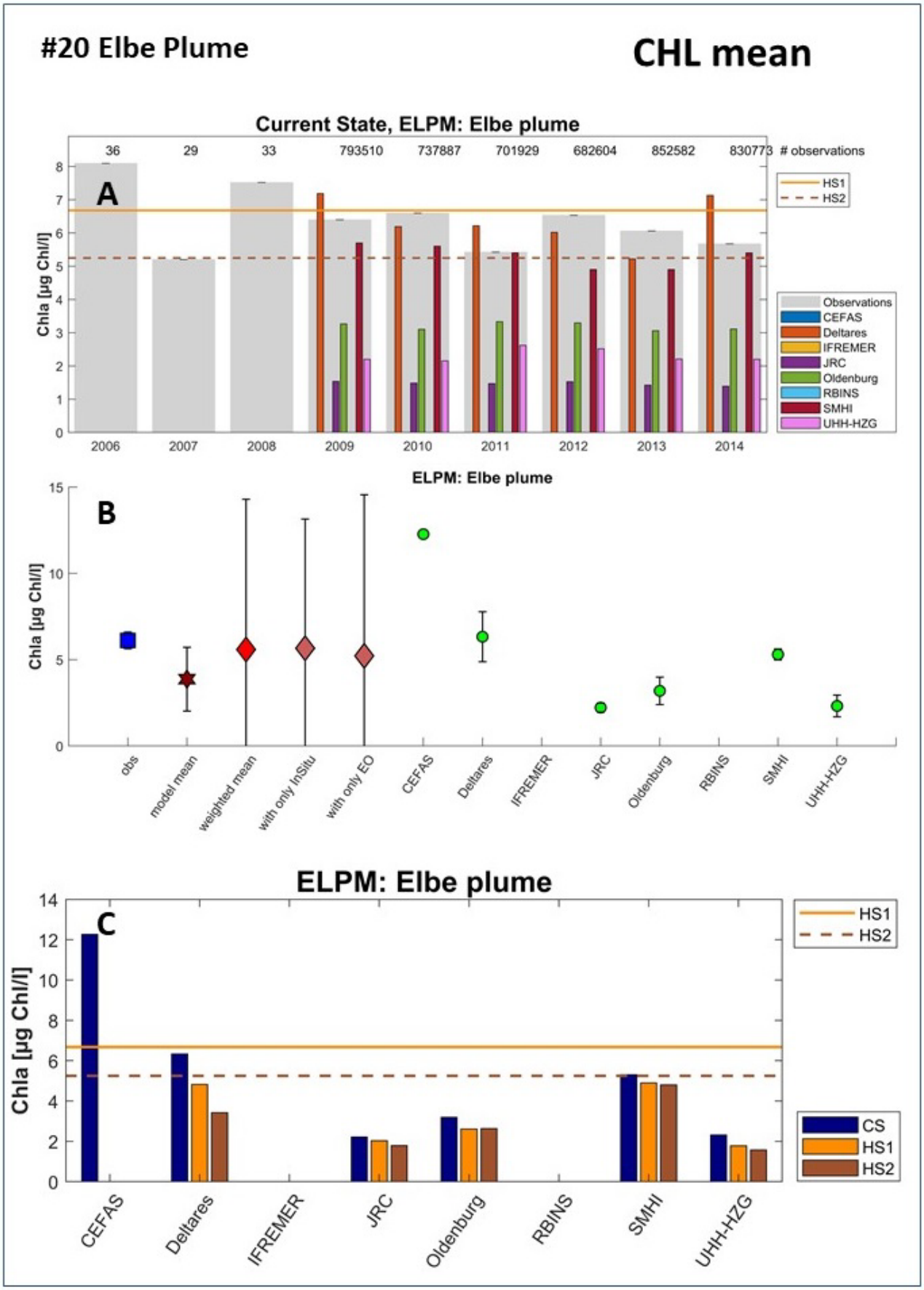


Figure 4.3A: Model analysis and derived threshold values for Chl mean for area #20 Elbe Plume.

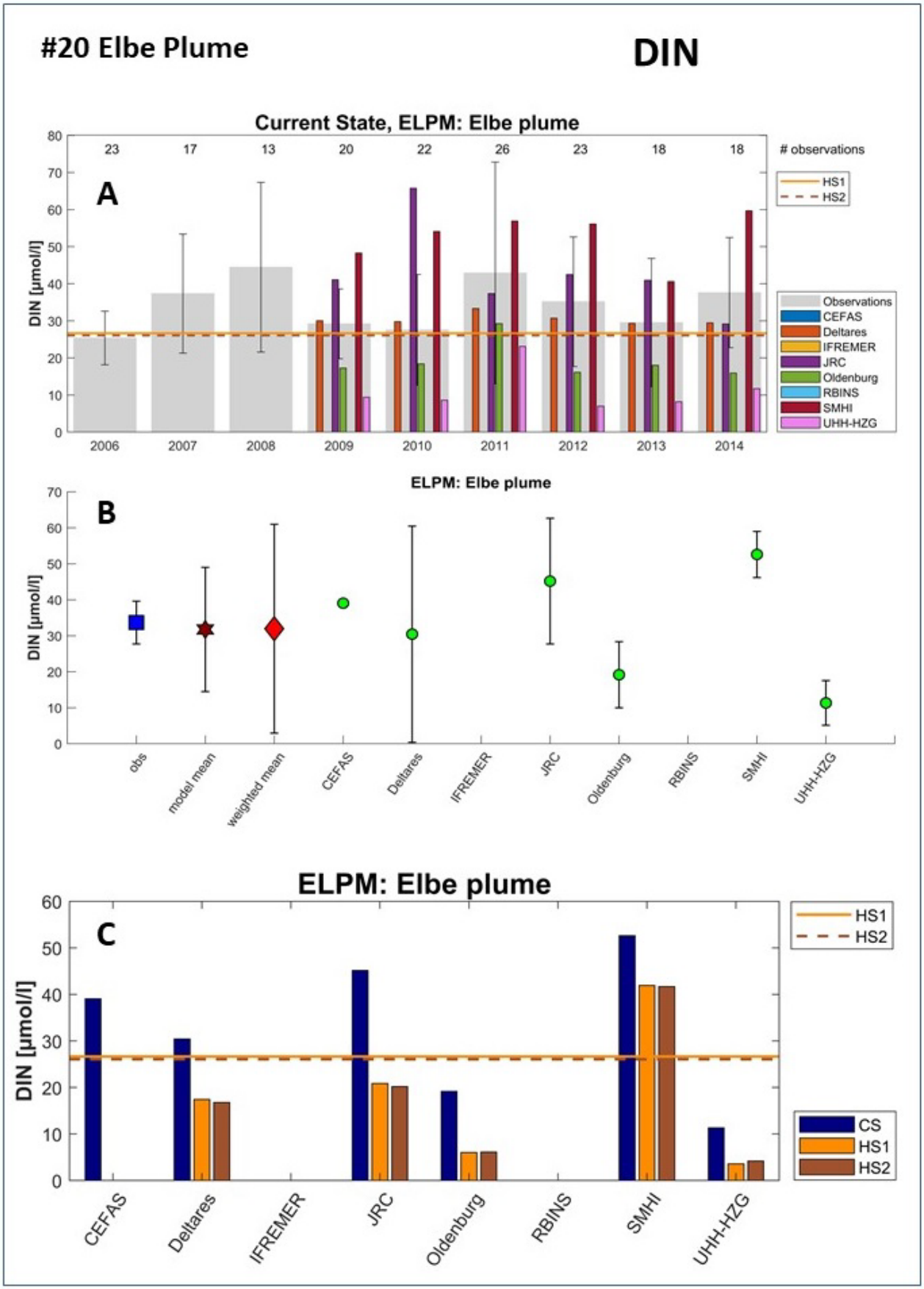


Figure 4.3B: Model analysis and derived threshold values for DIN for area #20 Elbe Plume.

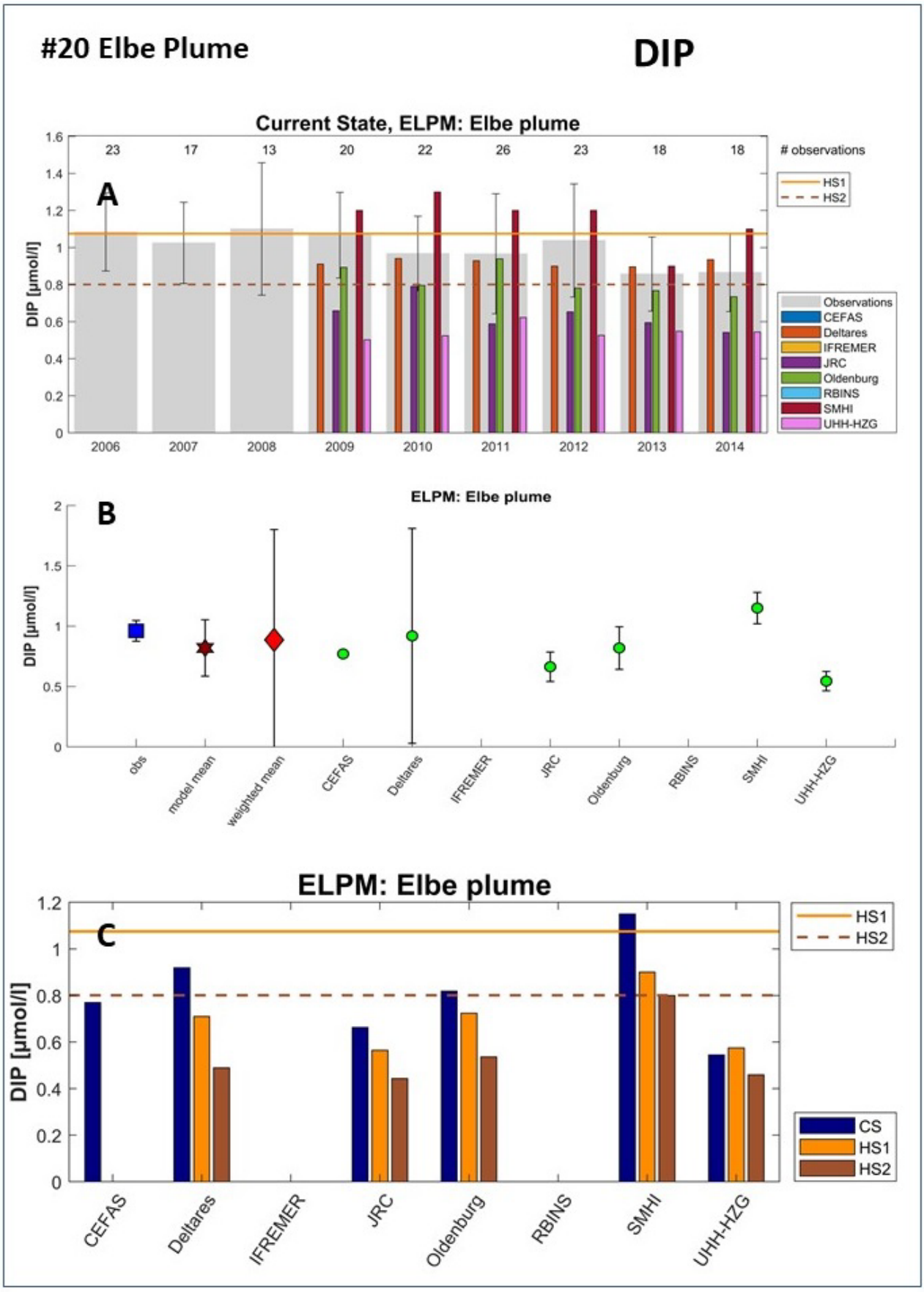


Figure 4.3C: Model analysis and derived threshold values for DIP for area #20 Elbe Plume.

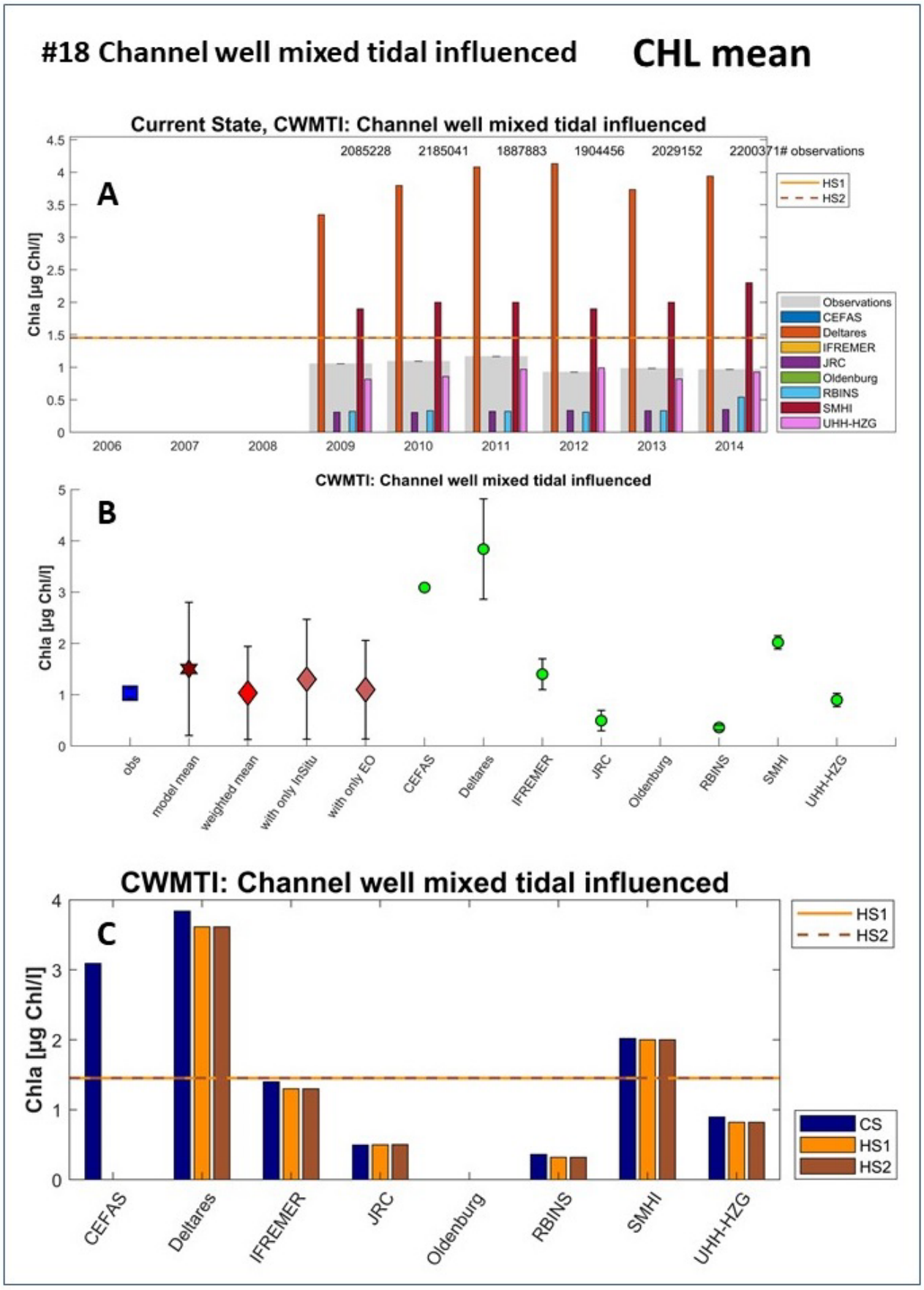


Figure 4.4A: Model analysis and derived threshold values for Chl mean for area #18 Channel well mixed tidal influenced.

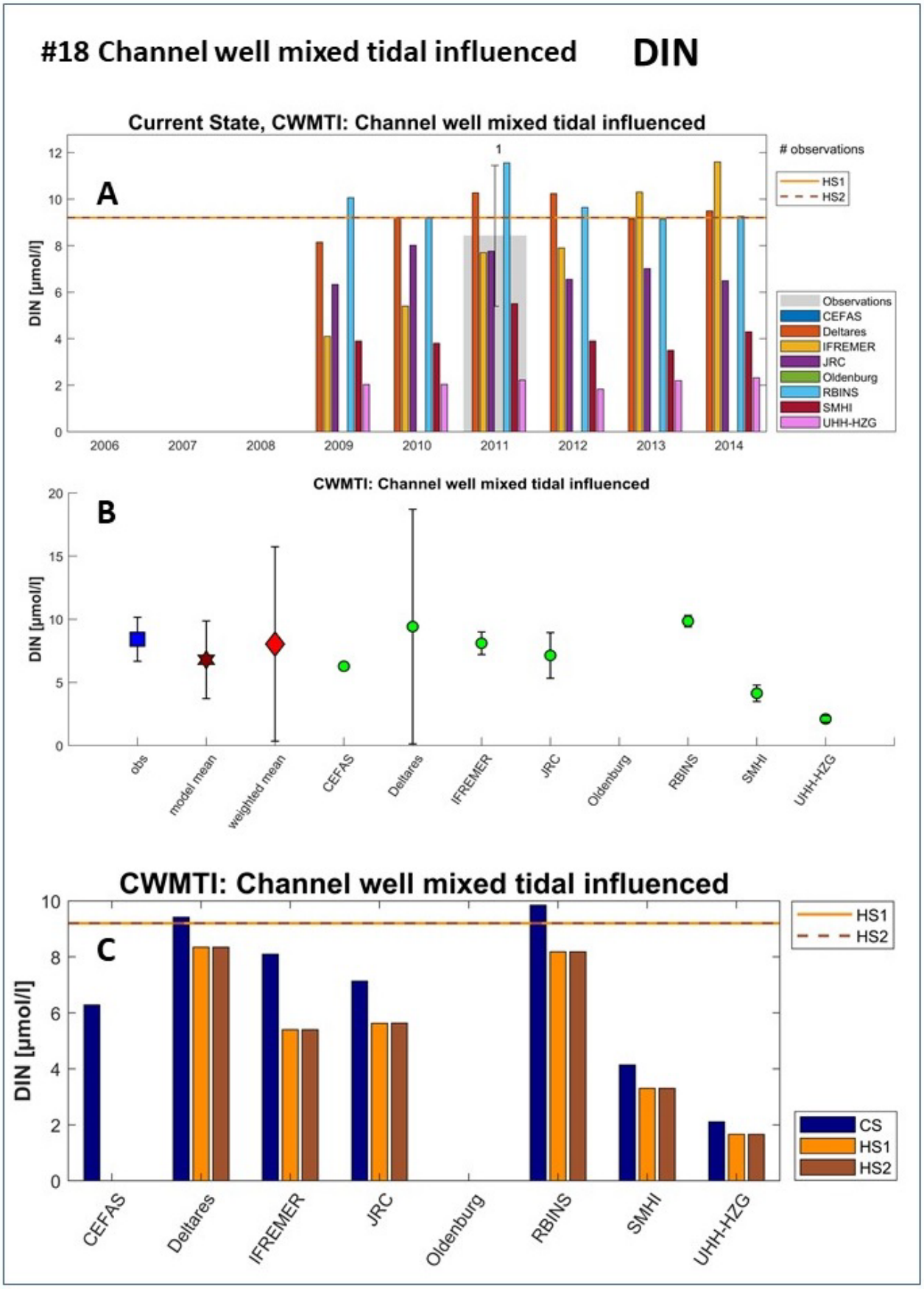


Figure 4.4B: Model analysis and derived threshold values for DIN for area #18 Channel well mixed tidal influenced.

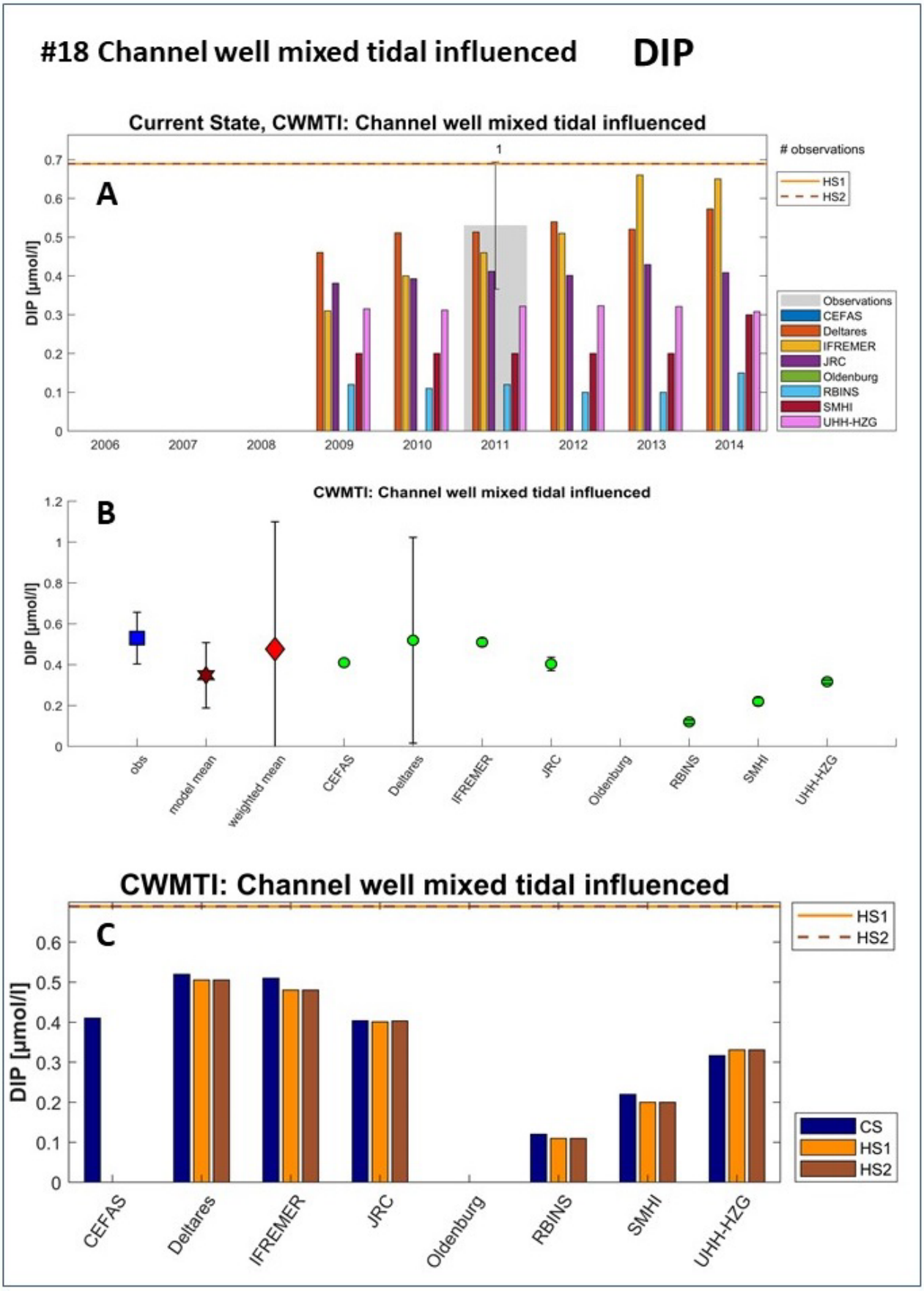


Figure 4.4C: Model analysis and derived threshold values for DIP for area #18 Channel well mixed tidal influenced.

The final overview example in Fig 4.4A-C represents the Channel well mixed tidal influenced area (#18), which reveals a poor data coverage for the nutrients with only one observation value (Fig. 4.4b and C, upper graph A). In contrast the Chl representation is very good as a result of the inclusion of EO data for the simulation period 2009 to 2014, as can be seen in Fig. 4.4A (upper graph A). While the simulated Chl concentration vary strongly between the models, the mean ensemble value overestimates the mean of the Chl observations. But the weighted model ensemble mean is reduced to the right level of the observation mean. While the area has a good model coverage, three model results for Chl from the CS run fall below the threshold, which do not differ between the two scenario. Only the simulated CS concentration from Cefas, Deltares and SMHI provide values way over the threshold values (lower graph C).

For DIN (Fig. 4.4.B) and DIP (Fig. 4.4.C) the results of the model simulation vary strongly (see middle graph B). Both for DIN and DIP the model mean underestimates the mean value of the observation. For both nutrient the the weighted ensemble mean value is slightly elevated to the level of the mean of the observation. For DIP (Fig. 4.4.C) all simulated SC run values (see lower graph C) fall below the threshold, while for the integrated DIN concentration (Fig. 4.4.B) the values from Deltares and RBNS reveal slightly higher values compared to the threshold ones.

Since the inclusion of the satellite EO data was intended to increase the basis of the Chl-a data for the weighted ensemble approach, it is worthwhile to demonstrate the impact of these additional data. In addition to the Fig 4.2A – 4.4A, where the mean values of the observation and EO data are displayed separately, we use the same graphical display to illustrate different results of the addition of the EO data. Fig. 4.5A shows a selection of areas (A: CCTI, B: CFR and C: THPM) where satellite Chl data increased the mean value in comparison to the observation, while in Fig 4.5B, the opposite is the case. While in Fig 4.5A the simulated Chl values from the models scatter a lot for the CCTI area #2 (upper graph A) and the CFR area #1 (middle graph B) the weighted ensemble lifts the mean value up to the mean of the observation. The reason is the impact of the EO data, which raise the number of observations from two *in-situ* in the period of 2009 to 2014 to above 40 000 EO values per year in the CCTI are, and from one *in-situ* observation in the CFR area to over 60 000 EO data that can be incorporated within the weighting approach. In both cases the EO lead to a good match between the weighted ensemble mean and the mean of the observation.

This is different for the Thames plume area THPM (area #25) in Fig. 4.5A (lower graph C), where the in-situ Chl observation vary between 1 to 12 between the years 2010 to 2014, and the basis is raised to above 20 000 for the whole simulation period. But the mean weighted ensemble concentration is only slightly higher than the ensemble mean, but the EO data related weighted value would be higher.

In Fig 4.5B we display a selection of areas where the incorporation of EO data result in a decrease of the weighted ensemble mean compared to the one based on *in-situ* data alone. For the Irish Sea IRS area #41 (upper graph A), there are 98 *in-situ* observation available in 2011 and in the years 2008 – 2010 they vary between 2-8. The EO data increase the data level to over 20 000 observation per year, which only leads to a small decrease in the weighted ensemble, simply by the fact that the model mean for the four models that cover this area was already near the mean of the observation.

For the IS2 area #29 (middle graph B) and the SS area #47 (lower graph C) the incorporation of the EO data lead to a considerably lower weighted ensemble mean in comparison to the simple model ensemble mean, which overestimates the observation mean considerably. This has to do with the

increase in observation, for IS2 from 3-25 in-situ observations to more than 20 000 per year and for SS from 9-24 in-situ observation to nearly 40 000 per year.

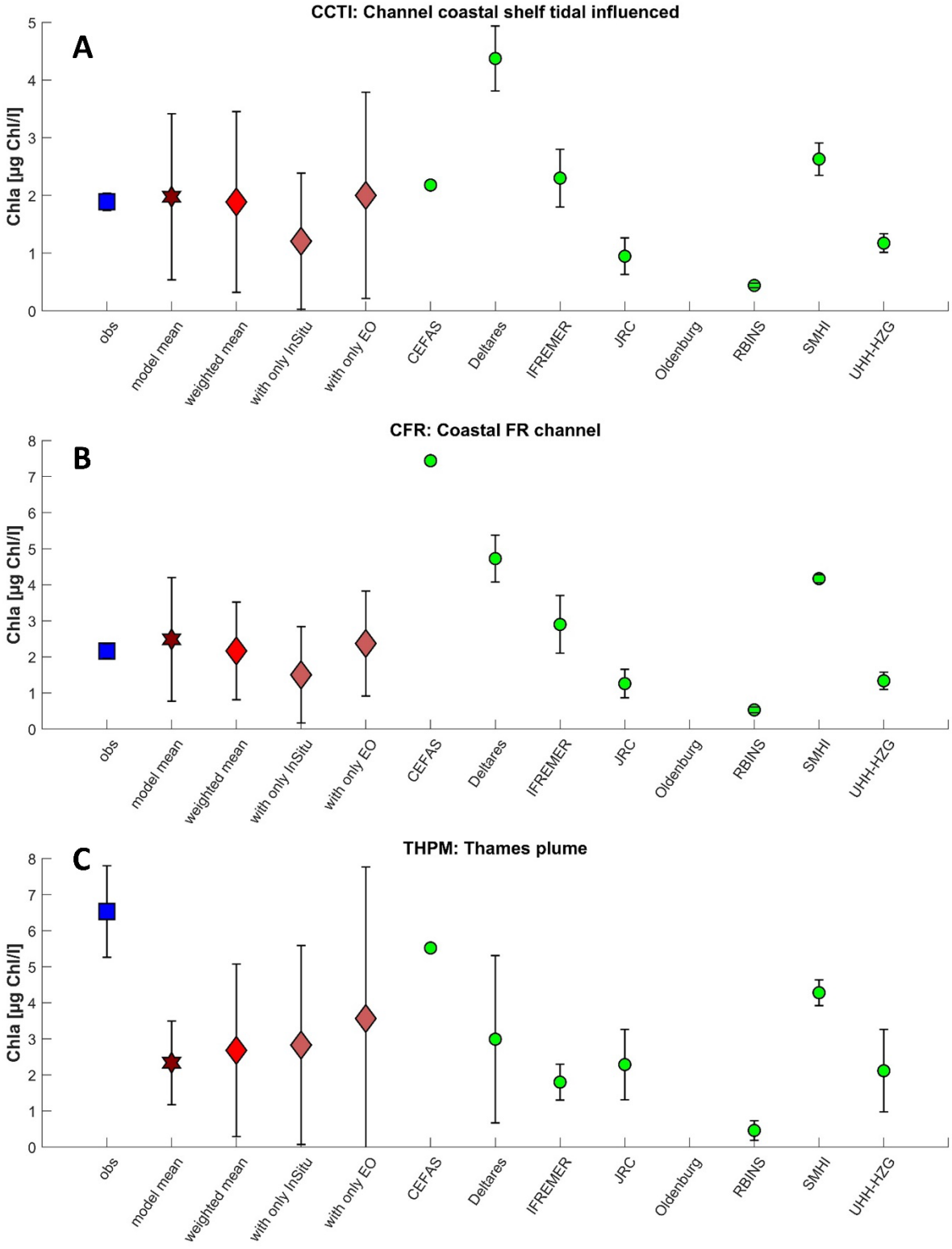


Figure 4.5A: areas which show earth observation Chl-a data that is higher than the local in-situ observations

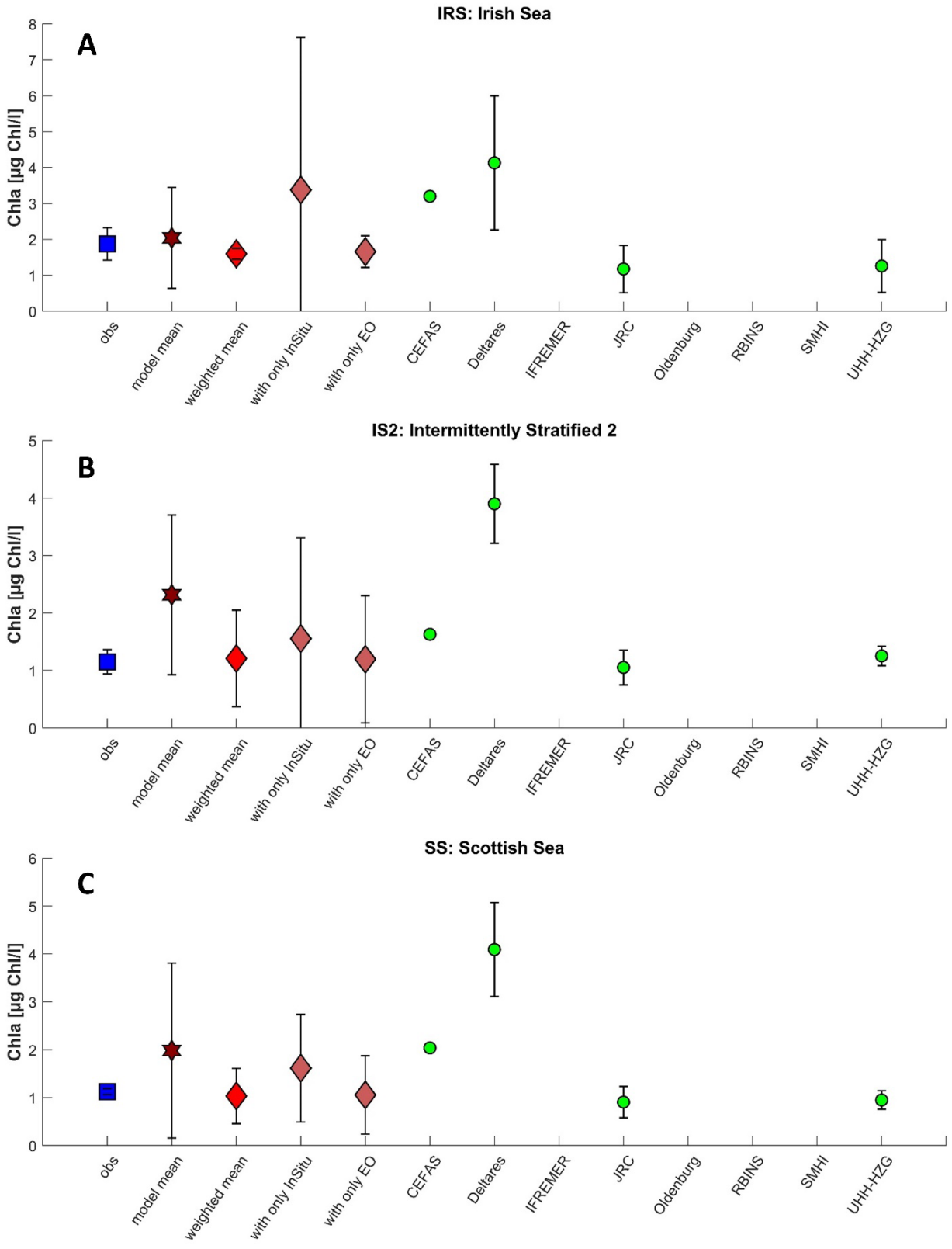


Figure 4.5B: Areas which show earth observation Chl-a data that is lower than the local in-situ observations

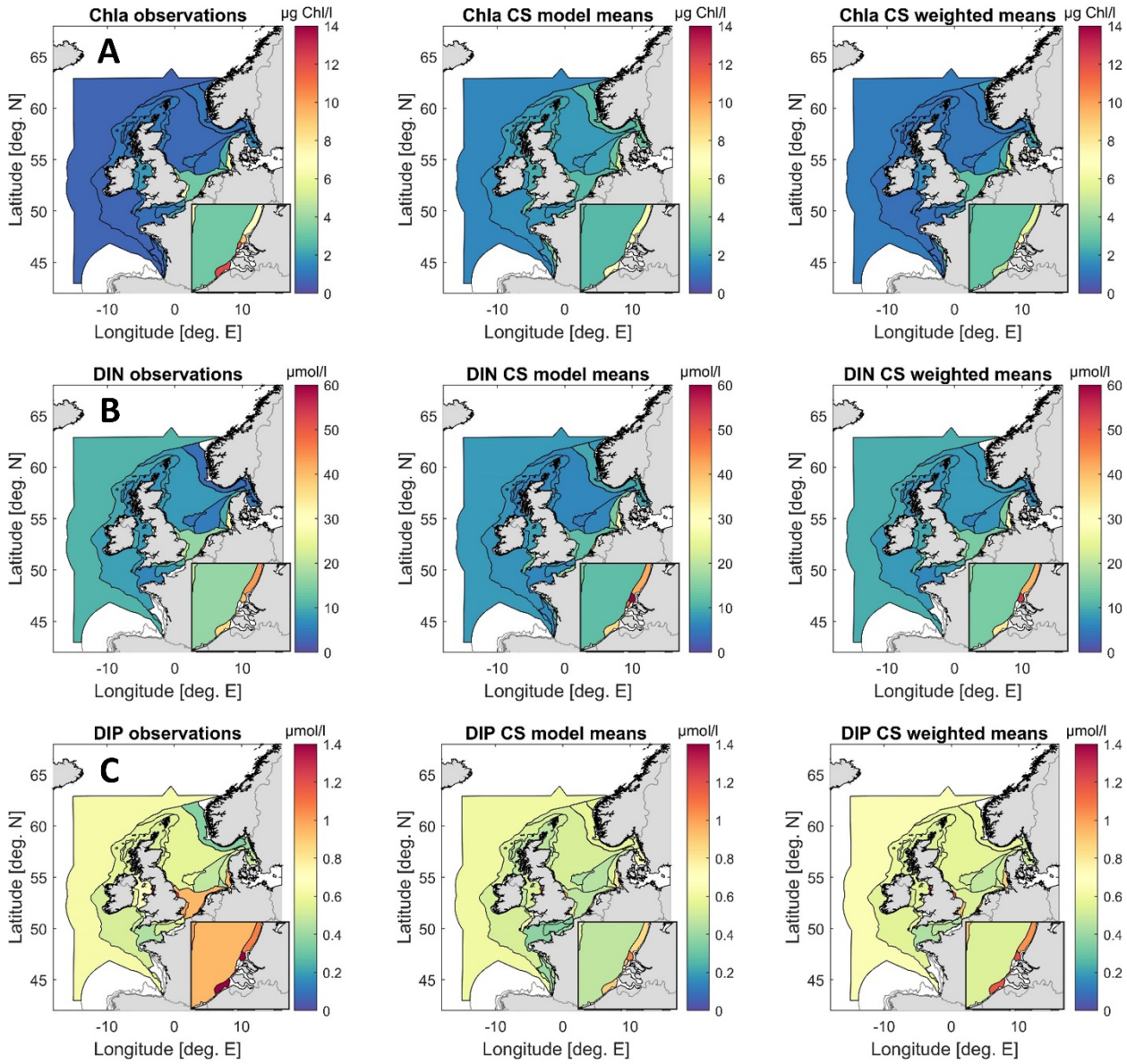


Figure 4.6: Improvement by weighted approach, showing observational mean (left), ensemble model mean (middle) and weighted ensemble model mean (right). For Chlorophyll-a (A), DIN (B) and DIP (C).

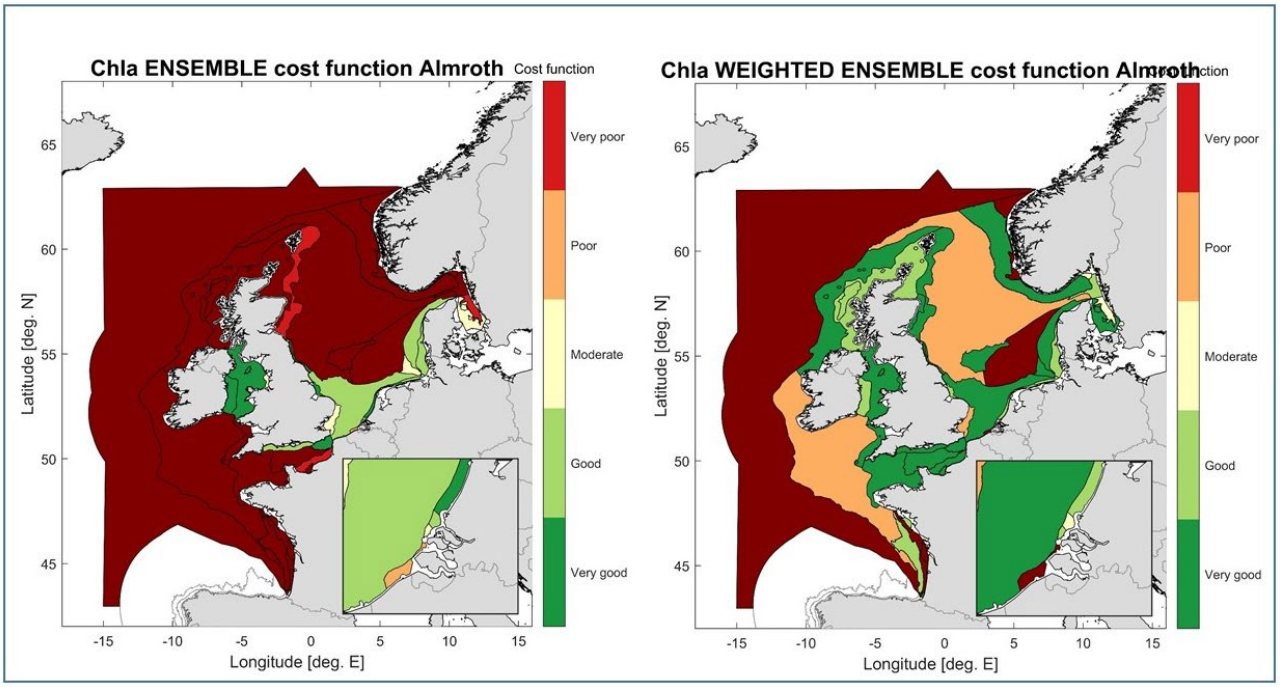


Figure 4.7A: Improvement by weighted approach cost function representation for Chl-a. Cost function values range from 0-5 with values < 1.0 raking "very good" (dark green), values between [1, 2) "good" (light green), values between [2,3) "moderate" (yellow), values between [3,4) "poor" (orange) and values between [4,5) "very poor" (red). Cost function values above 5 have been assigned the dark red colour.

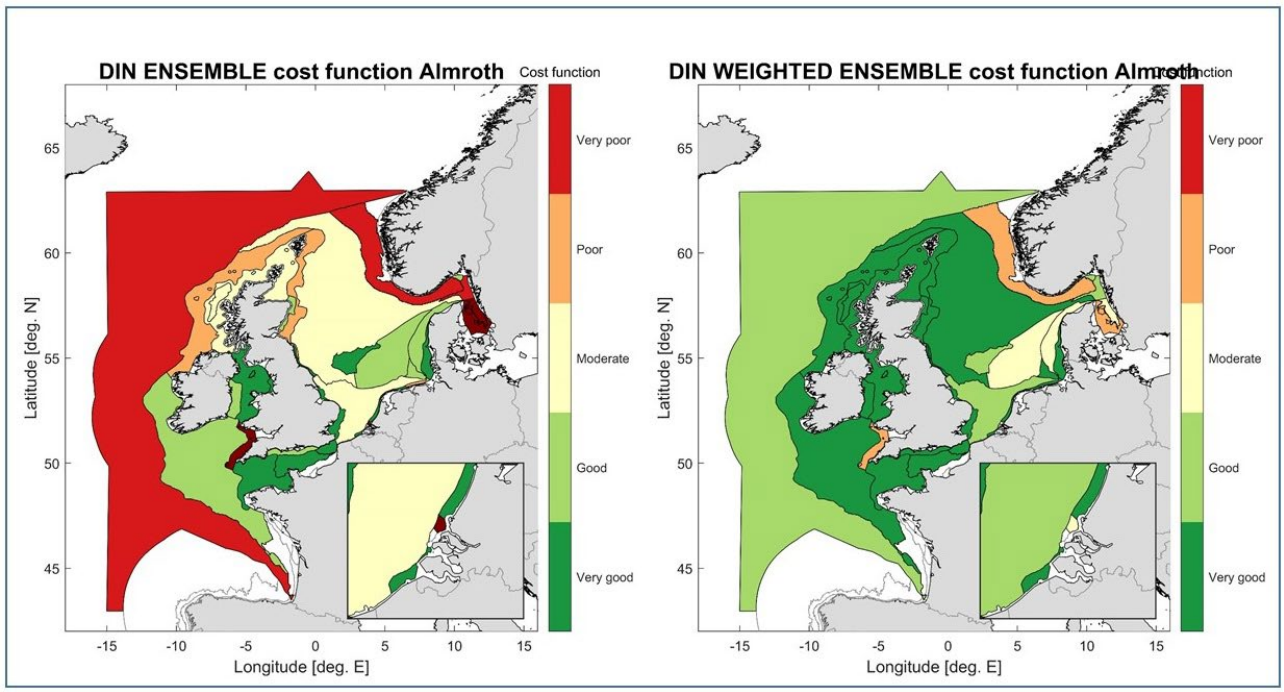


Figure 4.7B: Improvement by weighted approach cost function representation for DIN

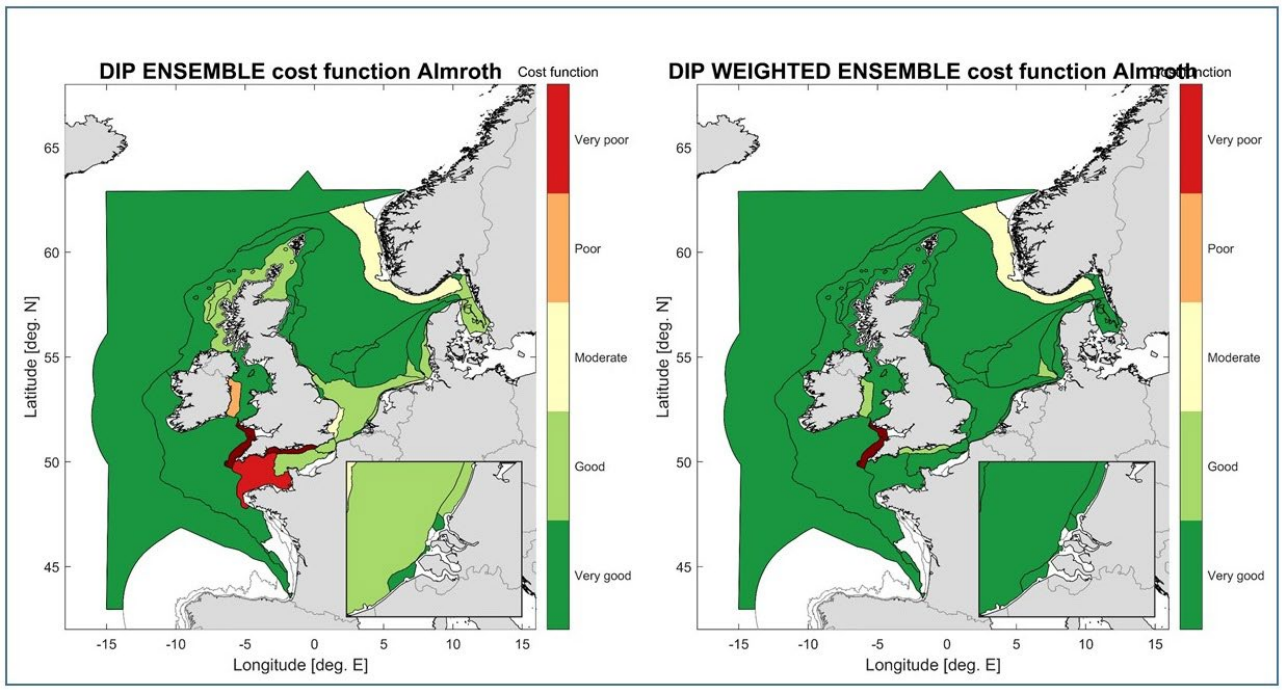


Figure 4.7C: Improvement by weighted approach cost function representation for DIP

Therefore, the incorporation of the EO data improves the quality of the model weighted ensemble mean considerably, either by raising this value as shown in Fig. 4.5A or by lowering it (Fig. 4.5B) towards the mean of the observation. In all cases the number of Chl-a observations has increased enormously compared to the in-situ observations available. The inclusion of the Chl-a EO data allows for the weighted ensemble approach to be used in areas where *in-situ* observations of Chl-a are lacking (GBCW, LPM, ECPM2, SCHPM2, SPM, GDPM, GBSW, ADPM, CNOR1, CNOR2). It has been shown by García-García et al (2019) that both *in-situ* and Earth Observation data are necessary for assessments, even in areas with high availability of *in-situ* data. It is therefore recommended to start Chl-a *in-situ* monitoring in the areas listed above.

Fig 4.6 illustrates the improvement of the weighted ensemble approach to bring the weighted model means closer to the observational means. In the upper row (A) this is demonstrated for Chl-a, in the middle row (B) for DIN and in the lower row (C) for DIP. The weighted ensemble model mean is closer to the observations than the unweighted ensemble model mean, but the adjustment varies per area due to different spatial and temporal coverage of the observations. For Chl-a this has been addressed by including earth observation data. The cost function calculated with the ensemble mean improves significantly between the unweighted and weighted ensemble model mean (Fig 4.7). However, it should be noted that the Chl-a cost function analysis shows much better results (for both individual models and the ensemble mean) if only *in-situ* Chl-a observations are used. This may be related to the fact that most *in-situ* locations are coastal and that high turbidity in coastal areas can upset the Earth Observation models (used to calculate Chl-a values from satellite wavelength observations), which are usually optimized for certain turbidity regimes. From space it can be difficult to distinguish between suspended particulate matter and Chlorophyll. From the observations itself it becomes clear that the EO Chl-a data has lower standard deviations than the *in-situ* Chl-a observations: as the cost function divides by the observational std this automatically increase the cost function value when Chl-a EO data is included.

Finally, the Fig. 4.7A to C display the comparison of the resulting cost function based (left) on the model ensemble mean concentration in comparison (right) to the weighted ensemble mean for the parameters Chl mean, DIN and DIP. Cost function values range from 0-5 with values < 1.0 raking “very good” (dark green), values between [1, 2) “good” (light green), values between [2,3) “moderate” (yellow), values between [3,4) “poor” (orange) and values between [4,5) “very poor” (red). Cost function values above 5 have been assigned the dark red colour.

For Chl (Fig. 4.7A) one can clearly see an improvement in the cost function in the weighted ensemble approach for a wide area within the North Sea or the Irish and Celtic Sea, which change from poor or even very poor on the east coast of Scotland to good or very good mainly in the coastal areas. There is no change in the outer Atlantic, and surprisingly enough for area #31, the Eastern North Sea. Here the cost function status remains very poor also for the weighted ensemble. In the Channel region and the Continental Coastal Region the status changes mainly from good to very good, only the Elbe plume region remains at the status good. There is one counterintuitive change in the small Scheldt plume region 1 (area #19) which decreases from poor to very poor.

For DIN (Fig. 4.7B) the shift in most COMP4 assessment areas in the North Sea and around Ireland goes from moderate or poor to good or very good by the application of the weighted ensemble approach. Even the wider North Atlantic and the Norwegian Trench region improved dramatically. Only for the Outer Coastal DEDK region (area #30) and the neighbouring Eastern North Sea region (area # 31) the cost function changes from good to moderate. Also, for the nearby Dogger Bank region (area #8) there is no improvement, but the status is reduced from very good to good. Somehow the south-eastern North Sea region off the coast do not benefit from the weighted ensemble approach by obtaining slightly lower cost function classification than in with the ensemble mean.

For DIP (Fig. 4.7C) there are no widespread changes between the two approaches, simply because most areas achieve already a good or very good status. There are only local improvements in the Channel region and the Coastal IRL region (area #38) the status changes from very poor to very good and from poor to good, respectively. The only regions which remains in a very poor status is Coastal UK1 region (area # 39).

5 Derived threshold values

This chapter provides the finally derived threshold values based on the aggregated results of the model contribution from the ICG-EMO partners, as delivered for analysis by 23.9.2021. Finally, we present an overview of the horizontal distribution of the threshold estimates for key eutrophication parameter.

5.1 Threshold values for all COMP4 assessment areas

Tables 5.1 (HS1) and 5.2 (HS2) show the resulting threshold levels. The applied model domains do not cover all areas: CWCC, OWCO, OWAO, IWCI, OWBO, CWBC, IWBI, CWAC, IWAI are not reported here. Thresholds for oxygen saturation, Kd (light attenuation coefficient) and Secchi disc depth have been omitted. HASEC has not provided clear guidance on how to calculate thresholds for indicators for which high values are related to less eutrophication. TotalN and TotalP were calculated with the weights from DIN and DIP respectively.

Table 5.1: Threshold levels for all areas covered by at least 1 model for Historic Scenario 1.

ICG EMO ID	Threshold levels	DIN	DIP	TotalN	TotalP	NP ratio	Chla	Chla 90th	O2	netPP
		µmol/l	µmol/l	µmol/l	µmol/l	–	µg Chl/l	µg Chl/l	mg O2/l	g C/m2/y
	Area	HS1	HS1	HS1	HS1	HS1	HS1	HS1	HS1	HS1
1	CFR	16.0	0.60	14.9	0.6	41.7	2.8	4.5	6.0	257.3
2	CCTI	12.0	0.64	16.7	0.7	33.1	2.3	3.8	6.0	192.6
3	ATL	15.3	0.98	20.9	1.4	22.1	1.8	4.2	6.0	262.8
4	SHPM	11.1	0.79	13.1	0.9	20.0	1.9	3.8	6.0	315.8
5	CNOR1	12.5	0.88	13.5	1.0	21.2	2.7	6.2	6.0	210.5
6	CNOR2	10.4	0.78	11.3	0.8	20.1	1.9	4.1	6.0	268.2
7	CNOR3	9.2	0.71	10.2	0.9	20.3	2.4	4.5	6.0	216.9
8	DB	7.2	0.76	6.6	0.7	16.8	1.3	2.9	6.0	191.4
9	KD	6.6	0.70	7.8	0.8	20.8	2.8	5.1	6.0	158.5
10	NT	10.9	0.87	14.2	1.1	20.2	1.7	4.2	6.0	291.6
11	SNS	13.1	0.71	10.0	0.7	22.2	3.8	4.8	6.0	198.0
12	GBC	7.3	0.72	7.6	0.8	25.2	2.8	4.4	6.0	185.4
13	ADPM	8.9	0.67	9.4	0.7	19.3	1.7	3.7	6.0	355.6
14	GBSW	8.7	0.69	9.4	0.7	18.5	0.9	2.2	6.0	263.2
15	SPM	38.5	0.92	30.4	0.8	55.7	5.1	7.2	6.0	302.6
16	GDPM	12.7	0.68	15.6	0.9	25.5	5.4	10.0	6.0	585.5
17	CUKC	12.8	0.74	13.8	0.8	32.4	2.3	3.9	6.0	186.9
18	CWMTI	9.2	0.69	11.2	0.7	30.7	1.5	2.7	6.0	192.5
19	SCHPM1	25.9	1.31	27.5	1.4	34.4	5.0	7.2	6.0	142.9
20	ELPM	26.7	1.08	21.6	1.0	34.2	6.7	10.3	6.0	193.6
21	SCHPM2	33.3	1.02	35.9	1.3	43.3	8.9	15.1	6.0	246.4
22	MPM	40.7	1.35	36.7	1.3	52.5	8.0	13.4	6.0	289.9
23	RHPM	29.7	1.15	23.9	0.4	37.9	6.8	7.1	6.0	225.8
24	EMPM	10.9	0.77	11.1	1.0	36.2	5.9	8.0	6.0	185.5
25	THPM	14.6	0.90	12.5	1.0	31.1	3.2	4.7	6.0	113.1
26	HPM	26.3	1.16	20.0	1.0	30.5	7.4	10.1	6.0	213.9

27	ECPM1	10.8	0.78	11.2	0.7	18.6	2.1	5.0	6.0	242.4
28	ECPM2	10.9	0.86	4.8	0.7	19.9	3.5	5.7	6.0	227.9
29	IS2	11.2	0.86	12.1	0.9	19.6	1.7	4.6	6.0	285.6
30	CO	13.5	0.72	11.6	0.8	30.7	2.9	5.2	6.0	186.0
31	ENS	8.1	0.68	8.0	0.7	20.8	1.7	3.8	6.0	166.5
37	ASS	11.6	0.83	12.6	0.9	19.9	1.4	2.9	6.0	313.9
38	CIRL	11.4	0.77	11.7	0.8	16.6	1.8	3.9	6.0	180.7
39	CUK1	11.7	0.82	13.2	0.9	21.1	1.7	3.6	6.0	247.5
40	IS1	13.7	0.89	14.2	1.0	20.1	1.6	4.0	6.0	307.8
41	IRS	9.9	0.78	11.0	0.9	18.1	2.0	4.2	6.0	138.0
42	KC	7.6	0.65	7.6	0.7	19.9	2.4	4.7	6.0	142.7
43	NNS	11.4	0.89	11.7	0.6	18.4	1.6	3.6	6.0	221.3
44	CWM	8.3	0.66	9.2	0.7	18.7	1.3	2.8	6.0	263.0
45	LBPM	29.3	1.37	28.9	1.4	31.5	7.4	13.3	6.0	241.3
46	SK	6.7	0.72	8.9	0.8	21.1	1.7	3.7	6.0	201.5
47	SS	9.6	0.80	10.4	0.8	18.7	1.5	3.7	6.0	254.2
52	LPM	19.5	0.79	18.2	0.9	29.2	3.3	5.7	6.0	466.2
53	GBCW	11.8	0.75	13.0	0.8	22.7	2.7	5.6	6.0	394.9

Table 5.1: Threshold levels for all areas covered by at least 1 model for Historic Scenario 2

	Threshold levels	DIN	DIP	TotalN	TotalP	NP ratio	Chla	Chla 90th	O2	netPP
		µmol/l	µmol/l	µmol/l	µmol/l	–	µg Chl/l	µg Chl/l	mg O2/l	g C/m2/y
ICG EMO ID	Area	HS2	HS2	HS2	HS2	HS2	HS2	HS2	HS2	HS2
1	CFR	15.8	0.60	14.7	0.6	41.5	2.8	4.5	6.0	257.4
2	CCTI	12.0	0.64	16.7	0.7	33.1	2.3	3.8	6.0	192.4
3	ATL	15.4	0.98	20.8	1.4	22.1	1.8	4.2	6.0	264.3
4	SHPM	11.1	0.79	13.1	0.9	20.0	1.8	3.7	6.0	315.3
5	CNOR1	12.5	0.87	13.5	1.0	21.3	2.7	6.4	6.0	210.7
6	CNOR2	10.3	0.77	11.4	0.8	20.3	1.9	4.0	6.0	267.1
7	CNOR3	9.2	0.68	10.3	0.9	20.7	2.4	4.4	6.0	216.5
8	DB	7.2	0.76	6.6	0.7	16.8	1.3	2.9	6.0	191.5
9	KD	6.6	0.69	7.8	0.8	21.4	2.8	5.1	6.0	158.2
10	NT	10.9	0.86	14.3	1.1	20.3	1.7	4.2	6.0	291.1
11	SNS	13.0	0.70	10.0	0.7	22.7	3.8	4.9	6.0	197.0
12	GBC	7.2	0.69	7.6	0.8	27.1	2.7	4.3	6.0	178.4
13	ADPM	8.9	0.67	9.4	0.7	19.3	1.7	3.7	6.0	355.6
14	GBSW	8.7	0.69	9.4	0.7	18.5	0.9	2.2	6.0	263.2
15	SPM	37.9	0.91	29.8	0.8	55.5	5.1	7.2	6.0	302.6
16	GDPM	12.7	0.68	15.6	0.9	25.5	5.4	10.0	6.0	585.5
17	CUKC	12.8	0.73	13.8	0.8	32.4	2.3	3.9	6.0	186.9
18	CWMTI	9.2	0.69	11.2	0.7	30.7	1.5	2.7	6.0	192.6
19	SCHPM1	25.7	1.16	27.4	1.2	36.0	5.0	7.1	6.0	139.7

20	ELPM	26.1	0.80	21.1	0.7	42.5	5.2	7.3	6.0	174.3
21	SCHPM2	33.1	0.86	35.5	1.1	49.8	8.3	13.8	6.0	230.6
22	MPM	40.6	1.14	36.5	1.0	61.6	6.7	10.6	6.0	266.0
23	RHPM	29.6	1.00	23.8	0.4	43.6	6.2	6.6	6.0	212.0
24	EMPM	10.8	0.68	11.0	0.9	40.9	5.3	6.9	6.0	174.6
25	THPM	14.5	0.90	12.5	1.0	31.1	3.2	4.7	6.0	113.1
26	HPM	26.1	1.16	19.9	1.0	30.4	7.4	10.1	6.0	214.2
27	ECPM1	11.0	0.78	11.4	0.8	18.7	2.1	5.0	6.0	242.4
28	ECPM2	10.9	0.86	4.8	0.7	19.9	3.5	5.7	6.0	227.6
29	IS2	11.3	0.86	12.2	0.9	19.7	1.7	4.6	6.0	285.9
30	CO	13.3	0.66	11.6	0.8	33.5	2.7	5.0	6.0	179.1
31	ENS	8.1	0.67	8.0	0.7	21.0	1.7	3.9	6.0	166.2
37	ASS	11.7	0.84	12.7	0.9	19.9	1.4	2.9	6.0	314.5
38	CIRL	11.4	0.77	11.7	0.8	16.6	1.8	3.9	6.0	180.5
39	CUK1	11.7	0.82	13.2	0.9	21.1	1.7	3.7	6.0	247.5
40	IS1	13.7	0.90	14.2	1.0	20.2	1.6	4.1	6.0	308.3
41	IRS	9.9	0.78	11.0	0.9	18.1	2.0	4.1	6.0	137.7
42	KC	7.5	0.64	7.6	0.7	20.5	2.4	4.7	6.0	142.3
43	NNS	11.5	0.89	11.8	0.6	18.5	1.6	3.6	6.0	221.5
44	CWM	8.3	0.66	9.2	0.7	18.7	1.3	2.8	6.0	262.8
45	LBPM	29.3	1.37	29.0	1.4	31.4	7.4	13.3	6.0	241.4
46	SK	6.7	0.71	8.9	0.8	21.5	1.7	3.7	6.0	201.6
47	SS	9.7	0.80	10.5	0.9	18.7	1.5	3.7	6.0	254.3
52	LPM	19.3	0.79	18.2	0.9	29.2	3.3	5.7	6.0	466.3
53	GBCW	11.8	0.75	13.0	0.8	22.7	2.7	5.6	6.0	394.9

5.2 Horizontal distribution of new threshold values

In this part we present the overall distribution of the threshold values, as derived in the previous chapter 4, for all COMP4 assessment areas. First, we display the horizontal distribution for the key eutrophication parameter DIN (Fig. 51), DIP (Fig. 5.2) and Chl (Fig. 5.3) for the two scenarios HS1 and HS2. As additional information we also provide the horizontal distribution for the Chl 90th percentile threshold values (Fig. 5.4), to enable a comparison with the previous COMP3 assessment. The differences between the two scenarios are illustrated for DIP and Chl for the North Sea (Fig. 5.5). In addition also the difference between HS1 and HS2 is provided in Fig. 5.6, in combination the number of contributing models per area. The model coverage per area is provided for Chl90th percentile, but it is representative for all other displayed parameters as well. More information on the model coverage can be found in Chapter 2.

Finally, the differences in the concentration for DIN, DIP and Chl 90th percentile between the new threshold values and the COMP3 thresholds are shown (5.8 and 5.10). For this comparison we add the information on the previous threshold level, but on the original COMP 3 assessment areas (Fig. 5. 7).

While Fig. 5.1 shows no difference between the two scenarios, one can clearly see the change in the DIP values (Fig. 5.2) around the German and Dutch coastline due to the lower P load in the second

scenario. These differences can also be observed in Fig 5.3 representing the resulting Chlorophyll-a threshold values. They are clearly displayed in Fig. 5.5 where the differences between the two scenarios are directly compared against each other for the North Sea. These results reflect the characteristics of the influence of the two scenarios, as HS2 only has a small local impact on the coastal area reaching from the Netherlands to Denmark.

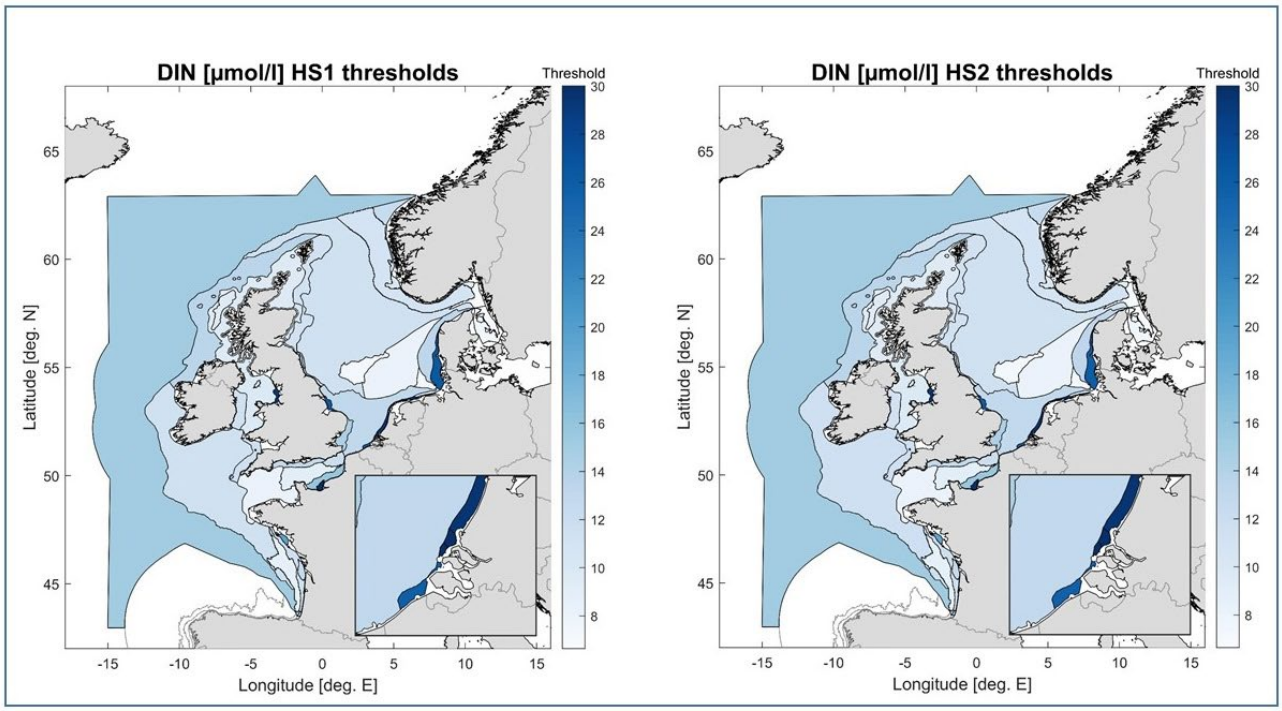


Figure 5.1: Horizontal distribution of DIN threshold values for the assessment areas for the two scenarios.

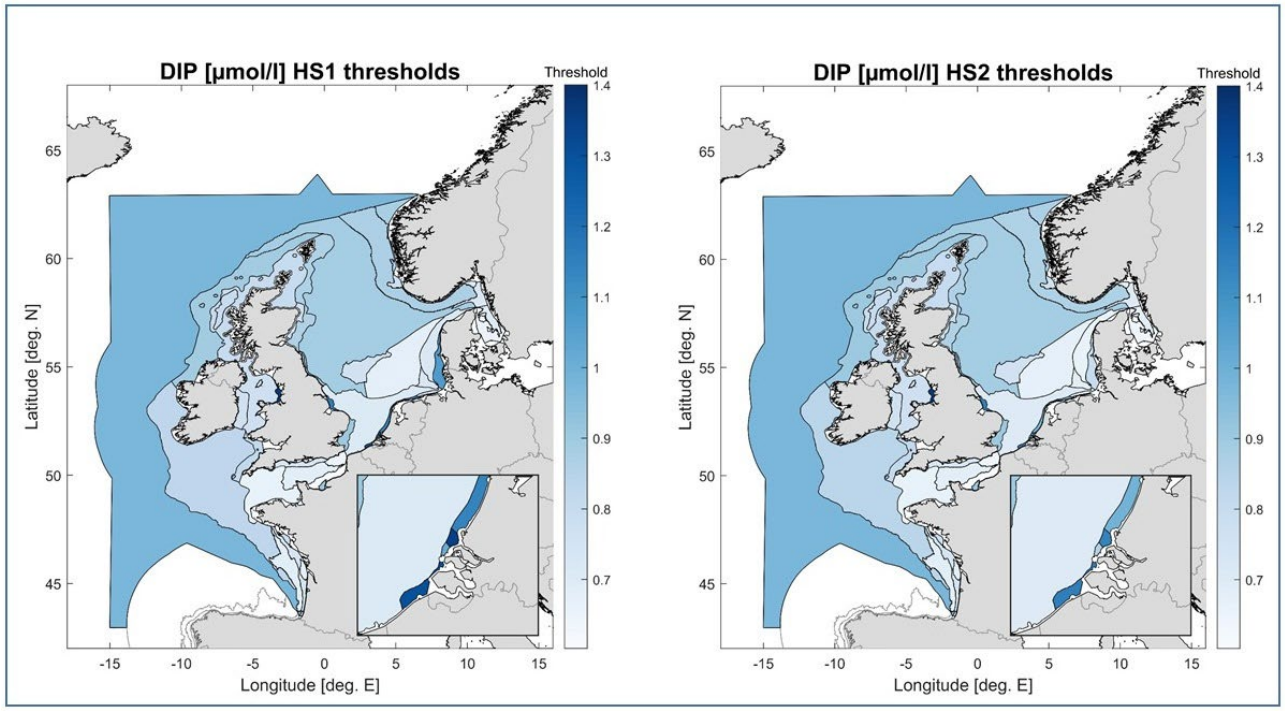


Figure 5.2: Horizontal distribution of DIP threshold values for the assessment areas for the two scenarios.

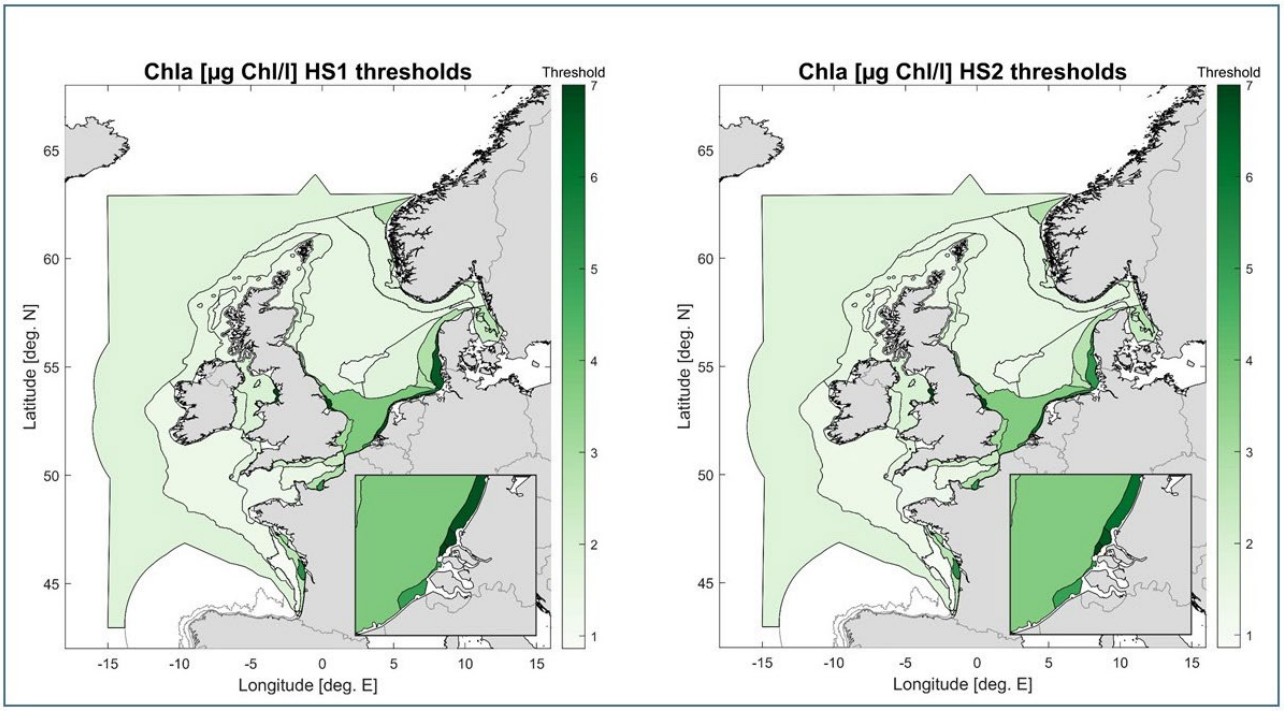


Figure 5.3: Horizontal distribution of Chl mean threshold values for the assessment areas for the two scenarios.

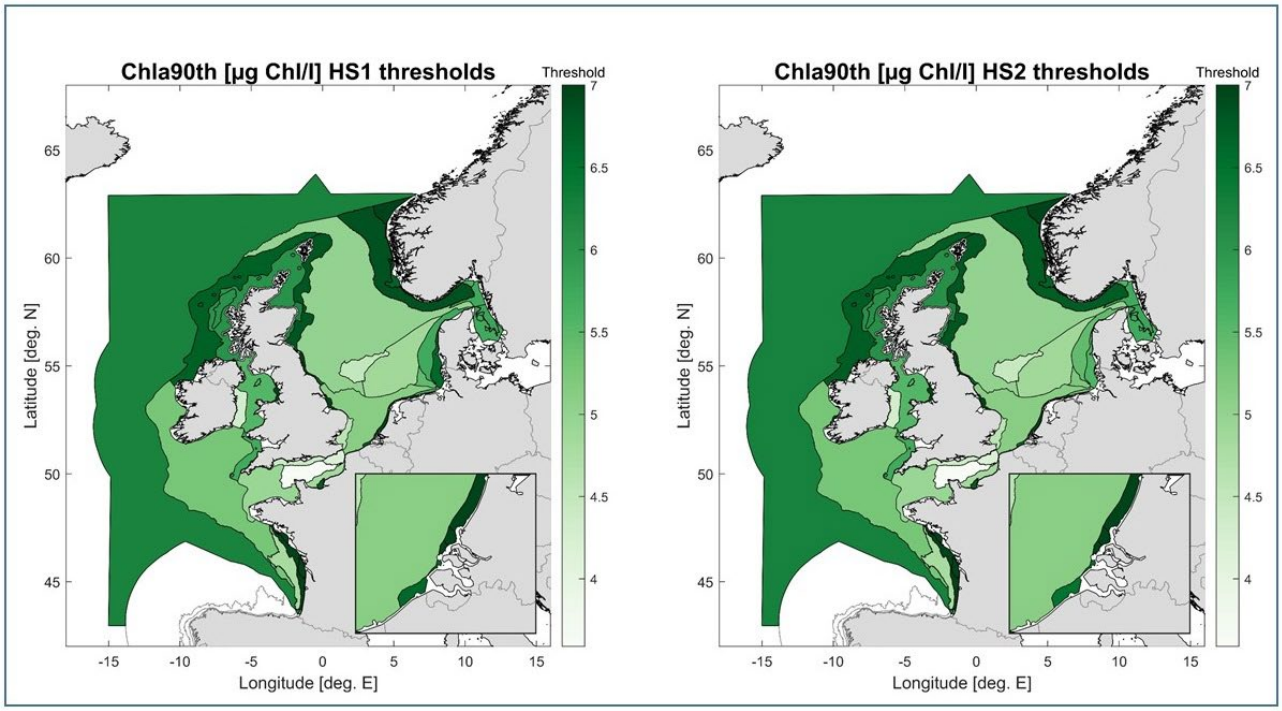


Figure 5.4: Horizontal distribution of Chl 90th percentile threshold values for the assessment areas for the two scenarios.

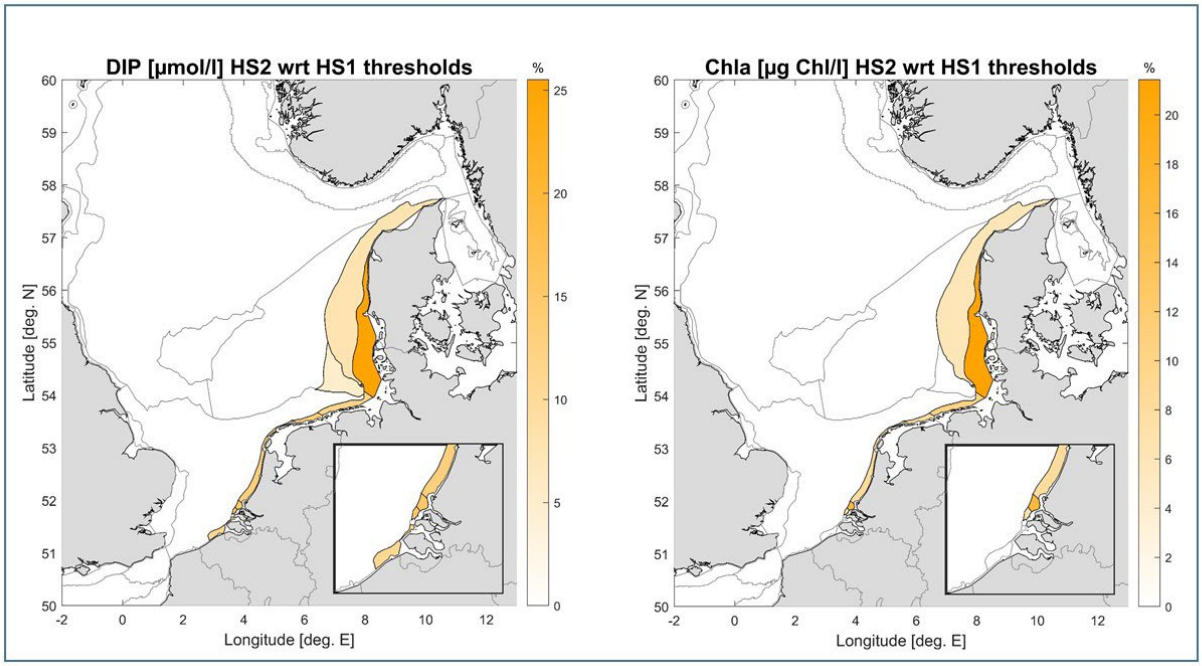


Figure 5.5: Horizontal distribution of the relative differences of DIP and Chl threshold values (in %) between the two scenarios with focus on the North Sea.

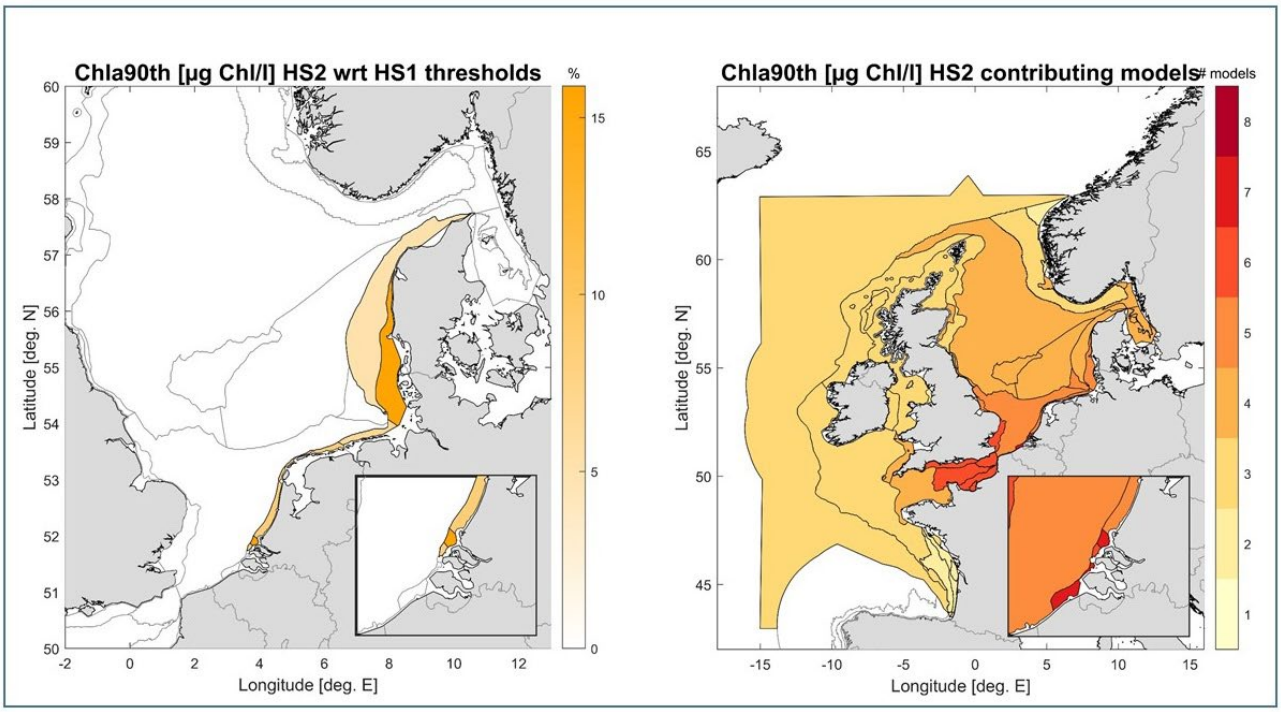


Figure 5.6: The left panel shows the horizontal distribution of the relative differences of Chl90 threshold values (in %) between the two scenarios with focus on the North Sea. The right panel shows the horizontal distribution of the number of contributing models for Chl values.

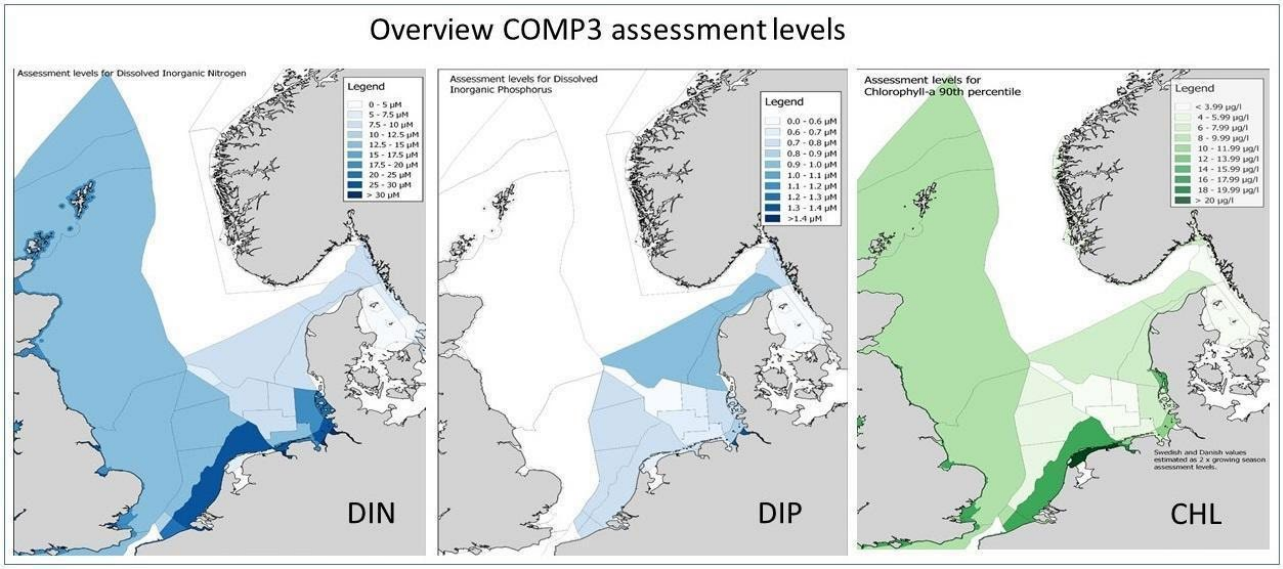


Figure 5.7: Overview on old COMP3 threshold values for DIN, DIP and Chl from OSPAR (2017)

For the comparison of the new threshold level with the old COMP 3 assessment level we provide an overview for DIN, DIP and CHL 90th percentiles (Fig. 5.7)

Figures 5.8, 5.9 and 5.10 show the horizontal distribution of the concentration differences between the newly define threshold values compared to the old COMP3 assessment levels, for DIN, DIP and CHL 90th percentiles.

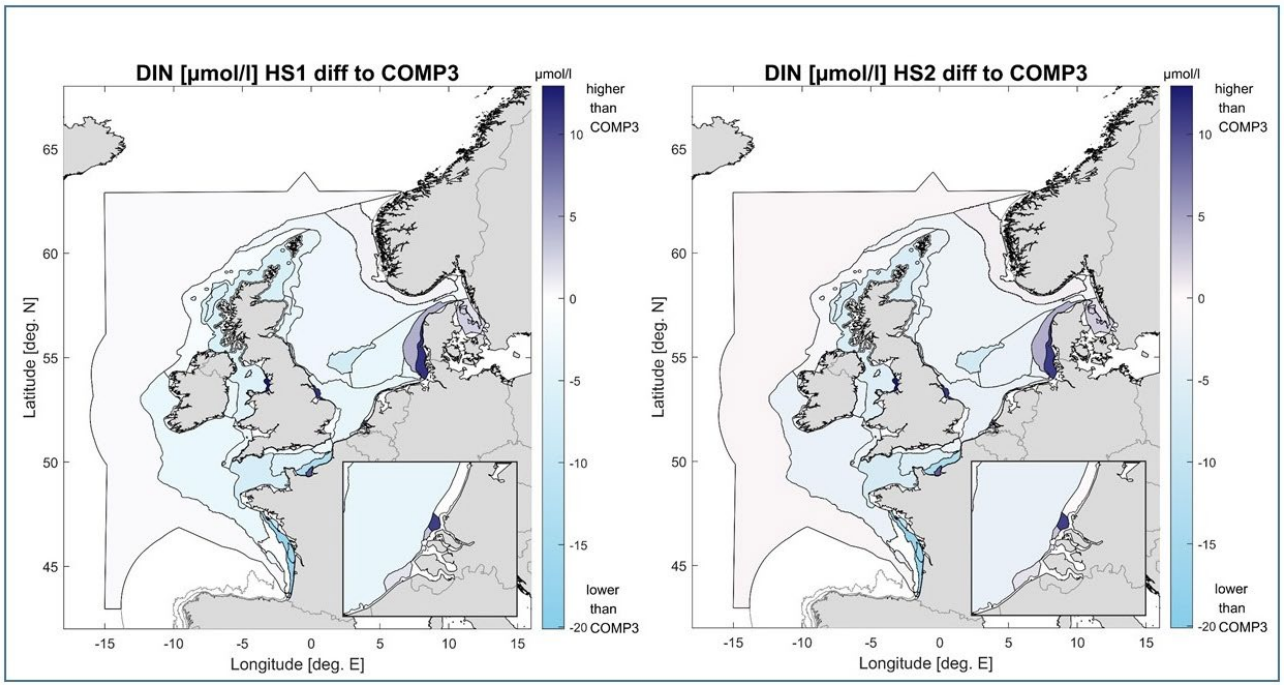


Figure 5.8: Horizontal distribution of the difference in concentration for the new DIN threshold values vs. the one from the COMP3 assessment.

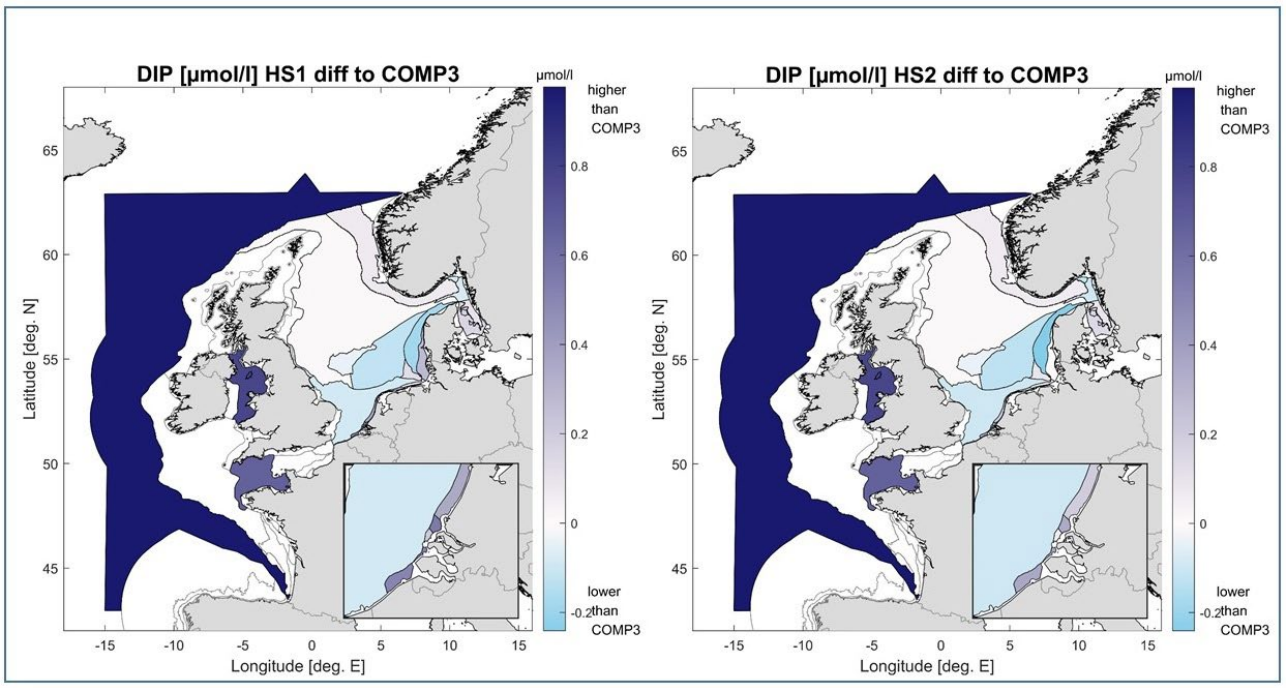


Figure 5.9: Horizontal distribution of the difference in concentration for the new DIP threshold values vs. the one from the COMP3 assessment.

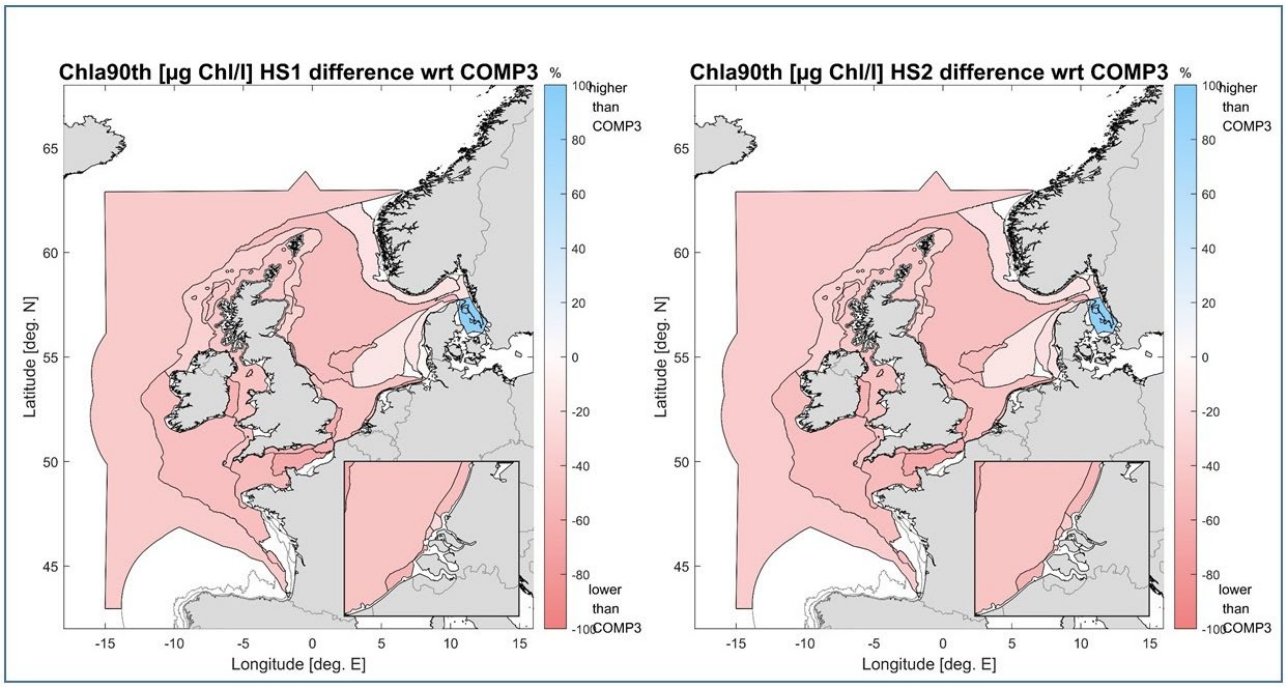


Figure 5.10: Horizontal distribution of the difference in concentration for the new CHL 90th percentile threshold values vs. the one from the COMP3 assessment.

6 Review of steps towards present assessment

6.1 Definition of steps towards present assessment

The newly proposed threshold levels for chlorophyll-a (Figure 5.3) are the combined result of different steps that have been taken to make threshold levels more coherent and more in line with natural gradients. Overall, two important steps have been taken: 1) new threshold levels are proposed for COMP-4, based on a coherent modelling approach to assess pre-eutrophic reference conditions and 2) new assessment areas are proposed for COMP-4 that better reflect natural ecosystem functioning and that are not affected by national borders. New coherent threshold levels have first been proposed by the JMP-EUNOSAT project. These have been presented to OSPAR-HASEC in March 2019. These were based on one model (from Deltares), which used a simplified approach. In March 2020 OSPAR ICG-EMO started an initiative to propose new threshold levels using a more advanced approach: based on a weighted ensemble of existing fully integrated physical-ecological North Sea models. Figure 6.1 illustrates how the threshold values for chlorophyll-a have been affected step-by-step by the above changes.

6.2 Evaluation of changes from previous COMP3 to newly proposed threshold values

Comparison of Figure 6.1 panels A and B shows the effect of the change from country specific approaches for estimating natural reference conditions to one coherent approach, without changing the assessment areas. This shows that new threshold levels (according to the JMP-EUNOSAT modelling approach) decrease in UK and Dutch waters. In German waters the new threshold levels are similar in offshore waters and are higher in near-shore waters. In Danish waters new threshold levels are lower than before in the North Sea area. In the Kattegat and Skagerrak threshold values increased.

Comparison of Figure 6.1 panels A and B shows the effect of the change from country specific approaches for estimating natural reference conditions to one coherent approach, without changing the assessment areas. This shows that new threshold levels (according to the JMP-EUNOSAT modelling approach) decrease in UK and Dutch waters. In German waters the new threshold levels are similar in offshore waters and are higher in near-shore waters. In Danish waters new threshold levels are lower than before in the North Sea area. In the Kattegat and Skagerrak threshold values increased.

Comparison of panels B and C shows the effect of a new definition of assessment areas. The new assessment areas exclude coastal waters covered by the WFD. Therefore, the assessment areas that had highest threshold levels in COMP-3 often do not have a threshold level for COMP-4. Highest threshold values are found in the river plume areas. Offshore waters have fairly similar threshold values between different countries and country borders do not affect the assessment result.

Comparison of panels C and D shows the effect of changing from a simplified modelling approach with the Deltares model (according to JMP-EUNOSAT) to a more sophisticated integrated physical-ecological modelling approach in the context of ICG-EMO and ensemble mean of several models.

Comparison of panels D and E shows the effect of using nutrient loads from historic scenario 2, which leads to lower threshold values compared to historic scenario 1 in the Elbe plume.

Comparison of panels E and F shows the effect of calculating a weighted ensemble mean compared to a regular ensemble mean. Although the assessment levels have clearly changed in some areas, (for example in Figure 4.5), these differences do not show up in the color scheme used for this plot.

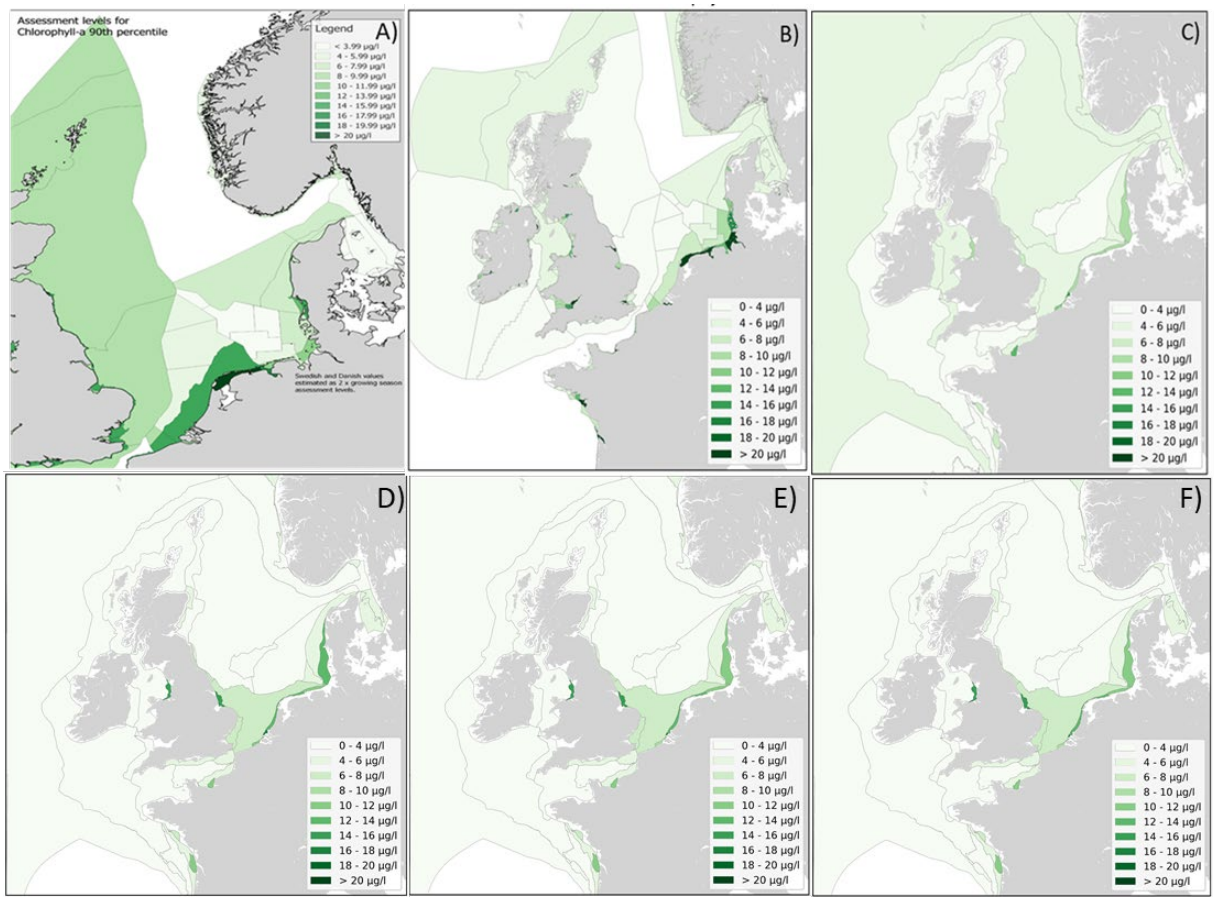


Figure 6.1: Stepwise changes of threshold levels of chlorophyll-a (90-percentile estimated as 2x growing season mean for comparability to panel A) COMP-3 assessment levels, averaged per COMP-3 assessment areas, B) JMP-EUNOSAT assessment levels, averaged per COMP-3 assessment area, C) JMP-EUNOSAT assessment levels, averaged per COMP-4 assessment area, D) ICG-EMO ensemble mean threshold historic scenario 1,E) ICG-EMO ensemble mean threshold historic scenario 2 F) ICG-EMO weighted ensemble mean threshold historic scenario 2.

7 Assessment based on newly derived threshold values

In chapter 7.1 we first present the COMPEAT assessment results which are based on the new threshold that are calculated with the weighted ensemble approach, which is described in detail in Chapter 4. In chapter 7.2 the problem of matching area averaged model data with the coastal biased in-situ data, which was the basic discussion point during the HASEC meeting in March 2021 and further TG-COMP meetings. In 7.3 we illustrate what ideas were discussed within the ICG-EMO group to further analyse the model output by the models or to find alternative approaches to moderate the thresholds, like the relative (Schernewski adopted) approach. Finally, in chapter 7.4 another weighted ensemble approach is described, the Taylor diagram analysis, which was not chosen for the final derivation of the thresholds!

7.1 COMPEAT assessment based on model derived threshold values

ICES has developed the COMPEAT tool to carry out the COMP assessment in a standardized way. With the caution that the model results and hence the threshold levels need further analysis, the results of the COMPEAT assessment give a first glance on the potential consequences of the new assessment areas and threshold levels. The threshold levels from Tables 5.1 and 5.2 were applied in the COMPEAT tool. An assessment was done using those threshold values and observations for the years 2006-2014. **The results of the COMPEAT tool for the new threshold values based on the weighted ensemble approach (see chapter 4)** from scenario 1 are shown in Figure 7.1. The results for scenario 2 derived with the same weighted ensemble approach are shown in Figure 7.2.

Those results show that there are only small differences between the two scenarios in the assessment outcome. Surprisingly there is a shift in the DIN assessment in the Southern North Sea area #11 from moderate to poor in the HS2 scenario. All other changes in HS2 in comparison to HS1 effect the Elbe Plume area #20, which changes from good to moderate for DIP, the combined *in-situ* and EO CHl-a and the final assessment.

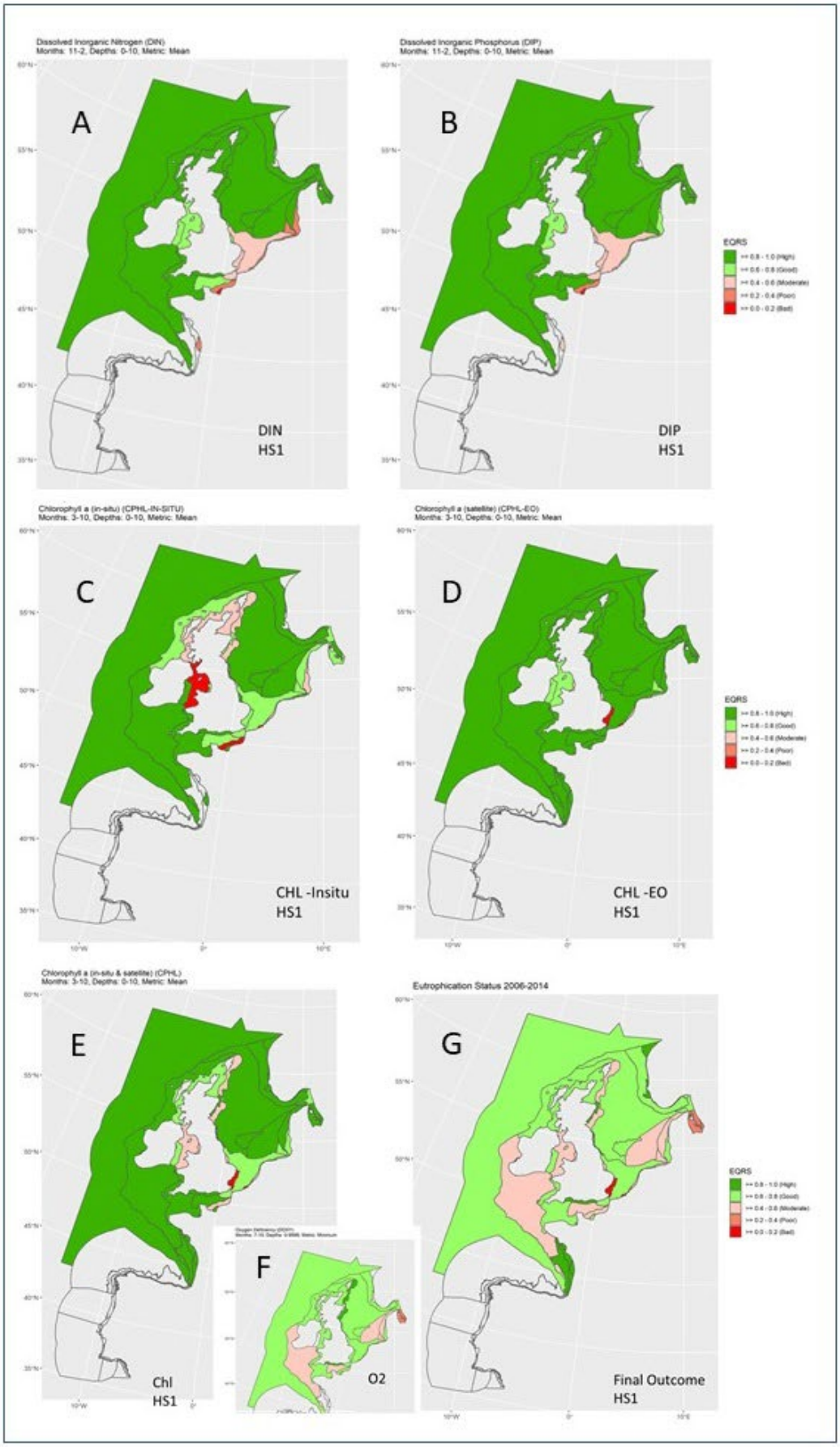


Figure 7.1 COMPEAT assessment result using HS1 thresholds for DIN (A), DIP (B), Chl Insitu (C) , Chl Satellite (D), Chl Combined (E), Oxygen (F) , Final Outcome (G). Note the oxygen assessment methodology is not finalised.

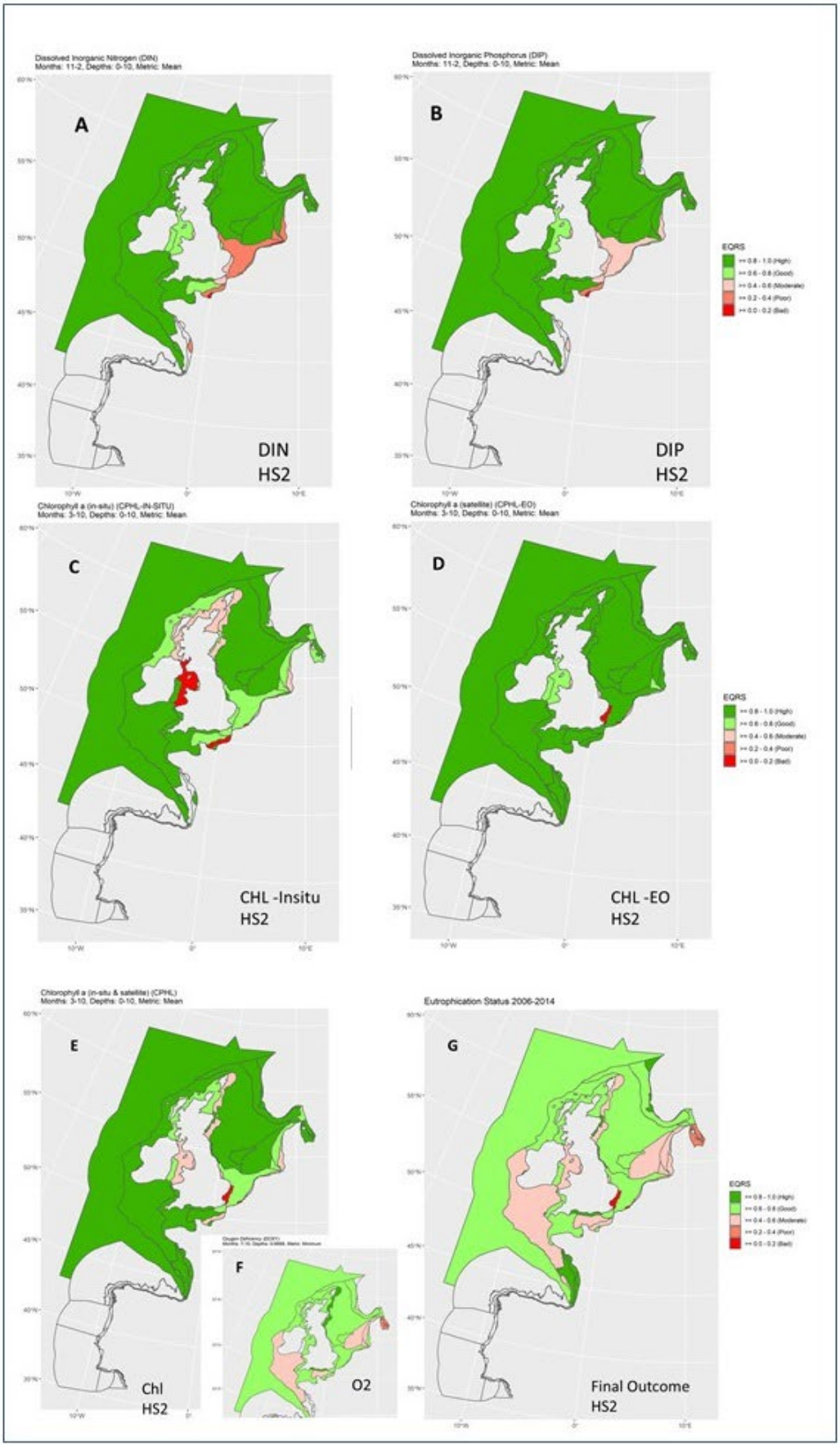


Figure 7.2 COMPEAT assessment result using HS2 thresholds for DIN (A), DIP (B), Chl Insitu (C) , Chl Satellite (D), Chl Combined (E), Oxygen (F) , Final Outcome (G). Note the oxygen assessment methodology is not finalised.

7.2 Problems of matching model derived threshold values with values based on in-situ data

In the COMPEAT assessment tool, average concentrations per assessment area are used to compare observations with threshold values. Those averages are based on in-situ data from one or more sampling stations. The spatial coverage of the assessment areas by *in-situ* sampling differs significantly and not all assessment areas are covered adequately by the current monitoring programs. The sampling frequency also differs between sampling stations. In addition to the *in- situ* data, satellite data can be used for chlorophyll-a, which would provide averages for assessment areas that are based on a more comprehensive data set, both in terms of spatial and temporal coverage.

The model results for the current state run and the historic scenarios are also converted to average values per assessment area. These averages are based on model output at a much finer spatial and temporal resolution than the observations.

In an ideal world, the average value per assessment area for the observations gives an accurate value for each of the variables that are used in the assessment. However, the probability of a bias in the observations has to be taken into account. This bias may be due to the position of monitoring sites in an assessment area that does not always show a spatially homogeneous distribution. The bias may also be caused by the timing of sampling that does not sufficiently cover all temporal variability in the concentrations.

As Figure 7.3 shows, the majority of observations in the ICES oceanographic database are found in areas close to the coasts. This means that most of the observations occur in areas that are more strongly influenced by freshwater discharges, have relatively high nutrient concentrations and to some extent also relatively high chlorophyll concentrations. Particularly the more offshore assessment areas are less well covered by *in-situ* monitoring and have a higher probability for biases in the average concentrations.

This indicates that there is a need to redesign monitoring programs to have a better and more representative coverage of the assessment areas, at least for those variables that can only be determined by *in-situ* sampling. For chlorophyll, the use of satellite data would be a solution to overcome this problem.

Alternatively, if monitoring sites cannot be changed or if data from past years are used in assessments, it should be considered to apply site-specific threshold levels instead of assessment area averaged thresholds. These site-specific threshold values can be derived from the model results.

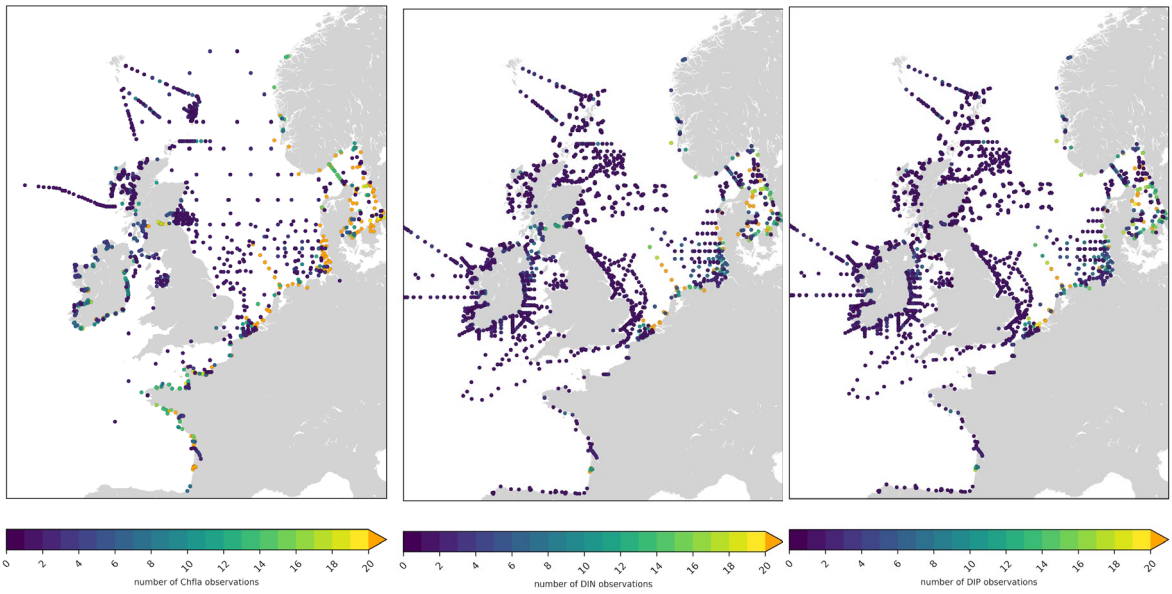


Figure 7.3: Horizontal distribution and number of observations of Chlorophyll, DIN and DIP in the ICES dataset in COMPEAT tool for 2006-2014.

COMPEAT produces estimates of confidence in the assessment, for each parameter. Below (Figure 7.4) are shown the spatial confidence for DIN, DIP and Chlorophyll as an example. The assessments are combined along with a methodological accuracy to produce an overall estimate. Evident from the below is that there is generally poor temporal and spatial coverage.

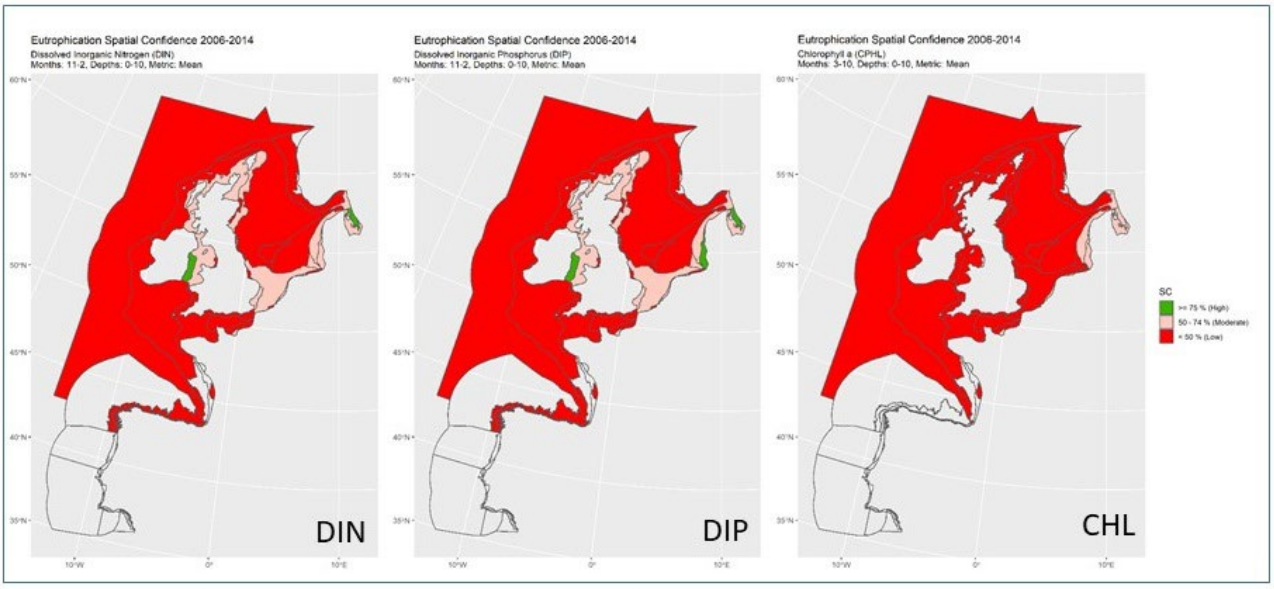


Figure 7.4: Horizontal distribution and of the spatial confidence rating of in-situ Chlorophyll, DIN and DIP in the ICES dataset in COMPEAT tool for 2006-2014.

7.3 Alternative approach of matching modelling and in-situ data based on threshold values

In the previous section, the consequence is discussed of using observation data that do not accurately describe an assessment area. In addition, the consequence of biases in model outcomes also needs to be considered. In an ideal world the model results for current state give an accurate description of the concentrations in an assessment area (and do not deviate from observations), and the model runs for the historic scenarios can be assumed to give a good description of concentrations under the conditions in those scenarios. However, there is also a possibility of bias in the model results. This bias can be caused by the fact that models are always a simplification of reality and do not include all biogeochemical processes determining the concentrations of DIN, DIP, chlorophyll and other variables. Consequently, it is unlikely that model averages for assessment areas exactly match the observation averages. If observations and models show random differences from the 'true' values for an assessment area, this just lowers the precision of the assessment result. But if there is a systematic deviation from the 'true' value, there is a risk of misclassification in the assessment.

If model values give an underestimation, the threshold values derived from the model results may be too low and this leads to the risk that the assessment is too strict and the area is misclassified as 'eutrophic'. In contrast, if the model values give an overestimation, the threshold values may be too high leading to the risk of a too loose assessment result and misclassification of an area as 'non-eutrophic'.

Since the HASEC meeting in March 2021, where the first, preliminary report was presented, a number of steps have been taken to overcome this and other problems. The ICG-EMO partners defined general topics out of the feedback that was aggregated from the request to TG-COMP to highlight problems of the previous report. The focus from ICG-EMO up to the present report are:

1. Quality of model results
2. Methods to derive thresholds for assessment

In terms of model validation, a huge effort was undertaken by comparing model time series with *in-situ* data. This effort is illustrated at the end of chapter 3.2.2. In the following sections we will describe further efforts that ICG-EMO has undertaken in digesting the model results in greater depth. The first analysis tool takes care of the responsiveness between the Current State run and the historic scenario. In view of modulating the threshold a number of issues where discussed within the ICG-EMO group, starting by the possibility of a salinity correction, followed by a deeper look into the Schernewski approach. and finally some more background on the weighted ensemble approach. While the later was chosen for the final derivation of the thresholds, and is discussed in detail in chapter 4, still another attempt was made based on Taylor diagram analysis from the time series data which is shown in chapter 8 (Fig. 8.1 – 8.3).

7.3.1 Comparison of model outputs under CS and HS1 conditions

The responsiveness of model results was analysed under the CS and HS1 conditions. All models involved in the study tend to show similar features such as the imbalance between winter Ntot(Total Nitrogen) and Ptot (Total Phosphorus) concentrations in coastal zones. This is illustrated by Figures 7.5 and 7.6 where model outputs follow a consistent curve across the range of nutrient concentrations.

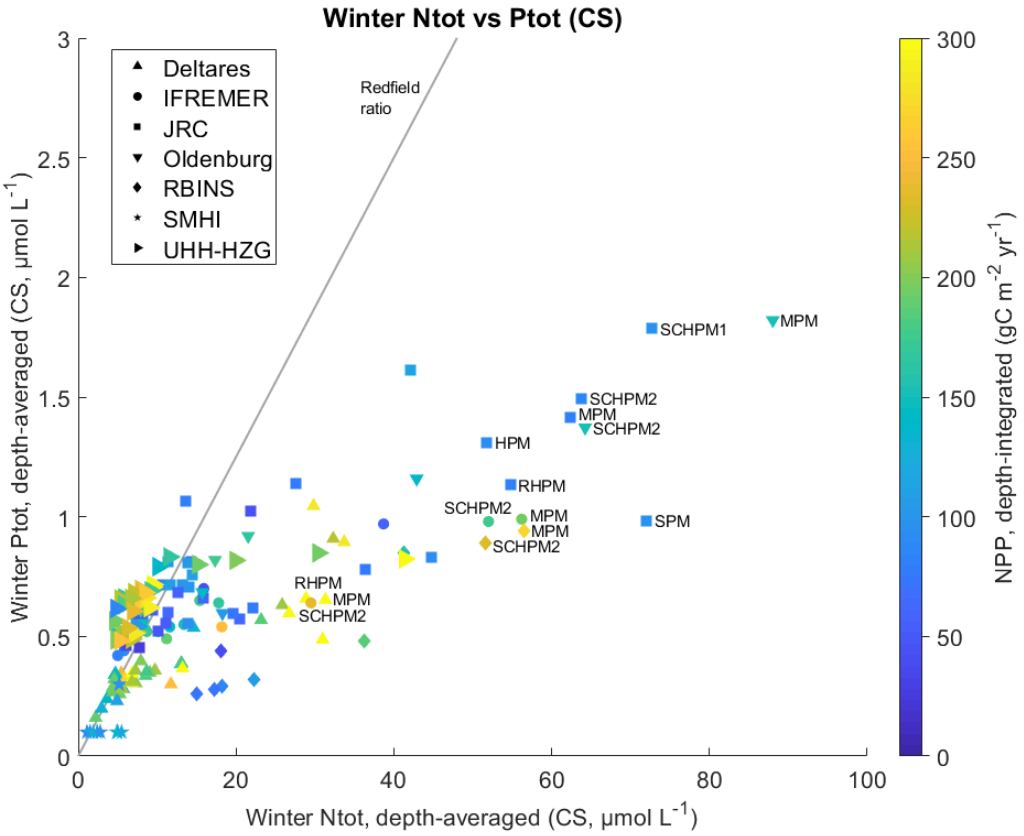


Figure 7.5: Model results of winter total phosphorus (TP) concentration (shown as Ptot) as a function of winter total nitrogen (TP) concentration (Ntot) in each area for each model under the CS conditions (period 2009-2014). The colours indicate the modelled net primary production (NPP) for the corresponding areas and models (see colour bar for values). The line indicate the global ocean Redfield ratio (N:P = 16).

In Figure 7.5, the relationship between modelled TN and TP tends to reproduce the Redfield ratio in offshore areas, characterized by low concentrations for both variables but not in coastal areas where TN is in large excess due to imbalance river input. Areas showing greatest imbalance in the N:P ratio are found close to the Delta rivers outlets (Rhine, Meuse, Scheldt), but also in the Humber plume and the Seine plume. This well-known feature is due to eutrophication, effects of which are most visible in the coastal areas and river plumes, and least in the offshore areas as the riverine material become diluted. Figure 7.6 shows the same model outputs for the pre-industrial scenario HS1, and illustrates how a nutrient load reduction could mitigate the excessive N input to the coastal zone while also lowering the P inputs. Comparing Figures 7.5 and 7.6, the highest N:P value (i.e., in the Meuse plume) drops from ~53 molN $molP^{-1}$ under CS conditions to 27 molN $molP^{-1}$ under pre-industrial conditions. It may be worth to point out that the pre-industrial conditions shown in this study are not identical to pristine conditions. In a previous study, Desmit et al. (2018) reported that the N:P ratio averaged across the North-East Atlantic coastal domain would be lowered from ~35 molN $molP^{-1}$ under current conditions to ~11 molN $molP^{-1}$ under pristine conditions. As a remark we want to mention that the Redfield ratio is only applicable to marine waters (more specifically to averaged global oceanic conditions). In freshwater systems the N:P ratio in pristine waters may be much higher than the Redfield value (Billen & Garnier, 1997; van Raaphorst & de Jonge, 2004). Nevertheless, the ratio between elements (N:P but also N:Si, not shown) in coastal bodies shapes the structure of the phytoplankton communities (Cloern, 2001). The example reported by Radach et al. (1990) in the

German Bight illustrates the issue: the decrease of the Si:N ratio below 0.1 led to a coincident shift in dominance from diatoms to flagellates/dinoflagellates.

The models perform similarly as they follow the same curve although they show significant differences in concentrations for winter TN and TP in coastal areas, as in the case of differences between Deltares vs. IFREMER or JRC vs. RBINS in the Delta River Plumes. This suggests that either the river loads are somewhat different between models (this is not likely as all modellers followed the same procedure) or the differences in hydrodynamics already impose differences in coastal winter nutrient concentrations. The proximity of boundary conditions may also influence the concentrations in an area, for instance the case of Oldenburg's model showing its southern boundary close to the Delta Rivers plumes. The differences between the ranges of nutrient concentration covered by models reflect the differences in application areas (e.g., RBINS is mostly a coastal model). Differences in the NPP response to nutrient concentrations (typically, Deltares and Hamburg exhibit relatively larger NPP than the other models) indicates that the biological modules also impose differences between model results.

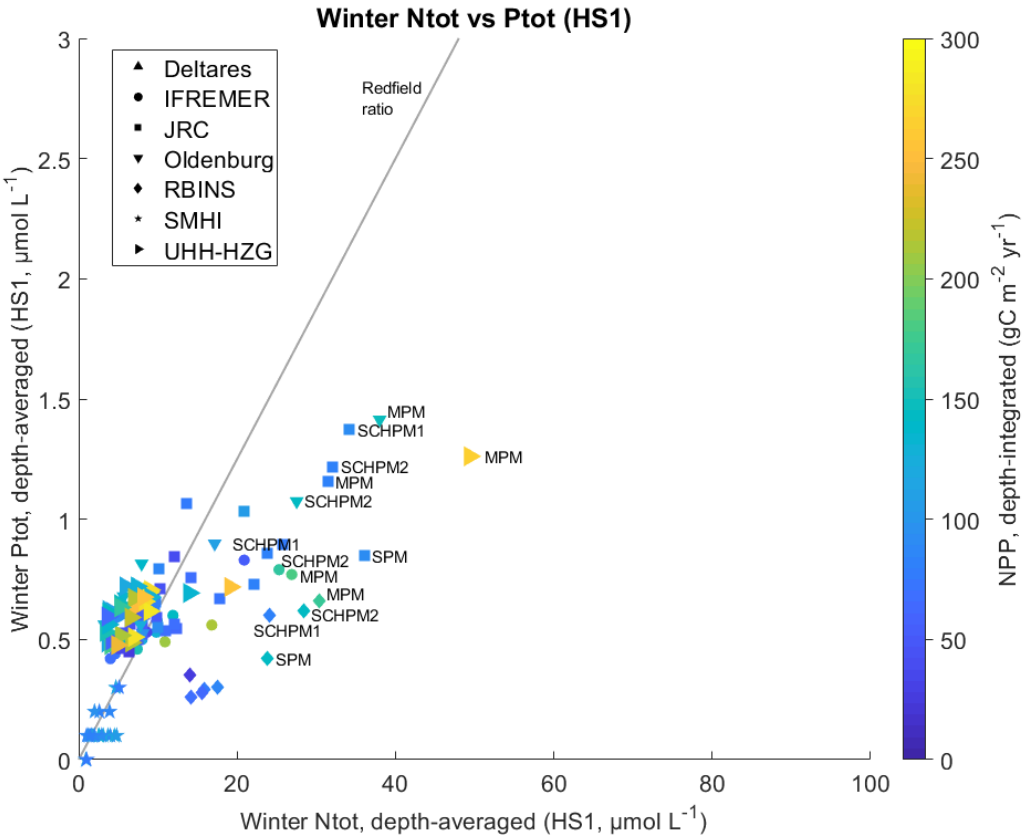


Figure 7.6: Model results of winter total phosphorus (TP) concentration (shown as Ptot) as a function of winter total nitrogen (TN) concentration (shown as Ntot) in each area for each model as output of the HS1 scenario (period 2009-2014). The colours indicate the modelled net primary production (NPP) for the corresponding area and model (see colour bar for values). The line indicate the global ocean Redfield ratio (N:P = 16).

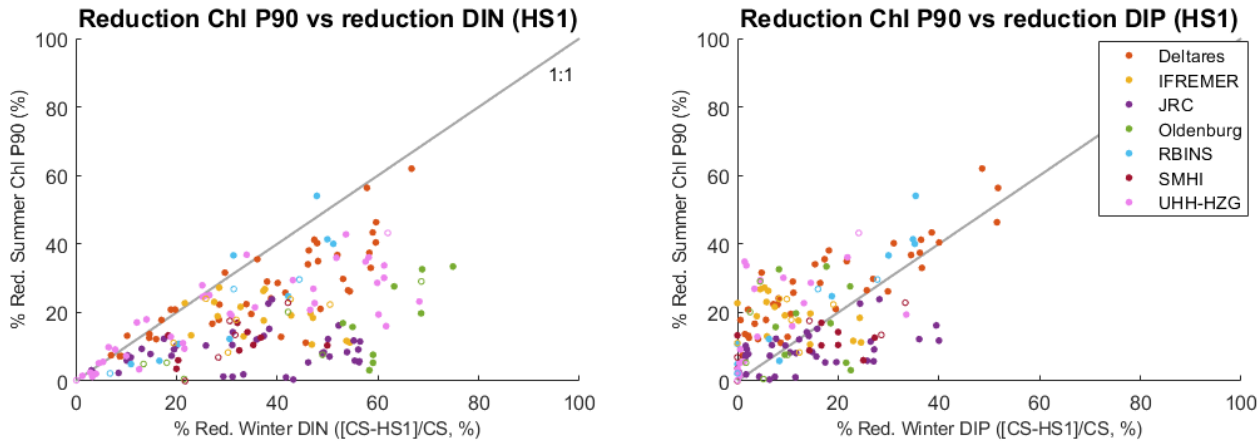


Figure 7.7: Model results illustrating the impact of the HS1 scenario compared to the CS situation. Relative reduction of summer Chl P90 as a function of the relative reduction of winter DIN (left) and winter DIP (right). Each dot indicates a marine area for one model and models are discriminated by their colour.

Figure 7.7 depicts how a relative reduction in winter nutrient concentration (HS1 compared to CS) may induce a relative reduction in Chl P90 (which represents the optimum of the Chl bloom). The relative reductions in DIN are in general stronger (mostly between 20-60%) than the reductions in DIP (mostly between 0-40%). A high amount of scatter occurs in the response of Chl P90 to winter DIN between the 40-60% reduction range. This suggests that Chl P90 can be relatively insensitive to DIN reductions in some areas, which are likely coastal bodies with too high background DIN (such that despite a significant DIN reduction, phytoplankton production and biomass accumulation remain controlled by other factors). On the contrary, the response of Chl P90 to the reduction in DIP seems to be more robust at a higher range of reduction range (>30%). At a lower range of reductions, we observe a mixed response: some regions show limited response (not surprising, as after a small reduction, P is likely not limiting), and other regions show relatively high responses (above the 1:1 line). This mixture of response under small P reduction may be due to concomitant N reductions or changes in the modelled phytoplankton community. This interpretation would require deeper analysis, which was not possible in the frame of this study due to time constraint.

7.3.2 Estimation of thresholds and assessment boundaries

Based on the threshold values for each parameter that are found by adding 50% to the reference values, the COMPEAT tool categorizes the observations according to the boundaries between various environmental status levels (reference and threshold values themselves known as the 'BEST' and 'ET' values, respectively, latter setting the boundary between good and moderate status). So for example if the historic Scenario estimated a DIN of 10 µmol/l, this value itself would be the maximum boundary for the BEST value, and adding 50% to it, it would set a threshold (ET) value of 15 µmol/l (50% added). The high/good boundary, would be 12.4 µmol/l, good/moderate 15.0 µmol/l, Moderate/Poor 19.0 µmol/l, Poor/Bad, 26.1 µmol/l. Observations < 12.4 µmol/l would be high and > 26.1 µmol/l Bad. Thus, an observation in the Atlantic region of 10.7 µmol/l for example, where the threshold is 15 µmol/l would be classified as high. In the Irish sea, observations of 8.2 µmol/l with a threshold of 9.5 µmol/l gives good, in the southern North Sea, observations of 15.1 µmol/l, with a threshold of 14.2 µmol/l gives moderate and in the Thames plume, observation of 20.9 µmol/l, with a threshold of 14.2 µmol/l and gives poor.

Relevantly, it should be pointed out that using the addition of 50% in all areas leads to counterintuitive results, such as threshold values of DIN and DIP being higher in some off-shore or oceanic waters such as the Atlantic region, or the Northern North Sea than the inner areas, such as the Eastern North Sea (Fig.5.1-5.2), which contrasts with the current values being naturally higher in the inner regions than in the outer regions as indicated by the CS run (e.g., Fig. 3.1). This apparent inconsistency was earlier pointed out by Kerimoglu et al. in the OSPAR meeting in Hamburg, 2019, and by Alain Lefebvre at a TG COMP meeting with actual model results used for the current exercise. Essentially, the reason for this counterintuitive outcome is the weak sensitivity of the assessment parameters in the off-shore areas to the riverine loadings, thanks to the 'flushing' by the ocean water. This weak sensitivity leads to historical estimates being very close to the current state. In turn, the threshold levels, obtained by elevating the historical estimates by 50% end up being higher in the outer regions than in the inner regions, as mentioned above, which can therefore be considered as a methodological artefact. It should be noted that the procedure (elevation of historical concentrations by 50%) was applied to the Baltic Sea by Schernewski et al. (2015), which did not give rise to any implausible results. This is probably because in the Baltic Sea, which is a semi-enclosed basin, there may be no regions that are rapidly flushed by the ocean, hence, insensitive to the riverine loadings.

Therefore, 50% addition is critically important for the whole assessment process, highlighting the need to understand the exact logic behind the chosen 50%. In the WFD CIS Guidance Document No. 5 (2003), OSPAR (2013), the 50% addition is grounded as 'reflecting natural variability and (slight) disturbance'. The ICG-EMO suggests that the amount of 'natural variability' and 'allowable level of (slight) disturbance' to be defined more clearly. The natural variability, and its temporal and spatial components can be determined on a technical basis, using observation and model products. For instance, due to the influence by the seasonal fluctuations in the riverine discharge rates, and steep coastal gradients, the natural variability in the near-coastal and plume areas are much larger than the offshore areas (Stegert et al., 2021). Moreover, DIN and DIP have a much lower variability (as a % of the mean) than chlorophyll. This would indicate that the 50% addition for DIN and DIP is too large an addition, whereas likely not sufficient for Chl. The allowed `slight' disturbance on the other hand, may need to be determined politically, at least in the near future. In the longer term, a scientific approach can be developed for this problem as well.

7.3.3 Realtive (Schernewski style) approach

To overcome the risk of misclassification due to systematic bias in the model results, an approach was developed by Schernewski *et al.* (2015) for the implementation of eutrophication targets for the WFD in the German coastal waters of the Baltic Sea. In this approach, the threshold values were not derived directly from modelled concentrations in the reference scenario, but rather by the relative difference between modelled concentrations in the reference scenario compared to the present scenario. The reference concentrations are then estimated by applying this relative difference to recent observations. This approach has also been used to estimate the riverine nutrient inputs under reference conditions based on E-HYPE modelling results and ICG-EMO observed data of present nutrient loads.

An example is shown in Figure 7.8. In Step 1, the model results for the current state run (CS) and the historic run (HS) are compared. In this case, the model result for the historic run (concentration of 60) is 40% lower than the current state run (concentrations 100). In Step 2, the relative difference between the two scenarios (-40%) is used to calculate a reference from the observed value

(concentration 85), which is then 51. In Step 3, the threshold value is calculated from the reference value of 51 by adding 50%, which is a concentration of 76.5. Applying this threshold value would show that the current concentration of 85 is above the threshold level. If simply the concentrations from the model runs would have been used, the reference value would have been the outcome of the HS model run (concentration 60), and the threshold value derived from that would have been 90. In that case, the current concentration would have been below the threshold value.

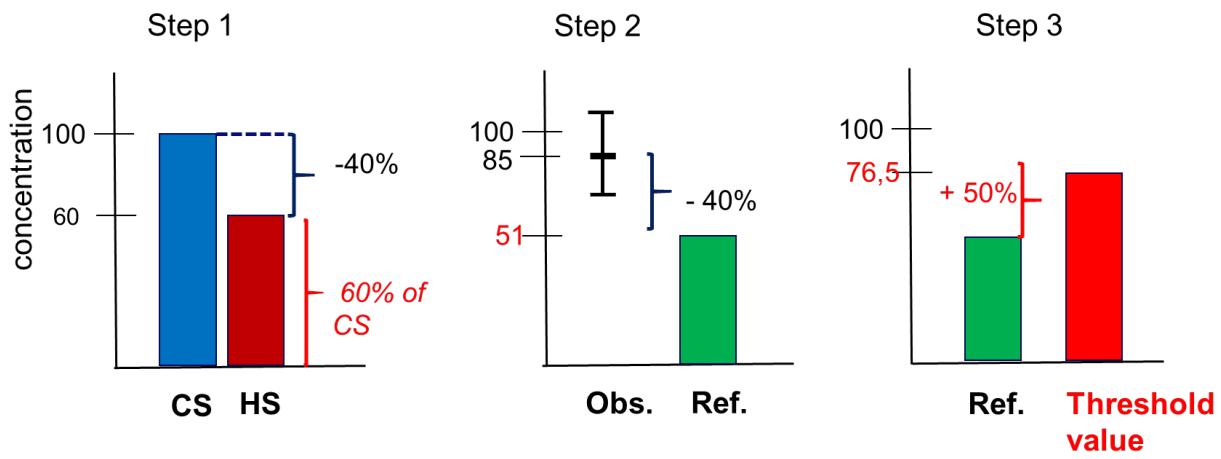


Figure 7.8. Example of alternative method to derive threshold values. See text for further explanation. In this example, the CS model run overestimates the concentration.

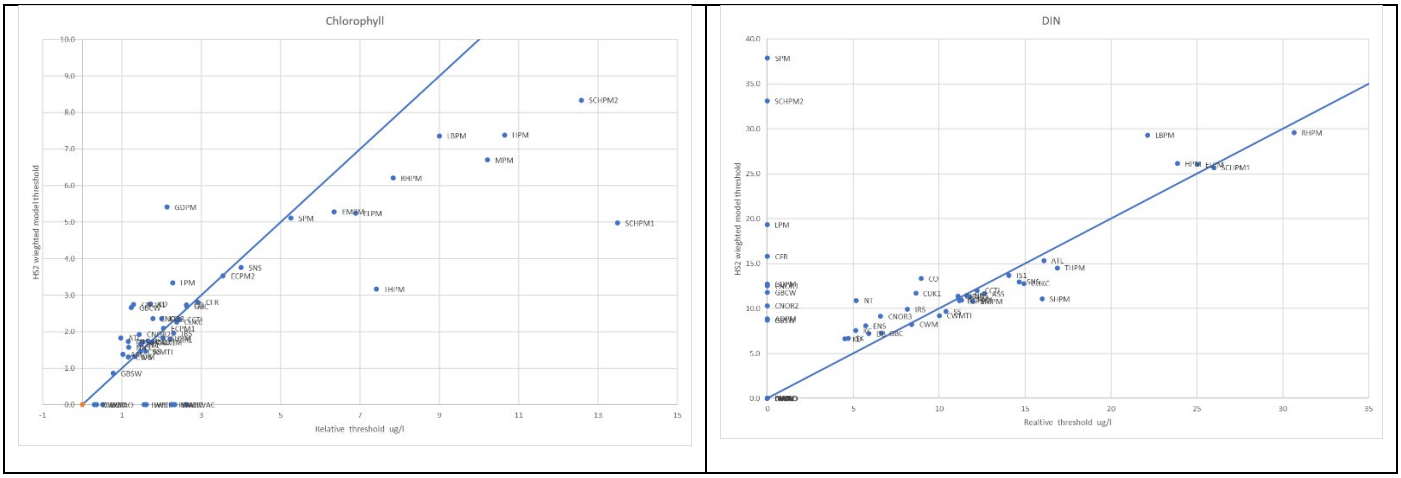


Figure 7.9 Threshold calculated using the weighted method versus the relative method (left for chlorophyll, right for DIN). Blue lines mark 1:1 ratio

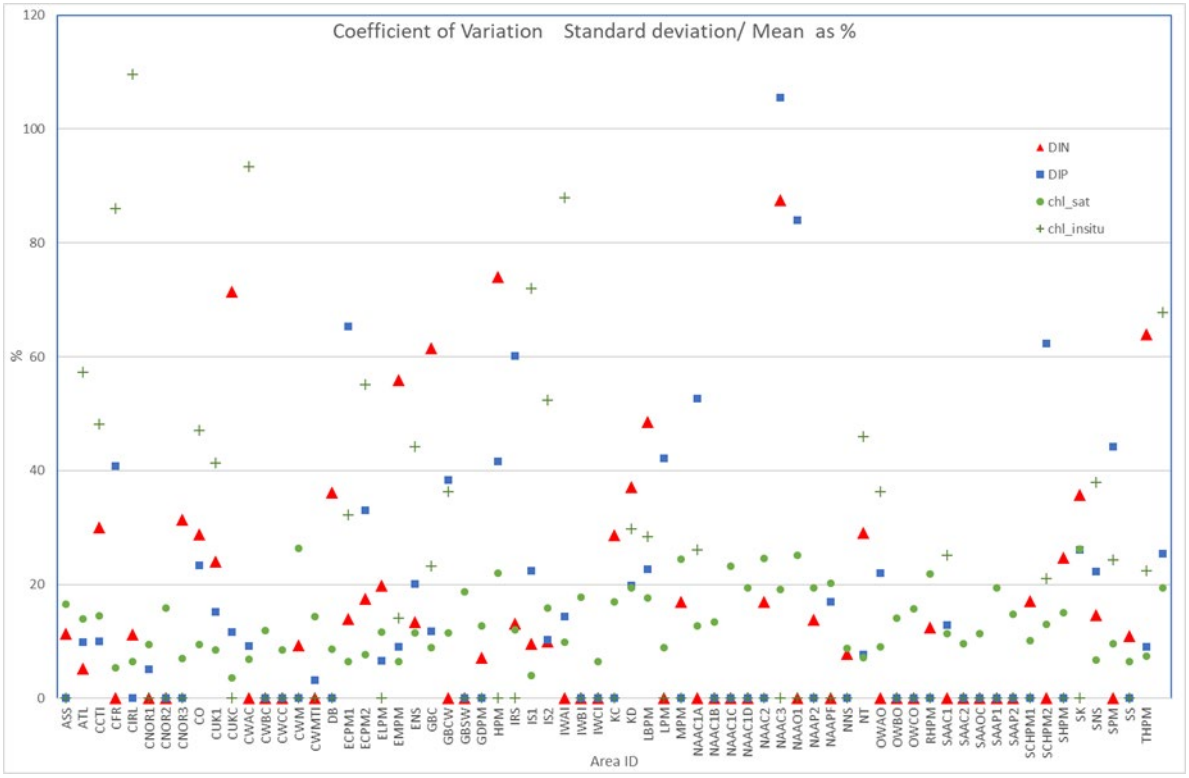


Figure 7.10. The variability of the parameters as the interannual standard deviation divided by annual mean.

Evident from Figure 7.9 is that for DIN there is generally very good agreement between the weighted ensemble method and the relative method. Also evident is the number of assessment areas that do not have DIN data and therefore cannot use the relative method to define a threshold. The figure for chlorophyll thresholds, shows a good relationship at low values of chlorophyll, but for many of the plume regions which have high values the relative method gives much higher thresholds. It is possible that in these plume regions the models have systematic bias, for chlorophyll, in these regions when compared with the satellite observations.

Analysis of the natural variability for each assessment area can help understand these results. Fig. 7.10 shows the relative standard deviation, for each assessment area, the inclusion of satellite derived chlorophyll significantly reduces the standard error and improves confidence in using the relative approach. However, DIN and DIP show high variability for some assessment areas and some areas have no value. Thus, it is appropriate to use the ensemble modelling for regions where there are no values and the weighted ensemble method for areas when confidence between the two methods is high. For regions where the two methods of threshold derivation are significantly different, then the relative method should be considered, as it may produce a threshold value (for chlorophyll) which is more consistent between similar regions, e.g. Thames plume and Rhine plume.

7.4 Taylor diagram analysis

In addition to the implemented weighted ensemble approach, alternatives were discussed and evaluated with respect to their individual advantages and shortcomings. We tried for example to compute the weights for each assessment area for each model system out of the time series analysis described in chapter 8 (Fig. 8.1 – 8.3). Thereby, we computed for each monitoring station the individual model quality by assuming that low RMSD and high correlation factors are indicating good model results. In a second step, these station-specific weights were transferred to whole COMP4 area and assessed for each unit individually. While this approach seemed to have a high potential benefit, it was hardly feasible, as sufficient observations were missing in large open sea areas, especially the Irish Seas, Bay of Biscay and the Atlantic Ocean (Fig. 7.11). Nevertheless, the authors will work on this approach further, as they see a high potential benefit, if suitable weights can be derived from the time series analysis.

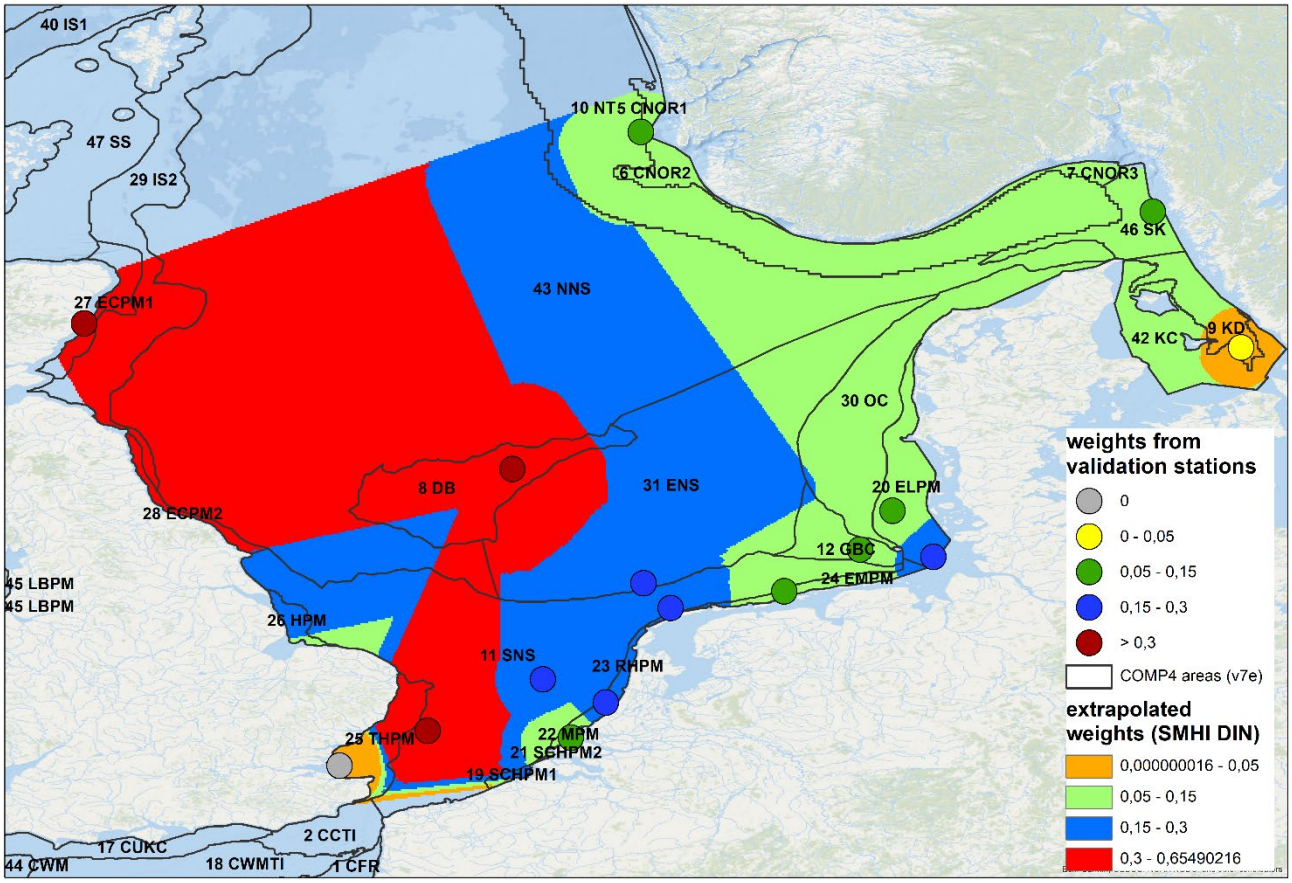


Figure 7.11: Example to matching weights based on the Taylor method onto the SMHI model results.

8 Reflection on ICG EMO approach; dealing with uncertainty

Ecosystem models are a simplified representation of the system that can help to gain mechanistic understanding of processes (Skogen et al. 2021), for example the relation between anthropogenic nutrient loading and the environmental state of marine systems. Data from observations are helpful to explore statistical relations, but in many cases statistical correlations cannot resolve ambiguity about cause-effect relations, like nonlinear responses between inorganic nutrients and primary production. The combination of models and observational data is essential to improve our system understanding.

Here, we have used a number of coupled hydrodynamic-biochemical models to estimate concentrations of nutrients and chlorophyll in a large part of the NE Atlantic, under current (2009-2014) conditions of anthropogenic nutrient loading and under a historic reconstruction of nutrient loading. The model exercise needs to provide a quantitative basis for the determination of coherent thresholds for the next assessment of eutrophication in the OSPAR area.

There are no straightforward methods to quantify model uncertainty, but as the model results need to support political decision-making, at least a qualitative assessment of uncertainty is necessary. Uncertainty stems from imperfect and simplified representation of the 'real' system in the models, uncertainty in model parameter values, forcing functions, etc. In addition, there is uncertainty due to variations in space, over seasons and between years (Stegert et al. 2021). In the context of the historic scenarios, which form the basis of the reference condition and therefore for the derived threshold estimates, uncertainties in relation to the input data for the scenario runs need to be added into the consideration.

For this report ICG-EMO has to deal with, at least, four types of uncertainty that have influence on the quality of the threshold estimates:

1. Plausibility of model results
2. Scientific background of the applied historic scenario
3. Model response to nutrient reduction
4. Weighted ensemble approach and threshold estimates

8.1 Plausibility of model results

The plausibility of the model results in the representation of the current state (CS) situation is a pre-request for any further work on reference condition and the final derivation of threshold estimates. Therefore, the first question is if the model results give a good representation of the current situation. There are several approaches to look at the plausibility of the model results.

We have compared the model results for the years 2009-2014 with observations for more than 30 *in situ* stations that cover a spatial gradient from stations near river mouths with high nutrient concentrations to offshore stations with low anthropogenic nutrient pressure. Both correspondence in the time series over the 6 years and correspondence in the 6-year averages were quantified. Model skill (does the model give an accurate prediction) can be quantified with various methods (Stow at al. 2009). One approach is to look at the correlation between model predictions and observations. The correlation coefficient "r" provides a quantitative metric of the performance of the model compared

to observations and measures the tendency of the modelled and observed values to vary together, for example whether the modelled values reflect the observed seasonal patterns. However, the correlation coefficient can be high while modelled and observed values differ with a consistent factor (e.g. modelled values are 2 times observed values). The root mean square error (RMSE) expresses the magnitude of the difference between modelled and observed values.

In a so-called Taylor diagram, two metrics are combined (Figure 8.1, 8.2, 8.3). On the horizontal axis the correlation coefficient is shown as 1-r. On the vertical axis, the RMSD normalized by the standard deviation of the observations is shown. This is equivalent to the cost function described in §4.2.1. The figures show the model results for individual monitoring stations and the overall model result (large symbols). The closer the values are to the origin (0,0) the better the model performance. The results for DIN, DIP and chlorophyll show that in general, model results for nutrients show a high similarity with observations, while chlorophyll shows more deviations. This is to be expected, as winter nutrient concentrations are to a large extent determined by transport processes, whereas for growing season chlorophyll many other processes are involved (primary production, sedimentation, grazing, mortality) that add variability.

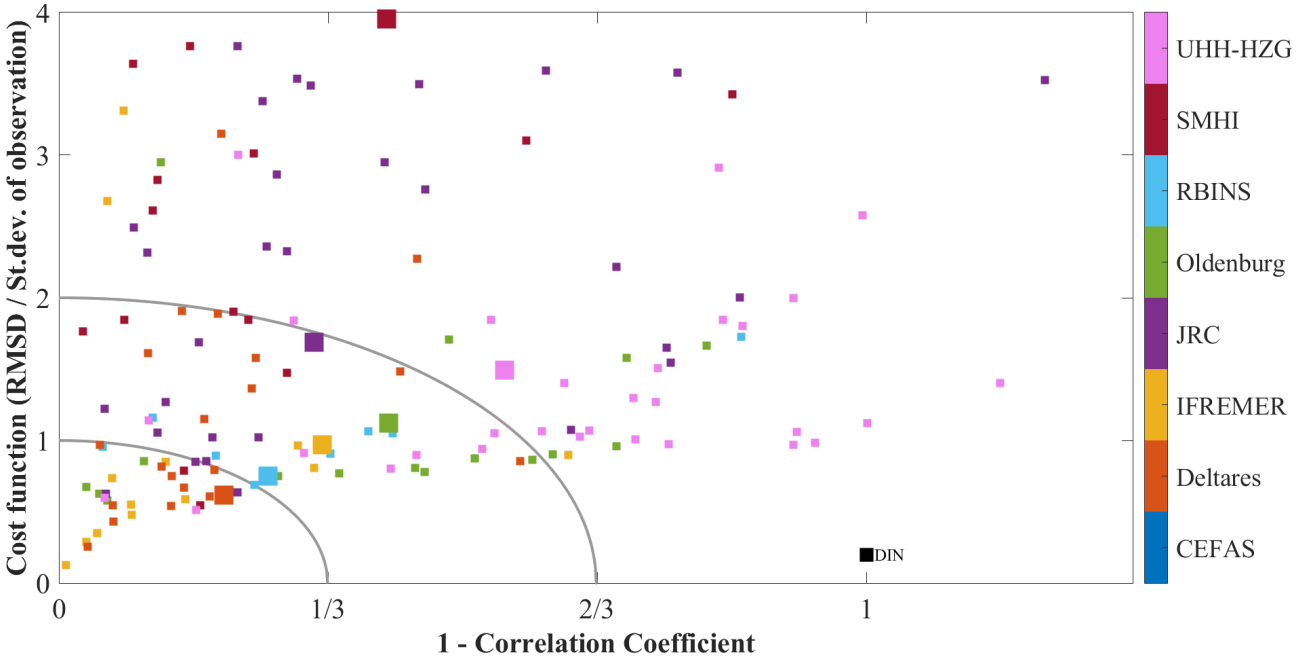


Figure 8.1 Taylor diagram showing the model skill for DIN. The horizontal axis shows the correlation coefficient between model values and observations, the vertical axis shows the cost function. The closer the values are to the origin, the better the model performance. Small symbols show the results of each model per monitoring station, the large symbols show the overall result for each model.

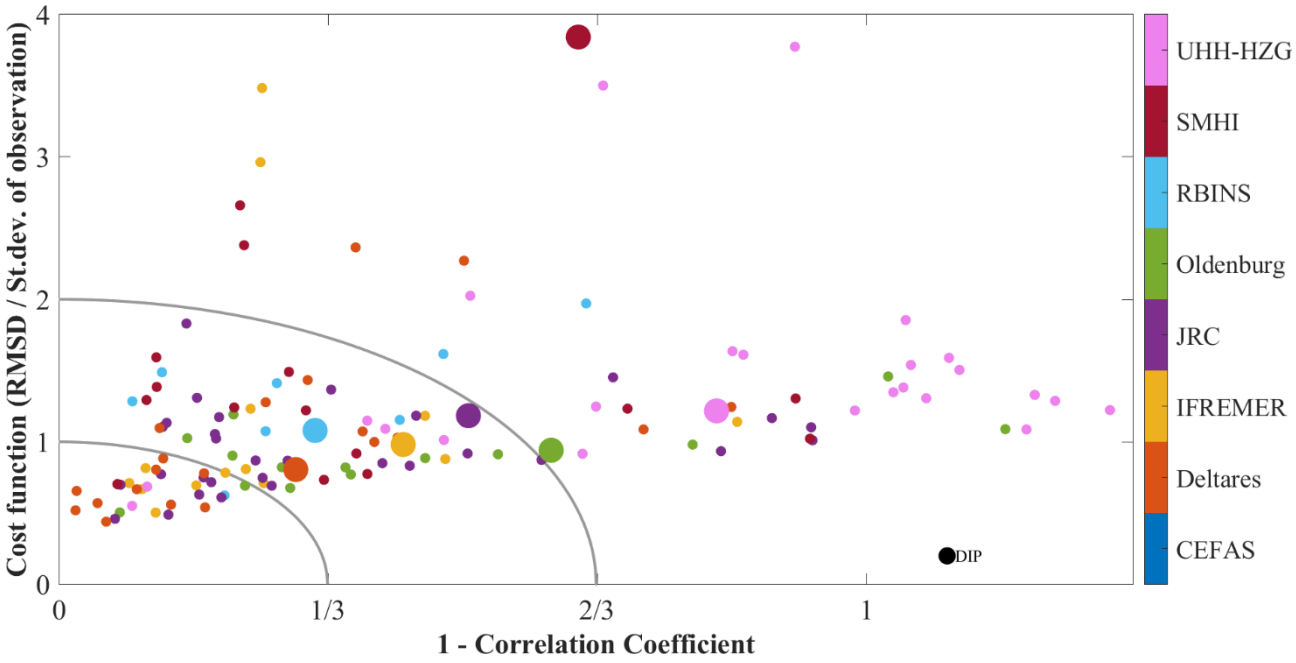


Figure 8.2 Taylor diagram showing the model skill for DIP. The horizontal axis shows the correlation coefficient between model values and observations, the vertical axis shows the cost function. The closer the values are to the origin, the better the model performance. Small symbols show the results of each model per monitoring station, the large symbols show the overall result for each model.

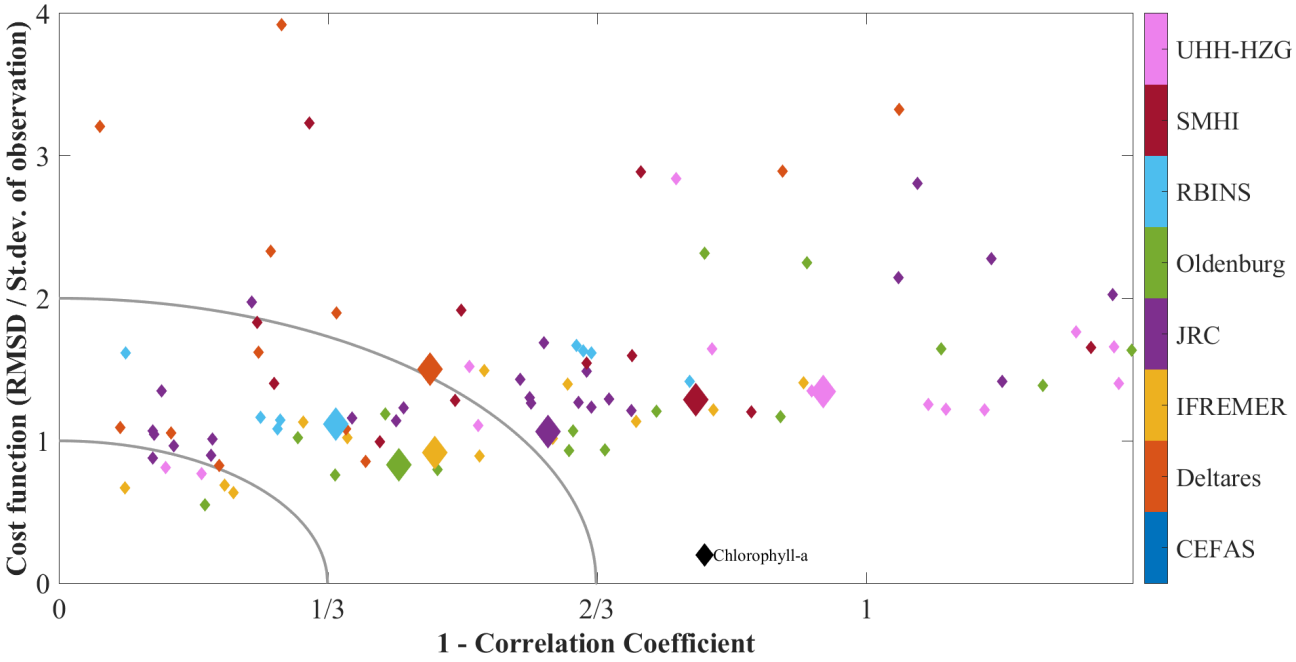


Figure 8.3 Taylor diagram showing the model skill for chlorophyll. The horizontal axis shows the correlation coefficient between model values and observations, the vertical axis shows the cost function. The closer the values are to the origin, the better the model performance. Small symbols show the results of each model per monitoring station, the large symbols show the overall result for each model.

The ensemble model approach reduces the uncertainty in model outcomes. As shown in §4.3, there can be considerable differences between model results for a specific assessment area. The average of the individual model results for an area generally shows a better correspondence with the area average of the *in-situ* data. Applying weights based on model performance to the model results resulted in weighted averages that in general show additional improvement (Figure 4.5, 4.6, 4.7). The results

presented in Figure 4.7 indicate that in some specific areas the cost function for DIN, DIP or chlorophyll is relatively high; in those areas the threshold values that have been derived may be more uncertain.

However, the latter (there is uncertainty due to variations in space, over seasons and between years) also is the case for observational data.

As discussed above, the correspondence of model results with observations shows that, in general, the models are able to reproduce the spatial gradient going from coastal waters and river mouths to offshore areas. Maps of the horizontal distribution (Chapter 3) also support this.

There is one additional aspect that needs to be taken into account. The newly defined, ecologically relevant, COMP4 assessment areas have a clear benefit against the previous, nationally defined, COMP3 assessment areas. For the modelling exercise there is the disadvantage that a direct comparison with previous threshold values is not possible. In addition, the new assessment areas can include spatial patterns or gradients that are not reflected when applying one threshold value for the whole area, this also requires more consideration of the spatial coverage by monitoring.

8.2 Plausibility of pre-eutrophic scenarios

One of the key problems of the present study is the fact that the input information for the historic scenarios is not precisely described. This starts with the fact that there is no fixed date which the historic reference condition should be settled at. Therefore the expert-group decided to define a time interpreted as „pre-eutrophic“, which is loosely related to the invention of the Haber-Bosch process at the end of the 19th century, as described in Annex 2. Knowing that, for example with the massive import of guano fertilizer from Latin America, the local nutrient budget has been affected previous to the Harbor-Bosch invention. The pre-eutrophic scenario can be considered to roughly describe the state at the end of the 19th century, with some level of anthropogenic disturbance (i.e. different from pristine conditions) but without the large and widespread use of nitrogen fertilizer, intensification of agriculture and emissions to surface waters (Galloway et al. 2003, Erisman et al. 2011, Sutton et al. 2011, Bouwman et al. 2013)

The second problem is the lack of a conceptual setup for the modelling for these pre-eutrophic conditions. There is a systemic problem included, as there are two types of catchment models involved. The MONERIS model incorporates very detailed process description (Hirt et al., 2014) which lead to a high data requirement for this model, which can only be solved by a restriction of modelling a few rivers. The Hype model has a more global approach and the European application of it, called E-HYPE, is therefore able to cover the catchment of the entire OSPAR Regional Convention Area. This wider area approach results in lower precision in representing individual rivers, like the Ems, where E-HYPE severely underestimated the P-load for the years 2000-2010, as presented at the ICG-EMO and ICG-EUT workshop in 2019 in Hamburg. Nutrient validation reports for E-HYPE are not yet available.

Because of this drawback in the E-HYPE model, especially within the P load estimates, a second hybrid scenario with corrections in the P load was agreed on. Here the P-loads were mainly corrected for the rivers where data from MONERIS were available (Germany, Netherlands). Only Denmark added local information on historic loads to the historic scenario (see detailed description in Annex 2).

While the MONERIS simulation was based on real data from Prussian archives around 1880, the historic simulation of the E-HYPE model was based on a well-thought-off conceptual setup with increasing reduction by including more and more historic assumptions, like forest coverage, historic agricultural activities of population density. The outcome was first applied in the JMP EUNOSAT project. However, with the available scientific expertise in the pre-eutrophic expert group it was not possible to compare and evaluate the assumptions for the historic scenarios in E-HYPE and MONERIS.

This implies that there are inconsistencies in the P-load representation from these countries that provided alternative historic P-load estimates, for example in relation to the N-load estimates. While the loads introduced by DE and NL based on the MONERIS model are based on the same historic assumptions and process formulation, the ones introduced by DK might be scientifically correct but do not match the setup of either the E-HYPE or the MONERIS model. However, given the circumstances described above it is the best estimate available for the ICG-EMO model exercise and the resulting threshold estimates. However, for the ICG-EMO members it is simply not possible to select one of the scenarios as the most reliable one and to justify this choice on a scientific basis.

Of course, ICG-EMO members would prefer to work with a consistent representation of pre-eutrophic load reduction. Given these conceptual and practical problems we had to work with, it seems logical that there is no way to setup any kind of error-propagation concept that would finally help to identify uncertainties in the threshold estimate related to uncertainties in the pre-eutrophic assumptions and the related load reductions.

8.3 Model response to nutrient reduction

Second, the question is if the model response to lower nutrient loads in the pre-eutrophic scenarios is plausible. It is to be expected that the lower nutrient loads in the pre-eutrophic scenario will result in reduced concentrations of nutrients and chlorophyll in areas with freshwater influence, while no effects are to be expected in offshore areas. It can also be expected that the response in chlorophyll (lower concentrations in pre-eutrophic scenario) will be more variable than the response in nutrient concentrations. This is due to the fact that in some areas, P-limitation or N-limitation determines phytoplankton biomass, while in some other areas other factors, like light limitation, may be dominant. One can clearly observe the differences in the response in relation to the two historic scenarios for the Delates model (Fig. 3.2), the Miro model (Fig. 3.2) and the GP model from Oldenburg (Fig. 3.4). The timeseries plots in Fig 3.7 and 3.8 further reveal a look into the resulting nutrient dynamic related to the two scenarios.

In addition, the results in Figure 7.7 reveal that the pre-eutrophic scenario shows both lower nutrient concentrations and lower chlorophyll concentrations. Between models, the responsiveness of chlorophyll to lower nutrient concentrations under pre-eutrophic conditions, differs as demonstrated in Chapter 3.1.

Memory Effect

In the expert group on pre-eutrophic condition it was discussed that phosphorus and to some degree nitrogen is accumulated in the sediments which will diminish the effect of reduction scenarios, since under lower nutrient condition in the water column this sediment reservoir is released. This is true for

reduction scenarios which look for the effect of nutrient reduction starting from the present environmental condition, like the ones based on measures under the WFD regulation. However, with the simulation that represents a reference state of the marine environment, and therefore by definition pre-eutrophic condition, there are no previous accumulations of organic matter and therefore possible sources of remineralisation, often referred to as memory effect, that need to be taken into account.

In addition, we used a period of 4 years as model spin-up, giving the sediments enough time to adjust to the changed nutrient inputs.

Baltic boundary condition

For the current state (CS) simulation the models used their own available Baltic boundary condition, if their model domain reaches so far. Much effort was put in the reconstruction of the historic estimates for the Baltic boundary condition, see the detailed description in the user guide (Annex 1, page 16) to be used for both historic scenarios HS1 and HS2. The Fig. 8.4 illustrates relative contribution to TN in relation to the Baltic Boundary condition. The figure was produced in the context of the publication Lenhart & Große (2018), based on TBNT method where all hydrodynamical as well as biogeochemical processes are taken into account.

It clearly shows that mainly the Kattegat, and to a lesser degree the coastal regions of the Skagerrak, are influenced by the Baltic boundary condition, whereas the Danish North Sea coast is mainly influenced by the nutrient loads that are transported by the Continental Coastal Current, as displayed in the difference plot for between HS1 and HS2 for DIP and Chl-a (Fig. 5.5).

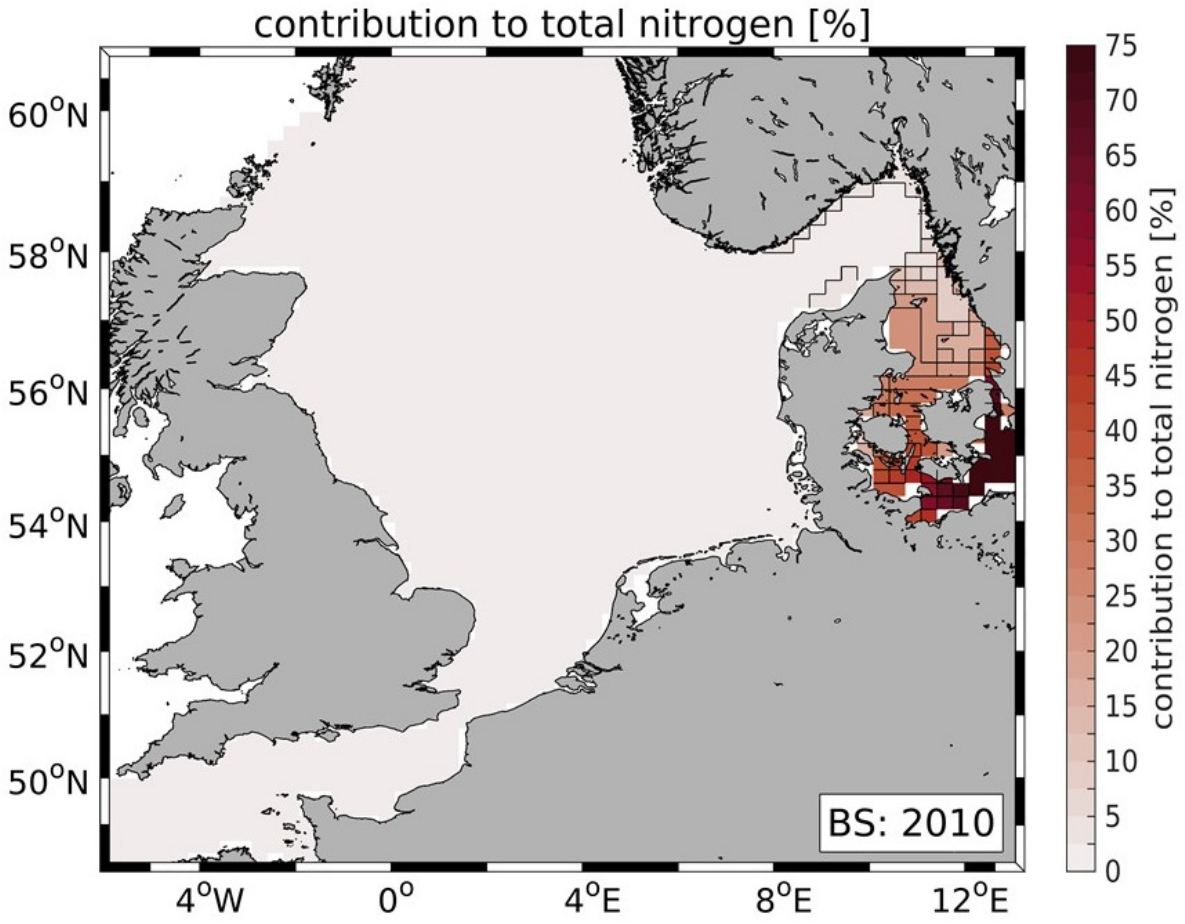


Fig. 8.4: Map of average relative contributions to TN related to the Baltic boundary condition based on a simulation with the ECOHAM model for 2010.

In addition, the changes of all assumed model variables along the Baltic Sea boundaries were assumed consistent to Baltic wide model simulations, conducted by Leibniz Institute for Baltic Sea Research Warnemünde and published i.e. at Meier et al. (2019).

8.4 Weighted ensemble application and threshold estimates

The approach to by Almroth and Skogen (2010) provides an opportunity to link the observation data provided as used in the COMPEAT assessment tool with the model results. In chapter 4 it could be clearly demonstrated that the weighted ensemble mean value is shifted towards the observation mean value in comparison to the simple ensemble mean of the models. The effect of the increase in quality could be demonstrated by applying the cost function again on both the ensemble mean and the weighted ensemble mean (Fig. 4.6 and 4.7).

Therefore, the application of the weighted ensemble approach also provides the solution to overcome the bias in the *in-situ* observation. Since the models provide continuous information in space and time, the model average for an assessment area takes into account the whole area in a consistent way (see chapter 7.2) whereas observations are usually dominated by coastal stations. In the ICG-EMO report for HASEC presented in May 2021 there was a mismatch between thresholds and observations caused by this bias in the observations. With the weighted ensemble approach we are able to overcome this

by including a weighting factor that relates the model results to the observational values and thus reduces the bias between modelled values and observations.

However, it was also shown that the result of this newly derived weighted ensemble mean is strongly dependent on the underlaying data coverage. This effect could be perfectly demonstrated by adding satellite Chl-a data, which increased the data coverage considerably in comparison to the mostly poor coverage of *in situ* Chl-a data and improved the weighted mean.

8.5 Concluding remarks

There is no method to define uncertainty for the threshold estimates derived from this model exercise and the weighted ensemble approach. There is no yardstick that would enable the ICG-EMO modelling even to compare their derived threshold estimates with previous COMP assessments. This has to do with the fact that we present the final product of a number of changes involved in this process to derive new threshold values for newly defined ecological relevant assessment areas. This starts already with the introduction of new assessment areas, that have the disadvantage that a direct comparison with previous threshold values is not possible.

Therefore, we are only able to address the uncertainty in the model exercise on different levels, like the representation of the models in comparison to observations, which was done in a considerable depth. The result is model performance is satisfactory for most of the assessment areas (Figure 4.7) and in general, better for nutrients than for chlorophyll due to the lower natural variability in winter nutrient concentrations (Figure 8.1-8.3).

The pre-eutrophic scenarios used in this study provide the best estimate available for the ICG-EMO model exercise and the resulting threshold estimates, given the existing uncertainties in the assumptions for the historic scenarios but also the constraints involved. However, the ICG-EMO group is not able to give preference to either of the two historic scenarios on the basis of their scientific expertise.

Concerning the responsiveness of the models in relation to the lower nutrient loads in the two historic scenarios one can conclude that these fall within the range of previous studies. Certain characteristics are reproduced, like the difference in the response of coastal vs. offshore regions and the lower, and more variable response of the phytoplankton.

Finally, the weighted ensemble approach has proven to improve the quality of the ensemble mean in comparison to observations, but it also links the quality of the ensemble mean from the models to the observations of each single parameter in each COMP4 assessment area. The use of Chl-a information from satellite data has proven to improve the quality of the weighted ensemble results considerably. For the other parameters, increasing the availability of in situ data is a prerequisite for improving model results.

9 Discussion

9.1 Definition of pre-eutrophic conditions

Determining pre-eutrophic conditions in the marine environment is a challenge for scientists and water managers. The need for such a definition is clear as it describes the undisturbed conditions ("reference" conditions) where marine ecosystems are hypothetically in a good status. However, the challenge is difficult to meet as very few data on water quality parameters exist before industrial times. This advocates for the use of marine ecosystem models and the development of historical scenarios, tailored on purpose to reproduce pre-industrial inputs to the sea, if necessary, by making assumptions on human activities in the watersheds. Two such scenarios have been prepared by the OSPAR "pre-eutrophication group", historical scenario 1 and historical scenario 2. These scenarios have generated pre-eutrophic inputs of nutrients to the North Sea (river loads, atmospheric deposition and Baltic outflow), which were in turn used in the marine ecosystem models. An evaluation of the assumptions underlying both historic scenarios is still necessary to assess which of the scenarios gives the best approximation of the 'pre-eutrophic' conditions, in a way coherent with other areas (see chapter 8.2).

Taking into account the importance of setting the right nutrient level of the river load for the reference simulation of the marine ecosystem models and the resulting threshold estimates, a sound conceptual setup for the catchment models is needed. This need to start with commonly accepted assumptions for a clearly defined historic time period. Afterwards a model simulation by one or, even better, a number of catchment models using these assumptions are put in practice and the results are discussed in an OSPAR or an EU frame.

9.2 ICG-EMO modelling approach

The ICG-EMO partners have applied these two historic scenarios into their respective model setup to establish coherent estimates of reference conditions in OSPAR regions II-IV. These reference conditions form the basis of the threshold values that were determined for each COMP4 assessment area. The definition of the pre-eutrophic scenarios, in collaboration with TG-COMP, took several months longer than expected.

Therefore, the model work started later than foreseen, with the additional problem that some model groups had to deal with long computation times or limited computational capacity caused by the delay, so that one of eight models could unfortunately not provide the historic reduction scenario. But also the logistics of aggregating the model output and to produce a consistent basis for the interpretation of the model results has proven to be time consuming.

The results that are currently available and that we presented in this report suggest that, overall, there is a reasonably good similarity between the models, both in reproducing a realistic spatial distribution of nutrients and chlorophyll in the OSPAR area and in reproducing comparable responses to nutrient reduction. The ensemble of model simulations provides a good base for coherently estimating reference levels for nutrients and chlorophyll with information on uncertainty ranges. It is worth to mention that there are more existing models that could be applied to these OSPAR regions. The base to establish reference levels could be strengthened if more national modelling institutes would partake

in the ICG-EMO exercise: currently only 8 of the 15 OSPAR Contracting Parties are represented in this ICG-EMO study (plus EU representation).

The results from the model applications and the calculated thresholds are presented in Chapters 3, 4 and 5. Uncertainties related to the approach in ICG-EMO were discussed in chapter 8. Even when two biogeochemical models are closely harmonized with regard to the most critical forcing conditions (like the nutrient inputs), and their performances are comparable, their predicted responses to changes in environmental conditions can differ substantially. This can be caused by differences in the spatial resolution, process description, hydrodynamical/transport models the biological models are coupled to, and other forcing factors, such as the open ocean boundary conditions (e.g., Skogen and Moll, 2005; Stegert et al., 2021). This motivates the use of multiple models, instead of relying on a single one, which is commonly recognized as the 'ensemble modelling' approach. In the context of eutrophication, an earlier ensemble modeling approach for understanding the response of the North Sea to reductions in riverine loadings was performed by Lenhart et al. (2010), where the authors focussed on comparing the model responses. For obtaining unified estimates based on an ensemble of model results, if each ensemble member (model) can be trusted to a similar degree, their contribution to the ensemble mean can be considered equal (e.g., as in Friedland et al., 2021). Alternatively, the relative contribution of each member to the mean can be based on their skill, i.e. how well do the models perform when compared to observations (Almroth & Skogen, 2010, Skogen et al., 2014; Meier et al. 2018). Here, we used the approach of Almroth & Skogen (2010), where weighing factors for ensemble members are determined separately for each assessment variable based on their skill in each COMP area.

By calculating the reference values as the weighted mean of an ensemble of model outputs in this report we have reduced the uncertainty in model estimates. For the reference values a comparison with observations is not possible, so we have no other option than to assume that the weighted approach can also be applied to the historic scenarios.

9.3 Definition of threshold levels

Discussion is still needed to address the uncertainty linked to the definition of the threshold value that switches an area from problem to non-problem status, as already mentioned in §7.3.2. The threshold value is now calculated as 50% above the reference value. The approach to add 50 % on the „natural background concentration" or reference concentration was developed to reflect natural variability and a slight disturbance. The deviation should be a justified area-specific % deviation from background not exceeding 50% (OSPAR 2013).

The problem with this formal approach is, that by applying 50 % on top of reference condition for historic nutrient and chlorophyll concentration alike a linearity between nutrients and chlorophyll is introduced which has no scientific basis. It would imply, that a 50 % increase in the nutrient load would result in the same increase in phytoplankton, and the same logic in the opposite direction by the reduction of the nutrient loads. This ignores the fact that, besides DIN and DIP, other limiting factors are determining phytoplankton biomass in particular under conditions of DIN and DIP enrichment.

The 50% is an arbitrary approach to determine a threshold that reflects the boundary between non-eutrophic and eutrophic conditions, due to the lack of a well-defined threshold for anthropogenic nutrient loads that represents the shift between a healthy ecosystem where eutrophication does not

occur and an ecosystem state where direct and indirect effects of eutrophication occur. As a proxy, chlorophyll concentrations are used as the main indicator for the assessment of eutrophication, as we lack quantitative dose-response relations between nutrient loads and other direct or indirect direct effects (phytoplankton indicator species, macrophytes and macroalgae, oxygen deficiency, etc.).

It might be useful to at least consider a more adaptive factor that would take into account the extent to which an area is influenced by human pressures (e.g. anthropogenic nutrient loads), such as application of an uplift percentage dependent on the natural variability of the parameter and a clearer definition of 'allowable level of (slight) disturbance'. That could result in different percentages (not necessarily 50%) depending on the characteristics of both the parameter and the assessment area. Application of different percentages for DIN, DIP and chlorophyll would also fit better with the non-linear relation between nutrient levels and chlorophyll concentrations.

The threshold levels that are provided in Tables 5.1 and 5.2, have been applied in the COMPEAT tool to carry out an assessment for the years 2006-2014 (Chapter 7). The results of the COMPEAT assessment give a first glance on the potential consequences of the new assessment areas and new threshold levels. A comparison between the results for the two scenarios shows that the impact of the low P loads from Dutch, German and Danish rivers in scenario HS2, leads to relatively small differences with the assessment outcome based on scenario HS1. Only for one or two assessment areas in the eastern part of the German Bight a slight shift in status for DIP and chlorophyll-a occur.

Comparison of the COMPEAT result in Chapter 7 with the COMP3 result (for the same years) shows that there are differences in the classification of problem areas. But it should be realized that COMP3 used different assessment areas including coastal waters that fall under the WFD and that are now excluded, and COMP3 assessment levels lacked coherence. The new threshold levels, based on the ensemble modelling approach, show a more plausible spatial distribution and a much improved coherence between assessment areas. In addition, the COMPEAT results are affected in some cases due to the fact that the spatial distribution of monitoring sites does not always match properly with the new assessment areas, i.e. averaging monitoring data over entire assessment areas can sometimes lead to biased estimates, as discussed in Chapter 7.2.

9.4 Results for specific areas

This modelling study reveals that in comparison to previous COMP3 thresholds, for which a transformation of the area was needed (see 6.2), in some assessment areas the new thresholds for winter DIN, winter DIP and Chl P90 have changed. Especially in coastal zones the new thresholds for winter DIN are often higher in this study than in COMP3. To some extent, this is caused by the fact that those coastal assessment areas are narrower and closer to river mouths, leading to higher nutrient concentrations. This may seem counterintuitive as the new thresholds for Chl P90 in coastal areas are mostly lower in this study compared to COMP3. To understand this apparent contradiction, we must point out the non-linearities in the natural system. Observations reveal that the response of chlorophyll to a reduction in winter DIN is not the same in the offshore as at the coast. In the offshore, N is limiting the phytoplankton bloom and, hence, a reduction in winter DIN will directly influence chlorophyll levels. Along the European shelf coast, P or light is often limiting the spring phytoplankton bloom and a reduction in winter DIN is not always followed by a reduction in Chl P90. The marine models reproduce these non-linearities. As a matter of fact, in coastal zones it is possible to get moderate Chl P90 values in the spring despite high winter DIN concentrations. The ensemble model approach considers this complexity. That explains why, in many coastal assessment areas, the new

thresholds for winter DIN (this study) are higher than the thresholds found in COMP3: a further decrease in the limiting nutrient (P) can cause an increase in the non-limiting nutrient (here N) as plankton growth is further curtailed and the non-limiting nutrient is used even less. Yet, it is paramount to point out that the new thresholds are only built on the link between the accumulation of biomass (Chl P90) and the winter DIN. They do not consider other adverse effects of high winter DIN, such as a shift in phytoplankton species towards undesirable species (as described by Radach et al. 1990 in the German Bight). This is because in our models (but also in OSPAR procedures) not enough attention has yet been given to the conditions of emergence of harmful phytoplankton species. Until now, most attention was allocated to phytoplankton total biomass (through Chl and Chl P90). We may thus expect that if we were to consider shifts in phytoplankton species assemblages, the thresholds for winter DIN in coastal areas would be lower than what was calculated in this study. This constitutes perhaps a point of improvement for future field and model studies. Note that some models already include colonial nuisance algae like *Phaeocystis*.

9.5 DIN versus DIP

Model results in Figure 7.7 suggests that in some areas a significant N reduction may not be followed by a proportional decrease in chlorophyll 90-percentile. Instead, the 90-percentile tends to offer a more linear response to large P reductions. This is consistent with previous findings that, under the current conditions, phosphorus is limiting the size of the spring chlorophyll bloom (i.e., chlorophyll 90-percentile) in some coastal areas (Billen et al., 2011; Desmit et al., 2015; 2018). This situation may remain as long as riverine loads have relatively high N levels, compared to P, resulting in high N:P ratios in coastal waters (Figure 7.5). Large accumulation of Chlorophyll (or phytoplankton biomass) may have serious consequences in coastal ecosystems, such as dead zones (Diaz and Rosenberg, 2008). In the Southern North Sea, *Phaeocystis globosa* has occasionally caused hypoxia in enclosed areas under adverse meteorological conditions (Peperzak and Poelman, 2008). Foam events are also known to be the result of an excessive accumulation of colonial *Phaeocystis globosa* in the spring (Jickells, 1998; Lancelot et al., 2014). However, massive chlorophyll blooms are not the only issue linked to marine eutrophication. If P limits chlorophyll levels in the coastal zones, excessive N relative to P or to Si may induce shifts in phytoplankton communities (Radach et al., 1990; Rousseau et al., 2000, Cloern, 2001; Rousseau et al., 2006) and may cause undesirable ecological disturbances, as pointed out in the definition of eutrophication by OSPAR, such as harmful algal blooms (Ménesguen, 1990; Fehling et al., 2004) or perturbations to higher trophic levels (e.g., Daro et al., 2006). Therefore, a dual reduction in N and P should be adopted in order to reach more acceptable levels of chlorophyll while avoiding problematic shifts in phytoplankton assemblages (Conley et al. 2009).

9.6 Other assessment parameters

In this exercise, we have focussed on deriving reference and threshold levels for DIN, DIP and chlorophyll. However, from the model results we can also derive estimates for light climate (extinction or Secchi depth) and oxygen concentrations, that are used as assessment parameters in at least some of the assessment areas. A decision is needed on how to derive threshold levels for these parameters, either by using reference values from the model results or by applying other methods.

9.7 Recommendations

For the application of the new thresholds in an assessment, we recommend a further analysis and evaluation. Considerations to be addressed are:

1. Improvement on the conceptual setup for pre-eutrophic condition, which includes clear definition and agreement on the assumptions applied for the historic scenario
2. Is 50% above reference concentrations a good definition of threshold levels, particularly in offshore waters?
3. What should the threshold definition be for oxygen and water transparency levels, which were higher in pre-eutrophic conditions?

9.8 Final remarks

With the work of ICG-EMO, a big step was set towards more coherent threshold levels for DIN, DIP and chlorophyll. The definition of reference levels is now based on commonly agreed scenarios for pre-eutrophic conditions. The application of the ensemble of models has resulted in a plausible description of concentrations of DIN, DIP and chlorophyll in the OSPAR maritime area under those pre-eutrophic conditions.

The reference values now form the basis for the definition of threshold levels. There are still issues around the way to derive those threshold levels that will need further discussion.

Disclaimer

HASEC 2022 noted that the ICG-EMO modelling work fulfilled the objective required by HASEC and delivered the proposed threshold estimates which form the basis for further consideration by ICG-EUT and HASEC. It should be noted that this publication reflects the views of its authors which might vary from those of the OSPAR Commission or its Contracting Parties.

9.9 Highlights

- OSPAR secretariat expressed their need for more consistent thresholds for eutrophication parameters (N, P, Chl and O_2) in the Greater North Sea.
- The new thresholds would be defined in new marine assessment areas that would be consistent with the geomorphology and ecology of the Greater North Sea.
- The new thresholds would be derived from historical (pre-industrial) conditions in the North Sea and its surrounding rivers.
- Previous freshwater model studies have been used to estimate the historical conditions in the rivers surrounding the North Sea.
- A modelling approach involving seven marine models was then used to derive the current and historical conditions in each marine assessment area.
- All marine models involved in this exercise received the same river loads under current and historical conditions.
- Marine *in-situ* observations in assessment areas helped determining the performance of the marine models under current conditions.
- Using these observations and model results, a weighted ensemble model approach allowed calculating the best estimate of eutrophication parameters in each area under historical conditions.
- These values were used to calculate the new thresholds for eutrophication parameters in each assessment area.
- The new thresholds display improved consistency at spatial level (natural marine continuum across assessment areas) and at biogeochemical level (biological response to drivers change).

10 Acknowledgements

ICG-EMO are grateful to Hjalte Parner of ICES, for his help in understanding the COMPEAT tool, his rapid response to questions and the updating of code in response to frequent additional requirements.

The German contribution from the ECOHAM and the GPM model were funded by the Umweltbundesamt (UBA, Project Nr. 3718252110) and partly by the Helmholtz-Zentrum Hereon. The model simulations were carried out on the mainframe MISTRAL of the German Climate Compute Center (DKRZ) in Hamburg, Germany. The hydrodynamical fields for the ECOHAM simulations and freshwater inputs have been kindly provided by Johannes Pätsch of Universität Hamburg.

The German contribution from the JRC model was funded by the Umweltbundesamt (UBA, Project Nr. 3720252020) and got supercomputing power provided by HLRN (North-German Supercomputing Alliance).

We would like to thank the Swedish Agency for Marine and Water Management for their support and financial contribution to this project.

We would also like to thank the Royal Belgian Institute of Natural Sciences for their support to this project.

We would also like to thank the IFREMER for their support to this project.

The CEFAS contribution was funded by the NERC/Defra project Campus - Combining autonomous observations and models for understanding shelf seas.

We thank the NIOZ Institute (the Netherlands) for supporting S.M. van Leeuwen in this work.

Parts of the Deltares results have been achieved using the DECI resource Cartesius based in The Netherlands at SURFsara with support from PRACE. The support of Maxime Mogé from SURFsara, The Netherlands to the technical work is gratefully acknowledged.

Deltares was supported by the European Union, European Maritime and Fisheries Fund

UK riverine data was processed from raw data provided by the Environment Agency, the Scottish Environment Protection Agency, the Rivers Agency (Northern Ireland) and the National River Flow Archive. French water quality data was provided by Agence de l'eau Loire-Bretagne, Agence de l'eau Seine-Normandie and IFREMER, while flow data was provided by Banque Hydro. German and Dutch riverine data was provided by the University of Hamburg (Johannes Pätsch, Hermann Lenhart, Dutch data from Rijkswaterstaat (waterinfo: https://waterinfo.rws.nl). Irish flow data was provided by Hydrodata and the Environment Protection Agency (Hydronet), while water quality data was obtained from OSPAR RID reports. Norwegian flow data was supplied by NVE's Anne Fleig (afl@nve.no), water quality data was obtained from NIVA (www.niva.no) and Tore Høgåsen (tore.hogaasen@niva.no). Danish water quality data was provided by the National Environmental Research Institute (NERI). Water quality data for Baltic rivers was provided by the University of Stockholm and the Baltic Nest (www.balticnest.org/bed). Spanish data was provided by Dr. Luz Garcia (IEO, Spain). Portuguese data was provided by Dr. Amelia Araujo (CEFAS, UK). Dr. S. M. van Leeuwen, NIOZ, Lansdiep 4, 't Horntje, Texel, the Netherlands, pers. comm. Please note that the riverine data for the UK has not been

formally checked or authorised by the UK and that any conclusions drawn from it need to be treated with caution.

11 References

Almroth, E., Skogen, M.D. A North Sea and Baltic Sea Model Ensemble Eutrophication Assessment. AMBIO 39, 59–69 (2010). https://doi.org/10.1007/s13280-009-0006-7

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. Aquat. Microb. Ecol. **13**: 3–17.

Billen, G., M. Silvestre, B. Grizzetti, and others. (2011) Nitrogen flows from European watersheds to coastal marine waters, p. 271–297. In The European nitrogen assessment : sources, effects, and policy perspectives. Sutton, M. A. (ed.), Cambridge University Press. ISBN 9781107006126.

Blauw, A., et al. (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 1 Report. 86 pp.

Bouwman, L., K. K. Goldewijk, K. W. Van Der Hoek, A. H. W. Beusen, D. P. Van Vuuren, J. Willems, M. C. Rufino, and E. Stehfest. 2013. Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900–2050 period. Proceedings of the National Academy of Sciences **110**:20882-20887.

Capell, R., A. Bartosova, K. Tonderski, B. Arheimer, S.M. Pedersen, and A. Zilans. 2021. From local measures to reginal impacts: Modelling changes in nutrient loads to teh Baltic Sea. J. of Hydrology: Regional Studies. **Vol 36**, https://doi.org/10.1016/j.ejrh.2021.100867

Cloern, J. E. 2001. Our evolving conceptual model of the coastal eutrophication problem. Mar. Ecol. Prog. Ser. **210**: 223–253. doi:10.3354/meps210223

Conley, D., H. Paerl, R. Howarth, D. Boesch, S. Seitzinger, K. Havens, C. Lancelot, and G. Likens. 2009. Controlling eutrophication: nitrogen and phosphorus. Science (80-.). **323 (5917)**: 1014–1015.

Daro, M.-H., E. Breton, E. Antajan, S. Gasparini, and V. Rousseau. 2006. Do Phaeocystis colony blooms affect zooplankton in the Belgian Coastal Zone?, p. 61–72. *In* V. Rousseau, C. Lancelot, and D. Cox [eds.], Current status of eutrophication in the Belgian coastal zone. Presses Universitaires de Bruxelles.

Desmit, X., K. Ruddick, and G. Lacroix. (2015) Salinity predicts the distribution of chlorophyll a spring peak in the southern North Sea continental waters. J. Sea Res. 103: 59–74. doi:10.1016/j.seares.2015.02.007

Desmit, X., V. Thieu, G. Billen, et al. (2018) Reducing marine eutrophication may require a paradigmatic change. Sci. Total Environ. **635**: 1444–1466. doi:10.1016/j.scitotenv.2018.04.181

Diaz, R. J., and R. Rosenberg. 2008. Spreading dead zones and consequences for marine ecosystems. Science **321**: 926–9. doi:10.1126/science.1156401

Dulière, V., N. Gypens, C. Lancelot, P. Luyten, and G. Lacroix. (2019) Origin of nitrogen in the English Channel and Southern Bight of the North Sea ecosystems. Hydrobiologia. doi:10.1007/s10750-017-3419-5

Edman, M., Omstedt, A. (2013) Modeling the dissolved CO2 system in the redox environment of the

Baltic Sea. Limnology and Oceanography, 58(1), 74-92, doi:10.4319/lo.2013.58.1.0074.

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

Erisman, J. W., H. Van Grinsven, B. Grizzetti, F. Bouraoui, D. Powlson, M. A. Sutton, A. Bleeker, and S. Reis. 2011. The European nitrogen problem in a global perspective. Pages 9-31 *in* M. A. Sutton, C. M. Howard, J. Erisman W., G. Billen, A. Bleeker, P. Grennfelt, H. Van Grinsven, and B. Grizzetti, editors. The European Nitrogen Assessment. Cambridge University Press, Cambridge.

Fehling, J., K. Davidson, C. J. Bolch, and S. S. Bates. 2004. Growth and domoic acid production by Pseudo-nitzschia seriata (Bacillariophyceae) under phosphate and silicate limitation. J. Phycol. **40**: 674–683.

Friedland, R., Macias, D., Coassarini, G., Daewel, U., Estournel, C., Garcia-Gorriz, E., et al. (2021) Effects of nutrient management scenarios on the marine eutrophication indicators: a Pan-European, multi-model assessment in support of the Marine Strategy Framework Directive. Front. Mar. Sci. 8:596126. doi: 10.3389/fmars.2021.596126.

Galloway, J. N., J. D. Aber, J. W. Erisman, S. P. Seitzinger, R. W. Howarth, E. B. Cowling, and B. J. Cosby. 2003. The Nitrogen Cascade. BioScience **53**:341-356.

García-García, L.M., Sivyer, D., Devlin, M., Painting, S., Collingridge, K. and Van Der Molen, J. (2019) Optimizing monitoring programs: a case study based on the OSPAR eutrophication assessment for UK waters. *Frontiers in Marine Science*, *5*, p.503.

HASEC (Hazardous Substances and Eutrophication Committee) HOD (Heads of Delegation), 21/5/1. Revision of the Common Procedure for the identification of eutrophication, presented by the Co-Convenors of ICG-EUT to the OSPAR Convention for the Protection of the Marine Environment of the North-East Atlantic.

Hirt, U., Mahnkopf, J., Gadegast, M., Czudowski, L., Mischke, U., Heidecke, C., Schernewski, G., Venohr, M. (2014). Reference conditions for rivers of the Ger- man Baltic Sea catchment: reconstructing nutrient regimes using the model MONERIS. Reg. Environ. Chang. 14, 1123–1138. https://doi.org/10.1007/s10113- 013-0559-7.

Jickells, T. D. 1998. Nutrient Biogeochemistry of the Coastal Zone. Science **281**: 217–21. doi:10.1126/SCIENCE.281.5374.217

Kerimoglu, O., Große, F., Kreus, M., Lenhart, H.-J., and van Beusekom, J. E. (2018) A model-based projection of historical state of a coastal ecosystem: relevance of phytoplankton stoichiometry. Sci. Total Environ. 639, 1311–1323. doi: 10.1016/j.scitotenv.2018.05.215.

Kerimoglu, O., Voynova, Y. G., Chegini, F., Brix, H., Callies, U., Hofmeister, R., Klingbeil, K., Schrum, C., and van Beusekom, J. E. E. (2020) Interactive impacts of meteorological and hydrological conditions on the physical and biogeochemical structure of a coastal system, Biogeosciences, 17, 5097–5127, doi:10.5194/bg-17-5097-2020.

Lancelot, C., Gypens, N., Billen, G., Garnier, J., Roubeix, V., (2007) Testing an integrated river-ocean mathematical tool for linking marine eutrophication to land use: the Phaeocystis-dominated Belgian coastal zone (Southern North Sea) over the past 50 years. J. Mar. Syst. 64, 216–228.

Lancelot, C., P. Passy, and N. Gypens. 2014. Model assessment of present-day Phaeocystis colony blooms in the Southern Bight of the North Sea (SBNS) by comparison with a reconstructed pristine situation. Harmful Algae **37**: 172–182. doi:10.1016/j.hal.2014.05.017

van Leeuwen, S.M., P. Tett, D.K. Mills, J. van der Molen, **2015**, *Stratified areas in the North Sea: long-term variability and biological and policy implications*, Journal of Geophysical Research - Oceans, Vol. 120 (7), pp 4670-4686, doi: 10.1002/2014JC010485

Lenhart, H.-J., Mills, D., Baretta-Bekker, H., van Leeuwen, S., van der Molen, J., Baretta, J., et al. (2010) Predicting the consequences of nutrient reduction on the eutrophication status of the North Sea. J. Mar. Syst. 81, 148–170. doi: 10.1016/j.jmarsys.2009.12.014

Loebl, M., Colijn, F., Van Beusekom, J.E.E., Barreta-Bekker, J.G., Lancelot, C., Philippart, C.J.M., Rousseau, V.,Wiltshire, K.H., (2009) Recent patterns in potential phytoplankton limitation along the Northwest European continental coast. J. Sea Res. 61, 34–43.

Meier, H. E. M., M. K. Edman, K. J. Eilola, M. Placke, T. Neumann, H. C. Andersson, S.-E. Brunnabend, C. Dieterich, C. Frauen, R. Friedland, M. Gröger, B. G. Gustafsson, E. Gustafsson, A. Isaev, M. Kniebusch, I. Kuznetsov, B. Müller-Karulis, A. Omstedt, V. Ryabchenko, S. Saraiva and O. P. Savchuk (2018) Assessment of eutrophication abatement scenarios for the Baltic Sea by multi-model ensemble simulations. Front. Mar. Sci. 5: 440, doi: 10.3389/fmars.2018.00440

Meier HEM, Eilola K, Almroth-Rosell E, et al (2019) Disentangling the impact of nutrient load and climate changes on Baltic Sea hypoxia and eutrophication since 1850. Clim Dyn 53:1145–1166. https://doi.org/10.1007/s00382-018-4296-y

Ménesguen, A. 1990. Eutrophication along the French coasts. In: Bart, H., Fegan, L. (Eds.), Eutrophication-related Phenomena in the Adriatic Sea and in Other Mediterranean Coastal Zones. Proc. Conf. 28–30 May 1990, Rome (Italie), C.E.C. Water Pollution Research Report 16, pp.

OSPAR (2013): Common Procedure for the Identification of the Eutrophication Status of the OSPAR Maritime Area. OSPAR Agreement, reference no. 2013-8, 66 pp.

OSPAR (2017): Third Integrated Report on the Eutrophication Status of the OSPAR Maritime Area. OSPAR, London, 694/2017, 166pp.

Peperzak, L., and M. Poelman. 2008. Mass mussel mortality in The Netherlands after a bloom of Phaeocystis globosa (prymnesiophyceae). J. Sea Res. **60**: 220–222. doi:10.1016/J.SEARES.2008.06.001

van Raaphorst, W., and V. de Jonge. 2004. Reconstruction of the total N and P inputs from the IJsselmeer into the western Wadden Sea between 1935–1998. J. Sea Res. **51**: 109–131.

Radach, G., J. Berg, and E. Hagmeier. 1990. Long-term changes of the annual cycles of meteorological, hydrographic, nutrient and phytoplankton time series at Helgoland and at LV ELBE 1 in the German Bight. Cont. Shelf Res. **10**: 305–328.

Rousseau, V., S. Becquevort, J.-Y. Parent, S. Gasparini, M.-H. Daro, M. Tackx, and C. Lancelot. 2000. Trophic efficiency of the planktonic food web in a coastal ecosystem dominated by Phaeocystis colonies. J. Sea Res. **43**: 357–372.

Rousseau, V., Y. Park, K. Ruddick, W. Vyverman, J.-Y. Parent, and C. Lancelot. 2006. Phytoplankton blooms in response to nutrient enrichment, p. 45–59. *In* V. Rousseau, C. Lancelot, and D. Cox [eds.], Current status of eutrophication in the Belgian coastal zone. Presses Universitaires de Bruxelles.

Schernewski, G., Neumann, T. (2005) The trophic state of the Baltic Sea a century ago: a model simulation study. Journal of Marine Systems 53 (2005) 109–124.

Schernewski, G., Friedland, R., Carstens, M., Hirt, U., Leujak, W., Nausch, G., Neumann, T., Petenati, T., Sagert, S., Wasmund, N. & von Weber, M. (2015) Implementation of European marine policy: New water quality targets for German Baltic waters. Marine Policy 51 (2015), 305-321. DOI:10.1016/j.marpol.2014.09.002.

Skogen, M.D., Moll, A. (2005). Interannual variability of the North Sea primary production: comparison from two model studies. Journal of Marine Systems 57, 289-300.

Skogen, M.D., Eilola, Hansen, J.L.S., Markus Meier, H.E.M., Molchanov, M.S., Ryabchenko, V.A. (2014) Eutrophication status of the North Sea, Skagerrak, Kattegat and the Baltic Sea in present and future climates: A model study. Journal of Marine Systems 132 (2014) 174–184

Skogen, M.D., Ji, R., Akimova, A., Daewel, U., Hansen, C., Hjøllo, S.S., van Leeuwen, S.M., Maar, M., Macias, D., Mousing, E.A., Almroth-Rosell, E., Sailley, S.F., Spence, M.A., Troost, T., van de Wolfshaar, K., 2021, *Disclosing the truth: are models better than observations?*, Marine Ecology Progress Series, DOI: 10.3354/meps13574

Stegert C., Lenhart H.-J., Blauw A., Friedland R., Leujak W. and Kerimoglu O. (2021) Evaluating Uncertainties in Reconstructing the Pre-eutrophic State of the North Sea. Front. Mar. Sci. 8:637483. doi: 10.3389/fmars.2021.637483.

Stow, C.A., J. Jolliff, D.J. McGillicuddy, S.C. Doney, J.I. Allen, M.A.M. Friedrichs, K.A. Rose and P. Wallhead (2009). Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems 76: 4-15

Sutton, M. A., C. M. Howard, J. W. Erisman, W. J. Bealey, G. Billen, A. Bleeker, A. F. Bouwman, P. Grennfelt, H. Van Grinsven, and B. Grizzetti. 2011. The challenge to integrate nitrogen science and policies: the European Nitrogen Assessment. Pages 82-96 *in* M. A. Sutton, C. M. Howard, J. W. Erisman, G. Billen, A. Bleeker, P. Grennfelt, H. Van Grinsven, and B. Grizzetti, editors. The European Nitrogen Assessment. Cambridge University Press, Cambridge.

Topcu, D., Behrend, H., Brockmann, U., and Claussen, U. (2011). Natural background concentrations of nutrients in the German Bight area (North Sea). Environ. Monit. Assess. 174, 361–388. doi: 10.1007/s10661-010-1463-y

WFD CIS Guidance Document No. 5 (2003). Transitional and Coastal Waters - Typology, Reference Conditions and Classification Systems. Published by the Directorate General Environment of the European Commission, Brussels, ISBN 92-894-5125-4, ISSN 1725-1087

ICG-EMO Report - Annex 1: Technical background information on modelling work

Model description

In this chapter a description of the models that take part in the model comparison study is provided. Please also have a look into Tab.. 1 with an overview of the model characteristic

MIRO&CO_3D (RBINS, Belgium)

MIRO&CO results from the coupling of the 3D hydrodynamic COHERENS model (Luyten, 2011) with the biogeochemical MIRO model (Lancelot et al., 2005). COHERENS is a three-dimensional numerical model, designed for application in coastal and shelf seas, estuaries, lakes, reservoirs. MIRO is a biogeochemical model that has been designed for *Phaeocystis*-dominated ecosystems. It describes the dynamics of phytoplankton (three functional groups), zooplankton (two functional groups), heterotrophic bacteria, organic matter degradation (dissolved and particulate) and nutrient cycles (N, P, Si) in the water column and the sediment. The current setup has been obtained by coupling MIRO with COHERENS v2 (MIRO&CO v2): details and validation are shown in Dulière et al. (2019).

The description above is extracted from Desmit et al., (2018). An extension of the hydrodynamical model is currently under work and the biogeochemical module might be improved in the coming future.

ECO_MARS3D (IFREMER, France)

The ECO-MARS3D ecological model is based on the MARS3D hydrodynamical code (Lazure & Dumas, 2008), a three dimensional model based on Navier-Stokes equations under the classic Boussinesq and hydrostatic assumptions within a sigma framework. The originality of this model is on the coupling between barotropic and baroclinic modes especially designed for the alternate direction implicit method (ADI). The time-step is adaptative and the model is forced by a barotropic sea-level oscillation (at the oceanic boundaries) and by atmospheric conditions (throughout the domain). It provides realistic descriptions of coastal hydrodynamics for research and operational interests. The fully coupled biogeochemical module ECO-MARS3D is a NPZD model type (Nutrient–Phytoplankton–Zooplankton–Detritus), that aims to simulate the fluxes of limiting elements such as nitrogen (N), phosphorus (P) and silicon (Si). The phytoplankton compartment is described by the following variables ‘diatoms’, ‘dinoflagellates’, ‘nanopicoplankton’ and the haptophyte ’Phaeocystis globosa’. The grazers are split into two groups ‘micro-zooplankton’ and ‘meso-zooplancton’ and the detritus coming from phytoplankton, zooplankton senescence and excretion are mineralized and contribute to the nutrient renewal. Moreover, variables ‘dissolved oxygen’ as well as ‘oxygen saturation’ are calculated taking into account air-water exchanges, primary production and respiration processes in the water column. The current application to the French Atlantic shelf is based on a regular grid with 4 × 4 km meshes and 30 sigma layers, which covers the Bay of Biscay, the English Channel and the southern part of the North Sea, up to the Rhine estuary. Detailed description and validation is provided in Ménesguen et al. (2018, 2019).

ECOHAM (University Hamburg & HZG Geesthacht, Germany)

ECOHAM (ECOlogical Model-HAMburg) is defined on 31 z-levels and a 1/5°x1/3°km grid, that covers almost the entire Northwest Continental Shelf. The biogeochemical model consists of two phytoplankton (diatoms and flagellates), two zooplankton groups (micro- and mesozooplankton) and a bacteria group with fixed stoichiometry, O2, and C, N, P bound to two detritus size classes, dissolved organic material, dissolved inorganic material and 2D (plate) sediment pools. At the open ocean boundaries, biogeochemical model variables are nudged to the values extracted from the World Ocean Atlas for all scenarios. Transport of modelled biogeochemical variables are calculated based on the diffusion and advection fields simulated by the hydrodynamical model HAMSOM (HAMburg Shelf Ocean Model). Details about the model setup and biogeochemical model can be found in Große et al., 2017 and references therein.

GPM (University Oldenburg & HZG Geesthacht, Germany)

GPM (Generalized Plankton Model) is defined on 20 σ-layers and a 1.5-4.5km curvlinear grid, covering only the southern North Sea. In the present implementation, the 'geochemistry' portion (O2 and C, N, P bound to dissolved inorganic and organic material, detritus and sediment pools) of the model is as in ECOHAM, but for the description of plankton growth and interactions, the variable chlorophyll content and C:N:P ratios of phytoplankton were taken into account. At the open ocean boundaries, geochemical variables are clamped to the ECOHAM results obtained for respective scenarios, whereas for plankton variables, zero-gradient conditions were assumed. The biogeochemical model is on-line coupled to the Generalized Estuarine Transport Model (GETM) as the hydrodynamical driver. A detailed description of the model setup and the biogeochemical model can be found in Kerimoglu et al. 2020 and references therein.

ECOHAM and GPM have been used recently in a similar nutrient reduction scenario study, where one of the historic scenarios considered was conceptually identical to the HS1 in the current study, although there have been technical differences with respect to the construction of river loadings (Stegert et al., in review).

Delft3D_GEM (Deltares, Netherlands)

The model used by Deltares for the model comparison is a combination of the Generic Ecological Model (GEM) for the water quality and ecological processes (Blauw et al., 2009) and a newly developed hydrodynamic model for the greater North Sea. The model uses the Delft Flexible Mesh (DFM) simulation software both for the hydrodynamics and water quality and ecological processes. The hydrodynamical part of the model is originally developed for flood forecasting and transport simulation purposes (Zijl et al., 2021). In combination with the GEM model it is used to study the effects of several anthropogenic impacts to the North Sea, such as eutrophication, climate change, aquaculture and wind farms. The GEM model simulates the nutrient cycles of nitrogen, phosphorus and silicate and the dynamics of phytoplankton and oxygen. Additionally, grazing by benthic filterfeeders is included based on Dynamic Energy Budget (DEB) modelling (Troost et al., 2010; 2018). Four groups of phytoplankton are modelled (diatoms, flagellates, dinoflagellates and *Phaeocystis*) and 2 groups of benthic filterfeeders (*Mytilus edulis* and *Ensis).*

JRC_ERSEM (JRC Ispra, EU)

The JRC-NWES is largely described by Friedland et al. (2020), including a variety of validation results. The coupled model systems consists of GETM and PML-ERSEM, which was coupled via FABM. For the present study, the model setup was adapted to the river inputs provided within ICG-EMO and enlarged by a more sophisticated sediment model, the atmospheric nitrogen deposition and an improved attenuation calculation method. ERSEM includes a sophisticated phytoplankton growth model and allows to distinguish the single nutrient cycles of N, P, C and Si, as all components of the pelagic and benthic model the stored nutrient amounts are considered separated.

GETM_ERSEM_BFM (Cefas, United Kingdom)

GETM-ERSEM-BFM, GETM (General Estuarine Transport Model) is a public domain, three-dimensional Finite Difference hydrodynamical model (www.getm.eu). It solves the 3D partial differential equations for conservation of mass, momentum, salt and heat, and was designed to handle drying and flooding (e.g. tidal flats). The ERSEM-BFM (European Regional Seas Ecosystem Model - Biogeochemical Flux Model) version is a development of the model ERSEM III (see Baretta et al., 1995; Ruardij and van Raaphorst, 1995; Vichi et al., 2007; van der Molen et al., 2013; www.nioz.nl/en/about/cos/ecosystem-modelling), and describes the dynamics of the biogeochemical fluxes within the pelagic and benthic environment. The ERSEM-BFM model simulates the cycles of carbon, nitrogen, phosphorus, silicate and oxygen and allows for variable internal nutrient ratios inside organisms, based on external availability and physiological status. The model applies a functional group approach and contains 6 phytoplankton groups, 4 zooplankton groups and 5 benthic groups, the latter comprising 4 macrofauna and 1 meiofauna group. Pelagic and benthic aerobic and anaerobic bacteria are also included. SPM concentrations are calculated as proportional to the local wave-induced bed-shear stress, varying linearly with depth, and with an exponential relaxation mechanism that represents delayed settling. The ERSEM-BFM model includes a 3-layer benthic module comprising 53 state variables. TEP production by diatoms is included, allowing for macro-aggregate formation and rapid sinking out of the spring bloom. The ERSEM-BFM model also has enhanced pelagic-benthic coupling compared other ERSEM-III based models.

The setup includes a spherical grid covering the area 46.4°N-63°N, 17.25°W-13°E with a resolution of 0.08° in longitude and 0.05° in latitude (approximately 5.5 km), and 25 non-equidistant layers in the vertical. The model bathymetry was based on the NOOS bathymetry (www.noos.cc/index.php?id=173). The model was forced with tidal constituents derived from TOPEX-POSEIDON satellite altimetry46, atmospheric forcing from ECMWF ERA-Interim.

SHMI model (SMHI, Sweden)

The Swedish Meteorological and Hydrological Institute (SMHI) model used here has been specifically configured for the Baltic and North Seas (NEMO-Nordic; Hordoir et al., 2019). Its ocean component is based on the Nucleus for European Modelling of the Ocean (NEMO) framework (Madec, 2010), version 3.6. It has 56 vertical levels with a resolution of 3 m close to the surface and decreasing up to 22 m at the bottom of the deepest part of the domain (Norwegian trench). The horizontal resolution is of approximately 2 nautical miles (~3700 m). NEMO-Nordic has two open boundaries located in the English Channel between Brittany and Cornwall and between Scotland and Norway (Hordoir et al., 2019). The biogeochemistry is simulated by the Swedish Coastal and Ocean Biogeochemical module (SCOBI; e.g., Eilola et al., 2009). It includes all major biochemical processes in the water column and sediments (for details in biogeochemical processes we refer to Almroth-Rosell et al., 2011).

All 3 runs where initiated in year 1992, with the same initial conditions created for year 1973 from previous runs where observations and model showed the best agreement. CMEMS boundary conditions could not be used, however near boundary profiles for nitrate, phosphate and oxygen are similar in the data used here and those from CMEMS.

The river nutrient forcing provided by the ICG-EMO group for all 3 scenarios was adjusted to our domain. Because our model domain includes the Baltic Sea, river points where no ICG-EMO data was available were added using e-hype data. These nutrient loads, as well as runoff in the entire domain come from e-hype version v.5.6.2, which showed a significant improvement on salinity in the Baltic Sea with respect to older e-hype versions. For the atmospheric NOx and NHx input from EMEP MSC-W 2018 data (with agreed reduction for the Historical runs) was used throughout the run.

References

Almroth-Rosell E, Eilola K, Hordoir R, Meier HM, Hall PO. Transport of fresh and resuspended particulate organic material in the Baltic Sea—a model study. Journal of Marine Systems. 2011 Jul 1;87(1):1-2.

Baretta, J.W., Ebenhöh, W., Ruardij, P., 1995. The European Regional Seas Ecosystem Model, a complex marine ecosystem model. J. Sea Res. 33, 233-246.

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), 175-198.

Conkright, M. E., Locarnini, R. A., Garcia, H. E., O'Brien, T. D., Boyer, T. P., and Stephens, C. (2002) WorldOceanAtlas 2001: Objective Analyses, Data Statistics, and Figures: CD-ROMDocumentation. (Accessed on November 16, 2016).

Dulière, V., Gypens, N., Lancelot, C., Luyten, P., Lacroix, G., 2019. Origin of nitrogen in the English Channel and Southern Bight of the North Sea ecosystems. Hydrobiologia 845, 13:33. https://doi.org/10.1007/s10750-017-3419-5.

Eilola K, Meier HM, Almroth E. On the dynamics of oxygen, phosphorus and cyanobacteria in the Baltic Sea; A model study. Journal of Marine Systems. 2009 Jan 1;75(1-2):163-84.

Friedland, R., Stips, A., Grizzetti, B., de Roo, A. and Lessin, G., Report on the biogeochemical model of the North-Western European Shelf, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-17866-8, doi:10.2760/78173, JRC120213

Große, F., Kreus, M., Lenhart, H.-J., Pätsch, J., and Pohlmann, T. (2017) A Novel Modeling Approach to Quantify the Influence of Nitrogen Inputs on the Oxygen Dynamics of the North Sea, Front. Mar. Sci., 4, 1–21, doi:10.3389/fmars.2017.00383, 2017.

Hordoir R, Axell L, Höglund A, Dieterich C, Fransner F, Gröger M, Liu Y, Pemberton P, Schimanke S, Andersson H, Ljungemyr P. Nemo-Nordic 1.0: a NEMO-based ocean model for the Baltic and North seas–research and operational applications. Geoscientific Model Development. 2019 Jan 21;12(1):363-86.

Kerimoglu, O., Voynova, Y. G., Chegini, F., Brix, H., Callies, U., Hofmeister, R., Klingbeil, K., Schrum, C., and van Beusekom, J. E. E. (2020) Interactive impacts of meteorological and hydrological conditions on the physical and biogeochemical structure of a coastal system, Biogeosciences 17, 5097–5127, doi:10.5194/bg-17-5097-2020.

Lancelot, C., Spitz, Y., Gypens, N., Ruddick, K., Becquevort, S., Rousseau, V., Lacroix, G., Billen, G., 2005. Modelling diatom and Phaeocystis blooms and nutrient cycles in the Southern Bight of the North Sea: the MIRO model. Mar. Ecol. Porgress Ser. 289, 63–78.

Lazure P., Dumas F. (2008). An external–internal mode coupling for a 3D hydrodynamical model for applications at regional scale (MARS). Adv. Water Resour., 31 (2), 233-250.

Loewe, P., 2003. Weekly North sea SST Analyses since 1968. Original digital archive held by Bundesamt fuer Seeschiffahrt und Hydrographie D-20305 Hamburg, P.O. Box 301220, Germany.

Luyten, P., 2011. COHERENS — A Coupled Hydrodynamical-Ecological Model for Regional and Shelf Seas: User Documentation. Version 2.0. RBINS-MUMM Report. Royal Belgian Institute of Natural Sciences.

Madec G. Nemo ocean engine, v. 3.3. IPSL Paris France. 2010.

Ménesguen A., Desmit X., Dulière V., Lacroix G., Thouvenin B., Thieu V., Dussauze M. (2018). How to avoid eutrophication in coastal seas? A new approach to derive river-specific combined nitrate and phosphate maximum concentrations. Science of The Total Environment, 628–629, 400-414.

Ménesguen A., Dussauze M., Dumas F., Thouvenin B., Garnier V., Lecornu F., Répécaud M. (2019). Ecological model of the Bay of Biscay and English Channel shelf for environmental status assessment part 1: Nutrients, phytoplankton and oxygen. Ocean Modelling, 133, 56-78.

Ruardij, P., van Raaphorst, W., 1995. Benthic nutrient regeneration in the ERSEM-BFM ecosystem model of the North Sea. Neth. J. Sea Res. 33, 453-483.

Sirjacobs, D., Alvera-Azcárate, A., Barth, A., Lacroix, G., Park, Y., Nechad, B., Ruddick, K., Beckers, J.-M., 2011. Cloud filling of ocean color and sea surface temperature remote sensing products over the Southern North Sea by the Data Interpolating Empirical Orthogonal Functions methodology. J. Sea Res. 65 (1), 114–130.

Stegert, C. Lenhart, H.J., Blauw, A., Friedland, R., Leujak W., Kerimoglu, O., Evaluating uncertainties in reconstructing the pre-eutrophic state of the North Sea. In review for publication in Frontiers in Marine Science

Troost, T. A., Desclaux, T., Vethaak, A. D., Leslie, H. A., & Meulen, M. D. Van Der. (2018). Do microplastics affect marine ecosystem productivity? Marine Pollution Bulletin, 135(May), 17–29. https://doi.org/S0025326X18303965

Troost, T. A., Wijsman, J. W. M., Saraiva, S., & Freitas, V. (2010). Modelling shellfish growth with dynamic energy budget models: An application for cockles and mussels in the Oosterschelde (southwest Netherlands). Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1557), 3567–3577. https://doi.org/10.1098/rstb.2010.0074

Vichi, M., Pinardi, N., Masina, S., 2007. A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory. J. Mar. Syst. 64, 89-109.

van der Molen, J., Aldridge, J.N., Coughlan, C., Parker, E.R., Stephens, D., Ruardij, P., 2013. Modelling marine ecosystem response to climate change and trawling in the North Sea. Biogeochem. 113, 213-236, DOI 10.1007/s10533-012-9763-7.

Walters, D., I. Boutle, M. Brooks, T. Melvin, R. Stratton, S.Vosper, H. Wells, K. Williams, N. Wood, T. Allen, A.Bushell, D. Copsey, P. Earnshaw, J. Edwards, M. Gross, S. Hardiman, C. Harris, J. Heming, N. Klingaman, R. Levine, J. Manners, G. Martin, S. Milton, M. Mittermaier, C. Morcrette, T. Riddick, M. Roberts, C. Sanchez, P. Selwood, A. Stirling, C. Smith, D. Suri, W. Tennant, P. L. Vidale, J. Wilkinson, M. Willett, S. Woolnough & P. Xavier, 2017. The met office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. Geoscientific Model Development 10(4): 1487–1520.

Zijl, F., Laan, S., Groenenboom, J., 2021. Development of a 3D model for the NW European Shelf (3D DCSN-FM). Deltares report 11205259-015-ZKS-0003.

Table 1: Overview on model characteristic

Model name	MIRO&CO-3D (RBINS, Belgium)	ECO-MARS3D (IFREMER, France)	ECOHAM5 (UHH-HZG, Germany)	GPM (Oldenburg, Germany)	Deft3D-GEM (Deltares, Netherlands)	PML-ERSEM (JRC Ispra, EU)	GETM-ERSEM-BFM (Cefas - United Kingdom)	(SMHI, Sweden)
General simulation characteristics								
Name hydrodynamic model	COHERENS (Luyten 2011)	MARS-3D	HAMSOM	GETM	Delft3D		GETM	NEMO
Name biogeochemical model	MIRO (Lancelot et al. 2005) Last validation MIRO&CO in Dulière et al. (2019)	ECO-MARS-3D	ECOHAM5	GPM	GEM	PML-ERSEM	ERSEM-BFM	SCOBI
Used model domain area	English Channel and Southern North Sea	Southern North Sea, Channel, Celtic Sea and bay of Biscay	European Shelf including Brittany	Selected area of Southern North Sea	European Shelf including the entire Bay of Biscay	European Shelf halfway down to the Bay of Biscay	European Shelf halfway down to the Bay of Biscay	North Sea, Channel, Baltic Sea
Spatial Resolution Δh (km)	*5.89 km (lon) x 4.63 km (lat)*	*4*	20	1.5 – 4.5 curvilinear grid	1-8 km: 1 km in waters < 100 m deep, 8 km in water > 400 m deep and 4 km in between	4.04-6.13 km (x-direction) and 5.56 km (y)	5.5	~3.7 km
Vertical resolution	5 sigma layers	30 sigma layers	31 z layer	20	20 layers	25 layers, dynamically adapting to density gradients	25 Sigma layers	56 layers, with a resolution of 3 m close to the surface and decreasing up to

Model name	MIRO&CO-3D (RBINS, Belgium)	ECO-MARS3D (IFREMER, France)	ECOHAM5 (UHH-HZG, Germany)	GPM (Oldenburg, Germany)	Deft3D-GEM (Deltares, Netherlands)	PML-ERSEM (JRC Ispra, EU)	GETM-ERSEM-BFM (Cefas - United Kingdom)	(SMHI, Sweden)
								22 m at the bottom of the deepest part of the domain
Longitude (degree)	[-4.0,5.0]	-7.922° – 5.104°E	15.°W – 14.°E	0.15.°W – 9.15°E	-15 to + 14	17.5° W-13.1° E	17.25°W-13°E	4.15278°W to 30.1802° E
Latitude (degree)	[48.5,52.5]	52.769°N – 43.267°N	47.85°N – 64.°N	51.35°N – 55.6°N	44 to 62	46.4° N-63° N	46.4°N-63°N,	48.4917– 65.8914° N
Spatial extent (km)	109 (lon) x 97 (lat) grid cells 642 km (lon) x 449 km (lat)	1060 x 976	82 X 88	137 x 94	?	383 X 334	338 by 390 Nodes.	
Temporal resolution Δt (sec)	Hydrodynamics: 60s, MIRO: 15 min.	200-240	180 hydrodynamic model 1800 biogeochemical model	Output: daily. Integrationt time step: 3D: 360 sec, 2D: 5 sec	Hydrodynamics 2 min (max or smaller where required), water quality: 10 minutes	12.87	15 seconds for hydrodynamics, 450 seconds for biology	Daily and monthly
Temporal range (years)	2000-2014	2006-2014	2006-2014	2009-2014	2009-2014	2006-2014 (2009 at the moment)	2009 – 2014 (CS) run	1992-2014
Spin up time	5 years	2006-2009	3 years	2006-2009	2006-2008	1 year	20 years	17 years
Meteo data	6-hourly surface fields of wind, atmospheric pressure, precipitation rate, cloud cover, humidity and air temperature	Météo France ALADIN model	ERA 5	COSMO-CLM	ERA5	ERA5	ECMWF	

Model name	MIRO&CO-3D (RBINS, Belgium)	ECO-MARS3D (IFREMER, France)	ECOHAM5 (UHH-HZG, Germany)	GPM (Oldenburg, Germany)	Deft3D-GEM (Deltares, Netherlands)	PML-ERSEM (JRC Ispra, EU)	GETM-ERSEM-BFM (Cefas - United Kingdom)	(SMHI, Sweden)
	provided by the Royal Meteorological Institute of Belgium based on the analysed/forecast data of the UK Met Office Global Atmospheric Model (Hi_Res, Walters et al., 2017) Sea surface temperature from weekly sea surface gridded temperature (20 × 20 km2) obtained from the BSH (Loewe 2003)							
Inclusion of tides	Yes (15 harmonics)	yes	yes	yes	Yes	Yes. Using http://volkov.oce.orst.edu/tides/AO.html along the 2d boundaries	Yes	no
Temperature & Salinity diagnostic or prognostic	??	prognostic	T: Prognostic S: prognostic	T: prognositc, S: prognostic	Prognostic i.e. simulated in the model	Prognostic ??	Prognostic	yes
Light	Irradiance model modulated by 6-hourly fields of cloud cover	yes	yes	Incoming+ Astronomically calculated irradiance + cloud cover + attenuation	Solar irradiance from ERA5 and extinction determined by SPM,	Using the ERSEM light model (light_iop)	Yes	no

Model name	MIRO&CO-3D (RBINS, Belgium)	ECO-MARS3D (IFREMER, France)	ECOHAM5 (UHH-HZG, Germany)	GPM (Oldenburg, Germany)	Deft3D-GEM (Deltares, Netherlands)	PML-ERSEM (JRC Ispra, EU)	GETM-ERSEM-BFM (Cefas - United Kingdom)	(SMHI, Sweden)
				by silt and organic material	phytoplankton, detritus and salinity (as proxy for CDOM from rivers)			
Oxygen dynamics	NA	yes	yes	yes	Yes	Using the ERSEM oxygen model with the ERSEM saturation formulation	Yes	yes
SPM dynamics	Forced with daily climatology of SPM derived from MODIS-Aqua images of SPM and interpolated with DINEOF methodology (Sirjacobs et al. 2011)	yes	Climatology	Climatology	External forcing, based on satellite data	No.	Yes	no
Pelagic description								
Pelagic matter cycle (C, N, P, Si)	Yes	N, P, Si	C,N,P,Si	C, N, P, Si	N, P, Si	Yes.	C,N,P.Si	yes
No. of Pelagic state variables	22	21	37	25	20 (each phytoplankton group has 3 phenotypes)	50	32	13
Pelagic Nutrients (bulk or explicit)	NO3 (being NO2+NO3), NH4, PO4, dissolved Si	explicit	explicit	explicit	explicit	Explicit.	Explicit	Explicit

Model name	MIRO&CO-3D (RBINS, Belgium)	ECO-MARS3D (IFREMER, France)	ECOHAM5 (UHH-HZG, Germany)	GPM (Oldenburg, Germany)	Deft3D-GEM (Deltares, Netherlands)	PML-ERSEM (JRC Ispra, EU)	GETM-ERSEM-BFM (Cefas - United Kingdom)	(SMHI, Sweden)
Types of Phytoplankton	Diatoms (functional group), *Phaeocystis globosa* (species, colonial form), autotrophic nanoflagellates (functional group, also acting as *P.globosa* in non-colonial form).	4	Diatoms and Flagellates	Diatoms and flagellates	Diatoms, flagellates, dinoflagellates, Phaeocystis	4 functional groups	Diatoms, Flagellates, PicoPhytoplankto, Dinoflagellates, Small diatoms, Phaeocystis colonies	3
Types of Zooplankton	Microzooplankton (functional group), Copepods (functional group)	2	Micro- and Mesozooplnkton	Micro and mesozooplankton	None	3 functional groups	Carnivorous mesozooplankton, Omnivorous mesozooplankton, Microzooplankton, Heterotrophic nanoflagellates	1
Types of bacteria	Heterotrophic pelagic bacteria (functional group)	0	Heterorophic bacteria	None	None	1 functional group	Pelagic bacteria, Nitrifying archaea	0
Pelagic POM	POC (refractory and non-refractory), PON (ref. and non-ref.), POP (ref. and non-ref.)	yes	Slow (C,N and P) and fast sinking (C,N,P,Si and $CaCO_3$)	Slow (C, N, P) and Fast sinking (C, N, P, Si)	POC, PON, POP, Opal	4 functional groups	Yes	0
Benthic description								
Benthic matter cycle (C, N, P, Si)	Yes	N, P, Si	C,N,P,Si and $CaCO_3$	C, N, P, Si	C, N, P, Si	Yes.	DIC, N03, HH4,S--, H+, P, SiO3	Yes

Model name	MIRO&CO-3D (RBINS, Belgium)	ECO-MARS3D (IFREMER, France)	ECOHAM5 (UHH-HZG, Germany)	GPM (Oldenburg, Germany)	Deft3D-GEM (Deltares, Netherlands)	PML-ERSEM (JRC Ispra, EU)	GETM-ERSEM-BFM (Cefas - United Kingdom)	(SMHI, Sweden)
No. of benthic state variables	10	3	5	4	4 detritus + 6 benthic filterfeeders (2 groups, 3 state variables per group)	36	53	4
Benthic Nutrients (bulk or explicit)	NO3 (being NO2+NO3), NH4, PO4, biogenic Si	bulk	bulk	bulk	Bulk (only detritus)	explicit	Explicit	Explicit
Types of Zoobenthos	NA	0	0	None	Mussels and Ensis	3 functional groups	Meiobenthos, Filter feeders, Infaunal predators, Deposit feeders, Megabenthos	0
DOM	Not in benthic system. Still, DOM is described in the pelagic system by DOC, DON, DOP (each having a ref. and non-ref. fraction)	0	0	None	None	1 functional group	Yes	0
Types of Bacteria	None as state variable. Implicitly, classical benthic bacterial processes are considered (OM degradation, nitrification, denitrification)	0	0	None	None	2 functional groups	Aerobic and anaerobic Benthic bacteria	0
Benthic POM	POC (refractory and non-refractory), PON (ref. and non-ref.),	yes	yes	None	yes	3 functional groups	Labile Organic Matter	0

Model name	MIRO&CO-3D (RBINS, Belgium)	ECO-MARS3D (IFREMER, France)	ECOHAM5 (UHH-HZG, Germany)	GPM (Oldenburg, Germany)	Deft3D-GEM (Deltares, Netherlands)	PML-ERSEM (JRC Ispra, EU)	GETM-ERSEM-BFM (Cefas - United Kingdom)	(SMHI, Sweden)
	POP (ref. and non-ref.)							
Participant to add further characteristics if required		yes					TEP production by nutrient stressed diatoms included.	

Model Setup for the ICG-EMO model comparison workshop for historical scenario

Author: The Convenor of ICG-EMO (Germany), based on feedback from ICG-EMO partner on virtual meetings incl. workshop 1. and 2. December 2020.

Definition of the scenarios

Goal: Derive new target values for OSPAR assessment related to new integrated assessment areas. This overall goal includes as a first objective to derive the „reference“ values related to the new assessment areas, where „reference“ refers to the level in a pre-eutrophication period.

The goal should be achieved by a model comparison study from a number of ICG-EMO partners in a combined effort. The basic idea is to run the different models in a setup which allows for maximum comparability by the use of common forcing, considering the constraints of time and the financial support that is offered to the different modelling groups.

Overall approach: The basic setup follows the modelling steps applied in the JMP EUNOSAT project. First the simulation covers the current state condition (CS) for the assessment period 2009-2014. The following scenario runs aim to represent so-called “historic” or “pre-eutrophic” environmental state of the marine environment by applying estimates for the river nutrient loads, atmospheric N deposition as well as boundary condition which represent a reference situation.

For the scenario runs the same assessment period 2006-2014 (incl. spin-up) should be applied, but the river loads, atm. N deposition and adopted boundary condition are adopted from the definition of the historic scenario setup. The scenario runs will consist of two different definitions for the historic scenarios. For both scenarios the same assumptions on the atmospheric N deposition and changes for the Baltic boundary condition are applied, whereas the river load information will be different and for each scenario the complete data for the river loads will be provided individually.

There is a decision from HASEC that the default scenario will the one by the JMP EUNOSAT project, which is based on estimates from the Swedish E-Hype catchment model. There will be slightly corrected to the original estimates by a non-deterioration approach, which simply guarantees that the pre-eutrophic river load applied in the model run can’t be higher than the current state load. Since the E-Hype model has some drawbacks within the P load estimates, a second hybrid scenario with corrections only in the P load are agreed on by the pre-eutrophic expert group. A report on the scenario definition is provided in Annex 1.

All data or other products for the model comparison are located on the cloud server of the University Hamburg under the link:

https://cloud.wr.informatik.uni-hamburg.de/s/CM3Gb3HfPec7ZL4

The password is ICG_EMO_2020

This includes an EXCEL workbook for each ICG-EMO participants which allows to gather the model output in a formal way for the aggregation into an ensemble product. Based on the outcome of this scenario the formal way to achieve target values is to add 50 % on top of the resulting concentration from this reference run in the marine environment, for the eutrophication indicators DIN, DIP and Chlorophyll-a. Finally, these new values need to be prepared as mean seasonal averages, for the upper 10 m, aggregated per assessment area that are just finalized in an effort from TG-COMP for the COMPEAT Tool provided by ICES.

Common model forcing

River load information: The idea is that Sonja van Leeuwen prepares the load data ready for use for the current state simulation as well as the **two** pre-eutrophic scenarios. These are based on the data she has compiled within her ICG-EMO database. The adaptation on the scenarios is another complex task which needs to match the information from the E-Hype input information (~900 discharges) with the ICG-EMO river loads (~ 300 discharges). All river daily load data cover the period from 1940 to 2022, which allows for a longer spin-up for models with a complex sediment chemistry.

Sonja van Leeuwen provides river files for the Current State run as well as for both scenario's, with all the reductions included. The reductio will be applied as follows

Q % : flow, Si, SPM

N % : TN, NO2, NO3, DIN, NH4

P %: TP, PO4

The link for the river load information is located in different directories under:

https://cloud.wr.informatik.uni-hamburg.de/s/CM3Gb3HfPec7ZL4?path=%2FRiver-data

The river loads are organized in directories that aggregate the river information for one country. The German river data (DE) cover the bigger rivers like Elbe, Weser, Ems and Eider plus three additional Wadden Sea inputs from Schleswig Holstein: Arlau, Bongsieler Kanal and Miele. Additional rivers are provided (Peene, Schwentine, Trave and Warnow) that could be used if your model domain covers the German part of the Baltic Sea.

For other organic compounds PON and POP the reduction should be linked to the same flow. Anouk Blauw provided climatological scaling factors which are available under the link:

https://cloud.wr.informatik.uni-hamburg.de/s/CM3Gb3HfPec7ZL4?path=%2FRiver-data%2FPON-POP-Scaling-Factors

An overview for the use of organic load within the individual models is provided in Tab. 1.

		Variables Directly Available in the ICG-EMO Database															
Models:	Institutes	Q	TN	NO3	NO2	NH4	DIN (1)	TP	PO4	Silicate	DIC	DOC	POC	TOC	Fe	TALK	SPM
ECOHAM	Uni:HH / HZG	Y	N	Y	N	Y	N	N	Y	Y	Y	Y	Y	N	N	Y	N
GPMEH	Uni-Oldenburg	Y	Y	Y	N	Y	N	Y	Y	Y	N	Y	Y	Y	N	N	N
MIRO&CO	RBINS	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	N	N	N	N	N
GETM-FABM-ERSEM (JRC)	JRC	Y	N	Y	N	Y	N	N	Y	Y	N	N	N	N	N	N	N
MARS3D-MANGA4	Ifremer	Y	N	Y	N	Y	N	N	Y	Y	N	N	N	N	N	N	Y
DFLOW-FM (Deltares)	Deltares	Y	N	Y	N	Y	N	N	Y	Y	N	N	N	N	N	N	N
NEMO-Nordic-SCOBI (smhi)	SMHI	y	N	Y	N	Y	N	N	Y	Y	N	N	N	N	N	N	N (8)
GETM-ERSEM-BFM	Cefas	Y	N	Y	N	Y	N	N	Y	Y	N	N	N	N	N	N	N

Table 1: Overview on variables use for river input by the models [(1) DIN = NO_3 + NO_2 + NH_4]

Boundary Condition:

Overall boundary condition:

It was decided from the partners that the boundary condition should be taken from the CMEMS dataset. The overall link to the data server is:

https://resources.marine.copernicus.eu/?option=com_csw&task=results

This link provides a general overview of all available datasets. For the use of the ICG-EMO model comparison the GLOBAL REANALYSIS BIO dataset should be used.

https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=GLOBAL_REANALYSIS_BIO_001_029

Deltares has developed a tool to convert the CMEMS model data to model input for the boundaries. The Python-script of this tool can be found at Github:

https://github.com/FineWilms/coastserv

There is also an online user interface for this tool available. The interface of the cloud tool can be accessed at: http://coastserv.westeurope.azurecontainer.io/ . (Input is bounding box, time interval, .pli file, CMEMS username and password).

An example of a pli-file (giving lat-lon coordinates of the boundary grid cells) is available under

https://cloud.wr.informatik.uni-hamburg.de/s/pYTnEDdgMwGTfiQ

One can adapt the pli-file format to your own boundaries and reformat the output of the tool to your own file format (or use another tool that you might have available).

Baltic boundary condition

In the expert group the question was raised if boundary condition for the Baltic need to be adopted for the Baltic outflow (see Annex 1). Anouk Blauw created these boundary conditions where the Baltic inflow into the North Sea for recent years is based on simulated discharge data by the Dars and Drogden sills, provided by DHI. They provided monthly total discharges for 2002 – 2019. We calculated monthly mean climatologies from these data as model input. The nutrient concentrations in recent years are based on monitoring data provided by Stiig Markager at three monitoring locations nearby the sills for the years 2009 – 2014. These were also converted to monthly mean climatologies as model input, for the variables: temperature, salinity, oxygen, NO3 (incl NO2), NH4, PO4, bioavailable DON, bioavailable DOP, POC, PON and POP. For Silicate an estimate of 10.4 µM was used, based on data in a paper by Bentzon-Tilia et al., (2014). It is generally believed that Si is not limiting phytoplankton growth in the Baltic. Reduction percentages for the historic scenario (Tab. 2) have been derived from a long model simulation (1850 – 2008) with the ERGOM model provided by Thomas Neumann (IOW, DE).

The detailed information on the export for the Drogden and Dars sills can be found in the EXCEL file “Drogden_and_dars_loads” on the cloud server under:

Tabel 2: Reduction percentage for historic scenario for Drogden and Dars sills

Variable (mol/kg)	Drogden sills	Dars sills
NO3	66 %	54 %
NH4	58 %	55 %
DON	53 &	53 %
PO4	33 %	33 %
POC	88 %	91 %

Atmospheric Forcing: EMEP atmospheric N deposition data will be used for the current state run. These data will be scaled with correction factors based on Schöpp et al. (2003) to represent pre-eutrophic condition around 1900. Onur Kerimoglu compiled the following information:

Reduction of atmospheric deposition rates were calculated based on the estimations by Schöpp et al. (2003). Previously, Markus Kreus had digitized the trajectories of NO_x and NH_3 estimates provided in their Fig. 2, and used these as normalization factors to project the spatially resolved nitrogen deposition rates provided by EMEP for the North Western Continental Shelf domain at 20km resolution (NWCS) back to 1880, as described by Große et al., 2016. For the current task, we simply calculated the average rates for the current (2009-2014) and historic (1890-1900) time periods for NO_x, NH_3 and N_{total} , and the respective ratios, as reported in Table 3. More information can be found in Annex 1.

Table 3: Average N deposition rates for the current (2009-2014) and historic (1890-1900) time periods and the respective ratios.

	Control [mg m^{-2} Y^{-1}]	Historic [mg m^{-2} Y^{-1}]	NX_H/NX_C
NO_x	183.574	26.389	0.144
NH_3	167.494	105.812	0.632
N_{total}	351.067	132.201	0.377

Model simulation period and spin up runs: The simulation period that will be compared between the models covers the year 2009 – 2014, since the period falls in line with the MSFD reporting interval of 6 years. The models should have a spin-up of 3 years. If a complex sediment representation requires more years this should be taken into account with the spin-up period and indicated later with the model results. Even for a longer spin-up period the river load data allow to use real data from individual years. Therefore, no re-run of on previous year of even 2009 data is needed for the spin-up process.

Assessment areas: The assessment will be carried out on the basis of the newly defined assessment areas. The OSPAR Hazardous Substances and Eutrophication Committee (HASEC) has agreed on adopting the subdivision of the North Sea based on ecological-relevant assessment areas, as established in the JMP EUNOSAT project. Some adaptations are still in process. We use the latest shapefile that represents the current setup of these assessment areas (Fig. 1) as provided by OSPAR in August 2020.

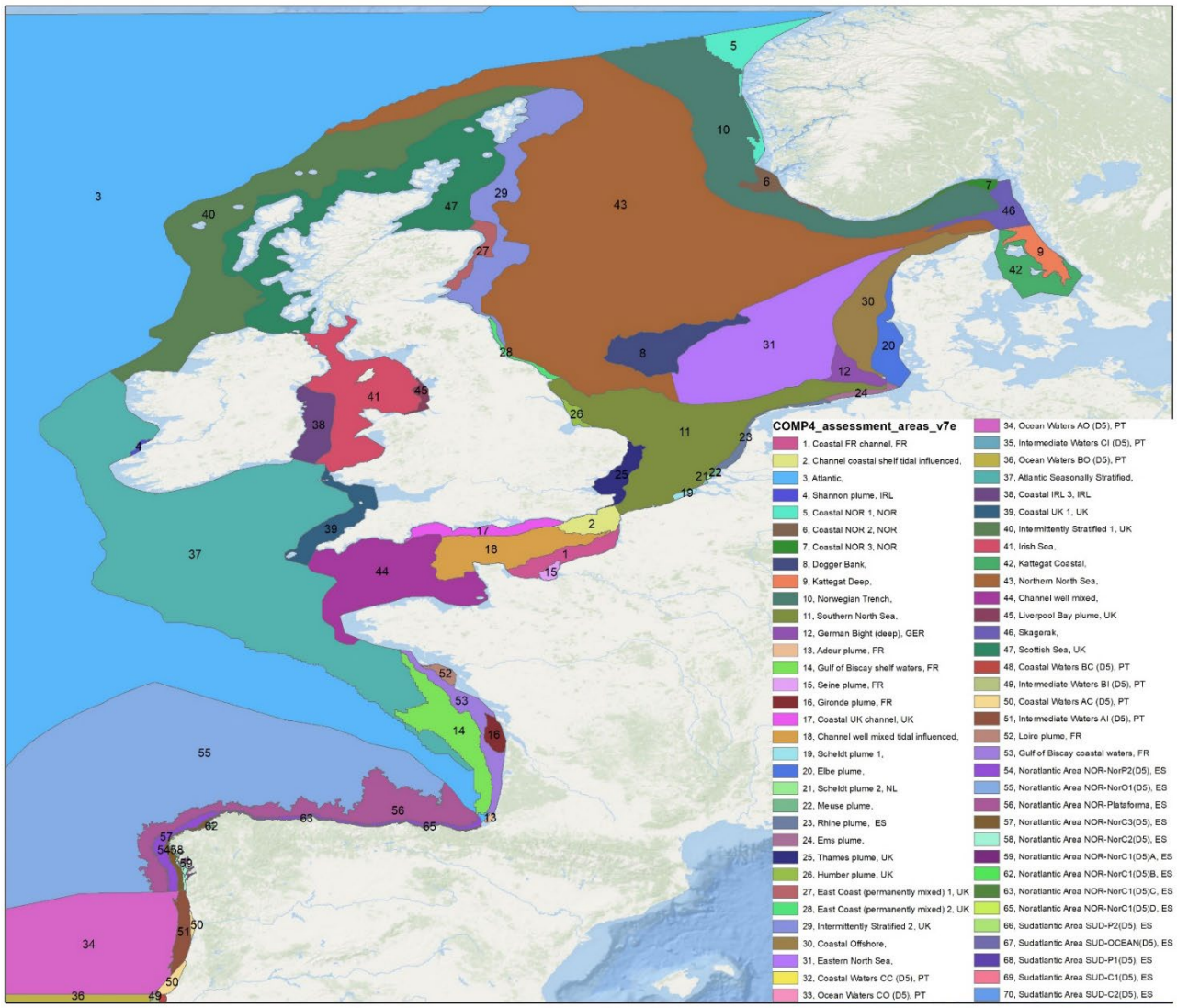


Figure 1: Overview on the assessment areas V7e (provided by René Friedland)

The link for the information, including the shapefile is:

https://cloud.wr.informatik.uni-hamburg.de/s/CM3Gb3HfPec7ZL4?path=%2FAssessment-area

The model output for the key eutrophication parameter will be provided as mean information for these assessment areas. This will be organized by means of an EXCEL workbook which will be introduced later.

Validation: René Friedland provided validation data which are already aggregated for these assessment areas. The data are quality checked from the ICES database which used these data within the newly developed assessment tool COMPEAT. Monthly climatologies of the observations were computed, basing exclusively on the quality-checked data. Two datasets are provided.

First, the observations were evaluated purely for the stations, where a high number of observations was available. Furthermore, the climatologies for the single COMP4 regions (version 7e) were computed by putting all data points from each region and month together. For more details René Friedland also provided a Readme which should be looked at before using the data. The data are available under:

https://cloud.wr.informatik.uni-hamburg.de/s/CM3Gb3HfPec7ZL4?path=%2FAssessment-area-validation

Scientific publication: Next to the OSPAR report we also aim for a scientific publication of the work from the model comparison. Sonja van Leeuwen will take the lead for this publication. For the post-processing of our model work this means that we have to take into account the output for the OSPAR eutrophication indicators, which will be the basis for the assessment and the resulting OSPAR report, as well as additional parameters for the scientific analysis for the publication.

Model output:

The list of parameter follows the discussed within the ICG-EMO group. Additional discussion took also place within the pre-eutrophic group. Points from these discussions from Philip Axe (HASEC chair) and Stiig Markager (Aarhus University, DK) are displayed for later discussion or interpretation of the model results.

OSPAR eutrophication indicator:

The parameters to be assessed are the eutrophication indicators like Winter DIN and DIP, Summer Chlorophyll-a (in the upper 10 m) and oxygen. For the spatial aggregation of the eutrophication indicators the newly defined EUNOSAT assessment areas, now officially addressed as "COMP" area, are used. This implies that shape files from the latest version of the assessment area definition v7e will be used.

The definition of Winter and Summer is linked to the use in EUNOSAT, simply for the reason to be consistent with the definition of the assessment areas. For clarification:

Winter: December – February

Summer (growing season): March – September the month are always inclusive.

For oxygen: Minimum O_2 concentration over the year for the bottom lay, and if possible the number of days below 6 mg/l. There was a discussion within the ICG-EMO group on how to represent oxygen with a minimum amout of post-pocessing involved.

In general, OSPAR also highlights the need the use of TN and TP as aggregated variables because of the importance within the MSFD. The TN and TP content could be aggregated by the inorganic and organic parts of the N and P compounds within the state variables, if this is possible from the model output. It is clear that we will miss the refractory part of the N and P compounds.

Stiig Markager commented: Regarding TN and TP: How will you handle the refractory part? You write that this is not included in the model, but it is a large fraction and must be included in order to compare with measurements.

The comments from Stiig are relevant, however we need to point out that this is a first testbed for the models to achieve a comparison for these parameters. So we have to learn how to deal with these parameters in perspective for the future use within the MSFD.

MSFD consideration

During the pre-eutrophic group meetings the delegates from DK pointed out their doubts on how representative the analysis with focus on the OSPAR eutrophication indicators will be. The main argument is that the MSFD asks for a description on higher trophic level for the definition of the GES. Therefore, the idea is to include these parameters, if possible in the model output. Then these parameters can be analysed for the CS run and the two reduction scenario as a basis for further discussion.

In relation to the MSFD Philip Axe provided the following information:

The Commission Decision

https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017D0848&from=EN

prioritises nutrient concentrations, chlorophyll concentrations and bottom oxygen. Secondary parameters are occurrence of HABs, photic limit, opportunistic macroalgae, macrophyte community structure, benthic macrofauna. Of the secondary criteria, the macroalgae/macrophytes are only appropriate in depths of less than about 20 metres (historic records suggest 25 metre depth limit for eelgrass in the Kattegat).

The ICG-EMO group decided on the following list of parameters as the basis for further discussion:

Primary Production: as defined as Gross primary production – respired carbon.

Philip information: "This would be useful guidance, in particular for setting the 50% limit. SE and DK have > 30 yrs of in-situ pelagic PP data from the Kattegat and Skagerrak. This is very variable, however as it seems phytoplankton like to be sampled at the same time every day, so using this for validation might be tricky."

Stiig commented in addition: "Regarding respiration it will be good to spilt it into Resp_phyto, Resp_bac and Resp_total. If you then also have GPP, all derived parameters can be calculated."

Photic zone and Secchi depth:

We will use the photic zone depth k_d as 1 % of the surface PAR, in the form k_d = 1 / secchi depth as referred to by Lee et al., (2015).

Philip information: I would definitely lift Secchi depth / photic limit as useful. There are some historic observations from the 1930s and earlier even in the North Sea. DK, DE, SE and I think NO also make use of Secchi depth. This is a useful integrative parameter. UK and IE avoid it as they have high suspended sediment loads (although only in the SW North Sea and central Irish Sea – so it could be helpful to have guidance here)."

Finally, there were some debates about the memory effect from the sediment within the pre-eutrophic group. For example, Stiig commented: *"What about pool of C, N and P in the sediments?"* . The discussion on this issue is very important and is therefore covered within annex 9 of the pre-eutrophic report which will be part of the ICG-EUT 2021documentation.

Output for scientific analysis:

Next to the eutrophication indicators we should also seek to define additional parameters used for a scientific analysis, for example in view of a later publication. Here is a short overview on the parameters presently under debate for this purpose:

1. The annual 90th percentile of Chl
2. 10th percentile of O2 and numbers of days below threshold
3. NPP
4. N:P ratio,
5. Phytoplankton species composition
6. Composition of the food web: phytoplankton: zooplankton ratio, phytoplankton:bacteria ratio and the zooplankton:bacteria ratio
7. If applicable: the diatom:flagellate phytoplankton ratio
8. Also if applicable, also the benthos:benthicbacteria ratio and the pelagic biomass:benthic biomass ratio

The idea that Sonja van Leeuwen proposed is that we want to be able to identify changes to the pelagic/benthic balance, the plankton/bacteria balance and the fast/slow phytoplankton balance.

Selection of assessment areas

Since there are about 70 assessment areas in total a choice was needed to focus on a number of most relevant areas. The option was given to the pre-eutrophic expert team to and the few feedbacks were taken into account. Fig. 3 provides an overview of the selected assessment areas for the focus of the OSPAR report.

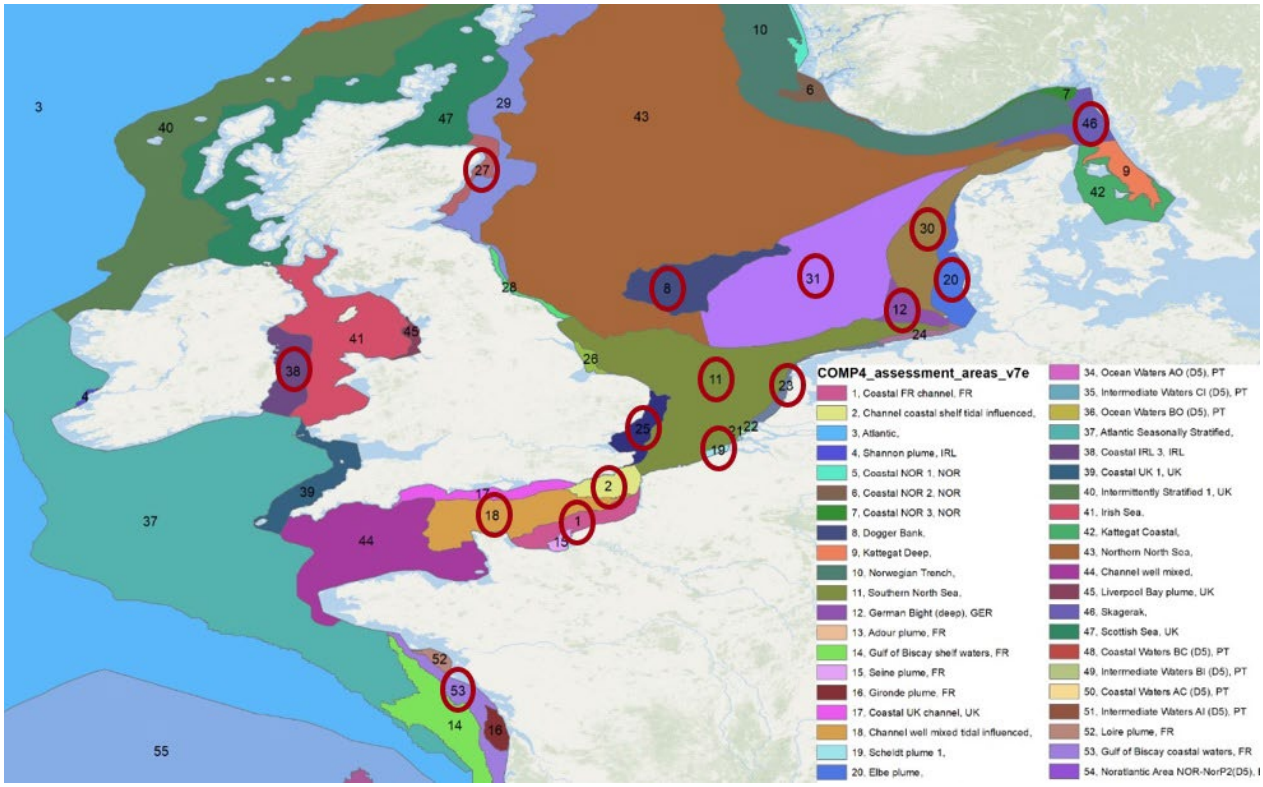


Figure 2: Overview on selected assessment areas

EXCEL Workbook

For the aggregation of the model results, each participant will be provided with an EXCEL workbook. This workbook has been developed by Sonja and should cover the whole assessment of his model exercise. Please transfer your results into the workbook for an overall assessment and sent the results to Sonja and myself.

The workbook contains the following sections:

- General definition of parameters
- Model information on assessment area coverage by your model
- OSPAR assessment sheet for the CS run and the first and second scenario (one sheet each)
- Scientific assessment sheet for the CS run and the first and second scenario (one sheet each)

The idea is that each participant fills out his copy of the workbook as progressing with the work. As soon as you have achieved on model run please transfer the post-processing results into the workbook and sent it to us so that we don't loose time to aggregate the results.

In a first step please identify the workbook for your model application or your institute. This means that you change in the very left upper corner the abbreviation for your application.

In a second step please indicate which areas are covered by your model domain and if possible to which extent (%)

IMPORTANT:

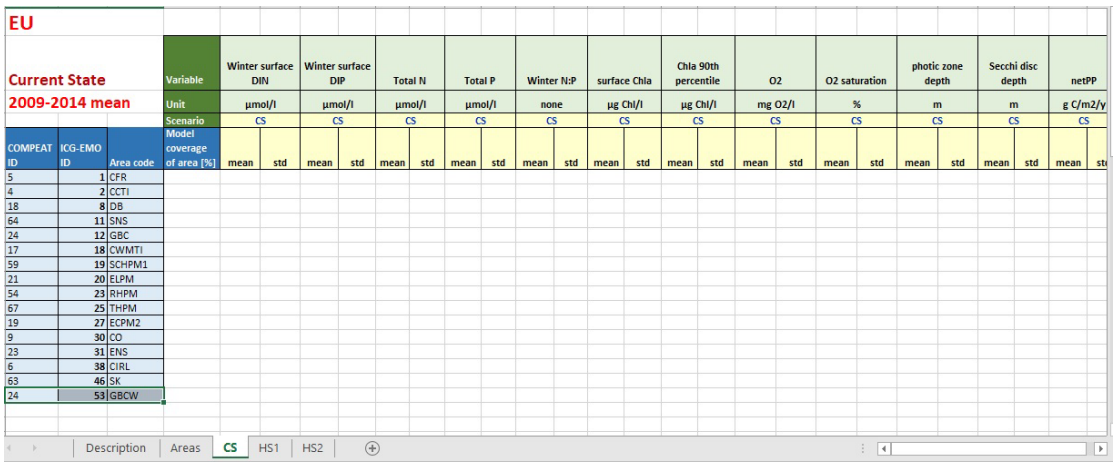
We need the ORIGINAL model results as information from the CS run and the 2 scenario, which means the concentration for the historic scenario. The derivation of the target concentration with the addition of 50 % on top of the historic load will be calculated by us.

OSPAR assessment sheet

For the workshop the first priority lies on the OSPAR assessment sheet. This covers the key eutrophication parameter which will form the basis of the OSPAR report we have to deliver. Please make sure the values you provide are in line with the definition of the units in the "Description" sheet. The idea is that here the information for all areas that are covered by your model domain should be provided. Therefore, we only ask for the statistical properties aggregated for the whole assessment period, which covers 2009 to 2014.

Fig. 3 provides an overview on the assessment sheet for the OSPAR key parameter for the CS current state simulation. The other two sheets for the different scenarios (HS1 and HS2 sheets) are similar. The example is not finished as it will finally cover all assessment area.

The "Descritption" provided the info on units, integration etc. The "Area" sheet will cover the information on the coverage of your domain for all assessment areas. As was pointed out during the workshop discussion by René Friedland some models with a lower grid resolution can not achieve a complete coverage even though the area lies completely within the model domain. This has to do with the match of the coastline for the individual grid cells. Therefore it was concluded that an 80 % coverage should be used as a threshold to take into account information from this assessment area for the overall aggregation.



EU

Current State 2009-2014 mean	Variable	Winter surface DIN		Winter surface DIP		Total N		Total P		Winter N:P		surface Chla		Chla 90th percentile		O2		O2 saturation		photic zone depth		Secchi disc depth		netPP	
	Unit	µmol/l		µmol/l		µmol/l		µmol/l		none		µg Chl/l		µg Chl/l		mg O2/l		%		m		m		g C/m2/y	
	Scenario	CS		CS		CS		CS		CS		CS		CS		CS		CS		CS		CS		CS	
COMPEAT ID	ICG-EMO ID	Area code	Model coverage of area [%]	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	st
5	1	CFR																							
4	2	CCTI																							
18	8	DB																							
64	11	SNS																							
24	12	GBC																							
17	18	CWMTI																							
59	19	SCHPM1																							
21	20	ELPM																							
54	23	RHPM																							
67	25	THPM																							
19	27	ECPM2																							
9	30	CO																							
23	31	ENS																							
6	38	CIRL																							
63	46	SK																							
24	53	GBCW																							

Description | Areas | CS | HS1 | HS2

Figure 3: Screenshot from CS assessment sheet in workbook (provided by Sonja van Leeuwen)

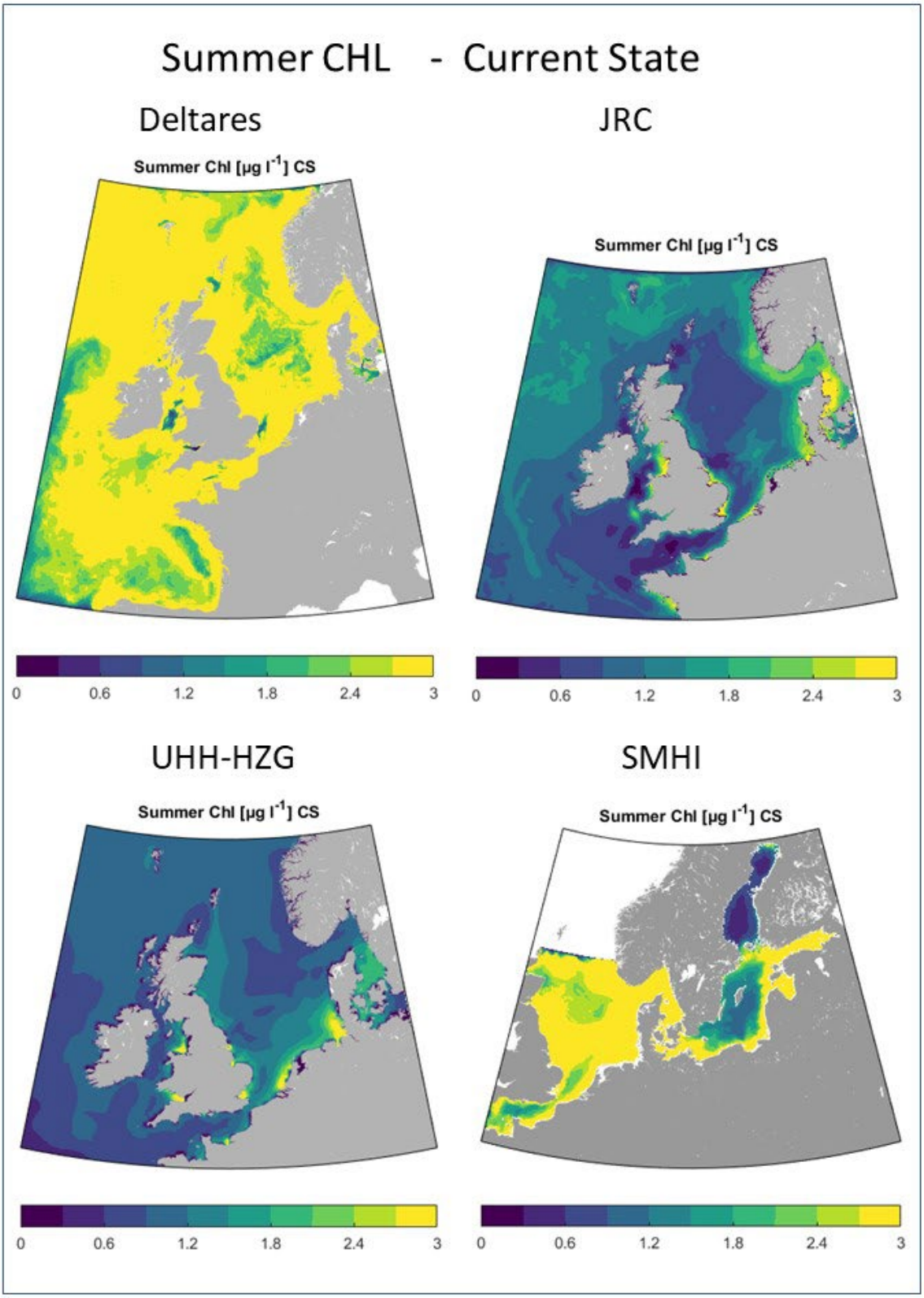
Scientific assessment sheet

These sheets will be added later since we have to focus first on the OSPAR report. They will cover a wide range of parameters for further scientific analysis of the model results. In contrast to the OSPAR assessment sheet, here we only ask the information for a number of selected areas, but this time split up for the individual years. Again, please make sure the values you provide are in line with the definition of the units in the "Description" sheet. If some of the parameters are not clear please ask.

Model results for Current State (CS) run and the tho scenario

Finally, the individual model results are presented for the key parameters Chlorophyll-a, DIN and DIP for the Current State (CS) simulation (figures 4 – 6), the HS1 (figures 10 – 12) and the HS2 scenario (figures 13 – 15). In addition, the percent differences between the historic scenarios and the CS run are presented for HS1 vs. CS ((figures 10 – 12) and HS2 vs. CS (figures 16 – 18). Included in all these figures are the resulting representation of the weighted ensemble concentration or the differences between the historic scenario and the CS run for the COMP4 assessment areas.

All tables and the graphs for the figures 1 – 3, as well as the graphs representing the COMP4 areas in figures 4 – 18 are provided by Sonja van Leeuwen (NIOZ, The Netherlands). All graphs for the individual models (figures 4 – 18) are supplied by Xavier Desmit (RBINS, Belgium).


Summer CHL - Current State
Deltares
Summer Chl [µg l-1] CS
0 0.6 1.2 1.8 2.4 3
JRC
Summer Chl [µg l-1] CS
0 0.6 1.2 1.8 2.4 3
UHH-HZG
Summer Chl [µg l-1] CS
0 0.6 1.2 1.8 2.4 3
SMHI
Summer Chl [µg l-1] CS
0 0.6 1.2 1.8 2.4 3

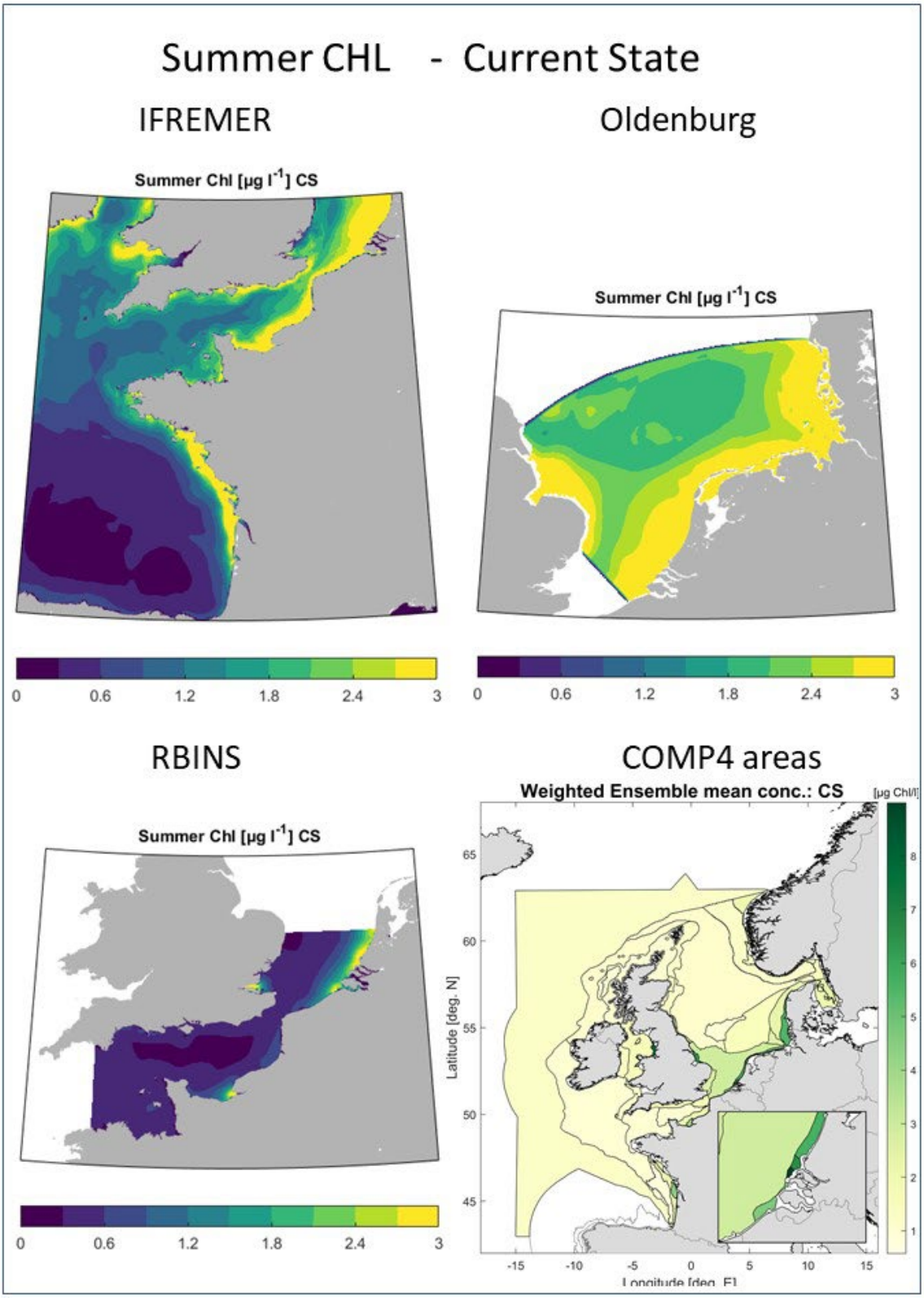
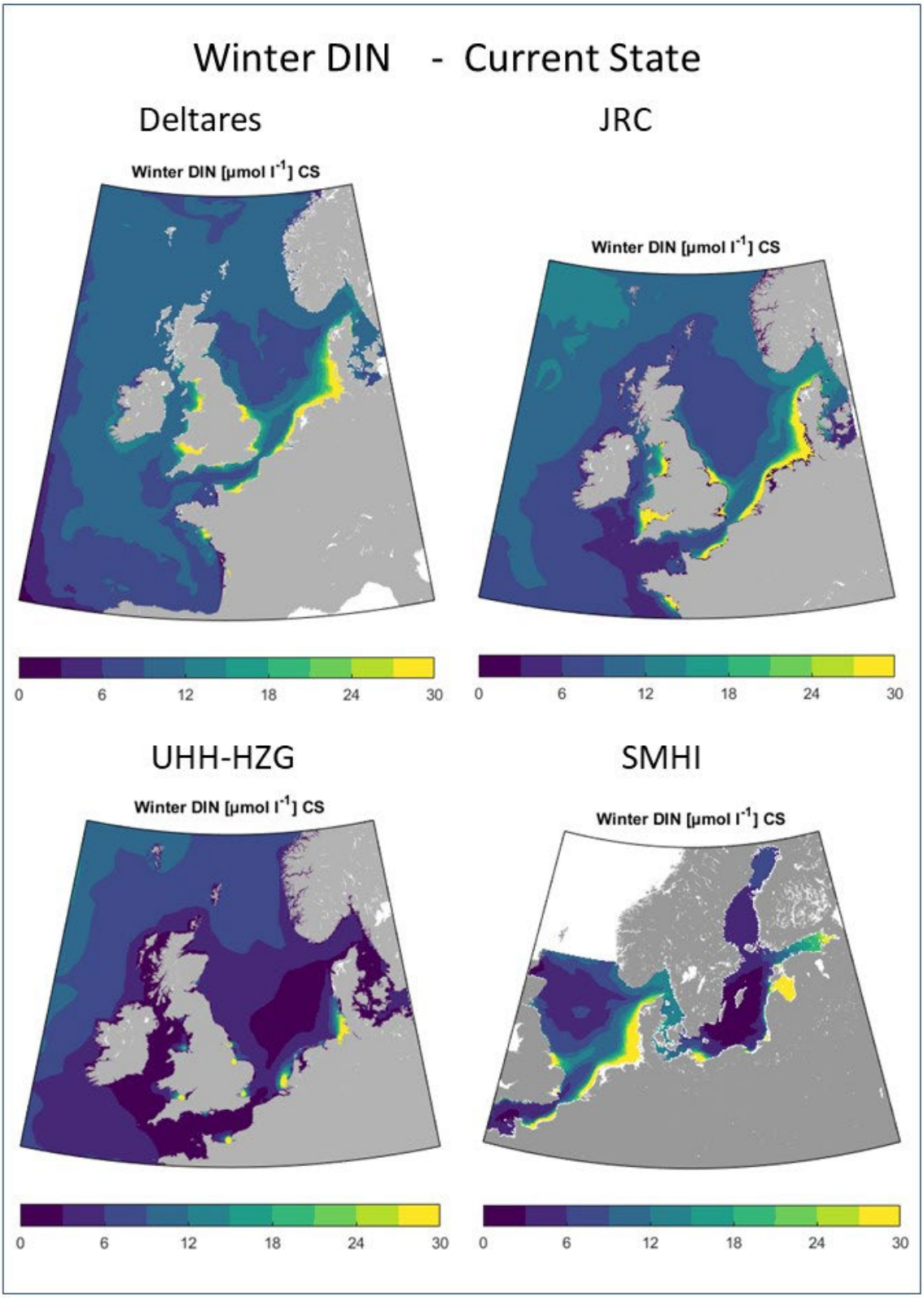


Figure 4: Chlorophyll-a concentration for the Current State (CS) simulation from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


Winter DIN - Current State
Deltares
Winter DIN [µmol l-1] CS
0 6 12 18 24 30
JRC
Winter DIN [µmol l-1] CS
0 6 12 18 24 30
UHH-HZG
Winter DIN [µmol l-1] CS
0 6 12 18 24 30
SMHI
Winter DIN [µmol l-1] CS
0 6 12 18 24 30

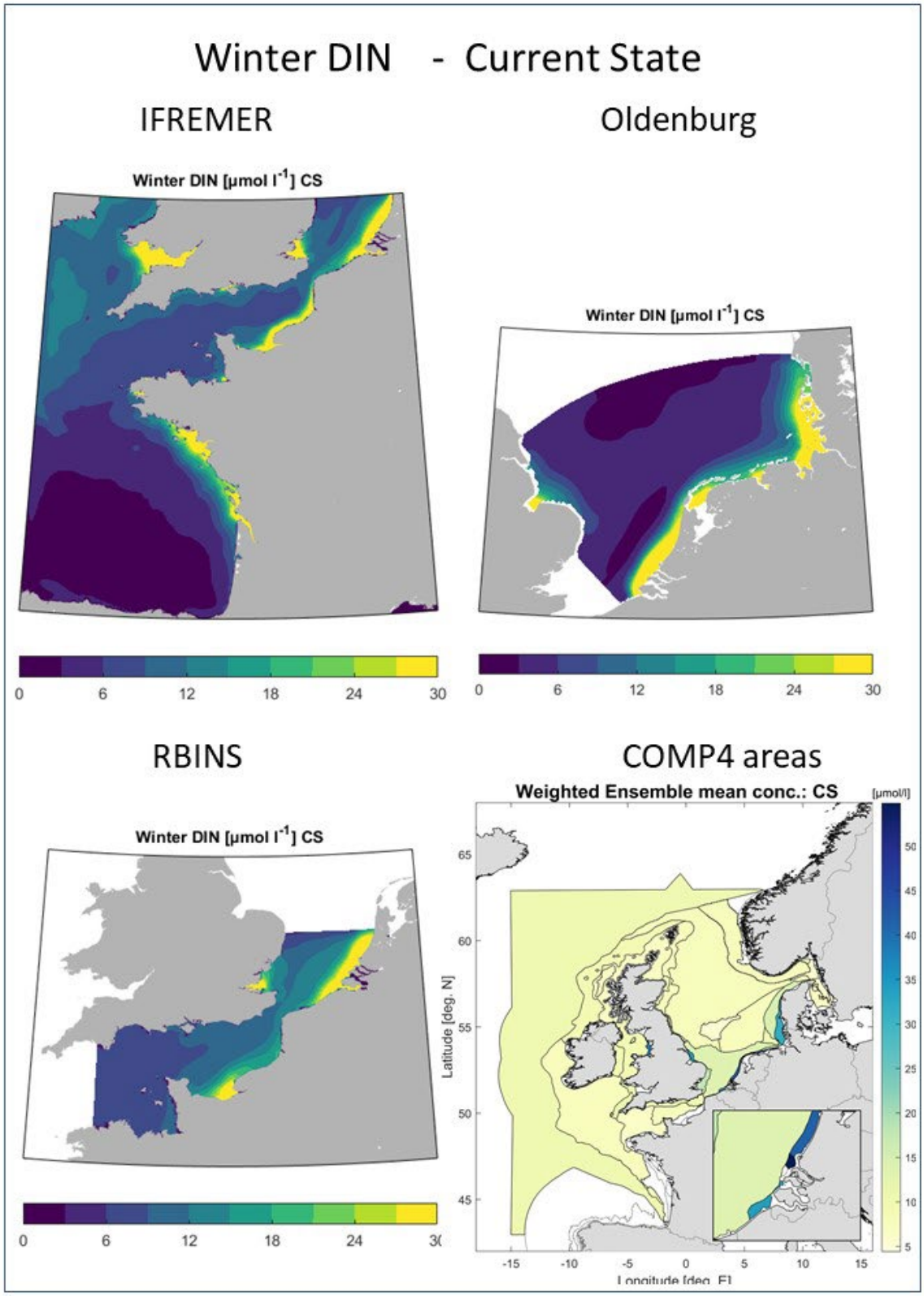
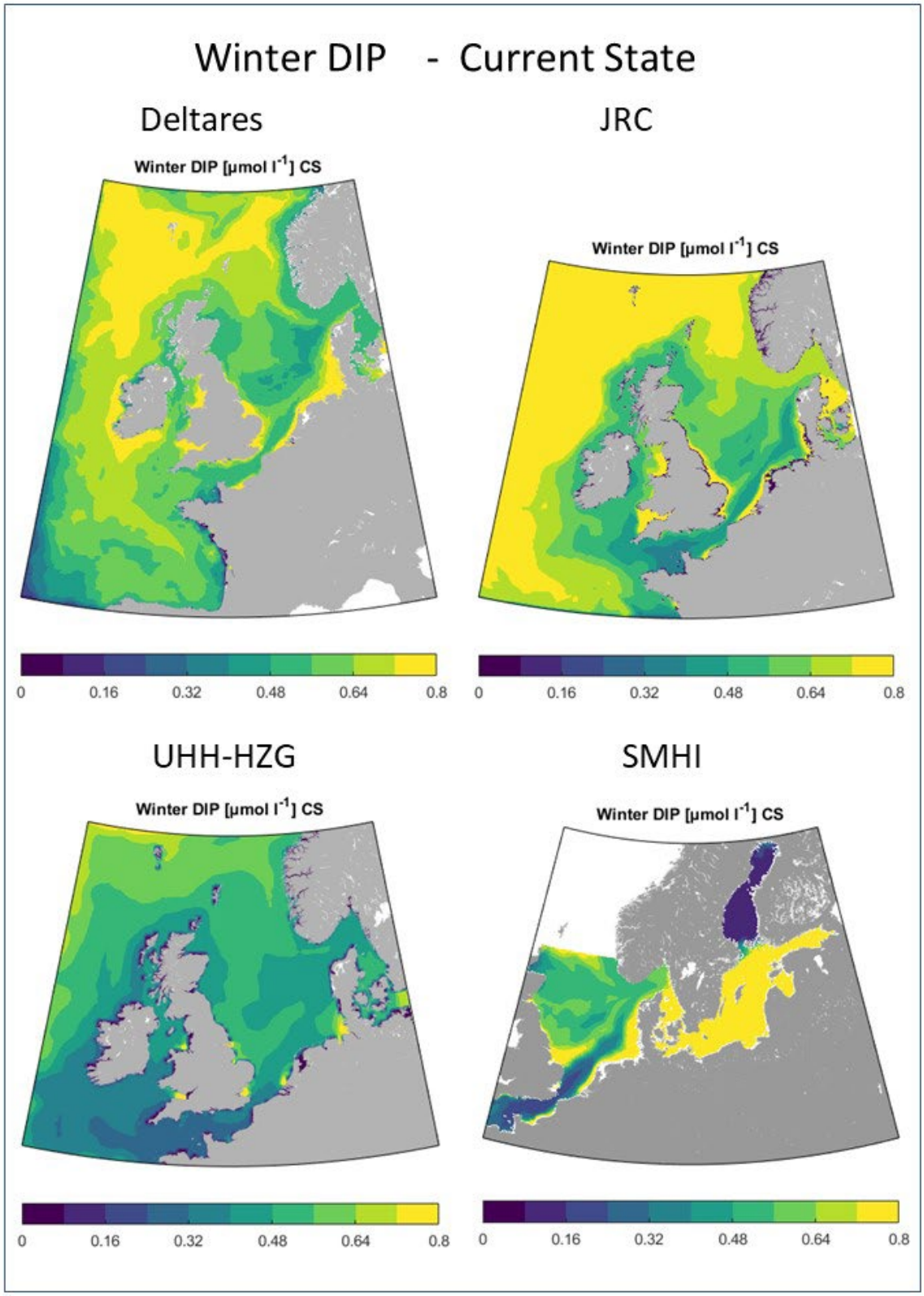


Figure 5: DIN concentration for the Current State (CS) simulation from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


Winter DIP - Current State
Deltares
Winter DIP [µmol l-1] CS
0 0.16 0.32 0.48 0.64 0.8
JRC
Winter DIP [µmol l-1] CS
0 0.16 0.32 0.48 0.64 0.8
UHH-HZG
Winter DIP [µmol l-1] CS
0 0.16 0.32 0.48 0.64 0.8
SMHI
Winter DIP [µmol l-1] CS
0 0.16 0.32 0.48 0.64 0.8

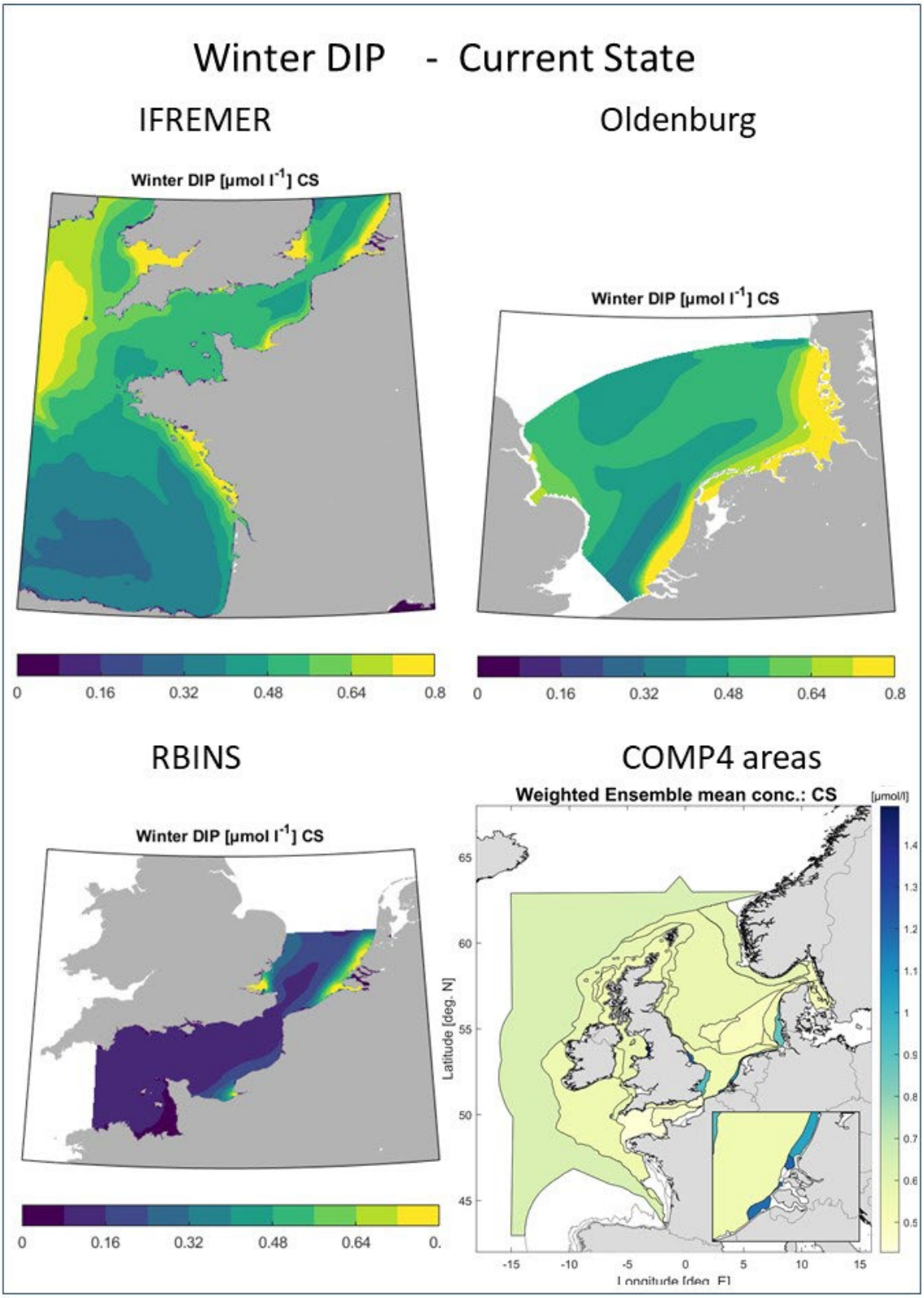
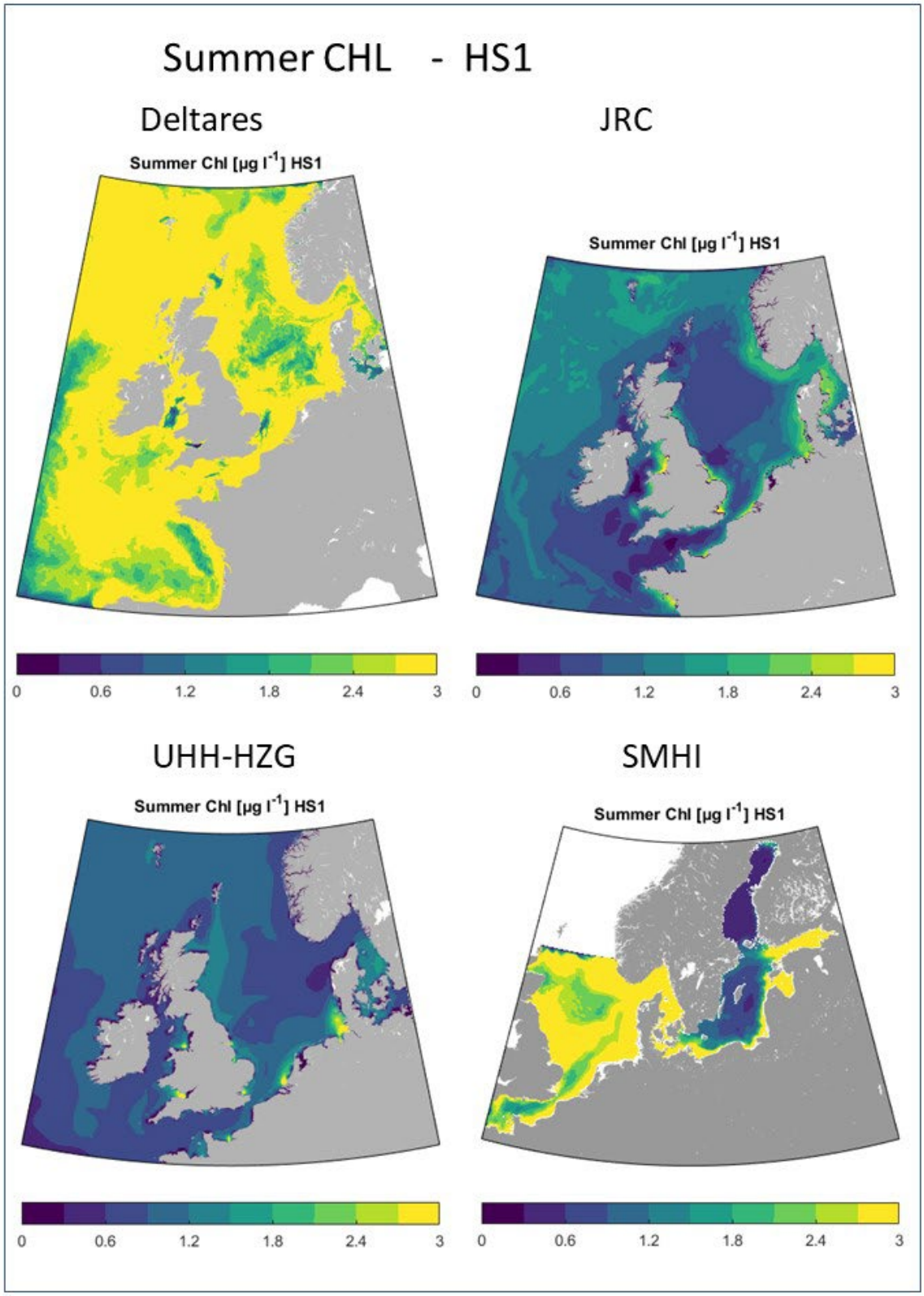


Figure 6: DIP concentration for the Current State (CS) simulation from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


Summer CHL - HS1
Deltares
Summer Chl [µg l-1] HS1
0 0.6 1.2 1.8 2.4 3
JRC
Summer Chl [µg l-1] HS1
0 0.6 1.2 1.8 2.4 3
UHH-HZG
Summer Chl [µg l-1] HS1
0 0.6 1.2 1.8 2.4 3
SMHI
Summer Chl [µg l-1] HS1
0 0.6 1.2 1.8 2.4 3

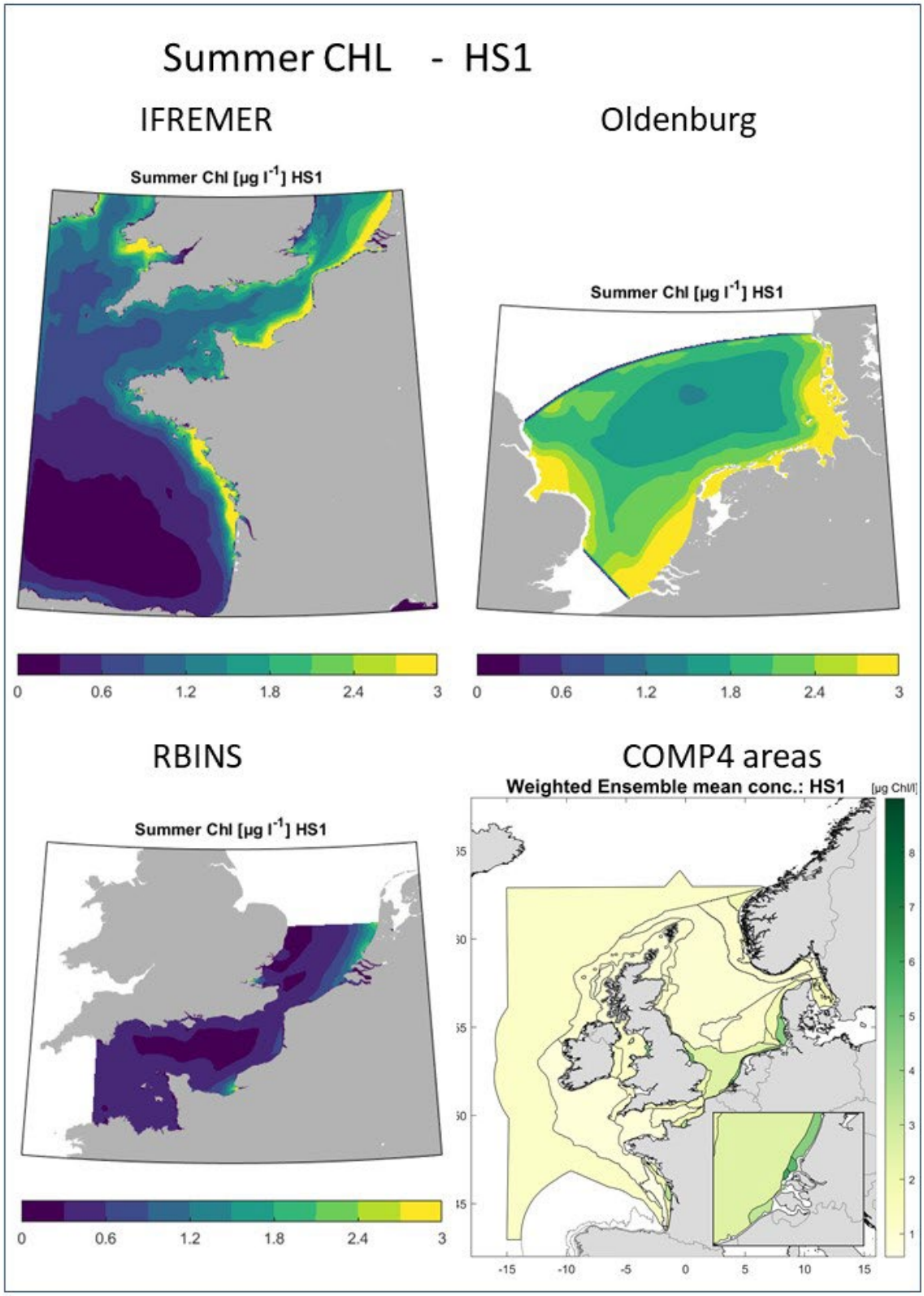
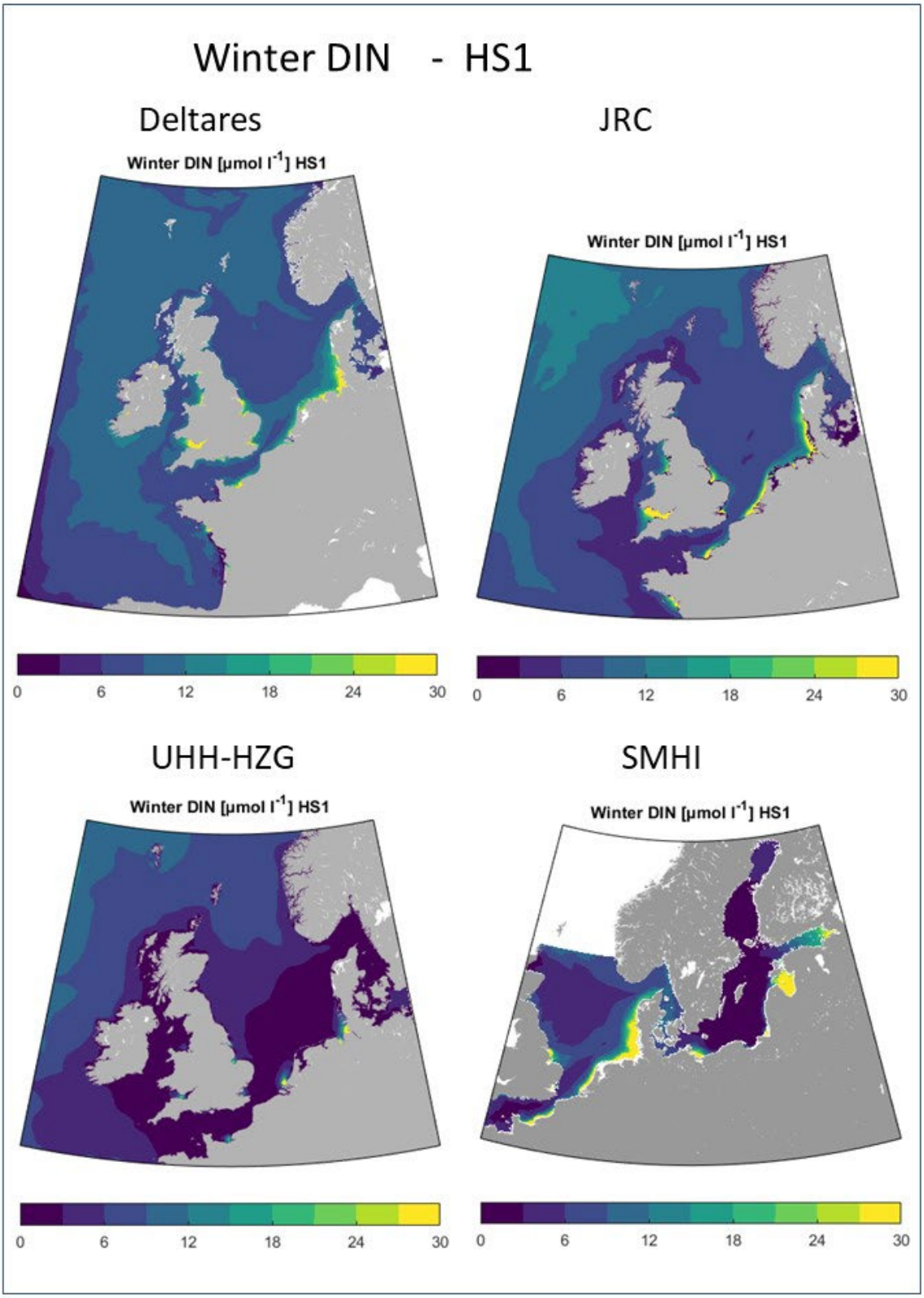


Figure 7: Chlorophyll-a concentration for the Historic Scenario 1 (HS1) simulation from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


Winter DIN - HS1
Deltares
Winter DIN [µmol l-1] HS1
0 6 12 18 24 30
JRC
Winter DIN [µmol l-1] HS1
0 6 12 18 24 30
UHH-HZG
Winter DIN [µmol l-1] HS1
0 6 12 18 24 30
SMHI
Winter DIN [µmol l-1] HS1
0 6 12 18 24 30

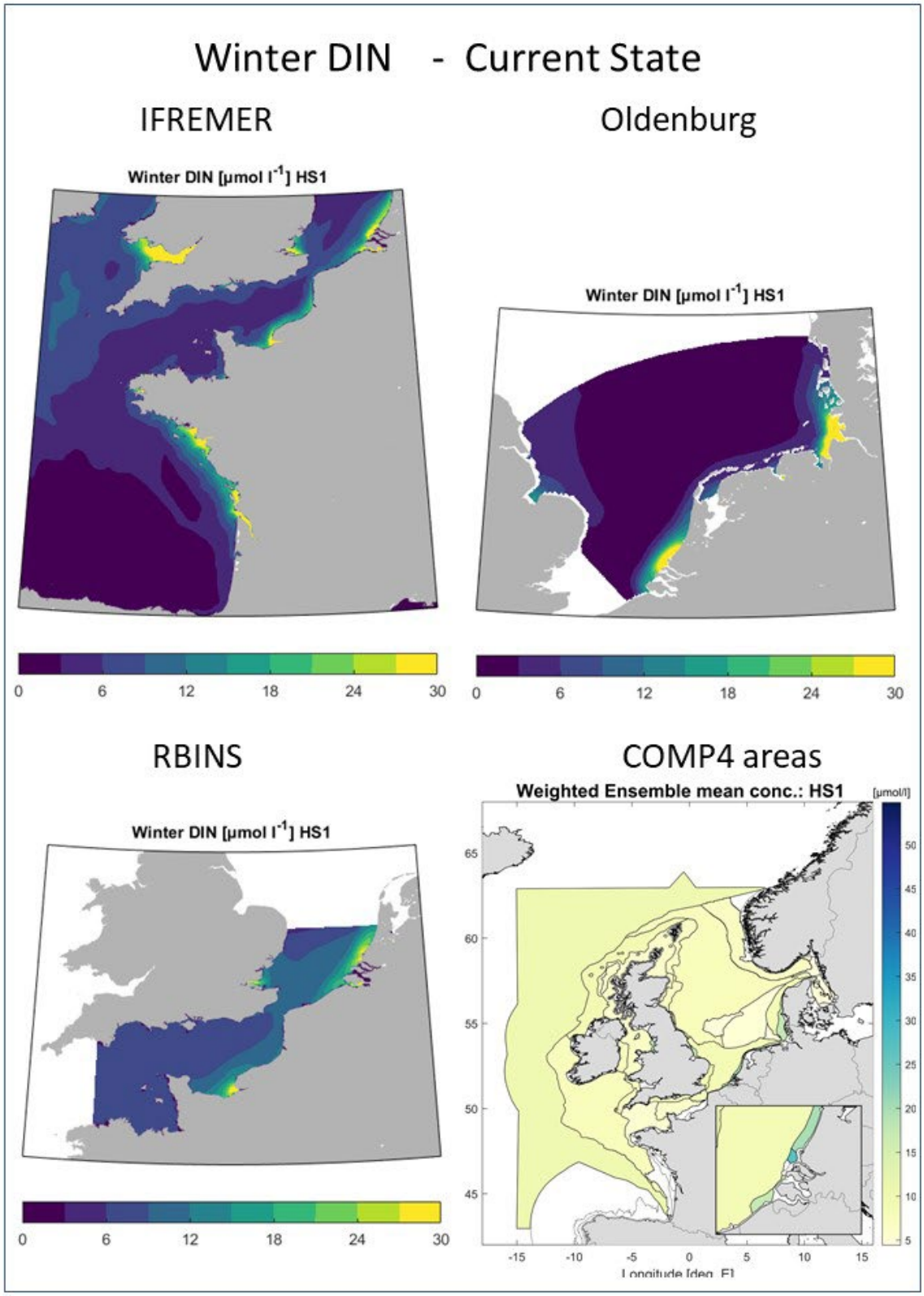
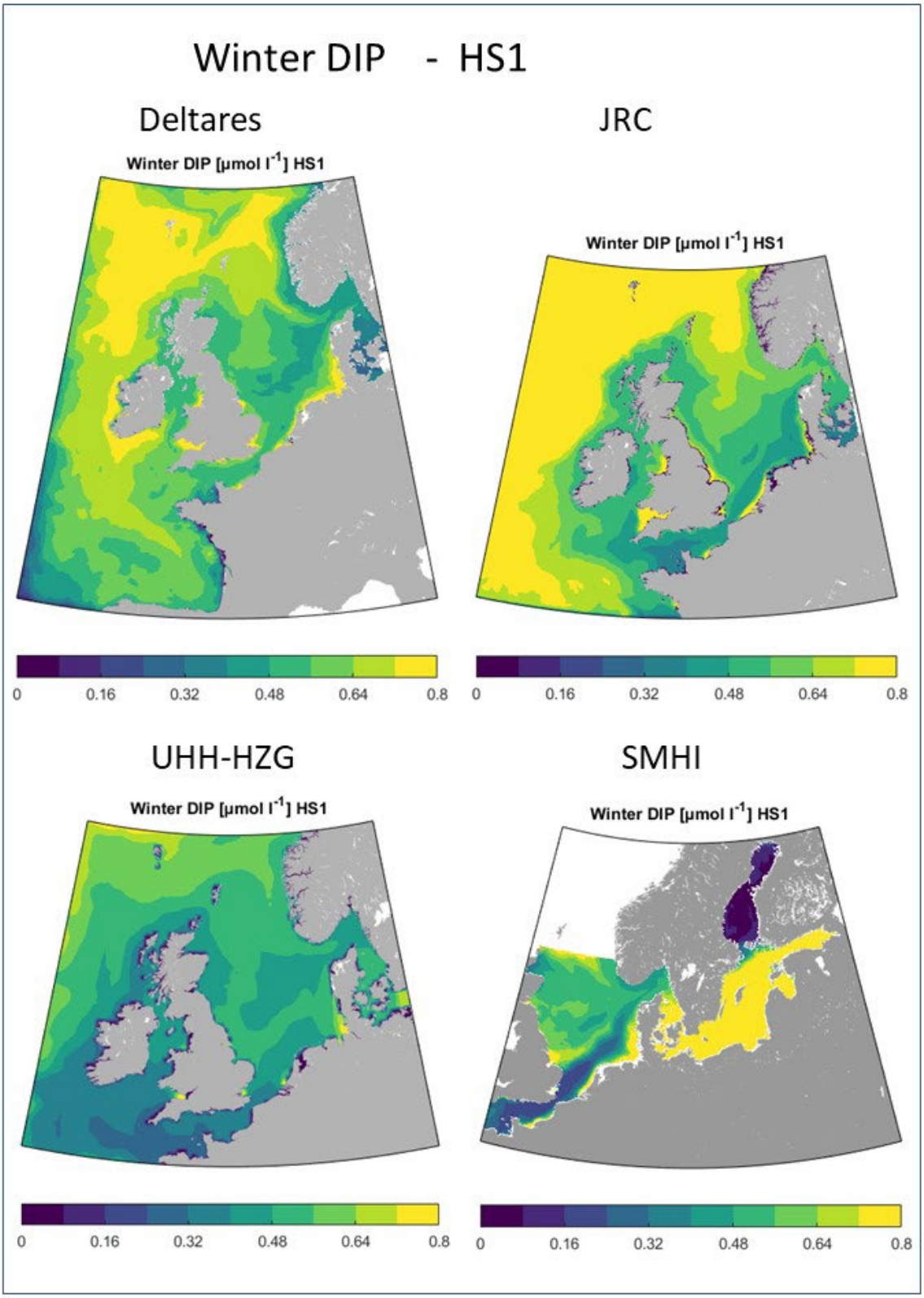


Figure 8: DIN concentration for the Historic Scenario 1 (HS1) simulation from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


Winter DIP - HS1
Deltares
Winter DIP [µmol l-1] HS1
0 0.16 0.32 0.48 0.64 0.8
JRC
Winter DIP [µmol l-1] HS1
0 0.16 0.32 0.48 0.64 0.8
UHH-HZG
Winter DIP [µmol l-1] HS1
0 0.16 0.32 0.48 0.64 0.8
SMHI
Winter DIP [µmol l-1] HS1
0 0.16 0.32 0.48 0.64 0.8

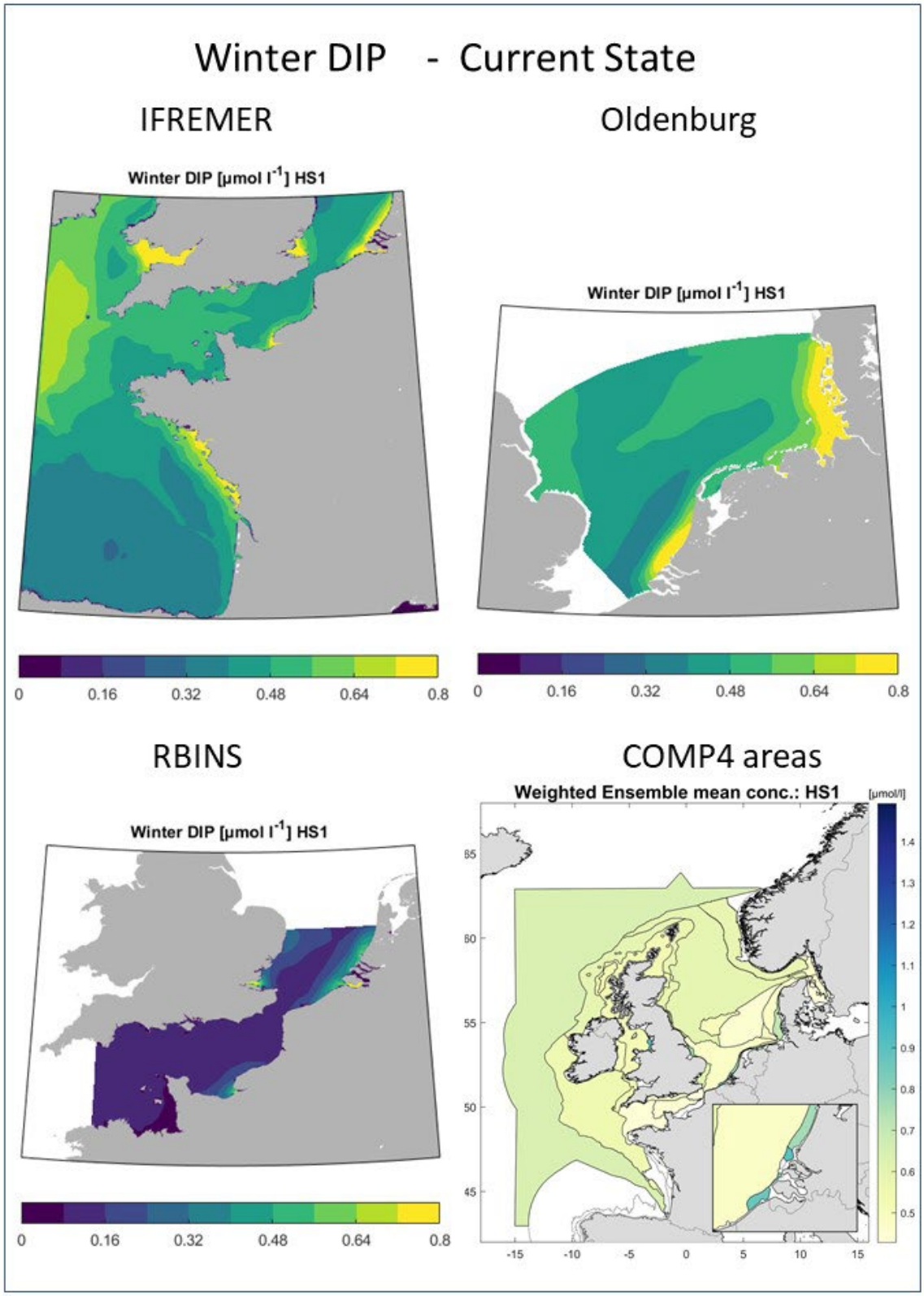
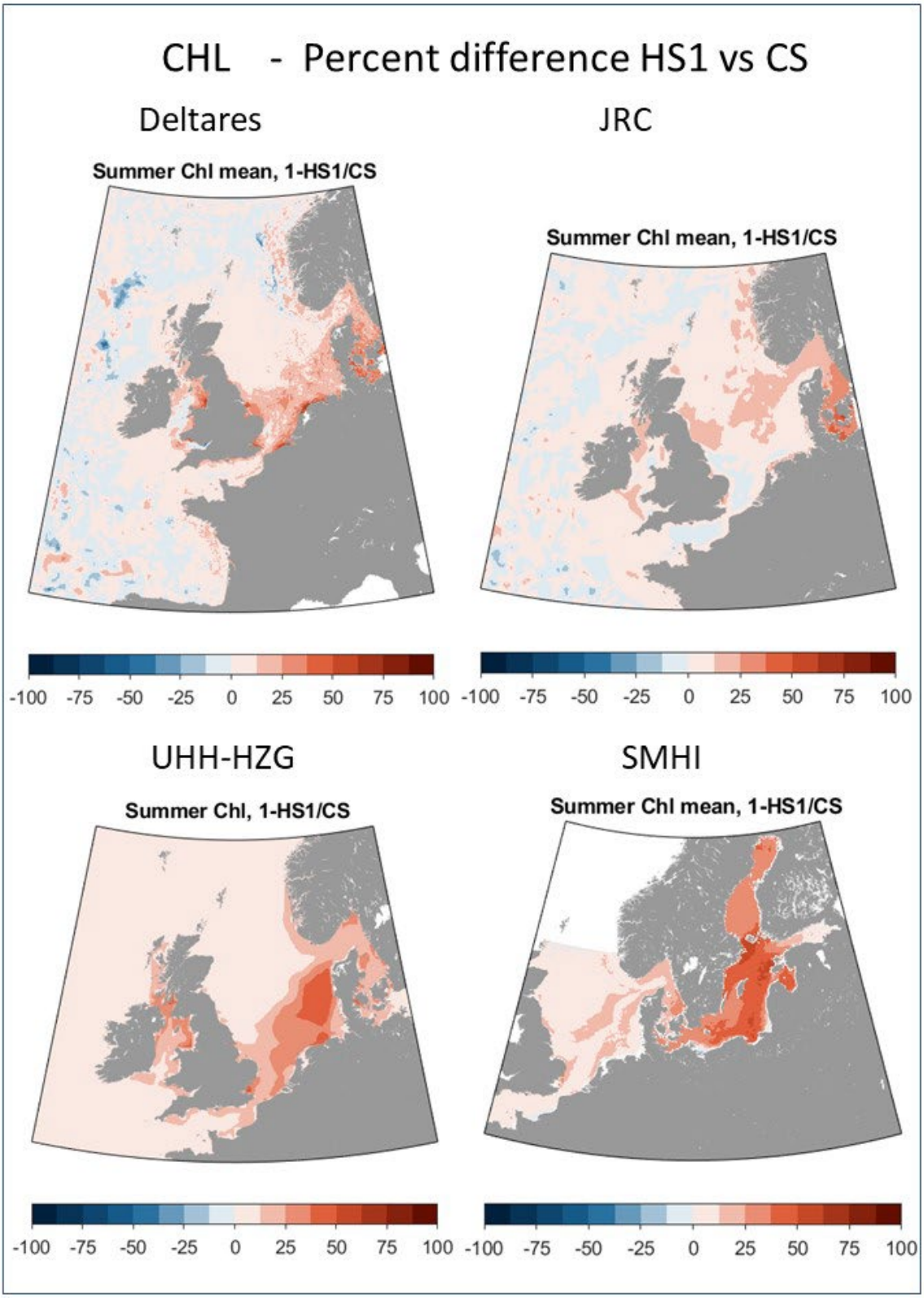


Figure 9: DIP concentration for the Historic Scenario 1 (HS1) simulation from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


CHL - Percent difference HS1 vs CS
Deltares
Summer Chl mean, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100
JRC
Summer Chl mean, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100
UHH-HZG
Summer Chl, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100
SMHI
Summer Chl mean, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100

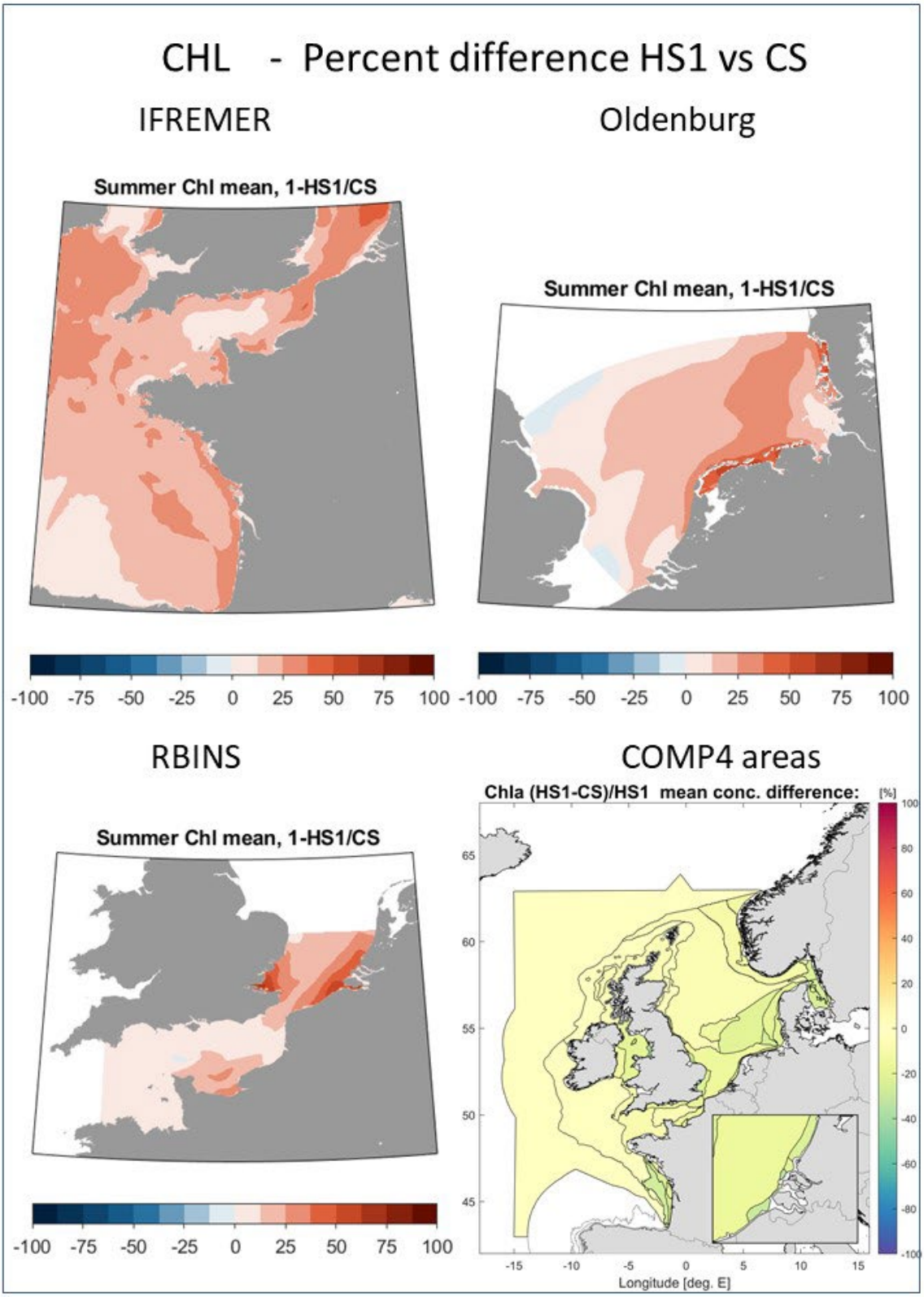
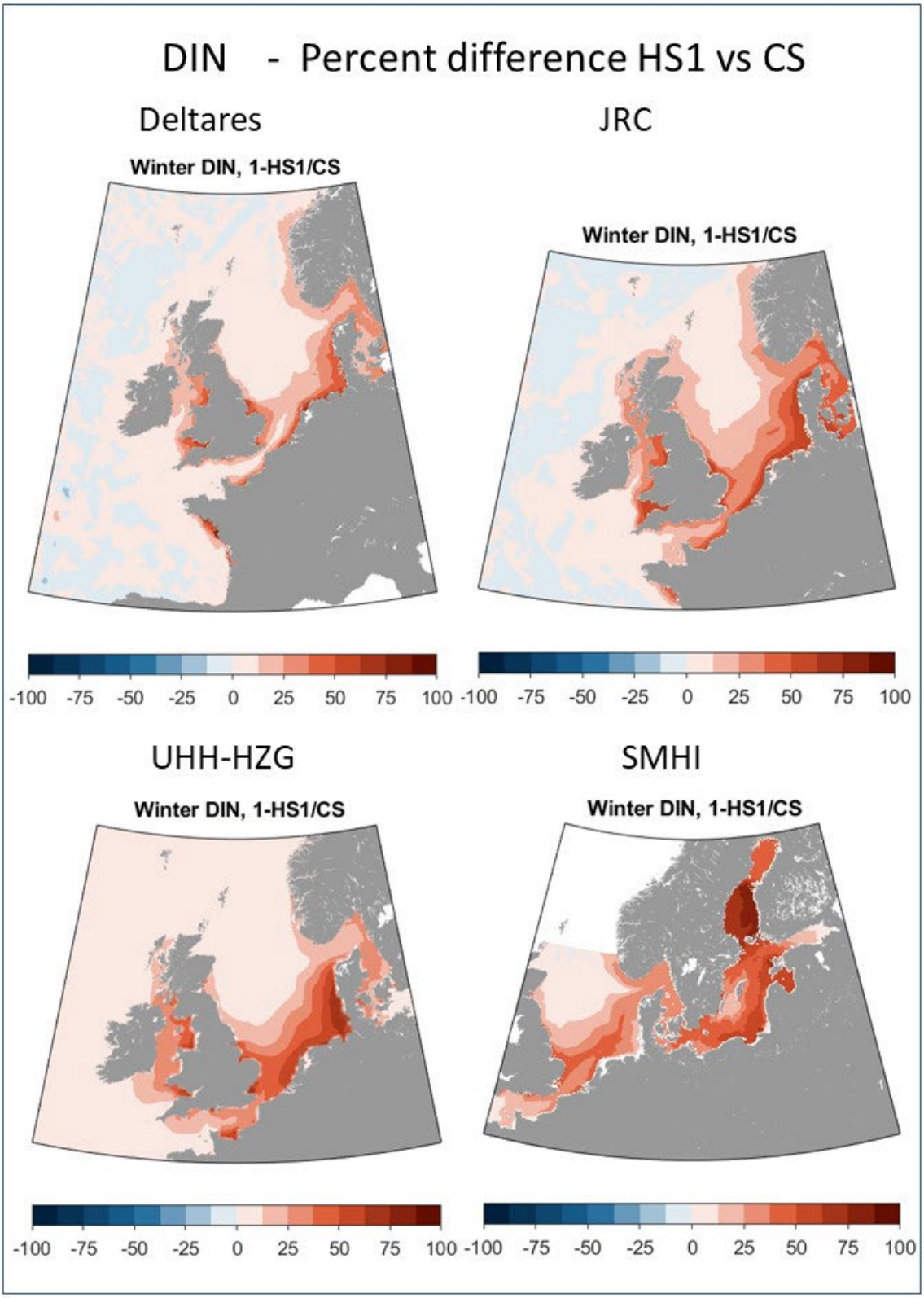


Figure 10: Percent difference between the HS1 and the CS simulation for Chlorophyll-a concentration from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


DIN - Percent difference HS1 vs CS
Deltares
Winter DIN, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100
JRC
Winter DIN, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100
UHH-HZG
Winter DIN, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100
SMHI
Winter DIN, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100

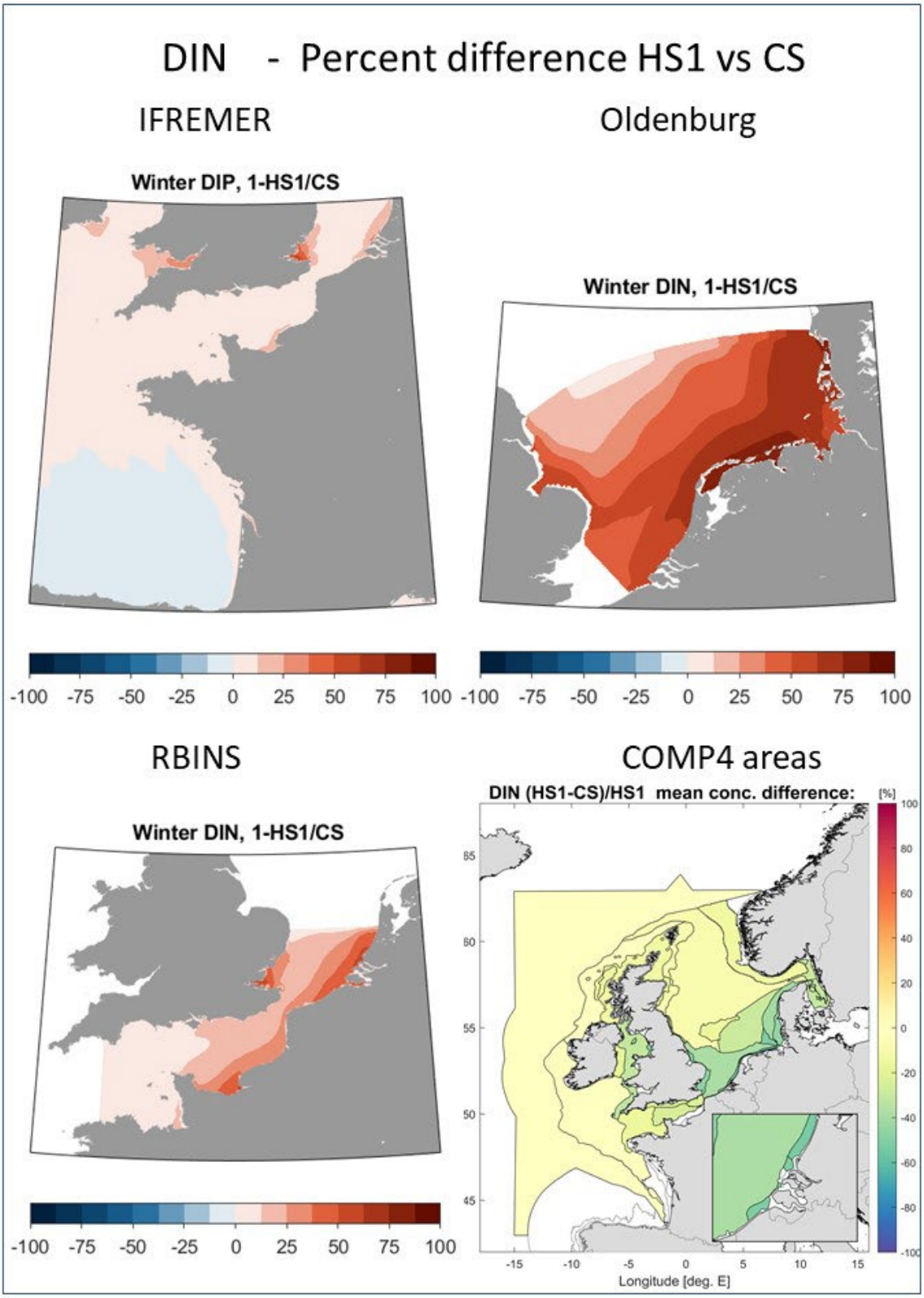
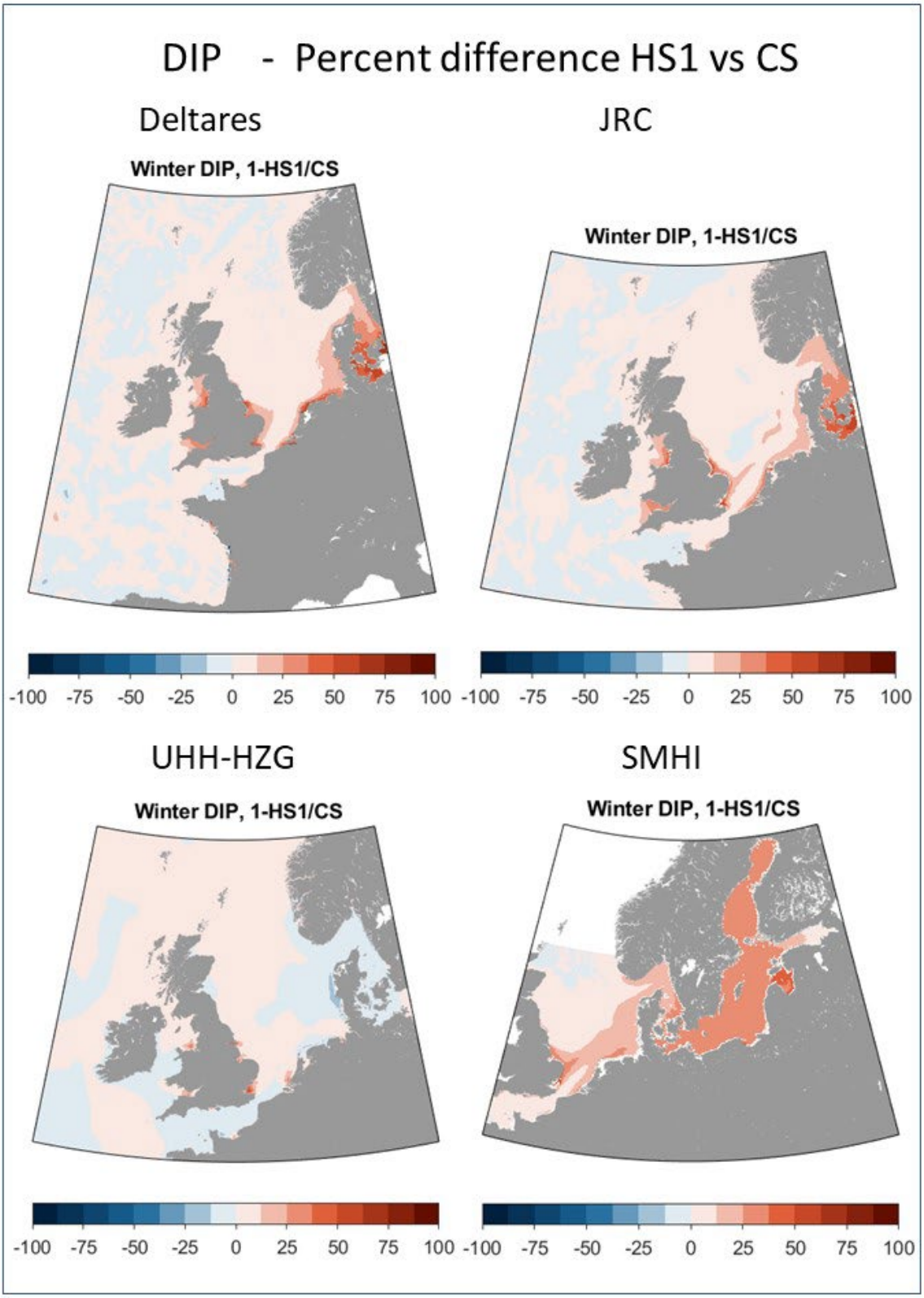


Figure 11: Percent difference between the HS1 and the CS simulation for DIN concentration from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


DIP - Percent difference HS1 vs CS
Deltares
Winter DIP, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100
JRC
Winter DIP, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100
UHH-HZG
Winter DIP, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100
SMHI
Winter DIP, 1-HS1/CS
-100 -75 -50 -25 0 25 50 75 100

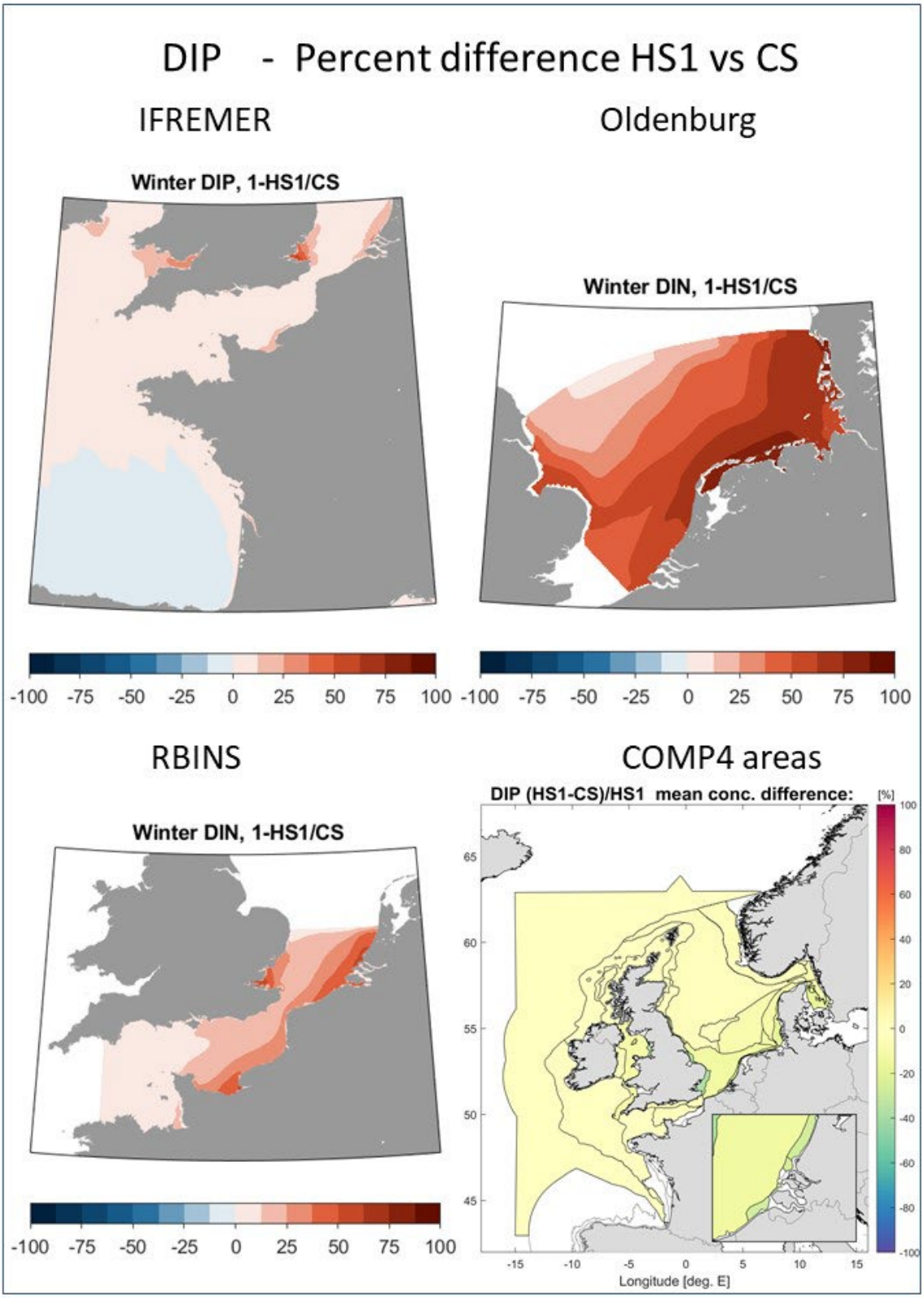
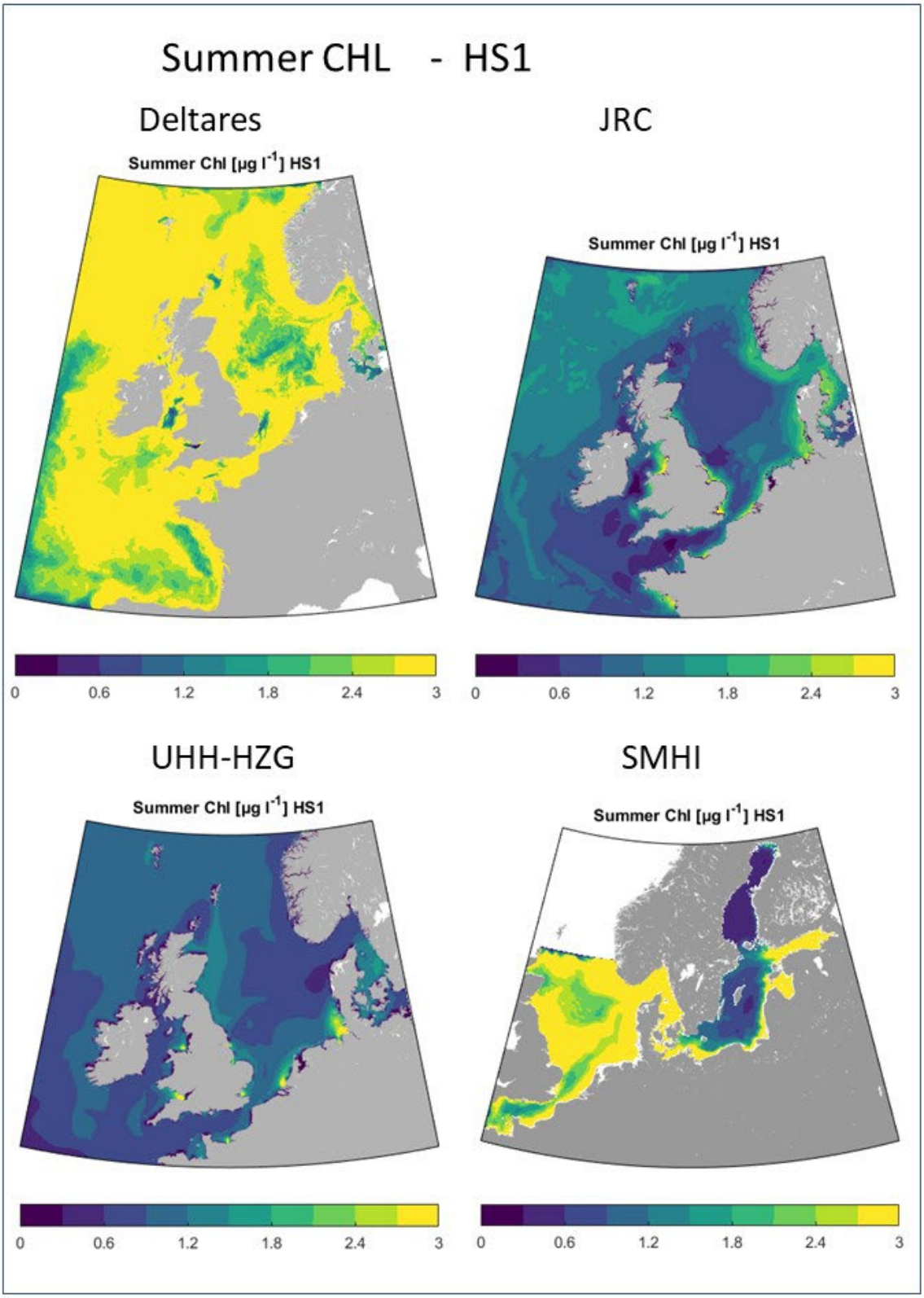


Figure 12: Percent difference between the HS1 and the CS simulation for DIP concentration from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


Summer CHL - HS1
Deltares
Summer Chl [µg l-1] HS1
0 0.6 1.2 1.8 2.4 3
JRC
Summer Chl [µg l-1] HS1
0 0.6 1.2 1.8 2.4 3
UHH-HZG
Summer Chl [µg l-1] HS1
0 0.6 1.2 1.8 2.4 3
SMHI
Summer Chl [µg l-1] HS1
0 0.6 1.2 1.8 2.4 3

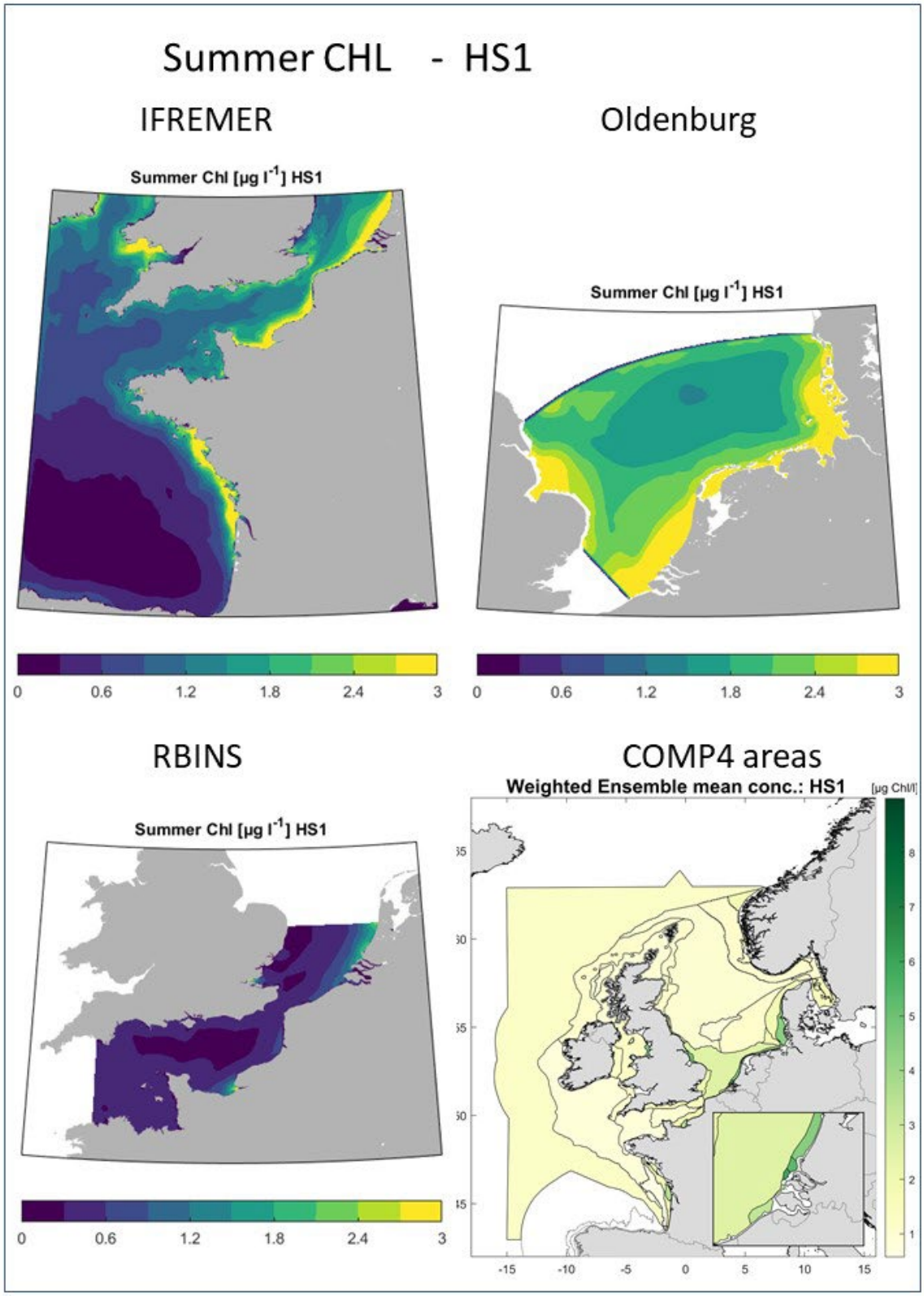
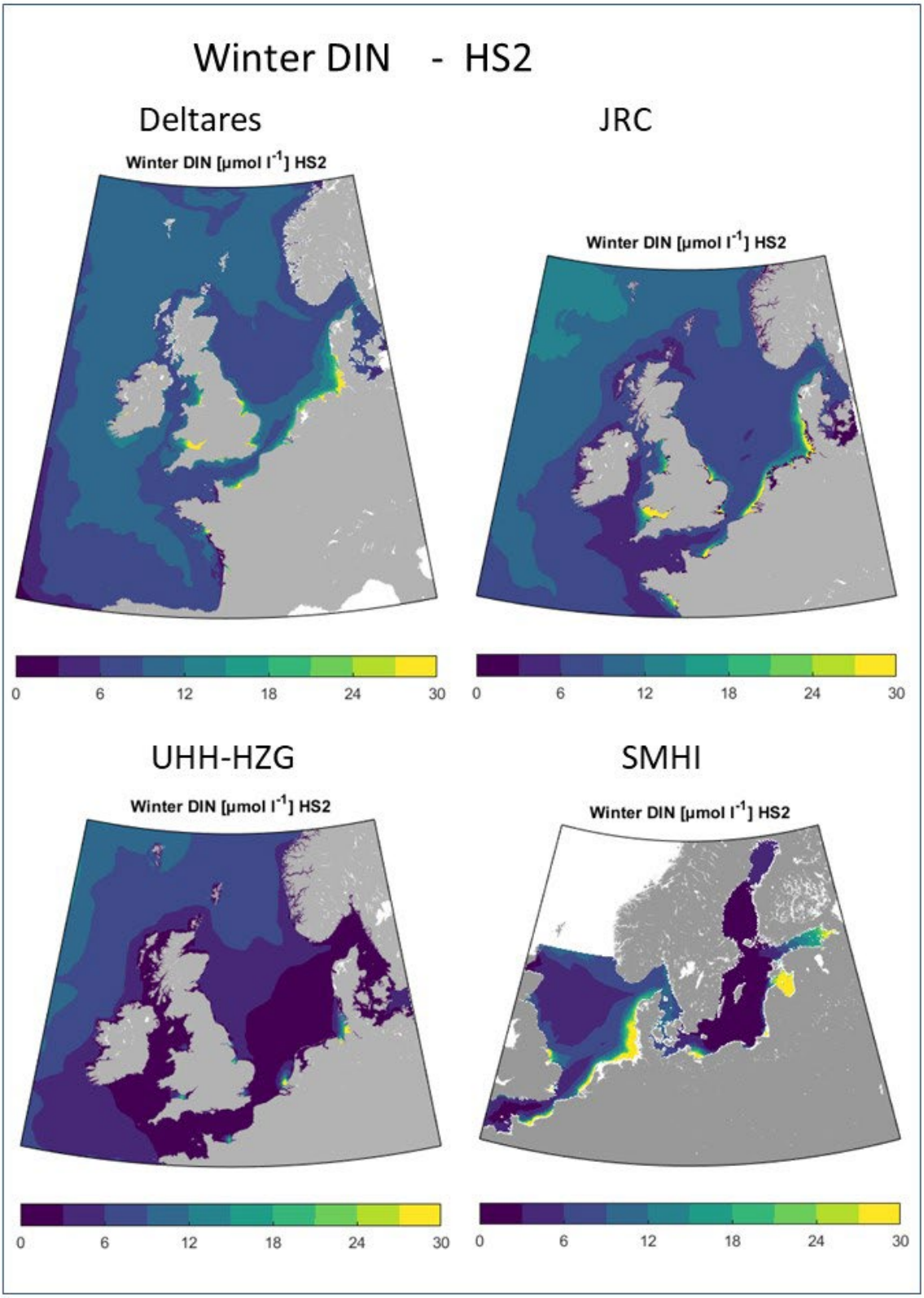


Figure 13: Chlorophyll-a concentration for the Historic Scenario 2 (HS2) simulation from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


Winter DIN - HS2
Deltares
Winter DIN [µmol l-1] HS2
0 6 12 18 24 30
JRC
Winter DIN [µmol l-1] HS2
0 6 12 18 24 30
UHH-HZG
Winter DIN [µmol l-1] HS2
0 6 12 18 24 30
SMHI
Winter DIN [µmol l-1] HS2
0 6 12 18 24 30

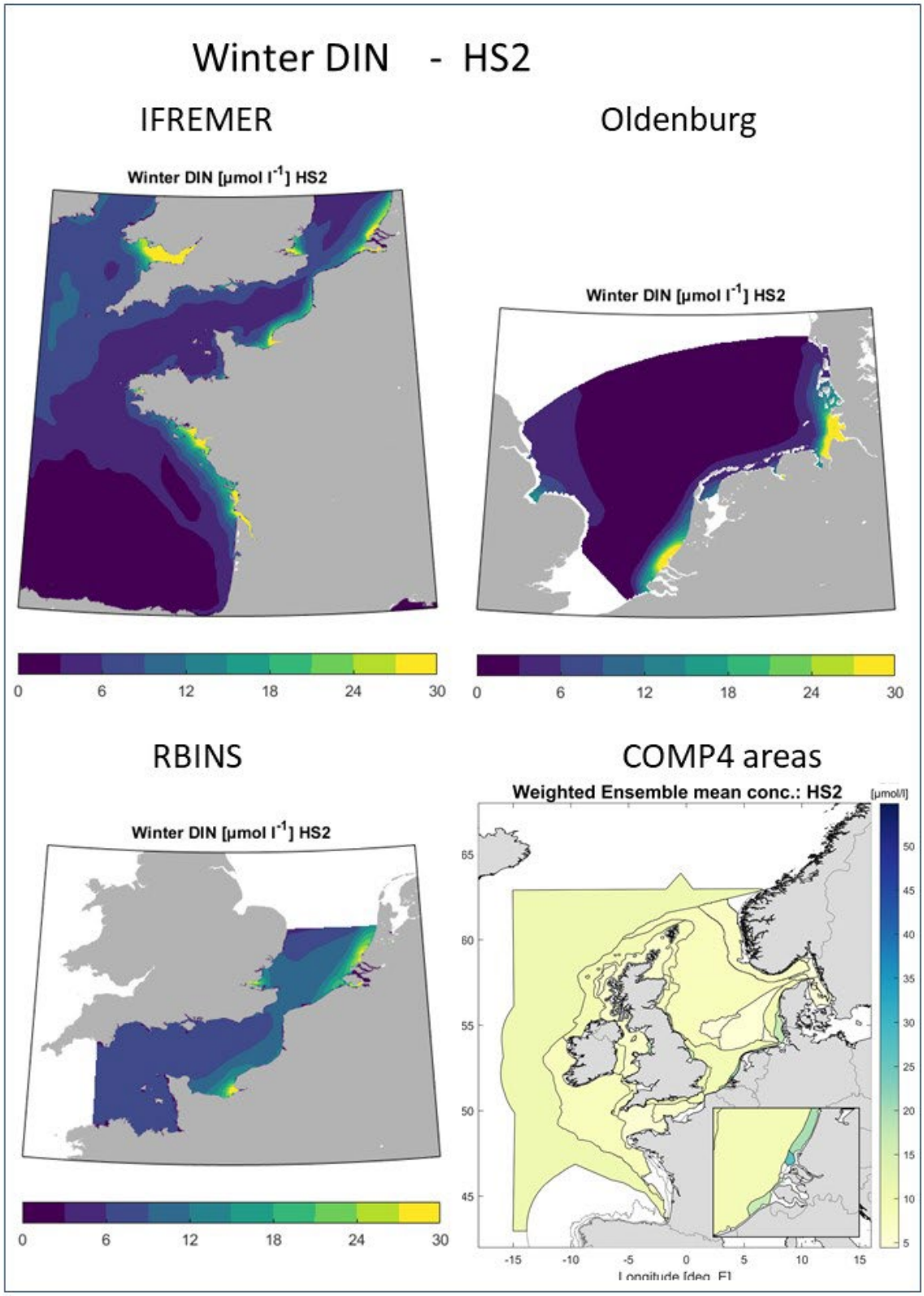
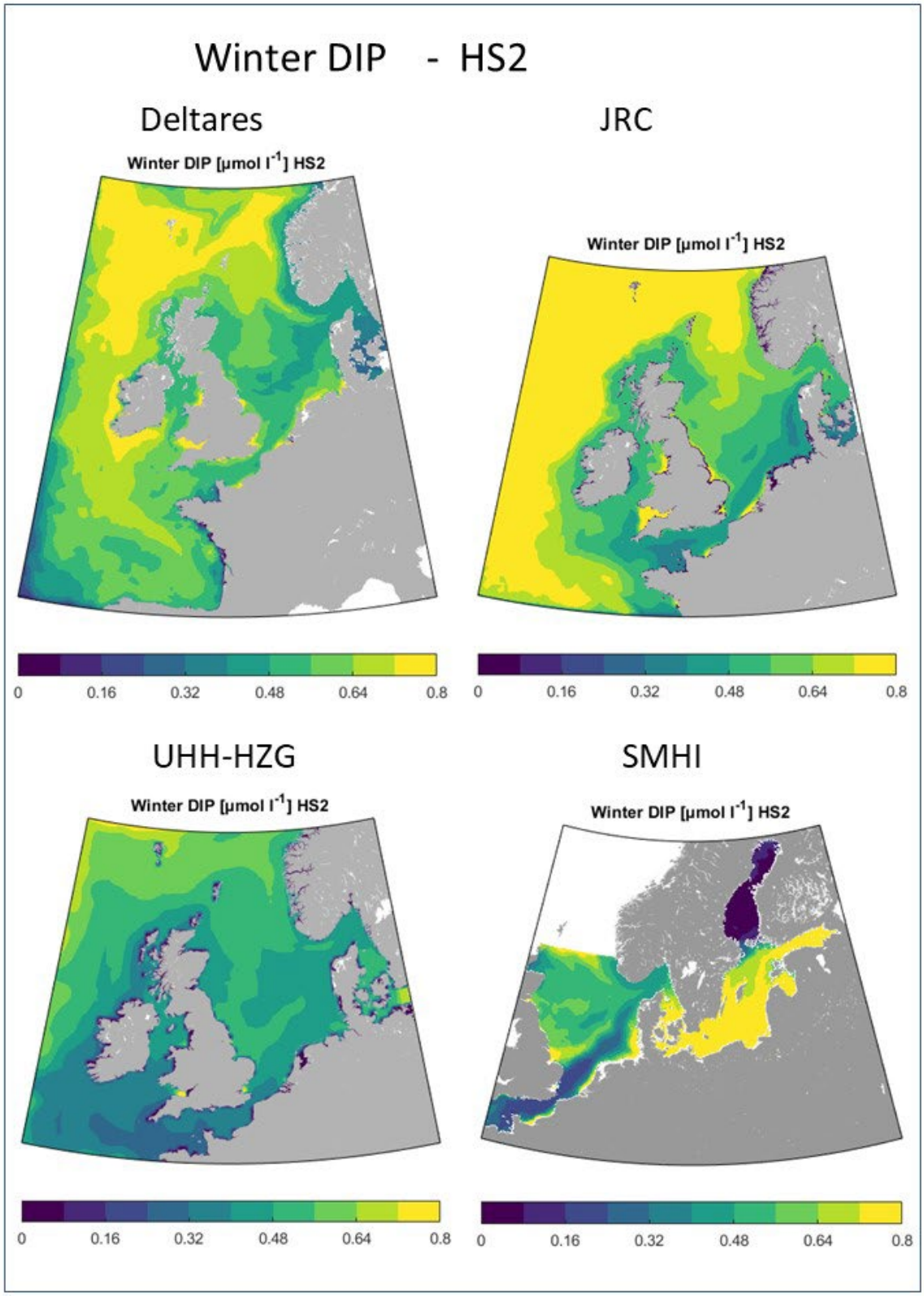


Figure 14: DIN concentration for the Historic Scenario 2 (HS2) simulation from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


Winter DIP - HS2
Deltares
Winter DIP [µmol l-1] HS2
0
0.16
0.32
0.48
0.64
0.8
JRC
Winter DIP [µmol l-1] HS2
0
0.16
0.32
0.48
0.64
0.8
UHH-HZG
Winter DIP [µmol l-1] HS2
0
0.16
0.32
0.48
0.64
0.8
SMHI
Winter DIP [µmol l-1] HS2
0
0.16
0.32
0.48
0.64
0.8

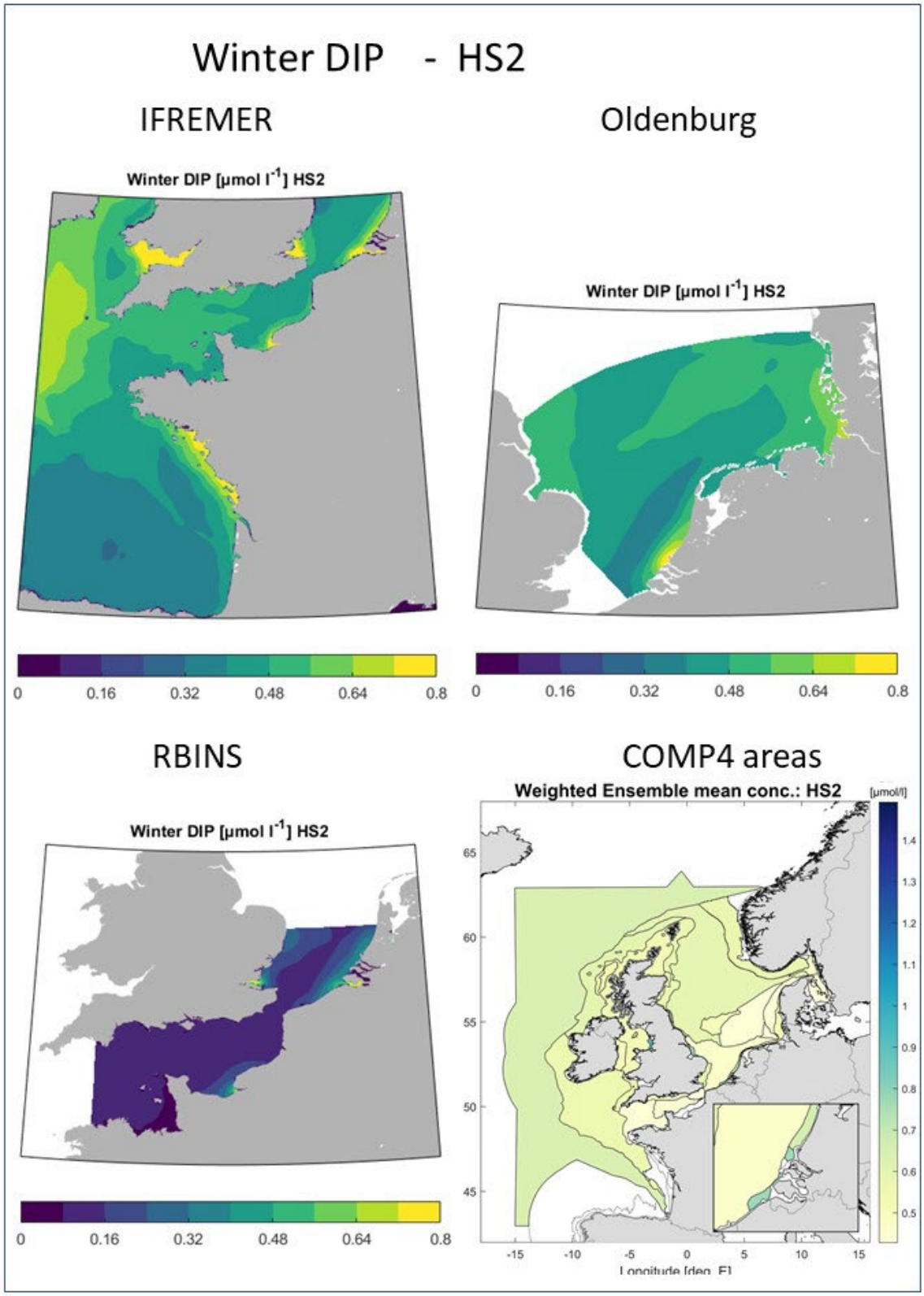
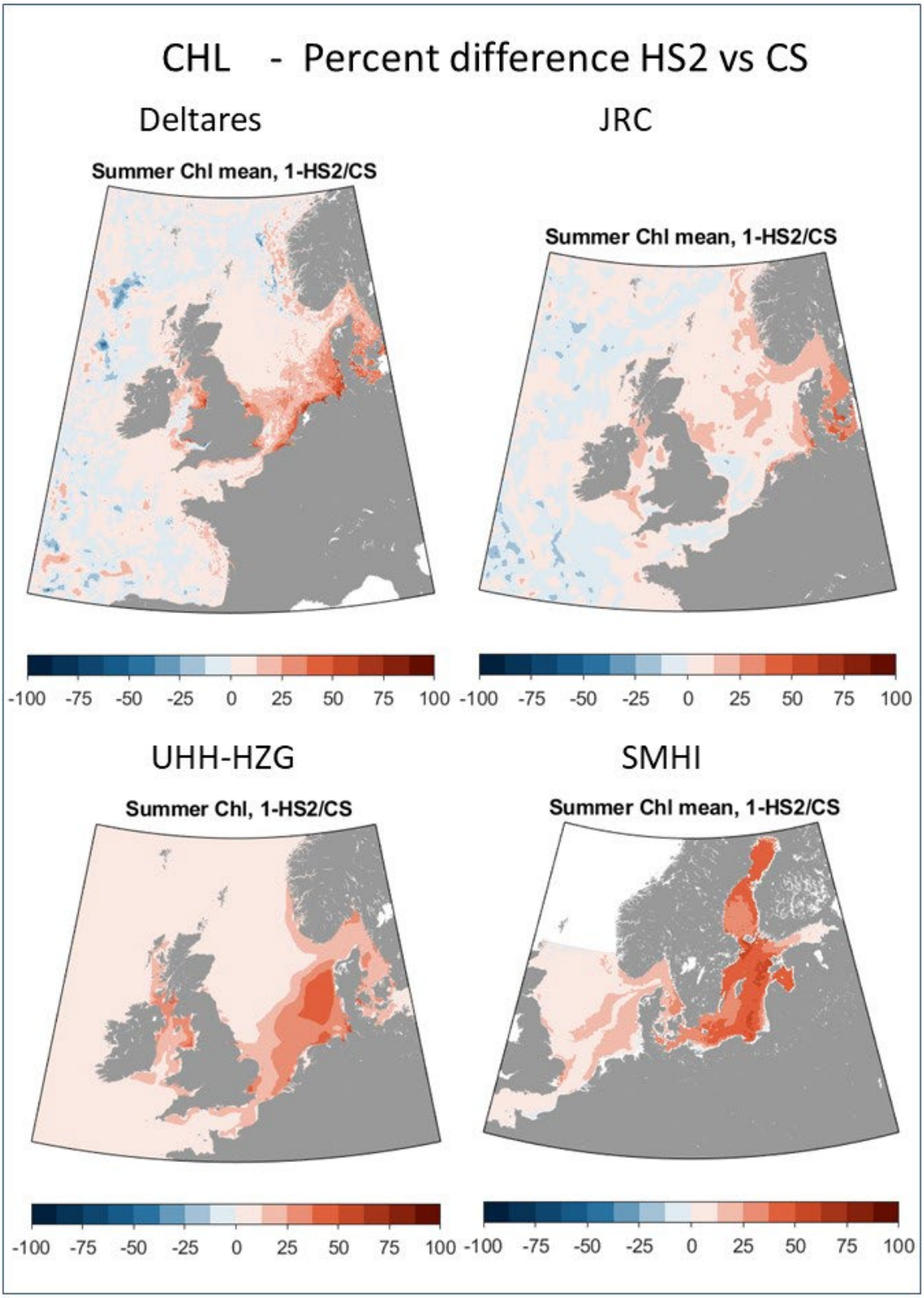


Figure 15: DIP concentration for the Historic Scenario 2 (HS2) simulation from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


CHL - Percent difference HS2 vs CS
Deltares
Summer Chl mean, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100
JRC
Summer Chl mean, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100
UHH-HZG
Summer Chl, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100
SMHI
Summer Chl mean, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100

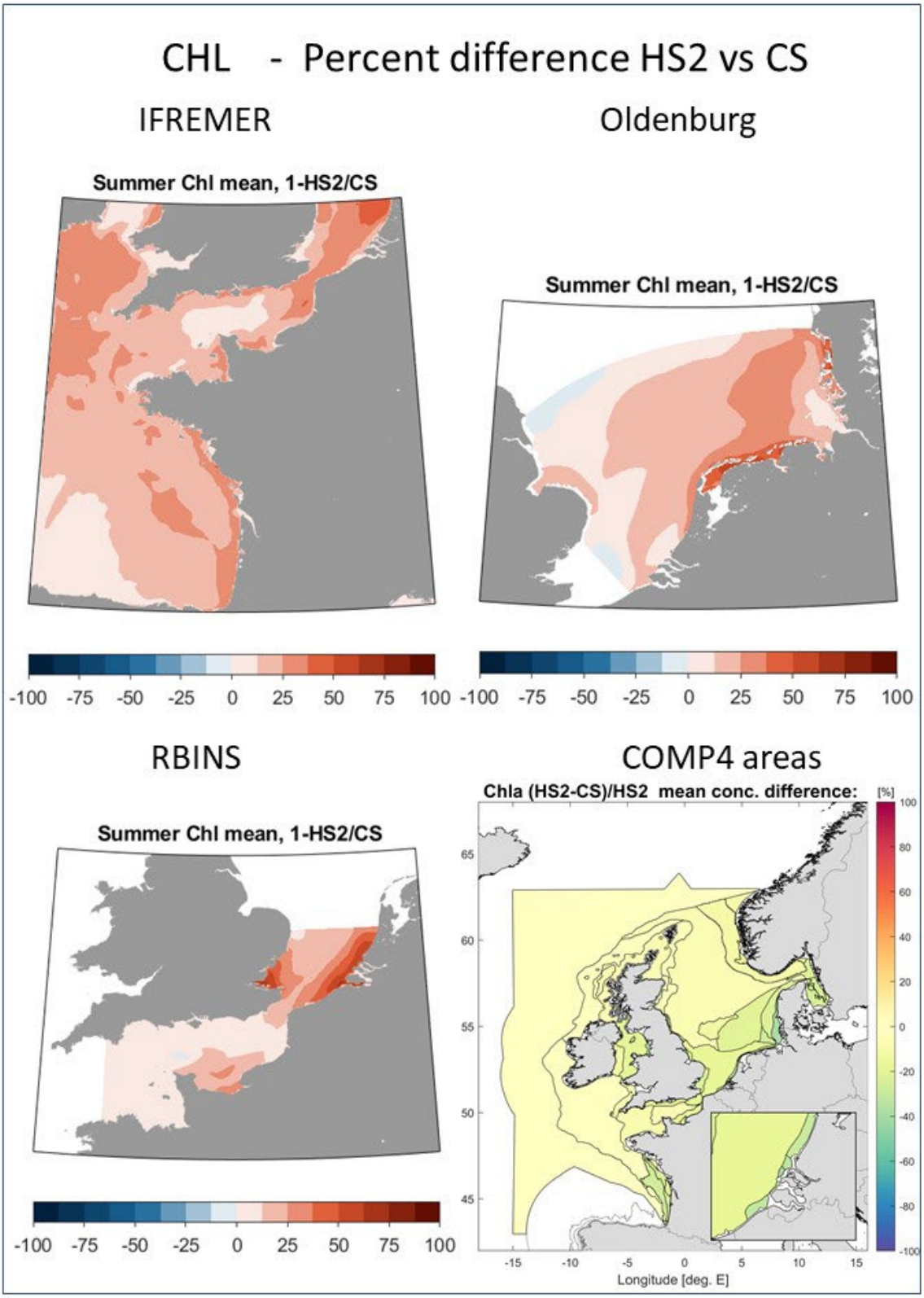
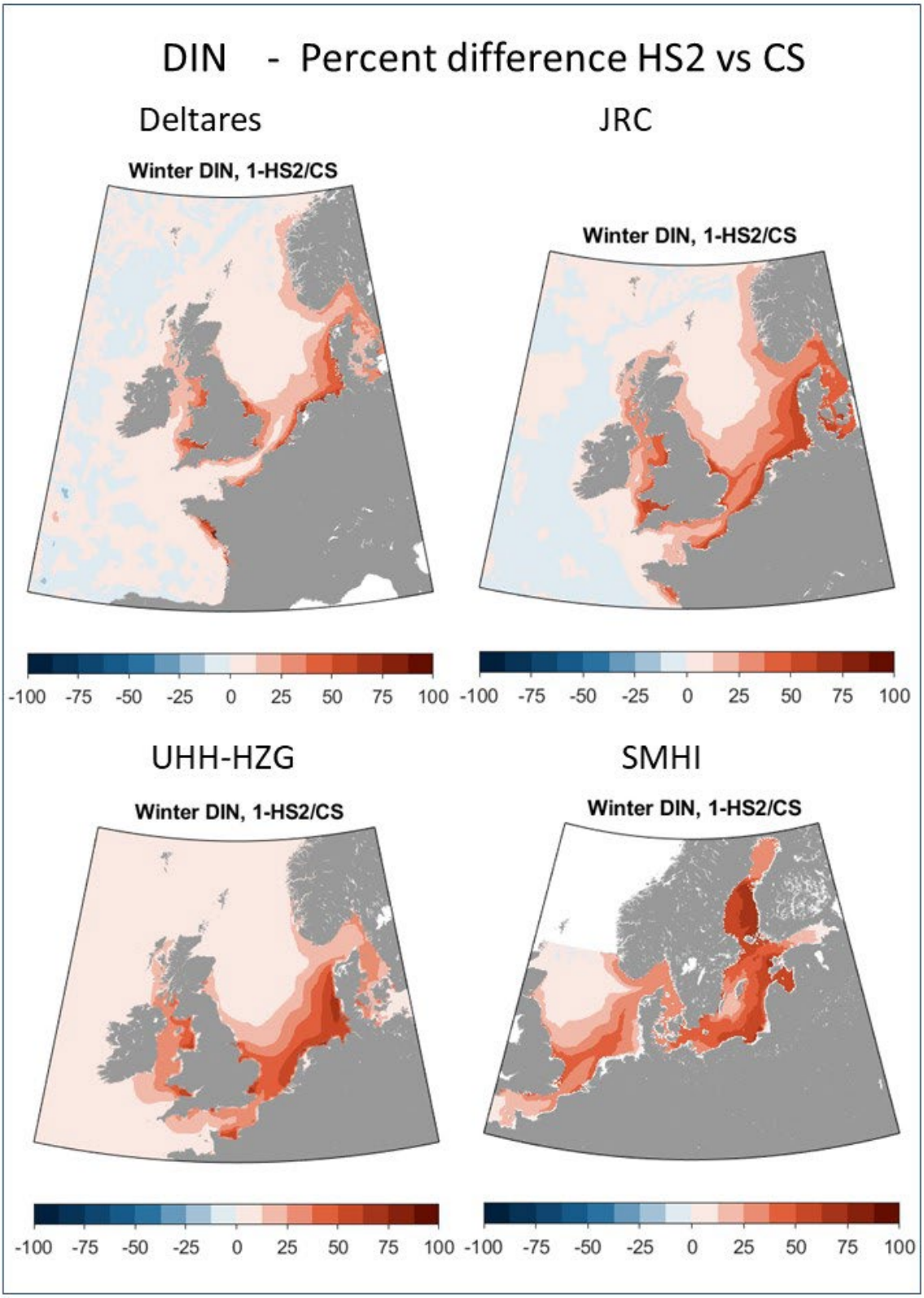


Figure 16: Percent difference between the HS2 and the CS simulation for Chlorophyll-a concentration from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


DIN - Percent difference HS2 vs CS
Deltares
Winter DIN, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100
JRC
Winter DIN, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100
UHH-HZG
Winter DIN, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100
SMHI
Winter DIN, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100

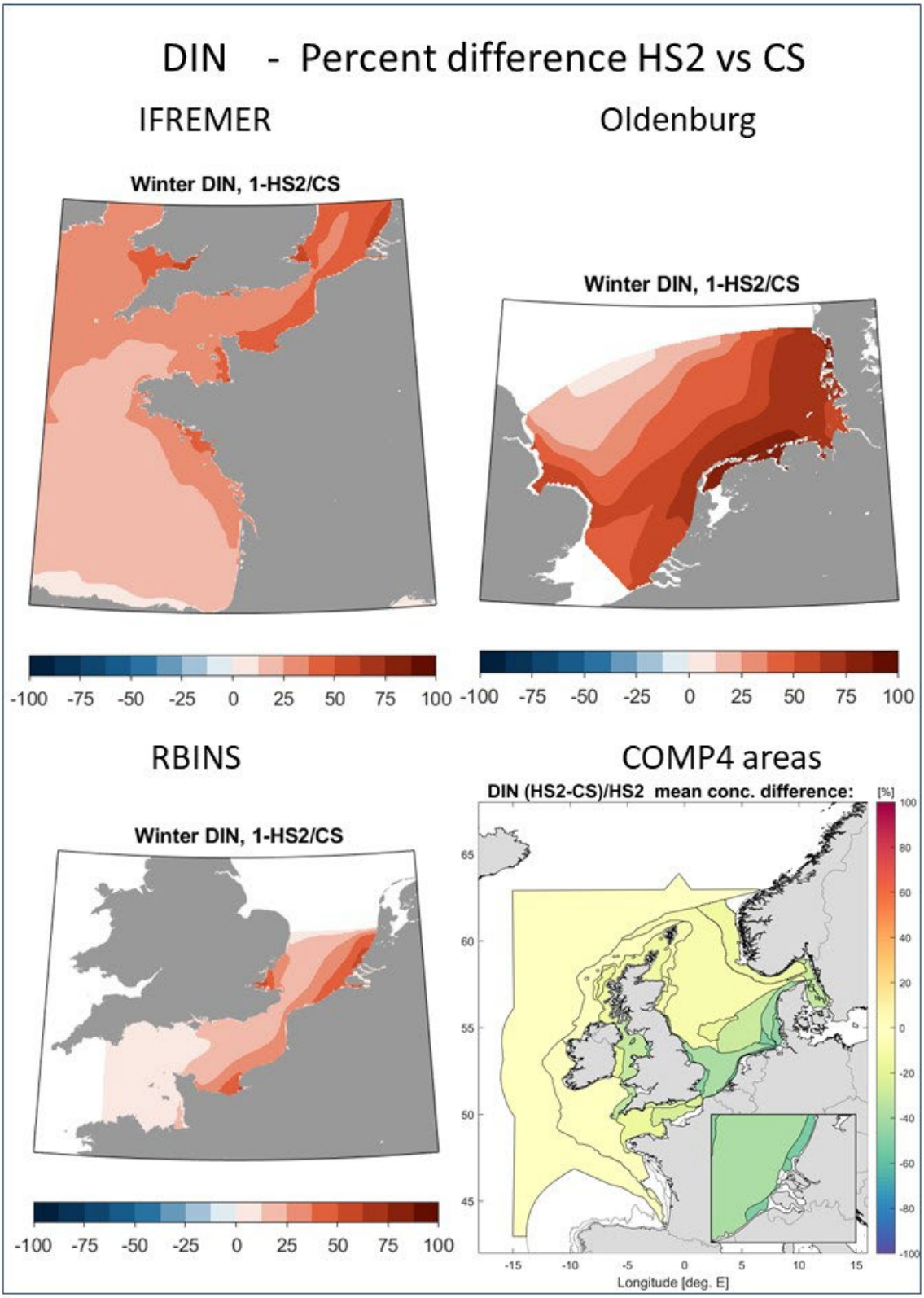
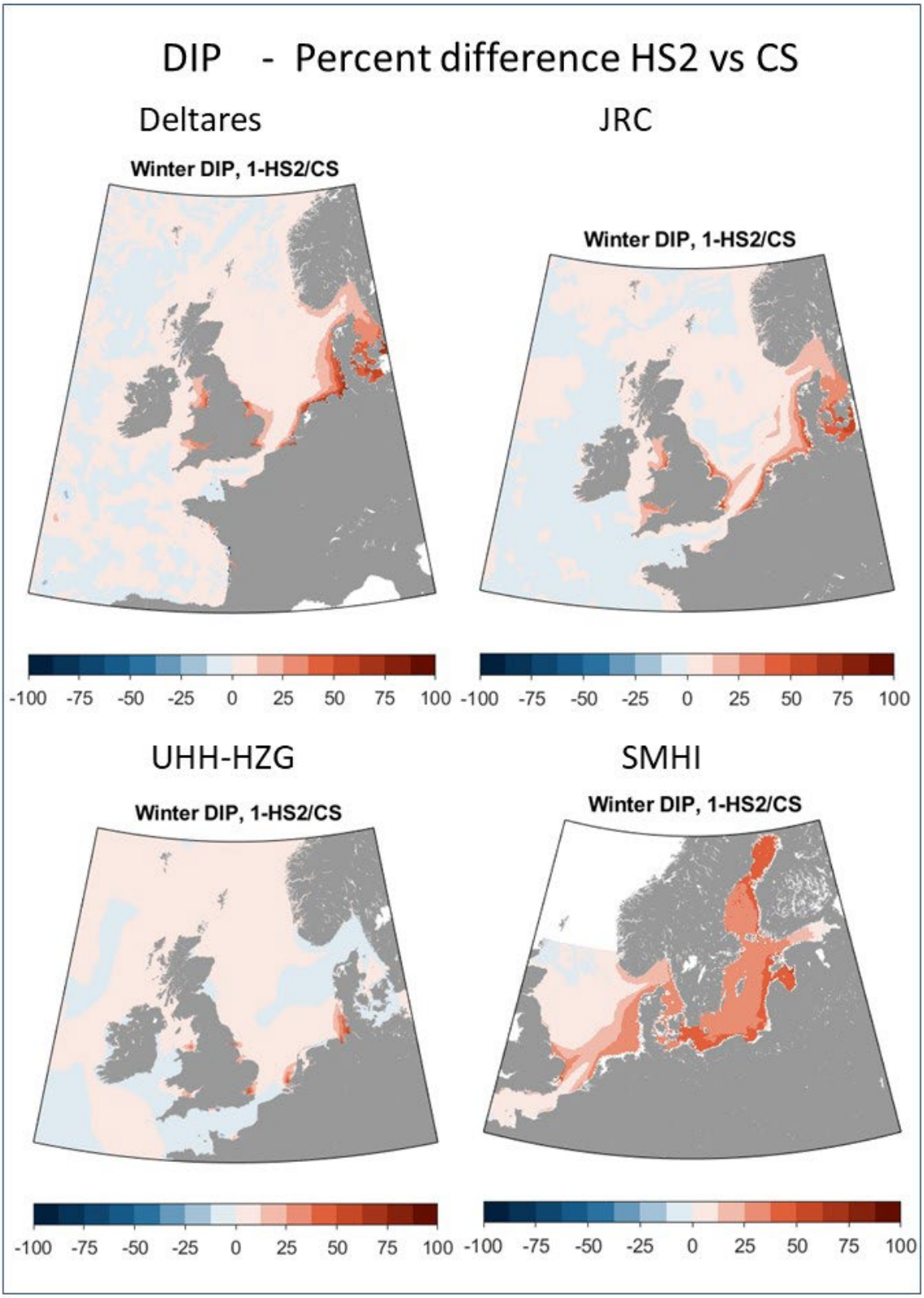


Figure 17: Percent difference between the HS2 and the CS simulation for DIN concentration from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).


DIP - Percent difference HS2 vs CS
Deltares
Winter DIP, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100
JRC
Winter DIP, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100
UHH-HZG
Winter DIP, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100
SMHI
Winter DIP, 1-HS2/CS
-100 -75 -50 -25 0 25 50 75 100

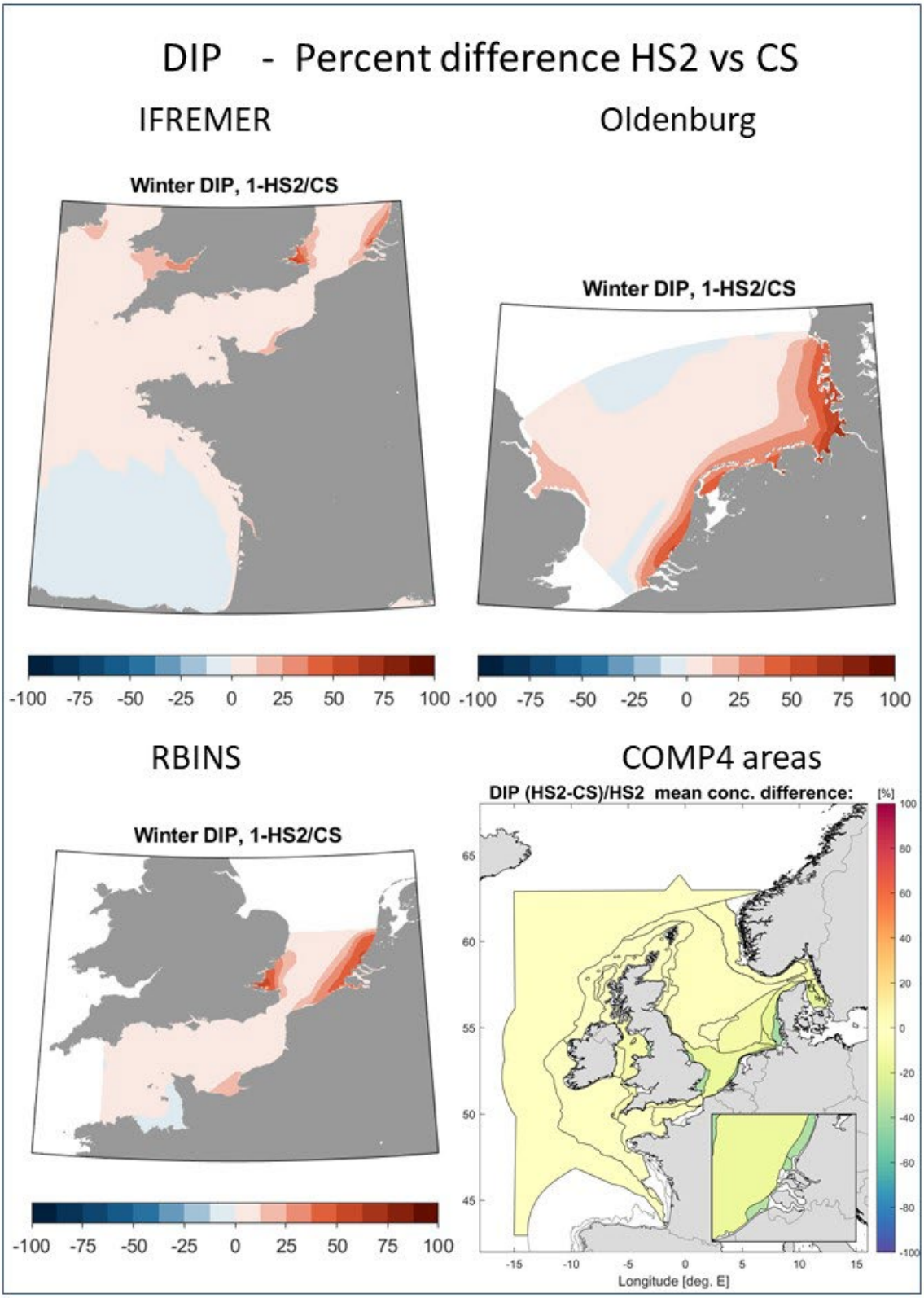


Figure 18: Percent difference between the HS2 and the CS simulation for DIP concentration from all models and the resulting weighted ensemble representation for the COMP4 assessment areas (different scale).

Reference:

Bentzon-Tilia, M., Traving, S. J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J. L., Markager, S., & Riemann, L. (2015). Significant N 2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. The ISME journal, 9(2), 273-285.

Große, F., Greenwood, N., Kreus, M., Lenhart, H.J., Machoczek, D., Pätsch, J., Salt, L.A., Thomas, H., 2016. Looking beyond stratification: a model-based analysis of the biological drivers ofoxygen deficiency in the North Sea. Biogeosciences 13, 2511– 2535. https://doi.org/10.5194/bgd-12-12543-2015.

Lee, ZP., Shang, S., Hu, C., Du, K., Weidemann, A., Hou, W., Lin, J., Lin, G.,2015. Secchi disk depth: A new theory and mechanistic model for underwater visibility. Remote Sensing of Environment Vol. 169, 139-149. https://doi.org/10.1016/j.rse.2015.08.002

Schöpp, W., Posch, M., Mylona, S., and Johansson, M. (2003). Longterm development of acid deposition (1880–2030) in sensitive freshwater regions in Europe. Hydrol. Earth Syst. Sci. 7, 436–446. doi: 10.5194/hess-7- 436-2003

Topcu, D., Behrendt, H., Brockmann, U., Claussen, U. (2011). Natural background concentrations of nutrients in the German Bight area (North Sea). Environ Monit Assess 174:36

ICG-EMO Report Annex 2 – Pre-eutrophic scenario definition

Rationale and description of a pre-eutrophication nutrient scenario to derive harmonised OSPAR assessment levels for eutrophication parameters in Region II

This report was produced by the pre-eutrophic expert group under the lead from Hermann Lenhart (Convenor of ICG-EMO) with the contribution from Philip Axe (Chair HASEC), Anouk Blauw (NL), Lisette Enserink (NL), Eivind Farmsen (NO), Liam Fernand (UK), Gunn Lise Hugstol (NO), Wera Leujak (DE), Stiig Markager (DK), Julian Mönnich (DE), Lasse Tor Nielsen (DK), Sonja van Leeuwen (NL), Theo Prins (NL), Itzel Baron Ruvalcaba (SE)

1) Background

Currently, eutrophication is assessed in OSPAR based on national assessment levels that are neither consistent nor harmonised between Contracting Parties. This has prevented the modelling community from setting nutrient reduction targets as agreed in the North-East Atlantic Environment Strategy 2010. To overcome this, a common approach is needed to derive regionally coherent assessment levels.

In principle, different approaches exist for deriving assessment levels. An approach successfully used in the Baltic Sea is the investigation of time series of eutrophication parameters to identify break points that mark the onset of significant eutrophication effects. One pre-requisite of using this approach is that time series reach back far enough to cover pre-eutrophic times. In the Greater North Sea (OSPAR Region II) such time series are sparse and therefore they cannot be used to derive assessment levels. Another approach that is suggested by the ECOSTAT working group is the derivation of nutrient assessment levels from correlations with biological quality elements, primarily phytoplankton. This approach is also not applicable in OSPAR, since the assessment levels of chlorophyll-a are not harmonised between Contracting Parties and can therefore not be used as a starting point to derive assessment levels for other eutrophication parameters.

An alternative approach is the use of a reference nutrient input scenario that can be derived by using a catchment model and that constitutes the basis for deriving reference conditions for eutrophication parameters in the North Sea. Assessment levels can be obtained by adding an allowable deviation of max. 50% to the reference conditions, which reflects the common practice under OSPAR and the Water Framework Directive (WFD). Denmark indicates this approach as too simplistic and argues that this is not in line with the WFD.

Table 1: Overview of definitions of "natural reference conditions" by countries around the North Sea (source: Blauw et al. 2019). These definitions have been used for the WFD and/or for deriving national assessment levels for eutrophication parameters in OSPAR.

Country	Choice of reference
Norway	Expert judgement and data from 1950s
Sweden	1930

Country	Choice of reference
Denmark	1900 (WFD)
Germany	1880
Netherlands	1930
Belgium	50% of loads in 1950
France	Geographical reference to relatively unimpacted sites
England & Scotland	Geographical reference: offshore waters

Natural reference conditions have been defined in different ways by different OSPAR Contracting Parties bordering the North Sea, either for use in the WFD, Marine Strategy Framework Directive (MSFD) or OSPAR. The JMP EUNOSAT project has reviewed these approaches and has compiled the overview shown in Table 1. From Table 1 it is evident that a wide range of approaches exist among OSPAR Contracting Parties that lack comparability. These differences gave rise to national assessment levels used in OSPAR to assess eutrophication that lack regional harmonisation.

According to the WFD "reference conditions" are "a description of the biological quality elements that exist or would exist …. with no or very minor disturbance from human activities" (WFD CIS Guidance Document No.5). This could also be the rationale for a reference nutrient input scenario that delivers reference conditions for nutrient concentrations in OSPAR. However, recent work by ECOSTAT has shown that the nutrient boundaries set for surface waters under the WFD lack comparability within a region. At the ICG-EMO workshop September 2019, some Contracting Parties informed the meeting that they had defined WFD Reference levels in order to be consistent with OSPAR Background Levels, which are now known to be regionally inconsistent as well. Therefore, using nutrient inputs under WFD reference conditions would not provide a regionally harmonised approach and a different rationale needs to be found.

The JMP EUNOSAT project has chosen a reference time period for the derivation of assessment levels that is characterised by little influence of anthropogenic activities. It presents a period before industrialisation and agricultural intensification and before the establishment of the Haber-Bosch process, that industrialised the production of inorganic nitrogen fertilisers (first demonstrated in 1909 with a first industrial-level production starting in 1913). Furthermore, a time period has been selected for which strong evidence exists that water transparency and macrophyte coverage were still high in the coastal waters of the German Bight, two major quality indicators used for describing eutrophication effects. For this pre-eutrophic time period, reference nutrient riverine inputs were derived using the pan-European catchment model E-HYPE, developed and maintained by SMHI. In addition to the definition of the pre-eutrophic river load estimates also the atmospheric N deposition and the boundary condition for the Baltic outflow need to be adopted for the historic scenario. The details to the later one can be found in Annex 11, while the focus of this reports is the definition for the historic scenario for the riverine input.

2) Results of the E-HYPE historic model run

A description and the assumptions of the E-HYPE model are provided in Annex 01. To validate the E-HYPE model v.3.1.3. a comprehensive model performance validation was undertaken based on today's river discharges and river concentrations taking into account the years 2009 – 2014 (for details see Blauw et al., 2019). The validation of the E-HYPE data showed that there are, in some cases, large inconsistencies with river loads estimated by OSPAR-RID. Therefore, a country-specific scaling factor was applied when using the

E-HYPE data in the JMP EUNOSAT project, so the total present river loads per country in the marine model equal those of the OSPAR-RID dataset. Table 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. The river loads for Germany are calculated with an additional correction factors for the contribution of the unmonitored areas catchment. For more details see chapter 12.

Table 2: Scaling factors applied to discharges and nutrient loads from the E-HYPE model, to calculate river inputs to the physical transport model (source: Blauw et al. 2019). *TN, total nitrogen. TP, total phosphorus*

Contracting Party	Source	Discharge	TN concentration	TP concentration
Netherlands	Nwaterweg	0.58	0.56	0.90
	Haringvliet	1.57	0.56	0.90
	Ijsselmeer	2.24	0.56	0.90
United Kingdom	All	0.98	0.75	0.73
France	All	0.97	0.85	0.58
Germany	All	0.77	0.73	1.21
Denmark	All	0.81	0.77	0.82
Sweden	All	1.15	1.10	1.23
Norway	All	1.00	1.00	1.00

Table 3: Nutrient loads of the historic scenario expressed as proportion of the current loads on a country-basis (source: Blauw et al. 2019). *TN, total nitrogen. TP, total phosphorus*

Contracting Party	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
United Kingdom	42	49
Average	**40**	**64**

The nutrient loads applied as inputs the marine physical transport model were the loads estimated by the E-HYPE (see Table 3), scaled by the country-specific factor (Table 2) based on the differences in nutrient loads between present day nutrient loads estimated by E-HYPE and by OSPAR-RID. In this way both spatial and temporal variability in river loads from the E-HYPE model are taken into account in the inputs to the physical transport model for nutrients in the marine environment

3) Task for OSPAR Contracting Parties

Based on this overview of deriving reference nutrient loads OSPAR Contracting Parties are requested to validate these loads by comparison to e.g. national modelling results or published literature. Once validated, the reference nutrient loads will be used by ICG-EMO as input data for setting up modelling scenarios.

At present there are two scenarios under debate:

1. The original JMP EUNOSAT scenario with correction for non-deterioration based on E-Hype estimates
2. A Hybrid Approach which is based on the previous scenario (1) but allows for adaptation of individual rivers based on a rationale which is expressed in the national annexes.

1) The JMP EUNOSAT scenario:

The TN and TP loads estimated for the reference scenario based on E-Hype were on average 40% and 64% of the current loads, respectively (see Table 3), but they differ between individual river catchments. For some river catchment (for example in Scotland) the estimated riverine nutrient loads may be even lower in recent years than in the reference year. Nevertheless, it is worthwhile to maintain this spatial (within country) and temporal variability in nutrient loads as input for the marine ecosystem models.

In a first step this requires a non-deterioration correction of the river catchment specific reduction percentages from E-HYPE, simply based on the rationale that no historic load is allowed to be higher than the present-day load. In a second step the corrected individual river estimates need to be mapped on the ICG-EMO database, which is used for the marine ecosystem model simulation. Since the ICG-EMO database has a lower number of river inlets to the North Sea and adjacent seas, the high number of 895 E-Hype loads need to be matched with the ones by ICG-EMO by aggregating river information to the amount of 391 rivers used within ICG-EMO.

Tab. 4 provides an overview on the reduction for TN and TP related for a selection of national rivers related to the scenario 1, the complete list for all rivers including more detailed information can be found within Annex 10.

2) A Hybrid Approach

This Hybrid Approach is based on the previous scenario (1) but allows for adaptation of individual rivers based on a rationale which is expressed in the national annexes. Since the E-Hype model has some drawbacks especially within the P load estimates, a second hybrid scenario with corrections only in the P load was agreed on. This setup was also supported by the ICG-Emo model community since the focus of only changing the P load estimates offers better comparability between the model studies.

Germany supports the use of a national model study based on the MONERIS catchment model (see detailed description in Annex 4). Since this study included the Rhine also adaptation for this load is needed within the Netherlands load estimates for this second pre-eutrophic scenario (see detailed description in Annex 5). While the changes in the P loads from Germany and the Netherlands are related to individual rivers, Denmark has provided estimates for historic P loads for the coastal areas that is covered by the Delatres model domain scenario.

The improved reduction estimates should be based on the year 2009 – 2014 for comparability with the JMP EUNOSAT scenario. Table 4 provides an overview of all changes in the TP loads (last column) only on the basis of the individual rivers for which different reduction estimates are proposed for this 2nd scenario.

Table 4: Estimates of pre-eutrophic condition for a selection of individual rivers for scenario 1 (TN and TP) and the 2nd scenario (TP only). When changes in the 2nd scenario occur the TP reduction values are highlighted in bold numbers.

Contracting Party	**River**	**TN load (%)** Scenario1	**TP load (%)** Scenario 1	**TP load (%)** **Scenario 2**
Belgium	IJzer	23	61	61
Belgium	GentOostendeCanal	17	76	76
Belgium	SchipdonkCanal	25	49	49
Belgium	LeopoldCanal	25	49	49
Denmark	Omme	30	38	**36**
Denmark	Skjern	30	38	**36**
Denmark	Stora	32	44	**36**
Denmark	Vida	30	30	**36**
France	Seine	45	71	71
France	Loire	50	92	92
France	Garonne	70	74	74
France	Dordogne	57	82	82
Germany	Elbe	51	95	**26**
Germany	Ems	26	60	**17**
Germany	Weser	37	74	**24**
Germany	Eider	23	73	**8**
Ireland	Blackwater	35	55	55
Ireland	Suir	34	57	57
Ireland	Barrow	34	57	57
Ireland	Boyne	31	50	50
The Netherlands	Meuse	38	44	**32**
The Netherlands	Rhine	43	72	**32**
The Netherlands	Lake IJssel East	22	34	**33**
The Netherlands	Lake IJssel West	21	21	**33**

Contracting Party	**River**	**TN load (%)** Scenario1	**TP load (%)** Scenario 1	**TP load (%)** **Scenario 2**
The Netherlands	North Sea Canal	30	27	27
The Netherlands	Schelde	46	81	81
Norway	Glomma	44	50	50
Norway	Skien	47	76	76
Norway	Otra	48	91	91
Norway	Kvina	37	80	80
Spain	Deba	44	34	34
Spain	Oiartzun	31	21	21
Spain	Urola	44	34	34
Spain	Urumea	31	21	21
Sweden	Gota alv	56	62	62
Sweden	Lagan	48	57	57
Sweden	Nissan	48	45	45
Sweden	Atran	48	66	66
United Kingdom	TWEED	56	83	83
United Kingdom	HUMBER	34	33	33
United Kingdom	THAMES	35	38	38
United Kingdom	TAY	63	100	100

* Estimates of historic percentage (with current values as 100%) by **loads in tons**

To derive at these load reduction for individual rivers as displayed in Tab. 4, a complete coverage is provided in Annex 10, the E-HYPE derived historic percentages were applied to the observations-based ICG-EMO riverine database organized by Sonja van Leeuwen. This database is based on original data, and uses interpolation and climatologies to arrive at daily values for both discharge and a range of nutrient loads for individual rivers. The procedure was a two step approach. First the E-HYPE input locations had to be matched with the ICG-EMO river location. And finally the river load reduction as defined for the two scenarios had to be adopted to the individual river load reduction, as presented in Tab. 4.

Within the ICG-EMO database, direct discharge data is only included for the UK (added to the nearest river), and catchment correction is only included for the UK and Germany, due to lack of data. Thus, the ICG-EMO database differs from the RID database (annual values only, per area) as aggregated values tend to be lower. Therefor, a process was started by the ICG-EMO conveners in 2019 to better align the two databases. Sonja van Leeuwen was invited to an INPUT meeting (January 2020, Gothenborg, Sweden) to explain the need for better alignment and access to the underlying data, and the OSPAR secretariat asked member states to

provide their original RID data to ICG-EMO. Some member states responded, and prior to the workshop new data for Ireland and Belgium was incorporated. Data from Germany and France has been received as well. However, the ICG-EMO database requires more nutrients than reported under RID, and so an effort still remains with ICG-EMO to find the missing nutrients (silicates, iron, carbon compounds, alkalinity). For the 2020 workshop the ICG-EMO riverine database provided discharge and nutrient load data for 364 rivers exiting onto the Northwest European Shelf, from Norway to Portugal, for the current state run CS and the two scenarios.

4) Rationale – Analysis from Belgium

Fluxes of N and P in the Scheldt River proposed in the rationale

Table 4 and 5 in the rationale propose a reduction in the Scheldt River by 77% for Tot N loads and by 44% for Tot P loads, to reach respectively 23% of Tot N and 56% of Tot P with regards to current Scheldt loads.

Previous modelling studies

We cannot reproduce the same approach and can only compare with proxies coming from our modelling studies. In these studies, there is no estimate of a "pre-eutrophication level" but there are other indications that may help defining a range of load reductions.

In Desmit et al. 2018, Pristine N and P loads were estimated for the river group (Rhine/Meuse/Scheldt), where it was hypothesized that there is no human activity in the watersheds:

- N fluxes: 7% of actual fluxes (reduction by 93%)
- P fluxes: 18% of actual fluxes (reduction by 82%)

In the LocOrgDem scenario, N and P loads were estimated in the case West-Europe Member States would comply with radical changes in agricultural, economic and social policies. Results for the river group (Rhine/Meuse/Scheldt):

- N fluxes: 49% of actual fluxes (reduction by 51%)
- P fluxes: 96.4% of actual fluxes (reduction by 3.6%)

Regarding N fluxes, we may hypothesize that the reduction to reach "pre-eutrophication" levels is between the Pristine and LocOrgDem reductions, i.e. between 93% and 51% reduction. Thus, a reduction of 77% - as proposed in the rationale - seems to fall into the range. But it will probably be unreachable in reality because the "best-case" scenario is a reduction of approximately 51% (LocOrgDem).

Regarding P fluxes, we may formulate the same hypothesis, i.e. "pre-eutrophication" levels impose a reduction between 82% and 3.6%. Thus, a reduction of 44% - as proposed in the rationale - seems to fall into the range. It is difficult to know whether such P reduction can be reached because the LocOrgDem scenario did not focus on all potential ways to reduce P river loads.

Conclusion

It is interesting from the academic point of view to consider "pre-eutrophic" concentrations. The approach of E-Hype is reliable and the proposed reductions for Scheldt N and P fall into the ranges we expect from previous studies. However, it is very probable that these levels (at least for N) will not be reached, even if Member States radically change their agricultural policies (Desmit et al. 2018). Therefore, if the intention in the future is to use these "pre-eutrophication" levels as a basis to define politically constraining thresholds (e.g. in the frame of the WFD), we might want to further discuss about it at that moment.

5) Rationale from Denmark

OSPAR Contracting Parties were requested to **confirm**, or where appropriate **revise**, **the reference nutrient loads** provided in Table 4 of the document "Rationale and description of a reference nutrient input scenario to derive harmonised OSPAR assessment levels for eutrophication parameters in Region II".

For Denmark, the proposed reference nutrient loads derived from E-HYPE were 28% of the current loads (2009-2014) for nitrogen and 43% of the current loads (2009-2014) for phosphorus.

Denmark has evaluated these percentages against the upcoming third generation of river basin management plans. We find that the proposed percentages are similar to our national work, but not identical. The proposed percentage for phosphorous for DK is somewhat higher than what we use nationally, whereas it is opposite for nitrogen.

We are therefore somewhat concerned regarding the deviations. We are also concerned that this might be the case for other CPs and if that is the case we would welcome if this could be addressed in the future work as also proposed by Germany. However, we recognise, and greatly appreciate, the importance of the work towards common assessment levels for eutrophication in the OSPAR region. Thus, we do not want to halt the modelling work at this stage.

For modelling scenario 1:

We therefore in the spirit of compromise can accept both percentages as proposed for Denmark (28% for nitrogen and 43% for phosphorous) for use in the further ICG-EMO modelling.

However, we take this opportunity to, once again, stress that the values in Table 4 of the document "Rationale and description of a reference nutrient input scenario to derive harmonised OSPAR assessment levels for eutrophication parameters in Region II" should not be perceived as reference conditions as defined in the EU water framework directive. Neither quantitatively, nor qualitatively. This owes to the assumptions of the E-HYPE model, which include significant sources of anthropogenic influence. This seems especially true for phosphorous. Our observations are apparently in line with the German request to reduce the percentages for reference loads of phosphorous from Germany.

For modelling scenario 2:

For the second modelling scenario, contracting parties have been asked to provide revised estimates based on national knowledge and approaches. In response, Denmark initially proposed to set the nitrogen reference loads to 32% of the current loads (2009-2014) and the phosphorous reference loads to 38% of the current loads (2009-2014). These estimates were based on the ongoing, comprehensive national work preparing the third generation river basin management plans. We expect thorough scientific documentation published by the end of year 2020[2]. A draft version of the report was distributed to OSPAR's reference level group along with this rationale.

OSPAR's group on reference levels agreed, based on Danish scientific advice, to change the ICG-EUT model domain bordering the Baltic Sea so as to follow the natural sills Drogden and Dars. This ensures the best

[2] *Methods for establishing Chlorophyll-a references and target values applicable for the River Basin Management Plans 2021-2027*. Prepared by DHI and Aarhus University for Ministry of Environment and Foods. In preparation

possible quantification of nutrient- and water fluxes in and out of the model domain. Resulting from this change, parts of the Danish run-off areas were excluded from the model domain, and Denmark was asked to revise the proposed national percentages accordingly. Simultaneously, it became clear that Denmark was the only contracting party asking for changes in percentages for reference loadings of nitrogen. OSPAR's group on reference levels therefore asked if Denmark would be willing to revoke the suggested change in reference nitrogen percentage from 28% to 32%. The resulting increased workload was deemed disproportionally large compared to the suggested small change from a single contracting party. The group also argued that simultaneously changing phosphorous and nitrogen would make it difficult to interpret the observed differences when evaluating the results of scenario 1 and 2.

We recognise, and greatly appreciate, the importance of the work towards common assessment levels for eutrophication in the OSPAR region. With this in mind, **we accept that scenario 2 sets the percentage reference loads of nitrogen to 28% as suggested by the E-HYPE model. The revised reference load percentage for phosphorous, considering the new model domain, is 36% of the current (2009-2014) loads.** This value is in accordance with the currently best available knowledge from the Danish national work on the upcoming 3rd generation river basin management plans as documented above.

We emphasise, however, that there are significant regional differences within Denmark in both reference loads and current loads. Therefore, the proposed percentages are valid only as bulk averages covering the entire Danish run-off area included in the model domain. We also emphasise that the exact percentages are sensitive to the choice of specific years for 'current' loads. The percentages would have been somewhat different if a different period than 2009-2014 was chosen.

Methods and assumptions for deriving thresholds

We also note that modelling threshold values from reference loads is a highly complicated task including numerous assumptions as well as technical and political decisions. Many of these issues are still unknown or undecided at this stage, and it is important that we ensure amble time to review, discuss and revise the thresholds proposed by the ICG-EMO modelling work. We find the currently proposed method of multiplying each reference conditions by 1.5 as too simplistic. Our acceptance of the proposed percentages for reference loads is thus not an *a priori* acceptance of the resulting threshold values.

6) Rationale from Germany

OSPAR Contracting Parties were requested to **confirm**, or where appropriate **revise**, **the reference nutrient loads** provided in Table 4 of the document "Rationale and description of a reference nutrient input scenario to derive harmonised OSPAR assessment levels for eutrophication parameters in Region II".

For Germany, the JMP EUNOSAT project proposed that the reference nutrient loads as an average of the major German rivers were 41% of the current loads for nitrogen and 85% of the current loads for phosphorus.

These percentages are acceptable for nitrogen but are far too high for phosphorus. Germany can accept the JMP EUNOSAT results for nitrogen but requests a change to the percentage for phosphorus to **22%** for conducting a second modelling scenario with this adjusted percentage.

The percentage for phosphorus that we propose to use is based on our historic MONERIS scenario and average RID data of 2009-2014 (without consideration of differences in discharge). Since reference nutrient loads do vary between the different rivers Germany suggests to use specific reduction percentages per river as provided in table 1.

Table 1 Nutrient loads of the reference scenario expressed as proportion of the current loads per river based on current nutrient loads of 2009-2014. The current nutrient loads for TP are based on RID data for 2009-2014 while the current nutrient loads for TN are based on the E-HYPE modelling output for 2009-2014 (without scaling). The percentages for TN are based on JMP EUNOSAT while the percentages for TP are based on MONERIS.

	TN (JMP EUNOSAT)	**TP (MONERIS)**
Elbe	**51**	**26**
Weser	**37**	**24**
Ems	**26**	**17**
Eider	**23**	**8**

One of the challenges is that our calculation for table 1 for percentages for TP is based on data from the RID database and that at least for the river Weser the RID data are based on different monitoring stations than used for the MSFD. The aim in the future is to determine one monitoring station per river that delivers data for OSPAR and MSFD, but this needs further national discussions with the Federal States and the process is likely not to be finished in 2020.

Concerning the percentages for TP the calculation in table 1 is based on recent data from 2009-2014 that were calculated based on the E-HYPE modelling results. These data do significantly differ from the RID data for the respective rivers. It needs to be further discussed how these differences can be taken into account before using the percentages for TN. In particular, it is unclear at this stage whether scaling factors, as used in the JMP EUNOSAT project, should be applied and how.

Furthermore, Germany also has a share in the nutrient loads of the river Rhine and for the Rhine Germany would like to propose a revision of the proposal made by JMP EUNOSAT for this river for the TP loads in order to arrive at a plausible scenario. Since the nutrient loads of the river Rhine dominate the inputs along the continental shelf and heavily influence the status of the German coastal waters it is important to use a consistent scenario for the German and Dutch rivers (a mixed approach could generate high nutrient reduction requirements for the German rivers in order to compensate for higher nutrient loads from the Dutch rivers). Our MONERIS modelling approach provides historic TP estimates for the Rhine, Ijsselmeer and Maas and we suggest to use those in the alternative scenario.

Rationale for the revision

Germany has used the catchment model MONERIS to derive reference nutrient loads for the major rivers (Gadegast & Venohr 2015). More specifically, a historic scenario was used that reconstructed the historic nutrient loads of 1880. The approach is very similar to the one used in the JMP EUNOSAT project. JMP EUNOSAT used the integrated rainfall-runoff and nutrient transport model E-HYPE version 3.1.3 to derive historic nutrient loads around 1900 Blauw et al. 2019).

It is assumed that there are two main reasons why the two catchment models arrived at different reference loads for the Germany rivers:

- The models worked with differing assumptions concerning the model input parameters (e.g. for the per capita production of N and P, atmospheric deposition etc.)
- The E-HYPE model performed poorly in particular for the Dutch and German rivers
- It seems that E-HYPE overestimates phosphorus loads in particular for rivers that around 1900 had already a high population density in the catchment area

While the reference loads for nitrogen were comparable between E-HYPE and MONERIS the reference loads for phosphorus were significantly overestimated by E-HYPE by a factor of 5 to 6 (for further information see document ICG EUT 20/4/2). The model comparison between E-HYPE and MONERIS did not provide a plausible explanation why these large differences arise especially for phosphorus. Germany trusts the MONERIS results for phosphorus and thinks that it is plausible that around 1880 phosphorus inputs were still close to the natural background. While population density around 1880 was about 100 inhabitants/km^2 and only 7% were connected to the sewerage system 93% of the population used pits for their sewage, so that the sewage had to pass the soil and groundwater and did reach the sea only after a long time. In addition, phosphorus fertilisation in agriculture was about 8.6 times lower than today.

Concerning the nitrogen loads Germany did not use the MONERIS estimates directly but assumed that 50% of the nitrogen loads were retained in the estuaries (Seitzinger 1988). This is plausible since the estuaries around 1880 were in a much more natural state (no deepening or straightening of watercourses). Such an assumption would lead to significantly lower TN loads compared to the JMP EUNOSAT results (see table 2 below). However, Germany suggests to not take account of the retention for the ICG EMO modelling scenario since it is unrealistic to assume that it can be re-established through measures and there is also considerable uncertainty concerning the size of its effect.

Table 2 shows the nutrient reference loads expressed as proportion of the current loads per river for TN comparing the JMP EUNOSAT data with MONERIS data with and without retention.

	TN JMP EUNOSAT	**TN MONERIS with retention**	**TN MONERIS without retention**
Elbe	51	21	42
Weser	37	24	48

Ems	26	18	35
Eider	23	17	35

It is also evident from table 2 that the estimates from JMP EUNOSAT and from MONERIS without considering retention are comparable. Hence Germany can accept using the JMP EUNOSAT data for TN. Furthermore, Germany would like to use the revised reference phosphorus loads as provided in table 1 for a second modelling scenario for the following reasons:

- The MONERIS data are the basis for national assessment levels currently used for WFD and MSFD
- The MONERIS modelling approach was specifically developed and tested for the Germany catchment area and is therefore believed to deliver better data than the E-HYPE model, that was developed for an operation on a European scale
- The German approach and the approach of the JMP EUNOSAT project are comparable in their level of ambition (historic nutrient loads of 1880 or 1900 as the basis for deriving assessment levels) and it is therefore justified to use the German reference loads for phosphorus
- The German reference loads for phosphorus are more ambitious compared to the reference loads suggested by JMP EUNOSAT

Correction factors for the unmonitored catchment areas

To consider the inputs from unmonitored areas of the major rivers in the German North Sea catchment area, factors were used to add these inputs to the measured inputs. In this case the major rivers are Elbe, Weser, Eider/Treene and Ems. For all five rivers, information on the size of the catchment area and the percentage of the catchment area covered by the respective monitoring station is available from the regional authorities and the river basin communities. The Elbe is a special case as the factor is intended to cover not only the inputs of the unmonitored area of the Elbe, but also the inputs from the Elbe tributaries. To calculate a factor for the Elbe tributaries, total nitrogen and total phosphorous data were used from the annual RID-reporting of the years 1996-2016. It was calculated how much the Elbe tributaries contribute to the total input compared to the Elbe estuary itself. This factor was added to the known factor of the unmonitored area of the Elbe estuary. Thus, the calculation resulted in a factor of 22% for the inputs into the Elbe including unmonitored area and tributary rivers inputs. Since this is only an estimation and the before used expert judgment (pers. Comm. Bergemann; ARGE) was similar with around 21% for the Elbe, we decided to keep the factor 21% for the Elbe.

7) Rationale from the Netherlands

No distinction between rivers in terms of % of current loads has been made. I wasn't sure whether the RID or ICG-EMO current loads data were used in Table 5, so I inserted both options. We also assumed that the 'current' period is 2003-2010 (this was used for JMP EUNOSAT).

The Dutch '1930 reference scenario' is described as 'river nutrient loads without the anthropogenic fraction'. This has also been used for the development of WFD standards. We currently have no further explanation of the assumptions made to derive these reference loads (but we could probably spend more effort on this).

The JMP EUNOSAT report 'Overview of approaches used for estimation of natural background concentrations of chlorophyll in the North Sea' was the inventory of methods/narratives used by North Sea countries pre-JMP EUNOSAT.

Rhine related inputs for 2nd scenario (Contributon from Anouk Blauw, Delatres)

This Hybrid Approach is based on the previous scenario (1) but allows for adaptation of individual rivers based on a rationale which is expressed in the national annexes. Germany supports the use of a national model study based on the MONERIS catchment model. This study included also the Meuse, Rhine and IJssel rivers that enter the North Sea in the Netherlands. Therefore the Netherlands load estimates for this second pre-eutrophic scenario are also adjusted. The adjusted reduction estimates, which differ from scenario 1 (see Table 3), are based on the years 2009 – 2014 for comparability with scenario 1. For an overview Table 4 will contain all changes on the basis of the individual rivers for which different reduction estimates are proposed for this 2nd scenario.

The rationale for the change in the 2nd hybrid scenario in relation to results from the MONERIS catchment model for the German rivers are attached.

The Dutch river loads for the Rhine and Meuse river branches are adjusted accordingly. There are 5 major discharges in Dutch coastal waters: the Haringvliet and Nieuwe Waterweg both discharge a mixture of Rhine and Meuse waters. These rivers get mixed through a network of interconnected river braches in the Dutch delta area. Therefore, we have assumed one common reduction factor for these two outlet points. Part of the Rhine water (before it gets mixed with Meuse waters) branches of to the north to discharge into Lake Ijssel. From Lake IJssel there are two sluices where the waters enter the North Sea, through the Wadden Sea. For these 2 discharges we assumed the same reduction percentage as for the Ijssel discharge into Lake Ijssel. A fifth major discharge is North Sea Channel, from Amsterdam to Ijmuiden. We have assumed that this only discharges local Dutch waters, so it is not affected by the MONERIS results.

We have stepwise calculate the reduction percentages for the Haringvliet and Nieuwe Waterweg and Lake IJssel sluices by first recalculating the total nutrient loads provided from the MONERIS study to nutrient loads for the present outlet points, assuming the river waters are distributed over the branches in the same way as in 2009-2014. I.e 23% of the Rhine load branches off to Lake IJssel and the remainer mixes with the Meuse waters. Secondly, we divided the resulting total nutrient loads from MONERIS by the 2009 – 2014 mean total nutrient loads for the IJssel and the combined Haringvliet and Nieuwe Waterweg discharges. This resulted in a reduction percentage on present nutrient loads to approximate historic nutrient loads.

8) Rationale – Norway reference nutrient loads

Table 1: reference nutrient loads

RID ID	River Name	2008-2018	2008-18	2008-18	Ref cond	Ref cond	Ref cond in % of pt	
		Mean flow*	TN	TP	TN	TP	TN	TP
		$m^3\ yr^{-1}$	$kg\ yr^{-1}$	$kg\ yr^{-1}$	$kg\ yr^{-1}$	$kg\ yr^{-1}$	%	%
2	Glomma	25257180602	14158201	462808	4798864	143966	34	31
15	Drammenselva	11273492568	4986886	115752	2141964	64259	43	56

18	Numedalslågen	3991411468	1757272	49763	758368	22751	43	46
20	Skienselva	9960182651	2710060	43820	1892435	56773	70	130
26	Otra	5032690865	1163253	18963	717158	23905	62	126
37	Orreelva	189287133.4	293413	13620	58442	1978	20	15
64	Vosso	3176449417	608808	11726	452644	15088	74	129
100	Orkla	1870574929	607450	11789	355409	10662	59	90
115	Vefsna	5066359678	628156	16444	721956	24065	115	146
140	Altaelva	3070735234	625302	27905	583440	17503	93	63
	Total	**68888364545**	**27538801**	**772589**	**12480681**	**380951**	**45**	**49**

* Mean flow in reference period assumed to be 5% lower than in 2008-2018.

Table 2: Present (2008-2018) vs. reference concentrations

RID ID	River Name	2008-18 TN mg m^{-3}	2008-18 TP mg m^{-3}	Ref conc TN mg m^{-3}	Ref conc TP mg m^{-3}
2	Glomma	561	18	200	6
15	Drammenselva	442	10	200	6
18	Numedalslågen	440	12	200	6
20	Skienselva	272	4	200	6
26	Otra	231	4	150	5
37	Orreelva	1550	72	325	11
64	Vosso	192	4	150	5
100	Orkla	325	6	200	6
115	Vefsna	124	3	150	5
140	Altaelva	204	9	200	6

Conclusion

Historic scenario expressed as proportion of the current loads (38% for TN and 49% for TP) seems realistic for Norway

9) Rationale from Sweden

OSPAR Contracting Parties were requested to **confirm**, or where appropriate **revise**, **the reference nutrient loads** provided in Table 4 of the document "Rationale and description of a reference nutrient input scenario to derive harmonised OSPAR assessment levels for eutrophication parameters in Region II".

For Sweden, the JMP EUNOSAT project proposed that the reference nutrient loads to the Kattegat and Skagerrak were 48% of the current loads for nitrogen and 58% of the current loads for phosphorus.

Using Water Framework Directive reference conditions (Swedish Agency for Marine and Water Management, 2019) which are thought to represent conditions around 1920 were used to estimate inputs from Swedish

rivers based on "pristine" nutrient concentrations at the limnic/marine limit. These were compared to observations in the RID database for the period 2008 - 2018). This suggests that pristine total nitrogen loads were 36% of present while total phosphorus loads were 58% of present day loads. These values are close to the EUNOSAT estimates, so **we accept the estimates from EUNOSAT.** Coastal reference levels, transformed to a salinity of 0 g/kg, were used as Sweden does not have generally applicable reference levels for nitrogen in fresh water.

Table x *Nutrient loads of the reference scenario expressed as proportion of the current loads per river. Current loads are an average of 2008-2018 and are taken from the RID database.*

RID ID	Name	Mean flow [m^3 yr^{-1}]	RID inputs 2008 – 2018 [kton yr^{-1}]		WFD reference inputs (~1920) [kton yr^{-1}]		WFD as a percentage of RID 2008-2018	
			TN	TP	TN	TP	TN	TP
96	**Rönne å**	7,41E+08	2.47	0.04	0.17	0.01	6.8	20.0
98	**Lagan**	2,70E+09	2.60	0.05	0.61	0.03	23.5	59.7
101	**Nissan**	1,35E+09	1.39	0.04	0.30	0.02	21.8	46.2
103	**Ätran**	1,85E+09	1.91	0.04	0.42	0.02	21.9	59.7
104	**Himleån**	1,20E+08	0.07	0.00	0.04	0.00	60.7	101.3
105	**Viskan**	1,01E+09	1.50	0.05	0.36	0.01	24.3	25.0
108	**Göta älv**	1,84E+10	14,07	0,34	6,68	0,23	47,5	67,5
109	**Bäveån**	1,28E+08	0,05	0,00	0,03	0,00	61,7	80,1
	Total	**2,63E+10**	**24,07**	**0,56**	**8,62**	**0,32**	**35,8**	**57,5**

10) Rationale from the United Kingdom

In the spirit of collaboration and the need for progress, the United Kingdom accepts the overall 42% and 49% reduction for nitrates and Phosphates, subject to certain caveats. There is some limited observational evidence which support these numbers; however, there are a number of areas in Northern Scotland, where the population was higher in 1900 than now and the associated nutrient loads was higher. These regions now broadly have very low population density and associate riverine concentrations of nitrate (and phosphate) levels are so low, the nutrient loads could not be further reduced. Therefore, it is not appropriate to assume that these reductions are applied directly across the whole of the UK. As in Northern Scotland, there may be other specific near-coastal waterbodies tied to particular discharges or rivers, where reductions would not be appropriate. To resolve this issue, the UK will examine the river by river input files for the E-HYPE modelling to see if there is agreement on how the reductions have been applied in the E-HYPE modelling, and if this is transferable to the new river database due to be distributed by the Netherlands (Sonja Van Leeuwen).

An additional caveat to the overall process. We also note that modelling threshold values from reference loads contains many steps, with numerous assumptions as well as technical and ultimately political decisions. Many of these processes and issues are still unknown or undecided at this stage, and it is important that we ensure ample time to review, discuss and revise the thresholds proposed by the ICG-EMO modelling work to bring them into line with our best knowledge and experience.

While accepting the broad collaborative approach, the UK's acceptance of the proposed percentages for reference loads is not an automatic assumption of acceptance of the resulting threshold values, as there may be specific issues, most likely near-shore, that are not adequately addressed by this approach.

Beyond the e-hype work there is a limited evidence base, with an observational data set from the Thames (Howden et al., 2010) below and a modelling study by Naden et al., (2016)

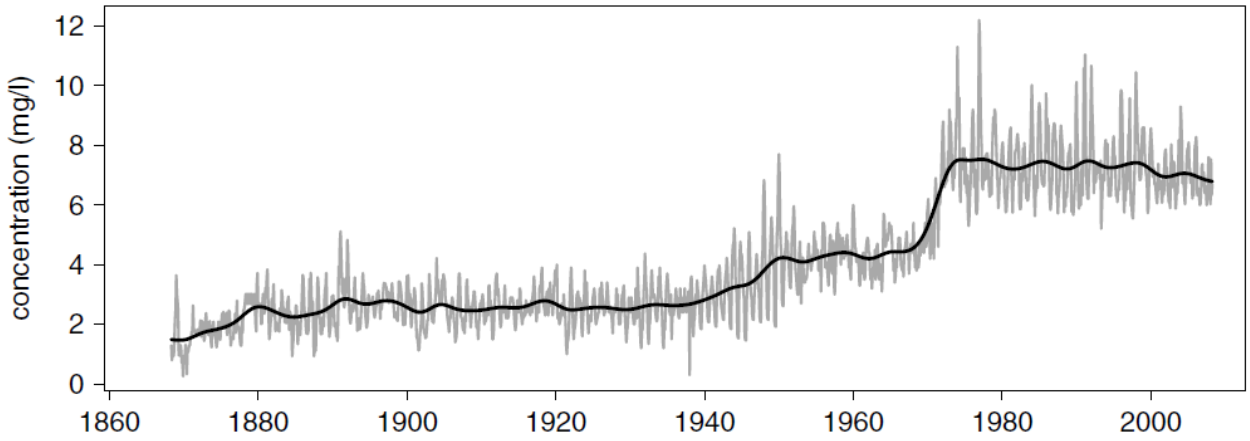


Figure 1. Time series plot of monthly nitrate concentration (mg NO3 –N/l), with an approximate 1-year moving average.

The paper makes some interesting points about reduction strategies for Land drainage results in drier soils,

"enhanced nitrogen turnover and reduced denitrification; moreover, field drains also transfer leached nitrate rapidly to the surface water network, reducing the potential for riparian zone denitrification (Burt and Pinay, 2005). Removal of other landscape barriers (e.g. hedgerows and ponds) has exacerbated this effect. Such fundamental alterations to the fabric of the land mean that the observed high nitrate concentrations will remain high unless very radical changes are made (e.g. 'setting aside' large areas; *cf* Johnson, 1991). Merely reducing nitrogen inputs will be ineffective, as indicated by the lack of a significant downward shift in fluvial nitrate concentrations since 1980".

However, the simple relevant point here is that 1900 Nitrate concentrations taken at Teddington wear 2-2.5 mg l^{-1} and for 2000 around 7.5 mg l^{-1} .

Or put as loads 15 K tonnes $year^{-1}$ in 1900 and 40 K tonnes $year^{-1}$ in 2000. It should be noted that this is record of the nutrient load at Teddington wear above the substantial loads input by citizens from London.

The great stink of 1858 led to plan for sewers and by 1874 Bazalgette's plan were operational. However, while these works regulated the discharge , raw sewage was discharged into the Thames estuary potentially at quite high loads as it was efficient at collating the sewage from the citizens of London.

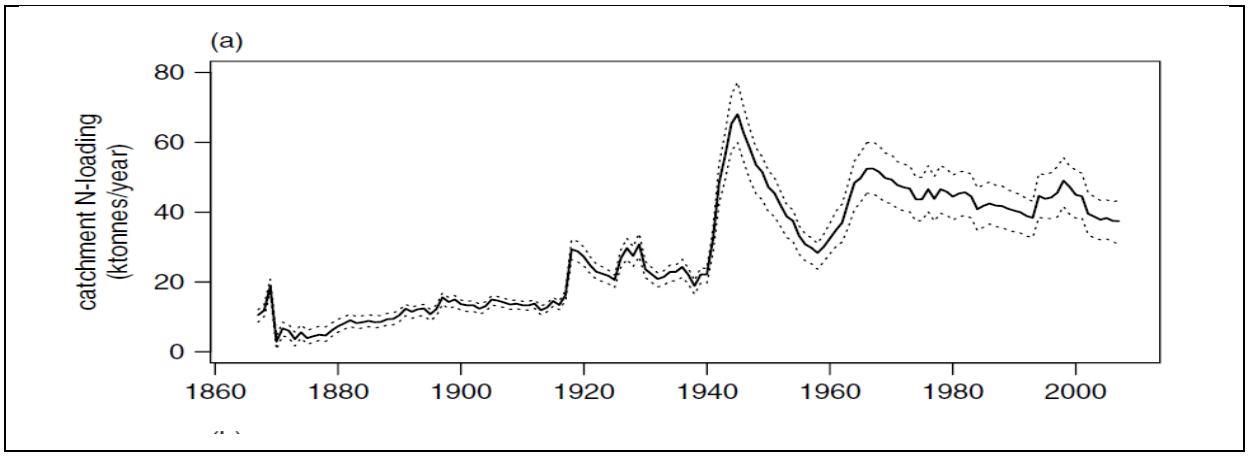


Figure 2: Modelling for the whole of the UK, a paper by Naden et al., (2016) has modelled human waste and was wastewater for the UK from 1800 -2010.

Nutrient emissions in human waste and wastewater effluent fluxes from domestic sources are quantified for the UK over the period 1800–2010 based on population data from UK Census returns. The most important drivers of change have been the introduction of the water closet (flush toilet) along with population growth, urbanization, connection to sewer, improvements in wastewater treatment and use of phosphorus in detergents. In 1800, the population of the UK was about 12 million and estimated emissions in human waste were 37 kt N, 6.2 kt P and 205 kt organic C/year. This would have been recycled to land with little or no sewage going directly to rivers or coastal waters. **By 1900, population had increased to 35.6 million and some 145 kt N were emitted in human waste but, with only the major urban areas connected to sewers, only about 19 kt N were discharged in sewage effluent.** With the use of phosphorus in detergents, estimated phosphorus emissions peaked at around

63.5 kt P/year in the 1980s, with about 28 kt P/year being discharged in sewage effluent. By 2010, population had increased to 63 million with estimated emissions of **263 kt N,** 43.6 kt P and 1460 kt organic C/year, and an estimated effluent flux of 104 kt N, 14.8 kt P and 63 kt organic C.

Key point i.e. in 1900 - 145 kT N compared to 263 kT N in 2000 but in 1900 only 19kt N actually makes it to the sea compared to 104 kt N in 2000.

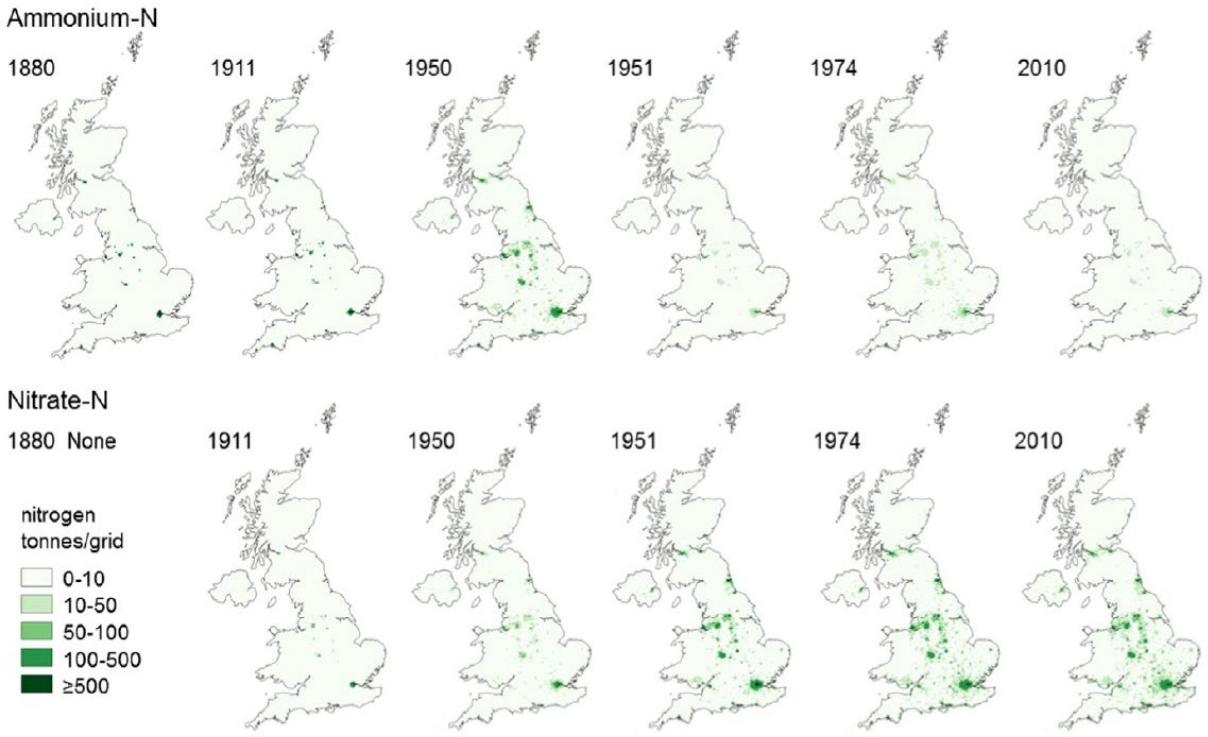


Fig. 8. Estimated annual dissolved ammonium and nitrate fluxes in effluent from domestic wastewater for the UK in selected years, based on gridded population data at a 5 km grid resolution.

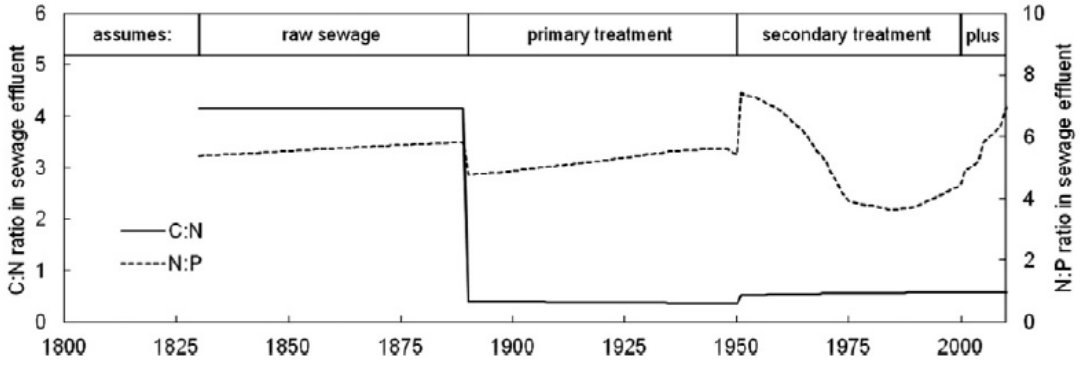
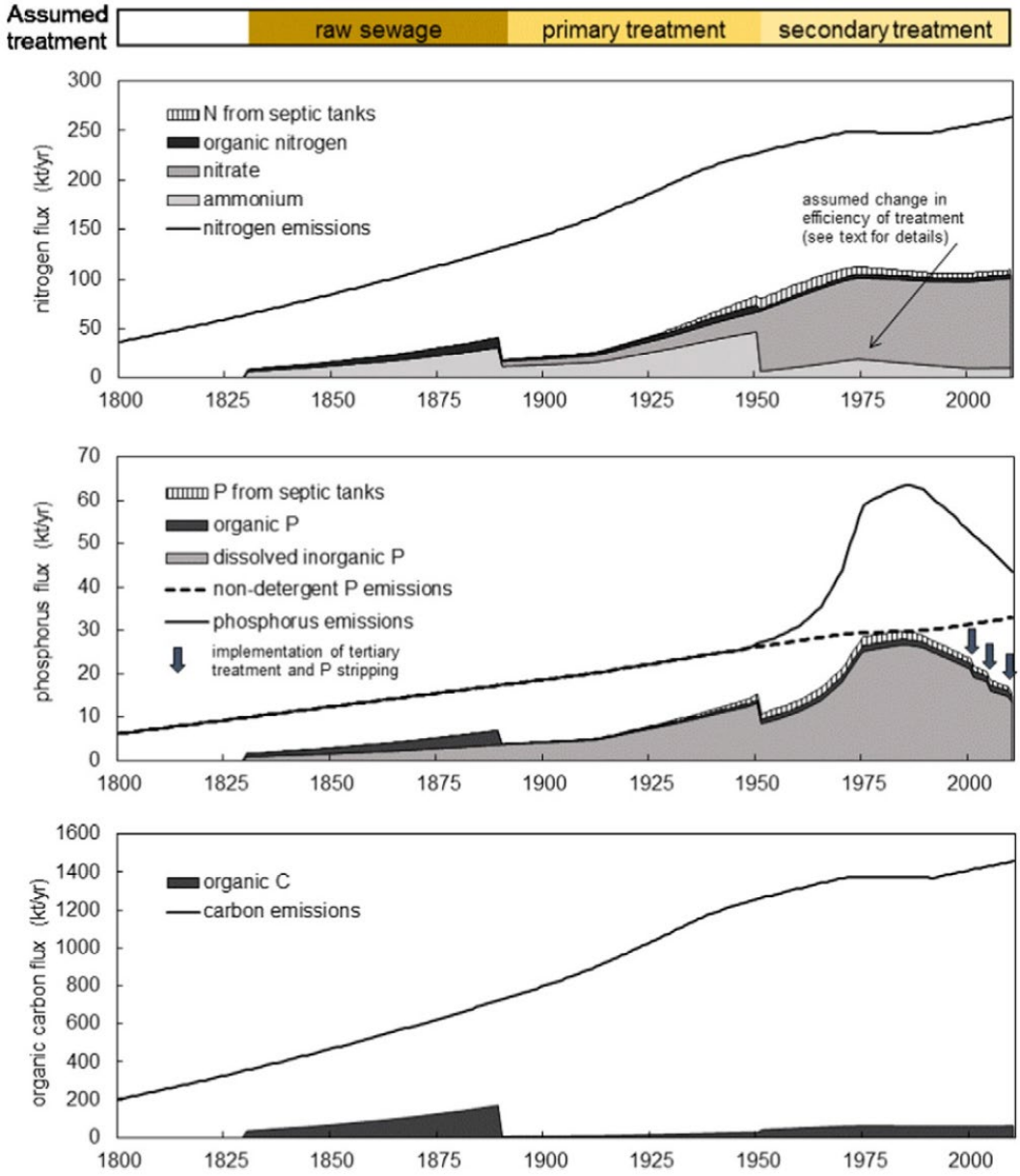


Fig. 6. Estimated changes in C:N and N:P ratios in sewage fluxes to river/sea: zero flux prior to 1830; raw sewage 1830–1889; primary treatment 1890–1950; secondary treatment 1951– 2000; tertiary treatment and P stripping for a proportion of effluent post 2000.

The key point here being that during the 1900s the C:N ratio is not significantly different from now and that the N:P ratio is not especially different either .



Estimated total nutrient flux in effluent from domestic wastewater for the UK 1800–2010 based on population (top graph). Total emissions of N, P and organic C shown by solid line; non-detergent P emissions by dashed line; amount going to river/sea by shaded area subdivided according to nutrient species. For septic tanks, the fluxes are ammonium N and dissolved inorganic P.

11) Description and assumptions of the E-HYPE model

HYPE (Hydrological Predictions for the Environment) is an integrated rainfall-runoff and nutrient transport model. E-HYPE version 3.1.3 (released in August 2016) was used for the model simulations (for further details see Blauw et al. 2019 and references therein).

The assumptions used for the historic scenario are detailed in Table 2.

Table 2: Assumptions used for the historic scenario in E-HYPE (source: Blauw et al. 2019). For all references see Blauw et al. 2019.

Subject	Assumption
Climate	• Today's climate
Water management	• Modern levels of channelisation, building and operation of reservoirs, dams, channels and dikes
Atmospheric deposition	• Nitrogen deposition rates reduced to 33% of the current deposition rates (based on Engardt et al. 2017) • Spatial distribution of nitrogen deposition corresponds to the recent distribution
Agricultural practices including fertilisation	Fertilisation: • maximum application rate was set to 100 kg N per hectare and 50 kg P per hectare (Smil 2000, van Grinsven et al 2015) • all nitrogen applied to fields was assumed to be in organic form (manure) (since Haber Bosch process was not yet operation) • phosphorus application rate was assumed to be applied to crops as manure for 80% of the application rate and as inorganic fertilisers for 20% of the application rate • the crops with the highest current fertiliser application rate still used the highest application rate in the historic scenario • application rates for phosphorus were 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 Crops cultivated: • crop fractions were left unchanged (today's conditions) and were only modified as a result of land use change; catch-crops were removed Nutrient balances: • nutrient balance in soils are calculated directly in HYPE; processes considered are plant uptake, removal through harvest, and decay
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 • Waste-water treatment technologies were in the early stages; 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) • septic tanks or cesspools were typically not well sealed and leaked significantly to soils • the use of "sewer farms" in cities (sludge harvested from sewers was put on fields) were not considered since there are no records on where they existed and what amounts of sewage was disposed

Subject	Assumption
	• based on urban population density, the urban population was divided into two groups: 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. • urban population divided into three groups by density: <100, 100-500, >500 persons/km^2; for these groups different treatment levels and connection to sewage system were applied 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. • 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 set to zero. • point sources are discharged directly to streams (net releases), while rural sources are partitioned into seepage to soils and streams; the partitioning remained unchanged at 25% (stream), 75% (soil).
Industrial sources	• no change was made to industrial sources as no reliable information was obtained • represent available point source releases from the European Pollutant Release and Transfer Register (EPRTR)
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.

Subject	Assumption
	• the population-related geospatial layers contain population density, population counts (urban population, rural population, and total population), and urban area • the following land uses were considered constant and were not modified for the historic simulation: lakes, glaciers, wetlands, and rivers • E-HYPE's basic units (HRUs) are defined for unique land use, soil, and crop combination • 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; if none of these were present, all such remaining change was allocated to mixed forest

12) Rationale for atmospheric Nitrogen deposition and Baltic Boundary Condition

In addition to the definition of the two historic riverine scenario further agreements are needed on two issues that are linked to the pre-eutrophic scenario narrative. These are the assumptions that are needed within the two scenarios for application of the atmospheric nitrogen deposition and the Baltic boundary condition. While there are two alternative scenarios for the historic riverine estimates, both assumptions are applied in the same form for both scenarios.

Baltic boundary condition

In the expert group the question was raised if boundary condition for the Baltic need to be adopted for the Baltic outflow. Anouk created these boundary conditions. The Baltic inflow for recent years is based on simulated discharge data by the Dars and Drogden sills, provided by DHI. They provided monthly total discharges for 2002 – 2019. We calculated monthly mean climatologies from these data as model input. The nutrient concentrations in recent years are based on monitoring data provided by Stiig Markager at three monitoring locations nearby the sills for the years 2009 – 2014. These were also converted to monthly mean climatologies as model input, for the variables: temperature, salinity, oxygen, NO3 (incl NO2), NH4, PO4, bioavailable DON, bioavailable DOP, POC, PON and POP. For Silicate an estimate of 10.4 µM was used, based on data in a paper by Bentzon-Tilia et al., (2014). It is generally believed that Si is not limiting phytoplankton growth in the Baltic. Reduction percentages for the historic scenario (Tab. 1) have been derived from a long model simulation (1850 – 2008) with the ERGOM model provided by Thomas Neumann (IOW, DE).

The detailed information on the export for the Drogden and Dars sills can be found in the EXCEL file "Drogden_and_dars_loads" on the cloud server under:

Tabel 1: Reduction percentage for historic scenario for Drogden and Dars sills

Variable (mol/kg)	Drogden sills	Dars sills
NO3	66 %	54 %

NH4	58 %	55 %
DON	53 &	53 %
PO4	33 %	33 %
POC	88 %	91 %

Atmospheric N deposition

EMEP atmospheric N deposition data will be used for the current state run. These data will be scaled with correction factors based on Schöpp et al. (2003) to represent pre-eutrophic condition around 1900. Onur Kerimoglu compiled the following information:

Reduction of atmospheric deposition rates were calculated based on the estimations by Schöpp et al. (2003). Previously, Markus Kreus had digitized the trajectories of NO_x and NH_3 estimates provided in their Fig. 2, and used these as normalization factors to project the spatially resolved nitrogen deposition rates provided by EMEP for the North Western Continental Shelf domain at 20km resolution (NWCS) back to 1880, as described by Große et al., 2016. For the current task, we simply calculated the average rates for the current (2009-2014) and historic (1890-1900) time periods for NO_x, NH_3 and N_{total} , and the respective ratios, as reported in Table 2.

Table 2: Average N deposition rates for the current (2009-2014) and historic (1890-1900) time periods and the respective ratios.

	Control [mg m^{-2} Y^{-1}]	Historic [mg m^{-2} Y^{-1}]	NX_H/NX_C
NO_x	183.574	26.389	0.144
NH_3	167.494	105.812	0.632
N_{total}	351.067	132.201	0.377

The complete technical details are explained by Große et al. 2016: 'Data for atmospheric N deposition were compiled using a hybrid approach. This was required since the overall simulation period (1977–2012) exceeds the period of data available from the EMEP (Cooperative pro-gram for monitoring and evaluation of the long-range trans- missions of air pollutants in Europe) model (1995–2012). First, the EMEP results for total deposition of oxidised (NOx) and reduced nitrogen (NH3) were interpolated to the model grid. Second, we calculated the average annual depo- sition rates for the NOx and NH3 for each grid cell, based on the 1995–2012 EMEP data. The resulting spatially re- solved arrays of average deposition rates were subsequently normalised by the spatial average of the entire domain to yield the spatially resolved anomaly fields. Finally, gridded deposition rates for individual years were obtained using (1) the gridded anomaly fields, (2) EMEP's spatially aver- aged (over our model domain) deposition rates for year 2005, and (3) long-term trends (normalised towards year 2005) for the temporal evolution of European emissions of NOx and NH3 (Fig. 2 in Schöpp et al., 2003).'

13) Constrains in model application to derive target values

The chapter comprises the constrains the modelling community wants to highlight as well as the discussion points that came up during the meetings of the pre-eutrophic expert group.

Denmark has repeatedly indicated that this approach as too simplistic and argues that this is not in line with the WFD. Stiig Markager commented that a hybrid approach can also be augmented from historic changes in agriculture practise and wastewater management. Although a major change in nitrogen use took place about 1910-20 with the invention of the Haber-Bosch process, systematic use of clover and other nitrogen fixing plants started at least fifty years before in some regions. This, combined with limited access to artificial fertiliser until after WWII, means that the increase in nitrogen inputs in agriculture, and hence loss to the environment, probably took place over at least hundred years from 1850 and with significant regional variation. Similar regional differences are likely for phosphorus since wastewater practice also developed during the nineteen century. In some densely populated areas sewage systems were implemented in mid-1800 , but without treatment before the sewage reached the rivers. Together, this means that a pre-eutrophication scenario cannot be linked to a specific year or period, but must be considered region-by-region. Hermann Lenhart adds that this reasoning is not only restricted to the hybrid scenario but applies for both scenarios.

Stiig Markager comments in addition to the above mentioned causes on land, marine ecosystems also respond to nutrients in a number of ways that hamper the idea of linking loadings from a specific year or period to a specific state, e.g. a pre-eutrophication state that can be used as reference conditions. The main effects are the accumulation of nutrients in the ecosystems and the filtering effects in the coastal zone.

In healthy marine systems, the estuaries and the coastal zone acts as a filter between land and sea where nutrients are processed and retained. Nitrogen and to some degree phosphor are transformed from inorganic to organic bound forms that are less bioavailable. In addition nitrogen is lost via denitrification. Macrophytes take up nutrients in the biomass during winter and spring and keep it over the growing season, making the nutrients unavailable for phytoplankton growth. Phosphorous and to some degree nitrogen is accumulated in the sediments where they stay as long as the sediment surface contains oxygen. Together, these processes make it likely that low or moderate fluxes of nutrients from land can have almost no impact in open marine areas, since most of the nutrients do not reach the open sea. Wera Leujak questions why estuaries are excluded in this overview and that one should better use the term “buffer” that filter.

We also know that nutrients accumulate over time with increased nutrient loadings in marine systems. This accumulation is largely in organic bound forms for nitrogen, both as dissolved (DON), particulate nitrogen (PON) and in the sediments. For phosphorous, the accumulation of iron bound phosphorus in sediments is also important. These processes combined with the estuarine circulation, which conserve nutrients within the estuary when water is passing through, means that the effects of nutrient loadings must be considered over periods of decades and maybe longer for large systems as e.g. the North Sea.

The consequence of the buffer effect, and a slow but steady accumulation of nutrients, is that marine systems might have stayed in an almost unaffected state from mid-1800 to mid-1900, despite nutrient loadings that slowly increased to a level that today is inconsistent with GES. The consequence for this work is that we should consider a regional approach accessing pre-eutrophication condition region by region without linking this to a specific year or period. Moreover, the model must be run with an intact coastal zone and the accompanying filtering effects and with pools of nutrients in all compartments that reflects a pre-eutrophication state. The later statement is commented by Wera Leujak that this needs to be discussed in which way the demand from Stiig Markager is realistic and can be achieved by the modellers.

Finally, Lisette Enserink reported about a MSFD Horizontal issues workshop on threshold values which showed that for the Mediterranean two different approaches were selected. In summary she stated, that in

the Mediterreanean there are still areas that can be regarded as unimpacted by eutrophication. If there is a time series of minimal 5 years an area can be selected for a baseline. The threshld value then is the upper limit of the 95% confidence interval of the range in naturally occurring concentrations of nutrients and chlorophyll a. This method takes into account characteristics of these areas with regard to salinity: lower threshold values in higher salinity areas. This approach is called a 'scientific threshold values' and is used by the majority of Mediterranean countries.

Spain and France however use threshold values that are background concentrations plus 50%. These threshold values are referred to as 'political' threshold values.

References

Blauw, A., et al. (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 1 Report. 86 pp.

Bentzon-Tilia, M., Traving, S. J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J. L., Markager, S., & Riemann, L. (2015). Significant N 2 fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. The ISME journal, 9(2), 273-285.

Gadegast, M., Venohr, M., 2015. Modellierung historischer Nährstoffeinträge und -frachten zur Ableitung von Nährstoffreferenz- und Orientierungswerten für mitteleuropäische Flussgebiete. Technical Report. Leibniz-Institut für Gewässerökologie und Binnenfischerei im Forschungsverbund Berlin e.V. Berlin.

Große, F., Greenwood, N., Kreus, M., Lenhart, H.J., Machoczek, D., Pätsch, J., Salt, L.A., Thomas, H., 2016. Looking beyond stratification: a model-based analysis of the biological drivers ofoxygen deficiency in the North Sea. Biogeosciences 13, 2511– 2535. https://doi.org/10.5194/bgd-12-12543-2015.

Howden, N. J. K. ., Burt, T. P. ., Worrall, F., Whelan, M. J. , and Bieroza, M (2010) Nitrate concentrations and fluxes in the River Thames over 140 years (1868–2008): are increases irreversible? . Hydrol. Process. 24, 2657–2662 (2010)

Naden, P., Bell, V., Carnell, E., Tomlinson, S., Dragosits, U., Chaplow, J., May, L., and Tipping, E., (2016) Nutrient fluxes from domestic wastewater: A national-scale historical perspective for the UK 1800–2010 Science of the total Environment 1471 – 1484.

Seitzinger, S. P. (1988) Denitrification in freshwater and coastal marine ecosystems: ecological and geochemical significance. Limnology and Oceanography 33:702-724.

Schöpp, W., Posch, M., Mylona, S., and Johansson, M. (2003). Longterm development of acid deposition (1880–2030) in sensitive freshwater regions in Europe. Hydrol. Earth Syst. Sci. 7, 436–446. doi: 10.5194/hess-7- 436-2003

Swedish Agency for Marine and Water Management, 2018, "Havs- och vattenmyndighetens föreskrifter om klassificering och miljökvalitetsnormer avseende ytvatten[3];", HVMFS 2019:25, 17 december 2019, (*in Swedish*)
https://www.havochvatten.se/download/18.4705beb516f0bcf57ce1c145/1576576601249/HVMFS%202019-25-ev.pdf

[3] Swedish Agency for Marine and Water Management's Regulations for classification and targets in surface waters

The Aspect
12 Finsbury Square
London EC2A 1AS
United Kingdom

t: +44 (0)20 7430 5200
f: +44 (0)20 7242 3737
e: secretariat@ospar.org
www.ospar.org

Our vision is a clean, healthy and biologically diverse North-East Atlantic Ocean, which is productive, used sustainably and resilient to climate change and ocean acidification.

Publication Number: 895/2022