Observing and modelling phytoplankton community structure in the North Sea: can ERSEM-type models simulate biodiversity?

Response to reviewers

We would like to thank the three anonymous reviewers and Dr Butenschön for their constructive comments. We will first give a general summary of our response, and then respond to each of the reviewers' points in turn.

The main concern raised by all the reviewers is that we did not do enough to address the impact that physical drivers could have on the differences in results, or to justify our conclusion that differences in phytoplankton community structure primarily arise from differences in biogeochemical model parameterisations. We accept this concern, and have expanded our discussion of these points, going into more details. Furthermore, we have included some validation against in situ salinity and nutrient data. However, we also stress that a detailed comparison of these processes was not our aim, and have made this clearer in the manuscript. Our aim was to provide an initial assessment of phytoplankton community structure against a new set of observational data, using pre-existing model configurations, with a preliminary discussion of the reasons behind any differences. To be confident of our conclusions will require further experiments in a more controlled manner, and so we have also expanded our discussion to propose targeted experiments which would lead on from our study. We have also revised the title to remove the reference to "biodiversity", added a figure showing the model domains and added two further validation figures, replotted other figures to be as clear, consistent and useful as possible, discussed and compared results from other studies in the literature, and addressed the various other points raised (detailed below).

Detailed response to Reviewer 1

"I suggest using more contrasted "pole" colors in the triangular representation used in figure 11."

We have replotted this with more contrasting colours and a discrete colour scale.

"As far as the models are concerned, a simple comparison between two versions of 3D-distributed ERSEM model is interesting, but rather difficult to be analysed, because both the physics and the biogeochemistry are different. As a caricatural example, the comparison between the assimilated temperature of NEMO-ERSEM and the freerunning temperature of GETM-ERSEM-BFM (figure 4) seems to me inapplicable. Secondly, when strong differences appear, the authors should try to explain more what process (physical or biogeochemical) can be thought to be possibly the main driver."

We accept that the many differences in the physics and other experimental configurations preclude a thorough assessment of the reasons behind the differences in results. This was not our aim, and we have added text to Sect. 3.3 and Sect. 5 to state this more clearly. Nonetheless, where differences in results are discussed, we have endeavoured to do so in more detail, adding text to Sect. 5. Figure 4 was included to demonstrate inter-annual and spatial variability, as well as for validation, so we have kept this, and added more analysis of inter-annual variability to Sect. 5.

"I disagree a little bit with the authors when they say that "both models were still able to capture the main observed features throughout the domain.""

We have reworded this accordingly.

"For nutrients, the too coarse maps of WOA climatologies don't allow a precise validation; but the authors could try to link the different behaviours of the two versions to their deposition and remineralisation processes : in the central North Sea, GETM-ERSEM-BFM seems to trap too much the detrital forms of N, P and Si. As far as biodiversity is concerned, GETM-ERSEM-BFM clearly gives a more realistic decomposition in phytoplanktonic functional types, while the weakness of diatoms in coastal nutrient-enriched zones seems to be strongly unrealistic in the case of NEMO-ERSEM, even in August."

We have added discussion regarding these points to Sect. 5.

"The discussion could be highly improved by replacing some general considerations by a comparison with results obtained in PFTs simulation obtained by other scientists, especially those using the DARWIN model (Follows et al., 2007)."

We have added a paragraph of discussion regarding other results in the literature, including works based on Follows et al. (2007), to Sect. 5.

"the results of the two models used don't support really the optimistic answer given at the end of the paper by the authors: for me the ERSEM model has to be better calibrated before being considered as able to simulate biodiversity."

We agree that better calibration is required, our conclusion was that such calibration appears to be achievable without requiring large fundamental changes. However, we accept this is still non-trivial, and have reworded the conclusion to come across more clearly and as more pragmatic.

"Nevertheless, I think that this paper is acceptable as a milestone one on the long way to biodiversity modelling, but only if the authors add a bibliography-based discussion and lower their optimistic evaluation of their present versions of the model."

We have added a bibliography-based discussion, and reworded various sections, particularly in Sect. 5, to be more pragmatic in our evaluation.

Detailed response to Reviewer 2

"1) The title is not representative:
a) The title suggests this work is about biodiversity, but actually it's not. Although the 'relative abundance' of PFT's (Fig. 2,10 and 11) is an index relevant for biodiversity, the results are not at all discussed or framed in the context of biodiversity (in fact only the very last short paragraph of the manuscript refer to biodiversity again after the introduction, and replacing 'biodiversity' with 'phytoplankton community structure' would probably improve this paragraph already).
b) Again, at the first glance 'ERSEM-type models' looks like more general than it is, whereas the manuscript is literally about 2 ERSEM versions, which could be more clearly indicated in the title. Accordingly, I invite co-authors to consider a more representative
title (as changing the content according to the current title would basically result in a new manuscript)"

Reviewer 3 also suggested changing the title, and we have done so accordingly, shortening it to simply "Observing and modelling phytoplankton community structure in the North Sea", so as to remove the references to "biodiversity" and "ERSEM-type models". We have also

replaced "biodiversity" with "phytoplankton community structure" in the last paragraph as suggested.

"2) Validation of PFT estimates is not a new endeavor:
A recurring notion in the manuscript (abstract, introduction, summary and discussion)
is that 'model estimates of PFTs are not validated', which is misleading. Quite a few
attempts have been made previously to assess model performance with respect to
reproducing the spatial distributions of PFTs, such as those making use of the data from
continuous plankton recorder surveys, monitoring data obtained at stations located at
different sites and estimates from satellite images, which should be acknowledged
(examples are too many that pointing to specific studies would be unfair to others)."

Our intention was not to claim that this has never been done before, more that it is not done nearly as routinely as validation of total chlorophyll and other variables. We have reworded the abstract, added text to the introduction referencing other works, and added a paragraph to the discussions section comparing our results with other works.

"3) 'Observations' section is incomplete:
In this section, only the procedures relating to the HPLC data are provided. Other
data sources, like the 'in-situ quantification of chl-a', OSTIA-SST, OC5-chlorophyll/SPM
and WOA data should be introduced also in this section, and not elsewhere (e.g., as
currently done in sections 4.1 and 4.2, see also the 'Technical Corrections')"

This section was intended purely as a description of the new set of observations introduced in this study. However, we have moved the description of the other data used for validation from section 4 to section 2 as requested.

"4) 'Models' section is inadequate:
a) 3.1 GETM-ERSEM-BFM: each item in the list of additional processes to 'make
[ERSEM as presented by Vichi et al. (2007)] suitable for temperate shelf seas' need
a reference to previous literature if possible (like in the list items iv and v), or if not, an
adequate description.
b) 3.2 NEMO-ERSEM: page(P)6,line(L)15-16: the employed version of NEMO appears
to be not described previously (although a full description is foreseen to be provided by
O'Dea at al., in prep), so here at least some hints should be provided
c) 3.3 Comparison: a figure showing the complete domains of the two models would
be helpful"

We have expanded the descriptions of both the models as requested, and included a figure comparing the domains (new Fig. 1).

"5) Fig.11 is difficult to perceive:
although I like the idea of overlaying measurements on the contour-plots using a
ternary color map, I found it difficult to distinguish the resulting colors. Although I admit
distinguishing colors is not my strength, I wonder whether using a more saturated and
bright color scheme would help (eg., as in Fig.4 of Arteaga et al. 2014). The authors
may also want to consider plotting the percentage histogram of each group across total
chlorophyll as in Fig. 3 of De Mora et al. (2016), which does not rely at all on the
color-perception abilities of the reader."

We have replotted Fig. 11 (now Fig. 14) using a more contrasting colour scheme and discrete colour scale. We have also added a histogram plot (new Fig. 12) as recommended.

"6) How to reproduce the spatial distribution of PFT's? :
as mentioned above, the results suggest that the ERSEM version previously tuned

more for coastal applications (GETM-ERSEM-BFM) performs better than the one used for simulating the margin of the Atlantic Ocean (NEMO-ERSEM). But considering that the performance difference wrt the PFTs basically boils down to the reproduction of the diatom blooms at the coast by GETM-ERSEM-BFM and not by NEMO-ERSEM, the obvious follow-up question is why the latter cannot reproduce the Diatom blooms at the coast, which I believe should be straightforward to figure out. It is hinted in section 3.3 and later in P15,L1-7 that the parameterization of the PFT's is likely to be the reason, but, it should be straightforward to test whether this explanation really holds. And I believe this would be a rather useful outcome of this study with a potentially wider relevance: can an open-ocean model transferred to a coastal system just by a reparameterization of the PFTs?"

We have added text to Sect. 5 discussing this issue.

"7) What drives inter-annual differences?:
Figures 4-11 suggest considerable differences between 2010 August and 2011 with respect to physical, chemical and biological characteristics. Although these differences are acknowledged throughout the manuscript, not much is done to elaborate the mechanisms driving these differences, apart from hinting 'a set of physical drivers' to be responsible for the difference (P15, L21). Again, I feel like more could be done with the available tools and data."

We have added more text to Sect. 5, describing how the physical variability appears to have driven the biogeochemical variability in the models. Although as noted in our response to Reviewer 1, this is difficult to assess thoroughly given the many differences in model configurations, and a thorough assessment was not our intention – that is left for future studies, and we have added suggestions for these.

"Technical Corrections
- P6, L34: sediments=? SPM?

SPM. We have clarified this.

- P7, L3: nutrients =? of all forms, or dissolved inorganic form only?

Nitrate, phosphate and silicate. We have clarified this.

- P7, L28,L30: first occurrence of the 'PML-ERSEM'. Although what this means should be clear to the careful reader, but better stick with the previous definitions.

We have removed references to PML-ERSEM, and simply used NEMO-ERSEM.

- P9, L11: first occurrence of MODIS (also consider my specific comment (SC)-3)

The acronym has been expanded, and included in Sect. 2.

- P9, L26-27: 'Picophytoplankton...' reference needed, and consider SC-3
- P10, L33-34: 'The largest PFT...' reference needed and consider SC-3

We have now detailed the pigments considered in Sect. 2, with references.

- P10, L9-14: consider SC-3

We have moved this to Sect. 2 as requested.

- P12,L15-21: belongs to section 4.2?

The comparison against IBTS chlorophyll was included in section 4.3 rather than 4.2 as we felt it fitted better with the validation of PFTs, which are a function of the IBTS chlorophyll data.

- P13, L24: NEMO-ERSEM chlorophyll is not so good really?

We have reworded this.

- P16, L6: specific issues =? please expand

We have reworded this to "results from this current study should further inform future model development."

- Table 2: are the question marks in the NEMO-ERSEM parameters typographic errors?

The question marks in Table 2 are not errors, but we accept their meaning is unclear and have modified the table accordingly. They were intended to indicate parameters for the four model variables, we have changed this to the notation (e.g.) P1 … P4.

- multiple occurrences of 'less good': not wrong, but wouldn't 'not as good' be more natural?

We have changed these as suggested.

- P24, L9: reformat Thorpe et al."

We have done this.

Detailed response to Reviewer 3

"The authors do not thoroughly investigate the differences in the physics or the nutritive environment and instead focus on the functional types chlorophyll, where as these aspects of the model are crucial to understanding the origins of the divergences between these two similar models. The two versions of ERSEM are structurally very similar, and so any diverges must be due to the physical environment, the river inputs, the SPM and light models or the parameterisations. The authors have not convinced me that the differences in community structure aren't due to the physical environment or the river influx, yet they conclude that the differences are due to the sensitivity to light and nutrients parameterisation."

As noted above, given the many differences in configurations, a thorough investigation of physical and other model processes was not our aim. This study was intended as an initial assessment of whether off-the-shelf model versions can reproduce these data, with more detailed comparisons between the models left for future studies – we have made this clearer in the manuscript. That said, we have expanded our discussion of the differences in results, and justified more clearly our conclusions. See below for details. We have also added text to Sect. 5 further considering the possible role of the physics, and suggesting further experiments to investigate this.

"For instance, the GETM-ERSEM-BFM has the productivity set much higher than NEMO-ERSEM, but this appears to be compensated by much higher river nutrient fluxes than NEMO-ERSEM. This extra riverine source of nutrients leads to the stable

diatom growth in the river-influenced regions. In addition, the central parts of the North Sea that do not feel the influence of the rivers have extremely low values for total chlorophyll, as the nitrate has been completely depleted there due to the high affinity of picophytoplankton. This suggests that the GETM-ERSEM-BFM community structure may be right, but for the wrong reasons. A thorough investigation of the river input, and the relationships between nutrients, chlorophyll and community structure for the two models may be needed to identify compensating errors."

We have checked the river inputs for the GETM-ERSEM-BFM model, and found these to be correct. More generally we agree, and have added a consideration of these points to the discussion in Sect. 5. However, we would prefer to leave such investigations for future studies, with dedicated experiments performed in a configuration designed for such a comparison, rather than using these pre-existing runs. We have therefore included in the discussion section a more comprehensive description of future work that should be performed as a result of this study.

"Otherwise, I find that this paper to be an adequate comparison of the data and the model, but I was hoping to see more depth in the discussion and conclusions section. For instance, this paper points out that the NEMO-ERSEM simulation struggles to reproduce the August diatom concentrations seen in both the data and the other model, but does not further investigate the root cause of the differences. After providing such an in-depth description of the divergences in between the model parametrisations in section 3.3 and Tables 1 and 2 , the authors are letting themselves down by not even speculating on how specific model features could bring about the observed differences in model output. For instance, what features are in one model but not the other that could cause this? If these questions are not answerable with the current set up, perhaps the authors could propose an extension to the work that would highlight the major cause of the differences, be it nutrient affinity, nutrient stress, river influx, stratification, sinking, lysis or light susceptibility. A sensitivity study in a 1D column may be a possible extension that could answer the remaining questions about the ERSEM parametrisations."

We have added to Sect. 5 a discussion of the root causes of the inter-annual variability mentioned, as well as further discussion of the other points raised. We have also added a paragraph suggesting further experiments which would extend the work accordingly.

"The title should be revised, removing Biodiversity and "ERSEM-type models". As the authors mention in the introduction, phytoplankton community structure is an important consideration when assessing marine biodiversity, but it certainly not the only indicator. Furthermore, ERSEM-type models is not strictly defined. Please consider something like: "Observing and modelling phytoplankton community structure in the North Sea: a comparison of two biogeochemical models", but I'll let the authors come up with something better."

We have removed "biodiversity" and "ERSEM-type models", shortening the title to simply "Observing and modelling phytoplankton community structure in the North Sea".

"There is only a shallow validation of the underlying physic model validation is shown. A single figure demonstrating that the SST matches is not particularly convincing by itself, especially as one of the models already assimilates SST. In the North Sea, it is important to demonstrate that the physical model that are capable of reproducing natural vertical mixing with appropriate sources of river influx (and river nutrients). This can be done by introducing a validation of the models stratification (perhaps Mixed layer depth) and surface salinity (as a proxy for influx of fresh water). The focus of the paper should remain on the PFT's, but a demonstration of the modelled physics would allow readers to gauge how realistic of an environment the two ERSEMs live in."

We have expanded the validation to include some assessment against in situ observations of salinity and nutrients from the IBTS cruises.

"I find the figure and table captions to be too brief. I would prefer having more information in them. Typically the rule of thumb is that it needs to be enough to describe the figure/table, were it taken out of context."

We have expanded these where appropriate.

"The text in the figures can be a little on the small side and the figures are inconsistent in terms of style, legends, axes and plotting range."

We have replotted many of the figures to address these points.

"L13-L16. "A comparison of the ...is key to capturing the observed biodiversity". While I agree with this statement (after replacing "biodiversity" with "Phytoplanton Community Structure"), I'm not convinced that the work establishes it as a fact. Some changes to the discussions section are needed for it to be the case."

We have both reworded this sentence and expanded the discussion section.

"P7-L18: "PML-ERSEM" is used from this point onwards, but should be explicitly defined in section 3.2."

For clarity, we have removed reference to "PML-ERSEM", and simply used "NEMO-ERSEM".

"P8-L8-L9: "More explicit approach to sinking". What does this mean? As the impact of sinking SPM underpins one of the important results of the paper, it would be good to have a more in-depth description of sinking in the two models. Also, the detrital sinking rates (if they are explicitly defined), should be presented here or in Table 2."

We have added this to the description of the models in Sect. 3, and added the sinking rates to Table 2.

"P8-L9-L11: The light susceptibility parameters in table 2 are not directly comparable, due the differences in model choice described here. Can they instead be converted into some other measure of light susceptibility in order to make them comparable? Alternatively, please include the equations to allow readers to draw their own conclusions."

We have included in Sect. 3 a more detailed description of the light susceptibility methods of each model.

"P10-L14. Does using the WOA data as validation and as a boundary conditions have an impact on the statistical independence of the validation? Similarly, figures 4, 5 and 7 have what looks like an edge effect in the GETM-ERSEM-BFM plots along the Northern Edge. The high chlorophyll growth there in figure 5 suggests that there is an influx of nutrients from the edge condition, especially as the central region of the North sea has been completely nitrogen depleted. Dp any of the IBTS measurements sit in a region influenced by edge effects?"

Yes, this will potentially affect the independence of the comparison, and we have added mention of this to the manuscript. However, we would not expect it to have too great an influence on broad-scale patterns across the North Sea, and the WOA values inside the

domain are not used by the model. There are edge effects, but the most northerly IBTS data (within the GETM-ESREM-BFM domain) are located just south of the main region of influence (see e.g. Fig. 14 for positions), so the effect on results should not be major.

"P11-L12: "both models were still able to capture the main observed features throughout the domain". I disagree. NEMO-ERSEM does not capture the high chlorophyll values along the German coastline, and otherwise appears relatively okay. GETM-ERSEMBFM also does not capture the high chlorophyll in the German coastline (but does a better job than NEMO-ERSEM), but has far too little chlorophyll by a factor 10x-20x over most of the North Sea. This model also has far too much chlorophyll at the mouth of the Great Ouse river, at least a factor of 10x, perhaps up to 100x, but it's hard to tell because of the colour scale. The colour scale is hiding quite a lot of the variability in high chlorophyll values. The legend is logarithmically even between 0.1 and 5., but then adds what should be three colour bands into one (ie 5-10, 10-20,20-50 are all hidden in 5-50.) My suspicion is that using a logarithmically even colour scale for this figure will result in a change the description in the section of the paper."

We have replotted the figure accordingly, although the general picture remains the same. Nonetheless, we have modified this line of the text accordingly.

"P11-L25-L28 and figure 7. The WOA datasets seem a little coarse for this analysis, and are already used as boundaries conditions for both models. ICES have extensive nitrate, phosphate and silicate datasets available for the North Sea, which could be converted into a climatological dataset for future works."

We have included some validation against in situ nutrient data also collected on the IBTS cruises. As far as we are aware, the WOA datasets already make use of ICES data.

"P11-L30: "Both [nitrate plots] are in broad agreement with the climatology. Likewise for phosphate and silicate". I do not agree with this statement. To me, it looks like GETM-ERSEM-BFM is entirely nitrate depleted over most of the model region, by a factor of between 10x-100x. In addition, the regions around the south-eastern North Sea have far too much nitrogen (factor 10x). These regions coincide with the largest diatom concentrations, and it looks like diatom growth is strictly governed by excess or deficiency of nitrate. This suggests the presence of compensating errors, and may imply that the diatom community structure is right for the wrong reasons. Similarly for phosphate, GETM-ERSEM-BFM has significantly lower by around 10x than WOA, and NEMO-ERSEM fails to capture the depleted area around the Norwegian Trench. In terms of silicate, NEMO-ERSEM has significantly depleted the silicate in the northern North Sea, which is surprising, as there are not enough diatoms in the community structure."

We have reworded this statement, and added further discussion to the manuscript to consider these points.

"P12-L14: "are likely to have arisen from the differences in parametrisations." Is this true? Could some of the differences in the models also arise from differences in the physical models? For instance, if one model's mixing or riverine input results in lower surface silicate, then a difference in diatom behaviour would be expected."

This is possible, we have expanded the discussion and justified further our reasoning. We have also added further text to Sect. 5 to acknowledge the possible role of physics, and proposed further experiments to investigate this.

"P12-L31: The ternary plots to describe the three population community structure are

unique to this work, and may deserve their own subsection (up to the authors discretion)."

We feel these are best kept where they are, but have improved on the colours used as noted above.

"P12-L32-33: "The observations form a distinct line in this space." Surely there are some interesting findings that are a direct result of this straight line? Can it be used to inform predictions? Could you fit a predictive function to this line and compare it with the three population model?"

We have added text to Sect. 4.3 discussing this. Essentially it demonstrates that Nano and Pico remain very similar to each other, when Micro changes. This differs to some extent from the three population model.

"P13-L5: Knowing about the distinct straight line in the in situ data of figure 10, would it be sensible to revise the colour scheme. How about something like Green: along the line, red: too much pico; Blue: too much nano? This would highlight the regions where the model sits along the accepted line seen the the data, and where it is wrong elsewhere."

We have considered this, but felt the results to be less clear, particularly in terms of highlighting spatial patterns.

"P14-L15: Why is August a challenging month to model the in North Sea?"

We have expanded on this in the manuscript.

"P14-L19-L31: Please add a clear link and some discussion about the differences in the light parametrisations of the two models in Table 2. The differences in light susceptibility between the two models may be quite significant, but its hard to tell because of the incomparable parameters. This should help support one of the major points that you make in the abstract (": : : changes in light and nutrients, is key to capturing the observed biodiversity.")"

We have added further description and discussion of the light parameterisations.

"P15-L3-L8: This is the most important part of the discussions section, and it would very informative to discuss how different model choices could have brought about the changes between these two similar simulations. (see general comments)."

We have expanded the discussion, see specific responses above.

"P15-L25: After this point, the conclusions risk becoming a little hand-wavey, nonspecific and generic. For instance "close communication is needed between modellers and decision makers, so that the potential of biogeochemical models to support decision making is not lost in translation." While true, is not directly related to the conclusions of the study."

We have deleted most of this paragraph and added/reworded other discussion.

"16-31 – p17-2. The closing paragraph will need to be changed to reflect the new title. "Can these two ERSEM variants simulate phytoplankton community structure? ...""

We have changed this accordingly.

"Regarding the figures, there is some inconsistency in the subplot titles and styles. Ie. Figure 2 has no sub plot titles, but figures 4,5,6,7,8 all do. The legends are in different places and have different styles. Ie Figure 5 uses discrete coloured circles, where as figure 2 uses rectangles. I prefer the larger rectangles or a color bar. (Congratulations are due for the consistent use of a single colour scheme.)"

We have replotted many of the figures to improve consistency.

"Please add a figure showing both model domains, the north sea domain used here and both cruise paths."

We have added a figure showing the model domains. The North Sea domain used for validation is the same as the GETM-ERSEM-BFM model domain. The cruise observation locations can be seen in Fig. 2 (formerly Fig. 1) and other figures.

"Figure 1: The axis labels, units, location markers are very small and hard to see. Please thicken the outline of the in situ data circles in a) and b) as in Figure 11. Also, can you please make the location makers (the capital letters) stand out more. You could do this by adding a white shadow to the letter edge or by adding an opaque text box. None of the subsequent maps include grid lines or latitude/longitude ticks, was there a reason why these plots do? If the goal is to introduce the readers to the domain, perhaps you would be better served by adding a map showing both model domains, the north sea domain used here and both cruise paths? The legend is above the plot axes, where as in all other plots it sits on the right of the plot. Elsewhere, the chlorophyll is shown using logarithmic scale, but c) and d) use a linear scale. Parts c) and d) are described but not referenced again. What was the motivation of including two times series plots?"

We have replotted this figure accordingly. c) and d) are intended to compare the different in situ sampling methods of chlorophyll, as well as variability.

"Figure 2: This plot has a very different style to the other plots. The sub-plots have no boarder, the coastlines are much higher resolution, they show rivers, the land is a different colour, and the coastline line is much thinner."

We have replotted this figure accordingly.

"Figure 3: The "closest to the reference data" point of a Taylor plot is on the x-axis at avalue of 1. This point is usually marked with a dot, or the axis label "ref". Please add that point or label to this figure."

We have added this.

"Figure 5: This figure shows the satellite Chlorophyll and both models, but it may benefit from showing the in situ data as well in another two subplots. The colour scale hides quite a lot of variability here. The legend is logarithmically even between 0.1 and 5, but then adds what should be three colour regions into one (ie 5-10, 10-20,20-50 are all hidden in 5-50.)"

We have modified the colour scale. We decided against including the in situ data as we felt that would duplicate the information in Fig. 1 (now Fig. 2).

"Figure 6: This figure and the associated description in p10-L20-23 may benefit from a logarithmic colour scale."

We have changed the colour scale accordingly.

"Figure 8: I find myself flipping back and forth between this figure, figure 2 and figure 5. There are two gaps in the 4x3 grid. Would it be possible to put total chlorophyll for each model in there? Alternatively, would it make more sense to add a second version of this figure with the three populations instead of the 4 and 5 PFTs and also the in situ data from figure 2?"

We have considered these different options, and feel the current layout is the best compromise of clarity, conciseness, and lack of duplication.

"Figure 9: The subplot shape is not the best here. Tall and narrow subplots seems to mask the overall shape of the data. If they were short and wide, it may make it easier to inspect the data. As mentioned above, have you considered overlaying the three population fits of of Devred 2011, Hirata 2011 or Brewin 2015 onto subfigures a) and d) of this figure, ? This would allow a comparison of how the North Sea compares with the rest of the global ocean, as in figure 11 of Holt 2014 or in de Mora 2016. With only 85 data points you may struggle to apply the three population model to your three data sets, but it should be possible to apply the three population model to the full (not subsampled) model datasets. This could be overlayed onto subfigure b), c) e) and f). Is it even necessary to sub-sample the model data to match the in situ measurements? Judging from figure 1, the cruises have done a great job at getting an spatially even distributed data set of the North Sea. Could you comment on the detection limit of the HPLC data? The GETM-ERSEM-BFM model seems to have values much lower than that seen in data. Is that a detection limit effect not present in the model?"

We have replotted these to be short and wide. We feel that sub-sampling the model data is important for a robust comparison with the observations. However, we agree that this limits the number of data points, so have included an additional figure using all the model grid points, showing a 2D histogram equivalent to de Mora et al. (2016). This will make it easier to compare our results with those from other studies. The HPLC data has a detection limit of 0.01 mg m^{-3}, but all observations were above this limit, hence why no cut-off was applied to the model data.

"Figure 10: Can you please change two of the circular markers to squares and triangles, like in figure 9 so that these plots work in grey scale too?"

We have done this.

"Figure 11: This plot is the only figure in the work with a continuous colour scale, but it is the figure that would benefit the most from a discrete colour scale. In grey-scale, the legend becomes very uniform. See above comments about an alternative colour scale."

We have changed this to be discrete, and also modified the colour scale.

"Table 1: Please add a "further reading" or "references" line with at least one reference for each model."

We have added this.

"Table 2: In the header line, please replace ERSEM's internal PFT's names (P1-P4) with the public facing names used elsewhere in the paper."

We have modified the headers accordingly.

Table 3: This table only shows total Chlorophyll, but it would be valuable to look at similar statistics for the rest of the community structure. ie pico, micro and nano in both models vs the IBTS data (but only for 2010+2011, you wouldn't need both years separately)"

We have added these statistics.

"Technical Corrections
Inconsistent use of "e.g." in the introduction section, especially when citing other works. My preference is to avoid it altogether. p2-L3: "e.g.", p2-L14: "e.g.".', p2-L21: e.g. p2-L33 e.g. appears twice.
p2-L2: the scientific community,
p2-L5: (MICE) acronym, while clever, is never mentioned again, so is not needed.
P2-L6: remove "though"
p2-L6: ...task at hand – different scientific.... The dash (-) should be an em dash, a semi-colon or even a full stop.
P2-L8: colloquial: "say, global-scale...".
p2-L14: van de Molen is spelt with a lower-case v in the authors list but with a capital V elsewhere in the document and in the reference list. Is this a conscious choice? Similarly for other names: de Mora, van der Woerd, van Raaphorst, and van Leeuwen are also spelt differently in various places.
p2-L25: I don't think you need to explicitly define the "(UK)" just as you correctly don't define USA later in the same sentence.
P2-L28-L30, please re-write this sentence with more clarity.
P3-L13: replace "too" with "as well".
p6-L2,4 ,6, 25: Are these web addresses the best way to include this information? If so, please cite them instead of in-line using the Copernicus standard styles: http://publications.copernicus.org/CopernicusPublicationsReferenceTypes:pdf
p3-L10, p3-L28: Would it be better to move both copies of the sentence "Details of the model configuration and forcing are given in Table 1." to the end of the introductory paragraph at p5-L9?
P6-L22: Please reverse FOAM definition and acronym: "Forecasting Ocean Assimilation Model (FOAM)"
P7-L6-L9: Long sentence, please split into two: "reanalysis period. The biogeochemical..."
P12-L3: Please add a reminder about how the model PFTs were aggregated. Something like: "as described in section 3.4".
P14-L17: change strange wording: "confidence must be had in".
p16-L21: The dash (-) should be an em dash.
P16-L27: Rephrase this sentence to re-word "for confidence to be had in them".
P16-L33: replace "current" with "these two".
p34 Figure 9: Phytoplankton Size Class instead of PSC in the figure caption.
P39 Table 33 caption: Statistical comparison of log 10 (chlorophyll) against IBTS observations."

We have addressed all these.

<u>Detailed response to M. Butenschön</u>

"However, with respect to the speculation on the reason of differences in comunity structure in the two models already discussed by the referees, I'd encourage the authors to take into consideration the insights of Sinha et al. 2010, demonstrating the sensitivity of the modelled ecosystem food-web structure to the physical driver alone. These findings are potentially in contrast to the hypothesis that the causes for differences in comunity

structure are mainly in the parametrisation of the biogeochemical model, which therefore requires a much more solid base (e.g. simulations using identical physical driver and external forcings)."

As suggested by the other referees, we have expanded our discussion to further consider the roles of different drivers, and have taken the work of Sinha et al. (2010) into account as part of this discussion. We have also made clear that further experiments are likely required to come to firm conclusions on this issue, and have outlined potential future work.

"Furthermore, it should be noted that one of the model versions used in this work (referred to as PML-ERSEM) is based on a parametrisation published in 2004. Since then, other versions and parametrisations of this ERSEM flavour have emerged (e.g. the recent works of Butenschön et al. 2016, Ciavatta et al. 2016 and de Mora et al. 2016) that show a distinctively different comunity structure than the parametrisation used in this work (even if for different configurations, see figures 2 and 3 in de Mora et al. 2016 and figure 10 in Butenschön et al. 2016). While I wouldn't expect the authors to repeat their work with a different parametrisation, the fact that the used parametrisation (which is the established one still in use in the operational suite of the UKMO) has evolved since its publication clearly deserves mentioning as these new parametrisations are openly available in the published, peer-reviewed scientific literature."

We have expanded the discussion about recent developments which have been made to ERSEM, particularly within the SSB framework, and the results that have been obtained from this work. We have also outlined future work that could address the points raised in our study, and which would likely best be done within the SSB modelling framework that you and others have developed.

Observing and modelling phytoplankton community structure in the North Sea: can ERSEM-type models simulate biodiversity?

David A. Ford[1], Johan van der Molen[2], Kieran Hyder[2], John Bacon[2], Rosa Barciela[1], Veronique Creach[2], Robert McEwan[1], Piet Ruardij[3], and Rodney Forster[2,4]

[1]Met Office, FitzRoy Road, Exeter, EX1 3PB, UK
[2]Centre for Environment, Fisheries and Aquaculture Science (Cefas), Pakefield Road, Lowestoft, NR33 0HT, UK
[3]Royal Netherlands Institute for Sea Research (NIOZ), Department of Coastal Systems and Utrecht University, PO Box 59, 1790 AB Den Burg, Texel, The Netherlands
[4]University of Hull, Cottingham Road, Hull, HU6 7RX, UK

Correspondence to: David Ford (david.ford@metoffice.gov.uk)

Abstract. Phytoplankton form the base of the marine food chain, and knowledge of phytoplankton community structure is fundamental when assessing marine biodiversity. Policy makers and other users require information on marine biodiversity and other aspects of the marine environment for the North Sea, a highly productive European shelf sea. This information must come from a combination of observations and models, but currently the coastal ocean is greatly under-sampled for
5 phytoplankton data, and outputs of phytoplankton community structure from models have therefore yet to be properly are therefore not yet frequently validated. This study presents a novel set of in situ observations of phytoplankton community structure for the North Sea using accessory pigment analysis. The observations allow a good understanding of the patterns of surface phytoplankton biomass and community structure in the North Sea for the observed months of August 2010 and 2011. Two physical-biogeochemical ocean models, the biogeochemical components of which are different variants of the
10 widely-used European Regional Seas Ecosystem Model (ERSEM), were then validated against these and other observations. Both models were a good match for sea surface temperature observations, and a reasonable match for remotely sensed ocean colour observations. However, the two models displayed very different phytoplankton community structures, with one better matching the in situ observations than the other. Nonetheless, both models shared some similarities with the observations in terms of spatial features and inter-annual variability. A An initial comparison of the formulations and parameterisations
15 of the two models suggests that diversity between the parameter settings of model phytoplankton functional types, along with formulations which promote a greater sensitivity to changes in light and nutrients, is key to capturing the observed biodiversityphytoplankton community structure. These findings will help inform future model development, which should be coupled with detailed validation studies, in order to help facilitate the wider application of marine biogeochemical modelling to user and policy needs.

1 Introduction

Marine biogeochemical model complexity has long been a subject of debate (e.g. Anderson, 2005). Simpler models require fewer, often better understood parameterisations, but omit processes which are known to be important. More complex models explicitly include these processes, but require an increased number of tuneable parameters, the ranges of which are often poorly constrained by observations and/or poorly defined where they aggregate over a range of species. No consensus exists within the scientific community, but recent studies have shown that simple to moderately complex models still do the best job of reproducing basic biogeochemical descriptors such as primary production and carbon fluxes (Kwiatkowski et al., 2014; Friedrichs et al., 2007; Ward et al., 2013; Xiao and Friedrichs, 2014). Further studies have suggested that "Models of Intermediate Complexity for Ecosystem assessments" are the most appropriate for fisheries management (Plagányi et al., 2014).

Ultimately, models are tools, and the most appropriate tool should be chosen for the task at hand; different scientific, societal and managerial questions will require models of different complexities. For instance, describing a complex coastal environment will likely require the explicit inclusion of processes which are less important when considering global-scale carbon budgets. Furthermore, some users require more detailed information about the marine environment than simple models can provide, necessitating the use of more complex models.

This demand for detailed information applies to the North Sea, with users and policy makers requiring information about topics including eutrophication and nutrient ratios (Skogen et al., 2014), productivity in relation to fisheries (Chassot et al., 2007), harmful and nuisance algal blooms (Blauw et al., 2010; Kurekin et al., 2014), water clarity (Dupont and Aksnes, 2013; Capuzzo et al., 2015), biodiversity (Brandsma et al., 2013), effects of climate change (van der Molen et al., 2013; Wakelin et al., 2015a), effects of trawling (Allen and Clarke, 2007; van der Molen et al., 2013), and impacts of marine renewable energy generation (van der Molen et al., 2014; van der Molen et al., 2016). In particular, indicators of Good Environmental Status (GES) are required in the context of the Marine Strategy Framework Directive (MSFD; Borja et al., 2013). These include descriptors of food-web structure, trophic status, and biodiversity, and elements of these can be assessed with various modelling approaches (Piroddi et al., 2015; Hyder et al., 2015).

Ecosystem models are central to the delivery of marine ecosystem-based management that is specified in existing legislation (MSFD (EU, 2008), CFP (EU, 2013), WFD (EU, 2000)). These models are important for the design of management measures and to assess the social, economic and environmental performance of management in relation to targets (Defra, 2014; Sutherland et al., 2006). This is done through improving our understanding of the links between pressures (human and environmental) and the response of the system to these pressures. However, ecosystem models are not used frequently in the UK and Europe in support of policy and management (Hyder et al., 2015), despite increasing use in the USA and Australia (Fulton and Link, 2014; Fulton et al., 2007). For models to have a larger impact on policy development and decision-making, modelling approaches need to be more transparent, verifiable, and repeatable than they are at present, as any outputs can be subject to legal challenge (Hyder et al., 2015). Poor uptake of ecosystem models by decision-makers is due to a lack of confidence in and understanding of models. This relates to how

models are used, terminology, ~~modelling terminology, outputs in the right currency~~type of outputs, treatment of uncertainty, ~~rigorous~~ required quality standards, and the presentation of model products (Hyder et al., 2015). The use of ecosystem models will become increasingly important as the complexity of marine legislation increases (~~e.g.~~ Boyes and Elliott, 2014). Hence, simple assessment of the skill of models in predicting outcomes (validation - ~~e.g.~~ Mackinson, 2014), model comparisons (e.g.

5   Kwiatkowski et al., 2014), and the clear treatment of the uncertainty associated with predictions (~~e.g.~~ Thorpe et al., 2015; Gårdmark et al., 2013; Stewart and Martell, 2015; Tebaldi and Knutti, 2007) are needed to increase the confidence in and uptake of models (Hyder et al., 2015).

At the base of the marine food chain are phytoplankton, and phytoplankton community structure is a fundamental consideration in any assessment of marine biodiversity (Garmendia et al., 2013). Changes in community structure can result

10   from large-scale environmental changes such as temperature rises or eutrophication, with different organisms favouring different conditions. Some organisms that favour changed conditions may be harmful to human health (Roselli and Basset, 2015; Bruggeman, 2009). Alternatively, top-down control by benthic or pelagic grazers can change the size-structure of phytoplankton by selective removal of larger species, resulting for instance in an increased proportion of pico-phytoplankton in areas with dense shellfish aquaculture (Smaal et al., 2013). Phytoplankton vary in size by up to nine orders of magnitude for cell volume

15   (Finkel et al., 2010), with variations in community structure reflected in the size and species of their predators, and the number of links in the food chain (Ryther, 1969; Chavez et al., 2011). Larger cells such as diatoms are consumed directly by copepod grazers, giving a higher transfer of energy and ultimately impacting commercial fish stocks (Jennings and Collingridge, 2015). As the physical structure of the North Sea becomes increasingly well understood due to advances in hydrodynamic modelling (~~Van~~ van Leeuwen et al., 2015) and availability of long-term observations (Greenwood et al., 2010; Núñez-Riboni

20   and Akimova, 2015), the potential to predictively model plankton population structure and distribution increases ~~too~~as well.

A common way to model plankton community structure is to take a phytoplankton functional type (PFT) approach, such as is done in variants of the European Regional Seas Ecosystem Model (ERSEM; Baretta et al., 1995). This approach groups phytoplankton into a number of PFTs, based on their general function within the ecosystem (Le Quéré et al., 2005). If information on phytoplankton community structure is to be modelled and provided to users, then it must be validated. ~~Commonly though,~~

25   Some studies have aimed to validate this against observations (Lewis et al., 2006, Gregg and Casey, 2007; Lewis and Allen, 2009; Hirata et al., 2013), but commonly validation studies go no further than total chlorophyll concentration (~~e.g.~~ Edwards et al., 2012; ~~De~~ de Mora et al., 2013). This is largely because there is a lack of observations that contain more detail about community structure against which to compare. Algorithms for deriving phytoplankton community structure from remotely sensed satellite ocean colour observations, either in the form of PFTs or phytoplankton size classes (PSCs) are being developed

30   (Brewin et al., 2011; Brito et al., 2014), but have not yet reached maturity and are not yet widely available to the general scientific community. Moreover, such remote sensing products require a similar level of validation (Brotas et al., 2013). In situ observations are sparse, particularly in shelf seas, and the measured variables may not be easily matched to model outputs, which do not always aggregate neatly over species or size classes.

This study presents a novel set of in situ observations of phytoplankton community structure in the North Sea using accessory

35   pigment analysis (Sherrard et al., 2006), noting that coastal seas are greatly under-represented in the existing global collection

of pigment data (Peloquin et al., 2013). Pigment data were analysed so as to give the relative distribution of different size classes, allowing a robust comparison with outputs from ERSEM-type models. Two variants of ERSEM, run by the Centre for Environment, Fisheries and Aquaculture Science (Cefas) and the Met Office, both public bodies in the UK, were validated against these and other observations. The aims of the study were to determine what these new observations add to current

5 scientific understanding of North Sea biogeochemistry, assess the extent to which the models can reproduce the observations, and discuss the implications for current and future user and policy needs, observing strategies and model development.

2 Observations

2.1 International Bottom Trawl Survey (IBTS)

The International Bottom Trawl Survey (IBTS) is a multi-national ecological research effort established by the International

10 Council for the Exploration of the Sea (ICES) in the early 1970s. Surveys using fisheries research vessels currently take place in the first and third quarter of the year and cover the entire North Sea, using standardised sampling gears and protocols. With cruise lengths of typically 6-8 weeks, each vessel undertakes a gridded survey of the North Sea, repeated each year, in which stations are sampled for groundfish (the primary target of the survey), but also secondary targets such as benthos, seabed litter, and hydrographic parameters. Individual station sampling is often accompanied by visual seabird and cetacean

15 estimates, underway acoustics, and online monitoring of near-surface water quality using FerryBox-type instruments (Petersen et al., 2008). The IBTS thus fits the needs of a multi-disciplinary survey capable of collecting data on human pressures and ecosystem responses for legislation such as the MSFD (http://www.jpi-oceans.eu/multi-use-infrastructure-monitoring). The open data policy of ICES has resulted in many significant publications in fisheries research (Jennings et al., 2002; Daan et al., 2005) and fisheries policy (Rombouts et al., 2013; Shephard et al., 2015).

20 Prior to 2010, phytoplankton had not been systematically sampled on the UK IBTS. Advances in the autonomous sampling and detection of particles in the water column (e.g. online flow cytometry (Thyssen et al., 2015)), and also the need for high-quality in situ data for validation of satellite remote sensing data, indicated that the addition of phytoplankton to the survey would be beneficial. Hence, sampling of PFTs using high-pressure liquid chromatography (HPLC) - pigment fingerprinting, and analytical flow cytometry (results reported elsewhere) were initiated on the third quarter IBTS cruise of the RV *Cefas*

25 *Endeavour* in August-September 2010 and subsequent years.

Seawater samples from depths of 4 m ('surface') were collected using 10 l Niskin bottles when weather conditions permitted, or from the ship's bow-intake flow-through clean seawater supply during adverse weather conditions. A known amount of water, typically 1000 ml, was passed through a 200 μm gauze to remove larger zooplankton and debris, then filtered within 1 h on 47 mm GFF filters, which were folded in half, wrapped in aluminium foil and snap frozen in liquid nitrogen dry shippers.

30 On return to shore, samples were transferred to a -80 °C freezer for a storage period of 1-2 months before shipping of samples on dry ice to an accredited HPLC laboratory (DHI Water Quality Institute; Horsholm, Denmark) for chlorophyll-a (chl-a) quantification and full accessory pigment analysis (Schlüter et al., 2011).

Pigment data from the surface stations were quality data controlled in several steps: first, with an initial comparison of HPLC chl-a against independent measures of chlorophyll fluorescence from the fluorometers on the ship's Ferrybox and CTD system. This step corrected a small number of mis-labelled samples. In a second step, anomalies within a sample were detected using methods described by Aiken et al. (2009), e.g. regression of total accessory pigments against chl-a concentration and search

5   for outliers.

Diagnostic pigment analysis was then used on the quality-controlled data set to relate the composition of specific accessory pigments to the relative contribution of different size classes to the total phytoplankton biomass. The designation of specific accessory pigments to algal taxonomic groups of different size, e.g. fucoxanthin and peridinin for large-cell diatoms and di-noflagellates, has been widely established in the biological oceanographic literature (Uitz et al., 2006, 2008). The equations

10 used to estimate the contribution of pico-phytoplankton (0-2 μm), nano-phytoplankton (2-20 μm) and micro- or net phyto-plankton (>20 μm) were later modified by Hirata et al. (2008, 2011) and Brewin et al. (2010). The various methods differ in the degree to which the marker pigments chlorophyll-b (chl-b) and 19-hex-fucoxanthin (19-hex) are attributed to the three size classes. Here, chl-b and 19-hex were assigned equally to the pico-phytoplankton and nano-phytoplankton size classes. Pico-phytoplankton are therefore represented by zeaxanthin, chl-b and 19-hex; nano-phytoplankton are represented by 19-hex,

15 19-but, alloxanthin and chl-b; and micro-phytoplankton are represented by fucoxanthin and peridinin. Results are expressed as a proportion of the total chl-a concentration for each station.

2.2   Other validation data

As well as the IBTS data introduced in this study, other observation-based products have been used for model validation. Sea surface temperature (SST) has been validated against OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis;

20 Donlon et al., 2012), which is an objective analysis product based on remotely sensed and in situ SST observations. Sea surface chlorophyll and suspended particulate matter (SPM) have been validated against remotely sensed ocean colour products from the Medium Resolution Imaging Spectrometer (MERIS) and Moderate Resolution Imaging Spectroradiometer (MODIS) sensors, developed by Ifremer using the OC5 algorithm (Gohin et al., 2002, 2005, 2008). Due to the limited availability of observations, nutrient concentrations have been validated against the 1° resolution World Ocean Atlas climatologies (Garcia et

25 al., 2010).

3   Models

Two different physical-biogeochemical modelling systems were used in this study: GETM-ERSEM-BFM, run by Cefas, and NEMO-ERSEM, run by the Met Office. Each is described in turn below, followed by a discussion of their differences and similarities. Existing configurations of each model were used, with no attempt made to increase their similarity. Details of the

30 model configuration and forcing are given in Table 1, and the model domains and bathymetries are shown in Fig. 1.

3.1 GETM-ERSEM-BFM

GETM (General Estuarine Transport Model) is a public domain, three-dimensional (3D) finite difference hydrodynamical model (Burchard and Bolding, 2002; available through http://www.getm.eu). It solves the 3D partial differential equations for conservation of mass, momentum, salt and heat. The ERSEM-BFM (European Regional Seas Ecosystem Model - Biogeo-chemical Flux Model) version used here is a development of the model ERSEM III (see Baretta et al., 1995; Ruardij and ~~Van~~ van Raaphorst, 1995; Ruardij et al., 1997; Vichi et al., 2003; Vichi et al., 2004; Ruardij et al., 2005; Vichi et al., 2007; ~~Van~~ van Leeuwen et al., 2013; ~~Van~~ van der Molen et al., 2013; ~~Van~~ van der Molen et al., 2014; ~~Van~~ van der Molen et al., ~~2016a; http://www.nioz.nl/northsea_model~~ 2016), 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 six phytoplankton groups, four zooplankton groups, and five benthic groups, the latter comprising four macrofauna and one meiofauna groups. Pelagic and benthic aerobic and anaerobic bacteria are also included. The pelagic module includes a number of processes in addition to those included in the oceanic version presented by Vichi et al. (2007) to make it suitable for temperate shelf seas: (i) a parameterisation for diatoms allowing growth in spring, (ii) enhanced transparent exopolymer particles (TEP) excretion by diatoms under nutrient stress, (iii) the associated formation of macro-aggregates consisting of TEP and diatoms, leading to enhanced sinking rates and a sufficient food supply to the benthic system especially in the deeper offshore areas (Engel, 2000), (iv) a *Phaeocystis* functional group for improved simulation of primary production in coastal areas (Peperzak et al., 1998, Ruardij et al., 2005), (v) a new ~~suspended particulate matter (SPM) resuspension module~~ resuspension module for inorganic fine SPM that responds to combined currents and surface waves, and uses a concentration-dependent settling velocity for improved simulation of the under-water light climate (~~Van~~ van der Molen et al., ~~2016b~~ 2017), and (vi) resuspension of particulate organic material~~coupled to the SPM model~~, in proportion to the resuspended inorganic SPM and the relative concentrations of organic and fine inorganic matter in the sea bed. The model includes a 3-layer benthic module comprising 53 state variables, which enables it to resolve a high level of detail of benthic processes and benthic-pelagic coupling. New features of the benthic model are: (i) benthic diatoms, and (ii) active oxygen uptake of deposit feeders from the water column. The first four additional pelagic processes listed above are related, and based on detailed implementation of the dynamic model of phytoplankton growth, explicit chlorophyll content and acclimation of Geider et al. (1997). In nutrient-enriched coastal zones, the competition between and seasonal succession of PFTs is influenced strongly by differences in their photosynthetic capability. The modelled photosynthesis and phytoplankton carbon and chlorophyll content follows Geider et al. (1997) closely, by first calculating light-saturated and nutrient-replete photosynthesis. In the second stage, light-adapted chlorophyll content is calculated and light limitation and nutrient limitation are applied. These result in changes in the chlorophyll:carbon ratio, and growth. Photo-inhibition is included explicitly in the chlorophyll calculcations, and carbon uptake is calculated before applying nutrient limitation. Under nutrient-limited conditions, diatoms excrete the excess carbon as TEP, which is modelled as a separate state variable. *Phaeocystis* cells, implemented as a simplified version of the model of Ruardij et al. (2005), excrete TEP within the colony. Implicit macro-aggregate sinking rates are calculated as a linear proportion

of a prescribed maximum sinking rate, governed by stickiness rates related to the concentrations of diatoms, TEP and the level of nutrient stress, and also induce sinking of other PFTs. Because (Geider et al., 1997) the initial slope of the P-I curve α_{chl}, the light-saturated carbon-specific photosynthesis rate P_m^C, the saturation parameter for the growth-irradiance curve K_I and the maximum chlorophyll:carbon ratio θ_m are related through:

$$\alpha_{chl} = \frac{P_m^C}{\theta_m K_I} \tag{1}$$

and K_I can be approximated by the light intensity at maximum photosynthetic rate (K_E), K_E was used to prescribe light sensitivity. Values were selected to simulate observed successions in the Dutch coastal zone (Table 2).

The model setup for the North Sea uses a spherical grid with a spatial resolution of approximately 11 km and 25 layers in the vertical (~~see http://www.nioz.nl/northsea_model and also~~ Lenhart et al., 2010; van der Molen et al., 2014, 2015). The model was forced with tidal boundary conditions from a shelf-scale model, temperature and salinity boundary conditions from a global hindcast (ECMWF-ORAS4; Balmaseda et al., 2013; Mogensen et al., 2012~~; http://www.ecmwf.int/products/forecasts/d/charts/oras4/reanal~~), climatological nutrient boundary conditions, observations-based river run-off and riverine nutrient loads (the National River Flow Archive (data available at http://www.ceh.ac.uk/data/nrfa/index.html) for UK rivers, the Agence de l'eau Loire-Bretagne, Agence de l'eau Seine-Normandie and IFREMER for French rivers, the DONAR database for Netherlands rivers, ARGE Elbe, the Niedersächsisches Landesamt für Ökologie and the Bundesanstalt für Gewässerkunde for German rivers, and the Institute for Marine Research, Bergen, for Norwegian rivers; see also Lenhart et al., 2010), and atmospheric forcing from the ECMWF ERA-40 and operational hindcast (ECMWF, 2006a,b). ~~Details of the model configuration and forcing are given in Table 1.~~

3.2 NEMO-ERSEM

The hydrodynamic component of the Met Office modelling system is NEMO (Nucleus for European Modelling of the Ocean; Madec, 2008). NEMO is an open source community model originally developed for global ocean modelling (e.g. Storkey et al., 2010), but which has also been recently developed for use in shelf seas (O'Dea et al., 2012). The version used in this study (CO5; O'Dea et al., in prep.) is based on NEMO v3.4, and is a development of that described in O'Dea et al. (2012) and Edwards et al. (2012). The main updates from O'Dea et al. (2012) are an upgrade from NEMO v3.2 to v3.4, an increase in vertical resolution from 33 to 51 levels and change of coordinate stretching function, changes to the river inputs and Baltic boundary condition, a change of data assimilation scheme from analysis correction to 3D-Var, and the use of bulk formulae to calculate the input atmospheric fluxes rather than direct forcing. These updates are described further below.

The version of ERSEM used is an alternative development of the original code of Baretta et al. (1995), led by Plymouth Marine Laboratory (PML), and is described in detail by Blackford et al. (2004) and Edwards et al. (2012). The pelagic component includes four phytoplankton and three zooplankton functional groups, and one bacterial group. The benthic component includes aerobic and anaerobic bacteria, suspension feeders, bottom feeders, and the meiobenthos. This version follows the photoacclimation model of Geider et al. (1997) in an adapted form, in which nutrient limitation is applied before the other

calculations, leading to much lower estimates of excess carbon, which is excreted as detritus. Photoinhibition is included as an additional parameterisation in the photoacclimation method. SPM is simulated as per Sykes and Barciela (2012).

As part of the ~~FOAM (~~Forecasting Ocean Assimilation Model (FOAM; Blockley et al., 2014) suite of models, NEMO-ERSEM is run operationally at the Met Office on a daily basis, providing five-day forecasts of physical and biogeochemical variables for the North-West European Shelf Seas. Analyses and forecasts are publicly available through the Copernicus Marine Environment Monitoring Service (CMEMS; http://marine.copernicus.eu), which is the operational service building on the MyOcean project. Physical and biogeochemical reanalysis products (Wakelin et al., 2015b) are also available through CMEMS, and results from the recent NEMO-ERSEM reanalysis were used in this study.

~~Details of the model configuration are summarised in Table 1.~~ The model was run on the 7 km resolution Atlantic Meridional Margin (AMM7) domain, covering the entire North-West European Shelf Seas, including the North Sea. There are 51 vertical levels, using a hybrid σ-S coordinate system with the stretching function of Siddorn and Furner (2013). This uses terrain-following coordinates whilst ensuring a fixed surface resolution of 1 m. Remotely sensed and in situ observations of ~~sea surface temperature (SST)~~ SST were assimilated using a 3D-Var implementation of the NEMOVAR data assimilation scheme (Waters et al., 2015; O'Dea et al., 2012). River inputs were taken from the E-HYPE model (Donnelly et al., 2015) for flow values, and from the same climatology as in Edwards et al. (2012) for nutrients and ~~sediments~~SPM. Lateral boundary conditions for physical variables were taken from a reanalysis of the GloSea5 seasonal forecasting system (MacLachlan et al., 2014) at the Atlantic boundaries, and from the IOW-GETM model (Stips et al., 2004) at the Baltic boundary. For biogeochemistry, lateral boundary conditions for ~~nutrients~~ nitrate, phosphate and silicate were taken from the World Ocean Atlas monthly climatology (Garcia et al., 2010) at the Atlantic boundaries, and zero flux boundary conditions were applied at the Baltic boundary. Zero flux boundary conditions were applied for all other biogeochemical variables at all boundaries. Surface forcing was from the ERA-Interim reanalysis (Dee et al., 2011). The NEMO-ERSEM reanalysis covers the period January 1985 to July 2012, but for practical reasons was run in three sections. The final section, which this study uses, started in November 2003, with physics initial conditions taken from the corresponding date of a non-assimilative hindcast of the entire reanalysis period~~, and biogeochemical initial conditions~~. Biogeochemical initial conditions were taken from a winter date of the run of NEMO-ERSEM described in Edwards et al. (2012).

3.3 Comparison of the two models

Even though both models used versions of ERSEM, it is reasonable to expect differences in the results. Such differences are inevitably a result of the accumulation of differences between the models. It should be noted that both models were run as usual, and no attempts were made to increase similarity. It is recognised that this means definitive conclusions cannot be reached here on the exact reasons behind differences in results, and this was not the aim of this study. A preliminary discussion is provided here, with more detailed follow-on experiments proposed in Sect. 5. To help understand the differences in model behaviour, this section summarises the main differences between the two models. We focus on two types of differences: general level differences (Table 1), and differences in phytoplankton parameters and parameterisations (Table 2). For the sake of readability,

and to limit repetition, the following summary is kept at a fairly basic level; for (numerical) detail the reader is referred to the tables.

The two hydrodynamics models were different, and in general used different domains, resolutions, and forcing data. The NEMO-ERSEM model had a larger domain (Fig. 1), at higher resolution, and used more advanced atmospheric forcing. More-over, in NEMO-ERSEM, SST was assimilated, while GETM-ERSEM-BFM had no data assimilation. NEMO-ERSEM's river runoff originated from a model, that of GETM-ERSEM-BFM from observations. The GETM-ERSEM-BFM model used time series of riverine nutrient inputs whereas the NEMO-ERSEM model used a climatology. The SPM model of NEMO-ERSEM contained explicit size fractions and cohesive interactions, but was only forced by flow velocities, while that of the GETM-ERSEM-BFM model was non-cohesive, with implicit size-related behaviour and included resuspension by both currents and waves (~~Van~~ van der Molen et al., ~~2016b~~2017). The models also used different initial conditions and spin-up sequences.

Both ERSEM versions share a common origin, both use the same base nutrients (N, P, Si, C), and are both based on a functional type approach. They share four phytoplankton types, three zooplankton types, and a basic bacteria type. Both have a three-layered benthic module, with similar nutrient regeneration mechanisms.

~~ERSEM-BFM~~ GETM-ERSEM-BFM had a number of additional functional types compared to ~~PML-ERSEM~~NEMO-ERSEM: *Phaeocystis* colonies, benthic diatoms, carnivorous zooplankton, filter feeder larvae, epibenthos, benthic predators, and benthic and pelagic nitrifying bacteria. Furthermore, it used a CO_2 module, whereas in ~~PML-ERSEM~~ NEMO-ERSEM this was switched off.

The models used different methods for nutrient affinity, nutrient stress and sinking, and light susceptibility. For nutrient affinity, ~~ERSEM-BFM~~ GETM-ERSEM-BFM used 10-100 times higher affinity values for nutrient uptake. There are two ways to measure phytoplankton nutrient uptake in experiments (Veldhuis and Admiraal, 1987): (i) a short-duration experiment in which nutrients are added to nutrient-deprived algal cultures and uptake rates into the internal nutrient buffer are measured; and (ii) an experiment that lasts a full day in which uptake rates needed for daily growth are measured. The parameters for GETM-ERSEM-BFM were based on short-duration experiments, whereas those for NEMO-ERSEM were based on long-duration experiments. The short-duration parameterisation allows for improved incorporation of the dependencies of cell properties such as cell size and buffer capacity. These features were needed to resolve the competition between diatoms and *Phaeocystis* colonies during excessive spring blooms in the Dutch coastal zone, which terminate through phosphate depletion. In ~~ERSEM-BFM~~GETM-ERSEM-BFM, nutrient stress of pelagic diatoms leads to excretion of all (new fixated) organic C that cannot be used for growth as carbohydrates (TEP). At high levels of diatoms, this excretion leads to the simulation of the effect of macro-aggregate formation through binding by these carbohydrates, through increases in the sinking rate of live and dead particulate matter. ~~PML-ERSEM~~ NEMO-ERSEM used a more implicit approach to sinking. For light susceptibility, both models used a photosynthesis-irradiance (P-I) curve approach, but ~~PML-ERSEM~~ NEMO-ERSEM defined it through the initial slope (alpha), whereas ~~ERSEM-BFM~~ GETM-ERSEM-BFM defined it through the light intensity at maximum photosynthetic rate (Ke). For several elements where both models used the same approach, parameter settings were different: maximum productivity, respiration, excretion, minimum quota for P, lysis, and C:Chl ratios. For these, there was typically more differen-

tiation in settings between phytoplankton functional types in ~~ERSEM-BFM than in PML-ERSEM~~GETM-ERSEM-BFM than in NEMO-ERSEM.

3.4 Aggregating model PFTs to match observed PSCs

To allow validation of phytoplankton community structure from the models against the IBTS observations, the four PFTs from NEMO-ERSEM and six PFTs from GETM-ERSEM-BFM must be appropriately aggregated to match the observed PSCs. Diatoms (both models), dinoflagellates (both models) and resuspended benthic diatoms (GETM-ERSEM-BFM only) were considered to be micro-phytoplankton. Flagellates (both models) and *Phaeocystis* colonies (GETM-ERSEM-BFM only) were considered to be nano-phytoplankton. The pico-phytoplankton PFT (both models) was directly mapped to the pico-phytoplankton PSC. For consistency with the IBTS observations, the PFTs and PSCs were expressed as fractions of total chlorophyll concentration, rather than biomass.

4 Results

4.1 Observations

Each year, the IBTS cruise starts in early August in the Southern Bight of the North Sea off the Thames estuary (51.5° N) and proceeds northwards via a series of longitudinal transects, with each transect taking between one to three days, depending upon the width of the North Sea at each point. The final transect between the Shetland Islands and Norway at 61° N was reached by early September for the 2010 and 2011 IBTS cruises. The spatially-averaged annual mean surface temperature for the North Sea was 9.9 °C in 2010 and 10.0 °C in 2011, which were very close to the long-term average of 10.0 °C for the 1985-2014 period. Hence, the years surveyed represent near-average conditions for temperature.

A continuous recording of chlorophyll fluorescence showed good agreement with the quantity of extracted chl-a determined by HPLC ($r^2 = 0.64$ for 2010 and 0.65 for 2011). The number of same-day match-ups between in situ chl-a and satellite-derived chl-a was low for both years, but a comparison of eight-day averaged surface chl-a from MERIS with in situ data showed an excellent qualitative agreement for both years (Fig. ~~1~~2). Satellite coverage was more complete in 2011 than 2010. Time series plots and maps of the two cruises showed a number of regularly-occurring features that can be observed at this time of year (labelled 'A' to 'J' on Fig. ~~1~~2).

A zone of high chl-a (> 2 mg m^{-3}) was observed with all methods in the coastal waters of Belgium, The Netherlands, Germany and Denmark. This zone extended between points 'A' and 'B' for the map of 2010, and points 'F' and 'H' for 2011. High chlorophyll values (> 2 mg m^{-3}) were observed in the outer Thames estuary and close to the English east coast as far north as the Humber estuary, but the English coastal zone was not as clearly demarked by high chl-a as the continental coast. The continuous recording of the first 7-10 days of the IBTS thus alternated between moderate chl-a (1 to 2 mg m^{-3}) and high chl-a as the vessel covered the southern North Sea between 51.5° N and 55° N. An exceptional bloom event with chl-a of over 6 mg m^{-3} was recorded at location 'G' in 2011, and was clearly visible in MERIS and MODIS images.

The central section of the North Sea between 55° N and 58° N was covered during the second and third weeks of the IBTS. This section showed low chl-a values (< 1 mg m^{-3}) across most of the zone (Fig. ~~1~~2), particularly in the region north of 'I', 56.5° N to 58.5° N, 0° E to 3° E, which was a large region with values < 0.5 mg m^{-3}. To the east, the Danish coastal waters ('B' and 'H') showed high chl-a. The inshore English coast north of the Humber, and Scottish coastal waters, are low in chl-a compared to those further south. A moderate chl-a bloom was evident in the chlorophyll fluorescence trace, MERIS image and extracted chl-a at position 'C' in 2010, and a high chl-a patch was evident close to the Scottish coast at Aberdeen at position 'D'.

The northern North Sea was sampled in weeks three and four (from 28th to 29th August onwards) and was similar in 2010 and 2011. An arc of high chl-a was detected from north of the Scottish mainland through the Orkneys and Shetlands, e.g. from 'D' to 'E' in 2010, with particularly high values at 'E'. In 2011, high values were observed from the Orkneys through to north of the Shetlands at 'J'. The ~~ferrybox~~ FerryBox chlorophyll fluorescence recorded a further large bloom on 6th September 2011, but this event was not sampled for pigments.

As described in Sect. 2.1, PFTs were determined on the basis of accessory pigment composition. In general, pigment diversity was lower in coastal waters and in the southern North Sea and reached peak diversity in the stratified central North Sea. Fucoxanthin was the dominant accessory marker pigment in the southern North Sea, and 19-hex was dominant in the northern North Sea. Pico-phytoplankton were represented in this analysis by the marker pigments zeaxanthin, chl-b and 19-hex; these pigments were rare in the southern North Sea below a line from East Anglia to the Wadden Sea, hence pico-phytoplankton contribution was estimated in this region to be less than 10 % of total phytoplankton biomass (Fig. ~~2~~3). The contribution of pico-phytoplankton increased with increasing latitude so that the area with highest contribution from the smallest PFT was found in both years to be located north of 57° N and east of 0° E. Nano-phytoplankton were represented by the pigments 19-hex, 19-but, alloxanthin and chl-b. The highest percentage contribution of nano-phytoplankton was found in both years to be located in the central and northern North Sea, including the high chl-a regions around the Shetland and Orkney islands. The largest PFT, micro-phytoplankton, were represented by the pigments fucoxanthin and peridinin. The distribution of this group showed highest abundance in the southern North Sea high chl-a regions near the continental coast, but also in location 'G' (2011) and between 'C' and 'D' in 2010.

The combination of continuous underway logging with autonomous instruments, high precision pigment measurements at selected stations, and good satellite earth observation coverage allowed the patterns of surface phytoplankton biomass and PFT distribution in the North Sea to be well understood. Together, this provided a solid observational base with which to test biogeochemical model accuracy.

4.2 Model validation - domain-scale

This section presents validation of physical and biogeochemical model variables against a range of observation-based products, in order to assess the models' skill at broader scales than the IBTS observations measured. ~~SST has been validated against OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis; Donlon et al., 2012), which is an objective analysis product based on remotely sensed and in situ SST observations. Sea surface chlorophyll and SPM have been validated against~~

remotely sensed ocean colour products, developed by Ifremer using the OC5 algorithm (Gohin et al., 2002, 2005, 2008). Due to the limited availability of observations, nutrient concentrations have been validated against the 1 resolution World Ocean Atlas climatologies (Garcia et al., 2010). Detailed validation of phytoplankton community structure against the IBTS observations follows in Sect. 4.3.

5   Since the focus of this study is on the phytoplankton community structure in August 2010 and 2011, most of the validation presented here is for these two months. For more general model validation the reader is referred to Edwards et al. (2012) and Wakelin et al. (2015b) for NEMO-ERSEM, and Lenhart et al. (2010), Aldridge et al. (2012), Van van Leeuwen et al. (2013), Van van der Molen et al. (2013), and Van van der Molen et al., (2016a, b 2016, 2017) for GETM-ERSEM-BFM in various configurations. However, some statistical assessment has been performed here for SST, chlorophyll and SPM over the period

10 March 2010 to October 2011. Two seasons have been defined for this assessment: the growing season and winter. The growing season is defined as March to October, and is averaged over 2010 and 2011. Winter is defined as November 2010 to February 2011. Statistics have been calculated in observation space by performing a bilinear interpolation of the daily mean model fields to the observation locations. Calculations have been performed for $\log_{10}$(chlorophyll) rather than for chlorophyll in order to provide a more Gaussian distribution (Campbell, 1995).

15   Taylor plots (Taylor, 2001) of SST, $\log_{10}$(chlorophyll) and SPM are shown in Fig. 3. 4. SST is a good match for the observations in both the growing season and in winter, although lower correlations are found for both models in August 2010 and 2011 than for the whole seasons. Slightly better statistics are obtained for NEMO-ERSEM than for GETM-ERSEM-BFM, reflecting the assimilation of SST data into NEMO-ERSEM. The statistics for $\log_{10}$(chlorophyll) differ more between models and between seasons, and the models are a less good not as good a match for the observations than with SST, as is common

20 in physical-biogeochemical models. With SPM, the two models show large differences in variability. GETM-ERSEM-BFM has a much higher standard deviation than the observations in both seasons, especially the growing season, whilst the standard deviation of NEMO-ERSEM is too low all year round.

    Maps of mean SST for August 2010 and 2011, the months during which most of the IBTS observations were collected, are plotted in Fig. 4 5, from GETM-ERSEM-BFM, NEMO-ERSEM, and OSTIA. There is a great deal of similarity between

25 NEMO-ERSEM and OSTIA, which is unsurprising since NEMO-ERSEM assimilates SST data, but both NEMO-ERSEM and GETM-ERSEM-BFM are able to simulate the spatial features seen in OSTIA, as well as the inter-annual variability between 2010 and 2011. Boundary effects can be seen in GETM-ERSEM-BFM, which has a smaller domain.

    Similar maps for Maps of mean sea surface salinity (SSS) for August 2010 from GETM-ERSEM-BFM and NEMO-ERSEM are plotted in Fig. 6. Overlaid in circles are the in situ SSS observations from the 2010 IBTS cruise. Both models show a good

30 qualitative match for the observations in most regions. The only area which differs substantially between the models is the Norwegian Trench and surrounding area in the north-east North Sea. In GETM-ERSEM-BFM the Norwegian coastal current disperses erroneously, spreading freshwater into the North Sea. This is not seen in the observations, and is an issue of model resolution: finer-resolution configurations of GETM do not suffer from this. The coarseness of the resolution also accounts for the Rhine freshwater plume being wider than observed.

Maps for sea surface chlorophyll concentration are plotted in Fig. ~~5~~7, from the two models and the OC5 products. Daily ocean colour coverage is incomplete due to cloud cover, so the observations plotted here are simply a composite of all observations available during the month, rather than a true monthly mean. To ensure a fair comparison, the daily mean model fields were bilinearly interpolated to observation locations, and equivalent composites plotted rather than the true model mean. Van der Molen et al. (~~2016b~~2017) presented a comparison of sub-sampled model results, accounting for cloud cover, of SPM with the true model mean, which suggested noticeable differences in winter, but only small differences in summer. The match between the models and the observations is ~~less~~ not as good for chlorophyll ~~than~~ as for SST, but both models were still able to capture ~~the main observed featuresthroughout the domain~~some of the observed features. In the central and northern North Sea, which has the lowest chlorophyll concentrations, values were generally under-estimated by GETM-ERSEM-BFM and over-estimated by NEMO-ERSEM. High coastal chlorophyll values were better simulated by GETM-ERSEM-BFM, whilst the Norwegian Trench is better represented by NEMO-ERSEM. GETM-ERSEM-BFM has more spatial variability than NEMO-ERSEM, despite having a lower model resolution. As with SST, there is notable inter-annual variability in the observations, with higher chlorophyll concentrations in 2011 than 2010. Both models captured this variability, although it is less evident in GETM-ERSEM-BFM, and over-pronounced in NEMO-ERSEM.

The models are similarly compared to the OC5 SPM products in Fig. ~~6.~~8. NEMO-ERSEM and GETM-ERSEM-BFM both under-estimate SPM in the central and northern North Sea, with NEMO-ERSEM also under-representing the plume of SPM off south-east England. Overall, the two models give very different results for SPM, and the reasons for and potential consequences of this are discussed in Sect. 5.

Maps of mean surface nitrate, phosphate and silicate for each model for August 2010 are shown in Fig. ~~7~~9, alongside the corresponding World Ocean Atlas climatology fields. Only 2010 is plotted because very similar patterns are seen in the models for both years, and the climatologies do not include inter-annual variability. It should also be noted that the climatologies are of relatively coarse 1° resolution, so provide only a basic representation of the North Sea, but are the only source of data with full spatial coverage available for such a comparison. The climatologies have been used as boundary conditions by both models, so are not strictly independent, but values within the North Sea domain have not been used as input to the models. For nitrate, the main limiting nutrient in the North Sea, GETM-ERSEM-BFM shows high coastal concentrations, and very low concentrations elsewhere. NEMO-ERSEM has a similar pattern, but with a much less extreme range of values. ~~Both are in broad agreement with the climatology.~~ Likewise for phosphate and silicate, NEMO-ERSEM and GETM-ERSEM-BFM show differing distributions to each other, ~~but nonetheless are both in broad agreement with the magnitudes and spatial patterns of the climatologies.~~ and match some of the climatological features but not others. Overlaid on the maps are in situ surface nutrient observations sampled on the 2010 IBTS cruise. These show near-depletion of nitrate and phosphate across most of the North Sea. The depletion of nitrate is captured well by GETM-ERSEM-BFM, but is not seen to the same extent in either NEMO-ERSEM or the climatological World Ocean Atlas fields. The in situ observations also show greater depletion of phosphate than either of the models or the climatology, but the models are a better match for the silicate observations.

4.3 Model validation against PFT observations

This section presents validation of the models against the chlorophyll and PFT observations collected on the IBTS cruises. Hereafter, Micro is used to refer to the fraction of total chlorophyll represented by the micro-phytoplankton size class, and similarly Nano and Pico. Model PFTs were aggregated as described in Sect. 3.4. The most northerly of the IBTS observations were located outside the GETM-ERSEM-BFM model domain; these have been excluded from the assessment to ensure the same observations were used to validate both models.

Maps of mean surface model PFTs for August 2010, as fractions of total chlorophyll, are plotted in Fig. 8. 10. These show very different distributions for NEMO-ERSEM and GETM-ERSEM-BFM, much more so than the difference in total chlorophyll might suggest. In particular, NEMO-ERSEM shows dominance by pico-phytoplankton in the southern North Sea, similar fractions of pico-phytoplankton and flagellates in the rest of the domain, and generally low concentrations of diatoms and dinoflagellates. In contrast, GETM-ERSEM-BFM shows dominance by diatoms in coastal regions, and by pico-phytoplankton in the centre of the domain, with generally lower fractions of the remaining PFTs. The two PFTs unique to GETM-ERSEM-BFM, resuspended benthic diatoms and *Phaeocystis* colonies, only show notable concentrations in certain coastal areas. The reasons for the differences between the two models are discussed in Sect. 5 ~~, but they are likely to have arisen from the differences in parameterisations detailed in Sect. 3.3~~ .

The bias, root mean square error (RMSE) and correlation of modelled versus observed total chlorophyll concentration are shown in Table 3. GETM-ERSEM-BFM has a slight bias towards too high chlorophyll, whilst the bias for NEMO-ERSEM is near-zero. GETM-ERSEM-BFM chlorophyll values are typically higher than those from NEMO-ERSEM, but also show a greater range. These features are consistent between the two years. However, whilst GETM-ERSEM-BFM has a similar correlation value for both years, the correlation for NEMO-ERSEM is much higher in 2010 than 2011. It should be noted though that these statistics are based on a relatively small number of points, so any conclusions drawn from this comparison are not guaranteed to be robust, particularly given the domain-scale spatial variability (see Sect. 4.2 and Fig. 5 7).

A comparison of phytoplankton community structure in the models and IBTS observations has been made by aggregating the model PFTs into the three observed PSCs, as described in Sect. 3.4. The bias, RMSE and correlation of the modelled versus observed total PSCs are shown in Table 3. Bias and RMSE are generally lower for GETM-ERSEM-BFM than for NEMO-ERSEM, particularly for Micro and Pico. NEMO-ERSEM has higher absolute correlations, but for Micro and Pico these tend to be negative, suggesting that NEMO-ERSEM is getting Nano approximately correct, but Micro and Pico are inversely distributed compared with the observations.

The distribution of relative PSC fractions with total chlorophyll is plotted for each data set in Fig. 9. 11. Consistent with results from previous studies (e.g. Devred et al., 2011), as observed chlorophyll increases, Micro tends to increase, and Nano and Pico decrease. This pattern is also seen to some extent in GETM-ERSEM-BFM, but less so in NEMO-ERSEM (and only in 2010), although NEMO-ERSEM has a smaller range of chlorophyll concentrations. In the IBTS data there is a clear overall dominance of Micro. This is well reproduced by GETM-ERSEM-BFM, but the opposite is found in NEMO-ERSEM. The exception to this is a group of observations at low chlorophyll concentrations, most notably in 2011, in which Micro is least

14

abundant, better matching typical NEMO-ERSEM results. These observations were all taken in the central North Sea, and this behaviour is discussed further in Sect. 5.

To explore the model behaviour further, and allow comparison with other works such as de Mora et al. (2016), histograms of the distribution of relative PSC fractions with total chlorophyll are plotted in Fig. 12, from each model grid point in the North Sea. These have used the mean model fields for August 2010 and August 2011. With this extended number of model points, a clear relationship is seen for GETM-ERSEM-BFM, with Micro increasing with total chlorophyll, and Pico decreasing. This matches the trend seen in the IBTS observations, as well as previous studies (e.g. Brewin et al., 2010; Devred et al., 2011). For NEMO-ERSEM, the range of chlorophyll concentrations remains small, making any relationship difficult to assess. When there are higher chlorophyll values a similar pattern of increasing Micro and decreasing Pico is seen, but there are too few points to draw a robust conclusion on the model relationship.

Three variables which always sum to one can be displayed in a single space, barycentric coordinates, using a ternary plot (e.g. Jupp et al., 2012). Phytoplankton community structure is plotted this way in Fig. ~~10.~~ 13. The observations form a distinct line in this space, from the centre of the plot to the corner representing dominance by Micro. At lower chlorophyll concentrations (not shown in Fig. ~~10~~13, but consistent with Fig. ~~9~~11) there are roughly equal fractions of Micro, Nano and Pico. As chlorophyll concentration increases, Nano and Pico decrease in roughly equal amounts, with Micro increasing accordingly. The fact that the observations form a line in this space shows Nano and Pico to change in roughly equal proportions when Micro changes with chlorophyll, which differs to some extent from other studies such as Brewin et al. (2010). GETM-ERSEM-BFM displays a similar pattern, with values in the same area of the plot as the observations, although with a less distinct relationship. NEMO-ERSEM values show a very different distribution however. There is some overlap with the observations in 2011, but otherwise a much less Micro-dominated regime is evident.

The ternary plot can also be used as a colour key to produce a map of phytoplankton community structure, as in Fig. ~~11.~~ 14. This plots the August mean community structure for each model and each year, overlaid with the IBTS point observations in circles. Plotting the community structure in such a fashion demonstrates that whilst GETM-ERSEM-BFM and NEMO-ERSEM give very different results in terms of the magnitudes of the PSC fractions, there are nonetheless some broadly consistent features in terms of spatial patterns, which are also evident to some extent in the observations. For instance, both models (although NEMO-ERSEM less so in 2010) show a distinct split in community structure between the southern and northern North Sea, and around bathymetric features such as Dogger Bank and coastlines. Such a split can be clearly seen in the 2011 observations, which show very little variation throughout the central North Sea, but is less clear in the 2010 observations. A difference between the years in the community structure in the central North Sea is also seen in NEMO-ERSEM, and to a lesser extent GETM-ERSEM-BFM, although the direction of change in the models is from Pico-dominated to Micro-dominated, the opposite of that in the observations. Although in most cases the community structure in the models does not match that of the observations, GETM-ERSEM-BFM is a very good match for the observations in the southern North Sea, an area particularly dominated by diatoms in the model. Silicate in this region is near-depleted in GETM-ERSEM-BFM (see Fig. ~~7~~9), but abundant in NEMO-ERSEM. In GETM-ERSEM-BFM, distinct blue patches can be seen off East Anglia, south Dorset and the German Bight, which are mostly areas where *Phaeocystis* colonies dominate in the model.

5 Summary and discussion

This study has presented a new set of in situ phytoplankton pigment observations for the North Sea, processed to give information on phytoplankton community structure. Two physical-biogeochemical models, the biogeochemical components of which are different variants of ERSEM, were then validated against these and other observations. Both models were a good match for SST observations, and a reasonable match for chlorophyll observations, but gave contrasting results for SPM. Furthermore, the two models displayed very different phytoplankton community structures. GETM-ERSEM-BFM was able to reproduce many of the features of the observations, particularly in the southern North Sea, whereas NEMO-ERSEM was a poor match for the observations, except at the lowest chlorophyll concentrations. Nonetheless, both models shared some similarities with each other and the observations in terms of spatial features and inter-annual variability.

The distribution of total phytoplankton biomass across the North Sea during summer of both years showed a high degree of consistency between three different observational methods: satellite remote sensing, high-frequency continuous measurement of chlorophyll fluorescence, and chl-a quantification at discrete stations. A similar set of spatial features can be observed in 2010 and 2011, which can be explained by the underlying hydrodynamics (~~Van~~ van Leeuwen et al., 2015). The central, strongly-stratified region of the North Sea has very low nutrient concentrations and correspondingly low chl-a. In areas where vertical mixing, riverine input or horizontal advection bring nutrients into the upper water column, phytoplankton biomass is elevated. As well as observing the total quantity of phytoplankton, deriving the composition of size- and functional types is important for a better understanding of ecosystem function and energy flows to higher trophic levels (Chavez et al., 2011). Accessory pigments have been widely used in biological oceanography to investigate community composition, but caution must be applied when interpreting results, and support from other methods should be used where possible (Schlüter et al., 2014). The original equations used by Hirata et al. (2008, 2011) to convert pigments to pico-, nano-, and micro-phytoplankton size classes underestimated the fraction of pico-phytoplankton compared to flow cytometric observations, and were modified by increasing the contribution of chl-b and 19-hex to this class. Results in both years showed a consistent pattern of decreasing micro-phytoplankton abundance with distance from the coast, and with increasing latitude, and this is supported by previous pigment-based studies. Work in the German Bight has also shown a change from a coastal, diatom-dominated community to a more diverse, small-celled community further offshore (Brandsma et al., 2013; Wollschläger et al., 2015).

Both NEMO-ERSEM and GETM-ERSEM-BFM have shown the ability to reproduce the physics and broad-scale biogeochemistry of the North Sea. However, results are more varied when considering specific aspects such as the phytoplankton community structure in August. In some ways this is to be expected, as this study has used existing model versions which have not been previously validated against or tuned to such observations. Furthermore, August is a challenging month to model in the North Sea, as evidenced by the reduced SST skill for this month compared with the seasonal average. This is because simulating the details of the stratification, nutrient concentrations and therefore phytoplankton concentrations is dependent on having successfully simulated processes in previous months, as well as the processes seen during August. This is more important than at other times of year, because, being at the height of summer, there are no strong temporal gradients driving the response of the system, as in spring and autumn. As a result, the internal biogeochemical dynamics can play out most freely

(both in reality and in the model). Nonetheless, this kind of specific information is in increasing demand, and ~~confidence must be had in the results if they~~ if results are to be provided to users then they should be understood and validated.

NEMO-ERSEM and GETM-ERSEM-BFM gave very different representations of SPM concentrations, which impacts ecosystem functioning through light limitation (see also the light susceptibility parameters, Table 2). NEMO-ERSEM uses the two-size class SPM model described by Sykes and Barciela (2012). This was implemented by Sykes and Barciela (2012) in the POLCOMS (Proudman Oceanographic Laboratory Coastal Ocean Modelling System; Holt and James, 2001; Holt et al., 2005) physical framework, in which it gave skilful results compared with observations. However, the model has not yet received the same degree of tuning and development since being implemented in NEMO-ERSEM, which may explain the consistent under-estimation of SPM found in this study. Furthermore, the high vertical resolution of NEMO-ERSEM means that the settling velocities must sometimes be artificially limited when used by the SPM model, in order to avoid breaking the Courant-Friedrichs-Lewy (CFL) condition (Courant et al., 1928), thus reducing resuspension. Changes to the settling parameters would be expected to lead to improvements. GETM-ERSEM-BFM uses an alternative SPM model (~~Van~~ van der Molen et al., ~~2016b~~2017), which only has one size class but includes resuspension by waves as well as currents, and which was developed within the GETM-ERSEM-BFM framework. This generally matches spatial distributions of SPM better, but often has concentrations which are extremely high or low compared with satellite data, leading to a degradation in some error statistics.

The starkest contrast between the model results presented in this study is in the simulated phytoplankton community structures, which differ far more than might be expected given the corresponding total chlorophyll concentrations. GETM-ERSEM-BFM gave a wider range of combinations of biomass in the three size classes resolved by the observations than NEMO-ERSEM did. This was also reflected in more spatial variability and stronger spatial gradients, which resulted in a better match of the coastal to offshore change in phytoplankton community structure evident in the observations, in which diatoms are particularly important. The limited biomass in the two additional PFTs in GETM-ERSEM-BFM (benthic diatoms and *Phaeocystis* colonies, Fig. ~~8~~10) suggests that these were not the primary cause of this difference in response. ~~It cannot be ruled out that the different physical models, including the different SPM methods, and differences in the model configurations, play a role (Table 1). However, it is most likely that this difference is caused in part by the greater difference in parameter settings in GETM-ERSEM-BFM between the four PFTs that the two models have in common, in combination with additional sensitivity introduced in GETM-ERSEM-BFM by the different methods of calculating nutrient affinity, nutrient stress, sinking, lysis and light susceptibility (Table 2).~~Two mechanisms are likely to play important roles in causing the differences: 1) the higher coastal nutrient concentrations in combination with the different nutrient affinity settings in GETM-ERSEM-BFM, allowing diatoms to out-compete other types, in contrast with identical nutrient affinity settings in NEMO-ERSEM; 2) the coincidence of diatoms with areas of high SPM concentrations in GETM-ERSEM-BFM (mostly absent in NEMO-ERSEM) in combination with greater light susceptibility of diatoms (again contrasting with uniform values in NEMO-ERSEM), giving them competitive advantage. Similar, but more subtle effects will modulate the response of the other PFTs in GETM-ERSEM-BFM.~~The~~Overall, the more uniform parameter settings of NEMO-ERSEM promote a more uniform response of the PFTs, as is evident in the results. Further work, with dedicated experiments performed in a configuration designed for such a comparison, could be carried out to further investigate these differences, and this is discussed below.

17

GETM-ERSEM-BFM provided a better match for the IBTS observations of phytoplankton community structure than NEMO-ERSEM did. The exception to this was in the low-chlorophyll waters of the central North Sea, the region of the domain with the weakest currents and largest residence times. Here the community structure of the observations more closely resembled that typical of NEMO-ERSEM. There are indications that recent versions of NEMO-ERSEM, applied to the global ocean rather than to the North-West European Shelf Seas, perform better at reproducing observed community structures (~~De~~ de Mora et al., 2016). Together with the results presented here, this suggests that NEMO-ERSEM may be more representative of an open ocean environment, whereas the settings in GETM-ERSEM-BFM are better suited to the complex coastal environment of the North Sea.

Whilst there were large contrasts in the corresponding ratios of PSCs between all three data sets, there was more agreement between the data sets about the spatial patterns of community structure. For instance, each had contrasting structures between the southern and central North Sea, and in coastal areas. Furthermore, inter-annual variability in the central North Sea was clearly evident in the observations, and also each of the models. This can be compared to differences in SST between the two years and the two models, suggesting a set of physical drivers which the models were able to capture. Across much of the North Sea the SST was cooler in 2011 than 2010, as seen in Fig. 5. Typical mixed layer depths were also deeper in the models in 2011 (not shown), likely in response to increased wind speeds. This deepening of the mixed layer cooled the SST and brought more nutrients to the surface, increasing production and chlorophyll, as seen in Fig. 7. In NEMO-ERSEM, which showed the most inter-annual variability in phytoplankton community structure, all PFTs increased in chlorophyll, but dinoflagellates increased the most in percentage terms, shifting the community structure more towards Micro-dominance. This could be because the changes in temperature and prey availability favoured smaller predators more than larger ones, and an increase in the ratio of nitrate to ammonium best suited larger phytoplankton. This implies that even if models are currently unable to accurately represent the exact community structures, they can still be used to assess the distribution of different habitats, and when and where these may change.

Due to differences in locations, timescales and datasets, a robust comparison with results from other PFT modelling studies in the literature cannot be made at this stage. Nonetheless, a consideration of the types of results obtained using different modelling approaches is of value. Lewis et al. (2006) and Lewis and Allen (2009) both validated a similar version of the PML-developed ERSEM, coupled with the POLCOMS hydrodynamic model, against in situ observations of PFTs in the North-west European Shelf. Lewis et al. (2006) validated against Continuous Plankton Recorder (CPR; Richardson et al., 2006) data, and found that whilst the model reproduced the main seasonal features of plankton succession, diatoms consistently bloomed too early, and there were some spatial differences, especially in the North Sea. This would seem consistent with the low diatoms seen in NEMO-ERSEM in August in this study. Lewis and Allen (2009) validated against a station in the Western English Channel, and found a poor match for phytoplankton variables when assessment was performed in observation space, as has been done here, but a better match when assessing more general seasonal variability. On a global open-ocean scale, Gregg and Casey (2007) validated the NASA Ocean Biogeochemical Model (NOBM) against both in situ and remotely sensed estimates of PFTs, with a focus on coccolithophores, and described their results as "mixed", noting contrasting results elsewhere in the literature. Hirata et al. (2013) validated the MEM-OU model against remote sensing estimates of large and small phytoplankton, and

found good agreement at basin scales, but which reduced at smaller scales. De Mora et al. (2016) compared PFT distributions with chlorophyll from the most recent ERSEM version of Butenschön et al. (2016) to those obtained using various remote sensing algorithms applied to the model data, and found the model displayed similar properties at global scales. These studies (not exhaustive) all used models with a relatively small fixed number of specifically-parameterised PFTs. An alternative approach is to initialise the model with a large number of randomly-parameterised PFTs, with the best-suited PFTs naturally dominating (Follows et al., 2007). In a global setting this approach has been found to successful reproduce large-scale patterns of phytoplankton diversity in terms of organism size (Follows et al., 2007) and number of co-existing species (Barton et al., 2010). The ability of such a model to reproduce the short-term variability of a complex coastal environment such as the North Sea would make for an interesting future inter-comparison. In general, results in the literature suggest some success of different approaches at reproducing large-scale patterns of phytoplankton community structure, but with more detailed skill yet to be properly demonstrated.

Careful thought needs to be given therefore to what products and information can be offered to users, which address user and policy requirements with a sufficient level of skill . There are many challenges related to the uptake and use of models (Hyder et al., 2015), and studies have looked at the potential for use of models for MSFD (Piroddi et al., 2015), and the issues of uncertainty in climate projections (Payne et al., 2015). The main drivers for decision makers relate to sustainable exploitation of resources whilst protecting the environment, so it is important to consider how best biogeochemical models can be used to support and underpin the evidence upon which decisions are based (Hyder et al., 2015). Biogeochemical models can provide insight into many of the key challenges for sustainable use and protection of our seas including eutrophication (Lenhart et al., 2010), harmful algal blooms (Shutler et al., 2015), non-native species, aquaculture (Aldridge et al., 2012), climate change and ocean acidification (Van Leeuwen et al., 2016), and marine spatial planning (Van der Molen et al., 2014). It is not possible to identify here exactly how biogeochemical models may be used in the future to support decision making, but it is clear that requirements will include insights into the attribution of changes in ecosystems to underlying drivers of environmental change, improving the efficiency of monitoring programmes, assessment of indicators, and economic impacts of management options (Hyder et al., 2015). Decision making timescales are often at odds with model development, so it is important to be able to adapt existing models to address decision makers' needs at short notice. To achieve this, close communication is needed between modellers and decision makers, so that the potential of biogeochemical models to support decision making is not lost in translation (Hyder et al., 2015). Continual model developments will be required to address these and other user needs. A comprehensive review of the challenges faced is given by Holt et al. (2014), but specific issues have emerged and results from this current study which should should further inform future model development. This will be particularly relevant in the context of the UK Shelf Seas Biogeochemistry (SSB) research programme (http://www.uk-ssb.org), in which Cefas and the Met Office are both participants. One of the aims of the SSB programme is to create a common version of ERSEM to be used by the UK research community, by combining features of the two versions of ERSEM used in this study. An initial combined version is described by Butenschön et al. (2015 2016), and this will be further developed within the SSB programme. The version of Butenschön et al. (2016) provides a major update to that of Blackford et al. (2004), which forms the basis of the NEMO-ERSEM version used in this study, and initial applications (Butenschön et al., 2016; de Mora et al., 2016; Ciavatta

et al., 2016) have shown different phytoplankton community structures to that obtained in this work. It is clear from this study that the details of the model components and parameterisations can lead to very different results, and validation against a range of data, using a range of methods, is vital throughout the model development cycle.

The assessment presented in this study suggests that the biogeochemical model parameterisations are important in controlling the phytoplankton community structure. However, due to the many differences in physical modelling environment and experimental configurations, it cannot be ruled out that differences in the physics are responsible for the differences in phytoplankton community structure, and previous studies have found these to be important. For instance, Sinha et al. (2010) coupled a single marine biogeochemical model with two different global physical ocean models, and found contrasting phytoplankton community structures due to differences in mixing. A number of further experiments could be performed to investigate the differences in PFT response between the models, and develop improvements. These can make use of the two ERSEM versions used in this study and the new SSB-ERSEM, along with NEMO, GETM, and GETM's 1D counterpart GOTM (General Ocean Turbulence Model). SSB-ERSEM is compatible with the Framework for Aquatic Biogeochemical Models (FABM; Bruggeman and Bolding, 2014), allowing it to be coupled with either GOTM, GETM or NEMO. Building on this framework could allow the different ERSEM versions to be run with identical hydrodynamics in 1D and 3D, and similarly allow individual ERSEM versions to be run with different hydrodynamics. This would help identify the differences in results introduced by different components of the system. Within this experimental framework, controlled differences to physical and biogeochemical model parameters could be made to investigate further. These would allow very efficient testing. However, the current field observations (quasi-synoptic spatial distribution for one month of the year) do not allow a fully robust model comparison. An observational time series resolving the seasonal cycle at one or more locations would be needed for this excercise. Recent developments of algorithms to derive PFTs from remote sensing data (Nair et al., 2008; Hirata et al., 2013) could benefit all these potential strands of further work.

A further development will be the assimilation of biogeochemical data. Ocean colour data assimilation is being increasingly utilised by the reanalysis and forecasting community (Gehlen et al., 2015), and has already been successfully demonstrated for ERSEM (Ciavatta et al., 2011, 2014, 2016). A suitable ocean colour assimilation scheme for operational purposes is being developed as a collaboration between the Met Office and PML, to be implemented in the SSB ERSEM version and run operationally as part of CMEMS. This will also give the opportunity to take advantage of the advent of remote sensing PFT/PSC products, incorporating such data into the assimilation and routine validation.

Information on the marine environment can come from three sources: in situ observations, remote sensing data, and models. These three sources are inter-linked and all are vital; sufficient scientific understanding of the North Sea and other environments cannot be obtained if any of these three sources are removed. Models provide complete 3D spatial and temporal coverage, can be used to simulate a range of hypotheses, and are relatively inexpensive. However, as demonstrated in this study, observations are necessary for the validation and development of models, and model data cannot be relied upon in isolation. Remote sensing data provide considerably greater observational coverage than in situ measurements, but this coverage is still limited to the sea surface and cloud-free conditions, and empirical algorithms based on in situ data are used in the construction of remote sensing products. These satellite data must be comprehensively ground-truthed against in situ observations for confidence to

~~be had in them~~if they are to be used with confidence. Continuing in situ observations are therefore required to under-pin model and remote sensing data, as well as to provide unique insights into the marine environment. In turn, modelling studies can be used to help inform sampling strategies of future observing programmes, to help provide value for money without sacrificing accurate scientific understanding.

5   Finally, it remains to answer the question ~~posed in the title~~ at the heart of this paper: "Can ERSEM-type models simulate ~~biodiversity~~phytoplankton community structure?" The evidence from this study suggests that ERSEM-type models ~~definitely~~ have the potential to accurately simulate ~~biodiversity~~phytoplankton community structure, but certain model formulations and parameterisations are required to do so, and ~~current~~ these two ERSEM versions do not reliably do so at this stage. Appropriate model development, informed by detailed validation studies, ~~will~~ appears to be a major but achievable challenge, and will help

10 facilitate the wider application of marine biogeochemical modelling to wide-ranging user and policy needs.

References

Aiken, J., Pradhan, Y., Barlow, R., Lavender, S., Poulton, A., Holligan, P., and Hardman-Mountford, N.: Phytoplankton pigments and functional types in the Atlantic Ocean: a decadal assessment, Deep-Sea Res. Pt II, 56(15), 899-917, 2009.

Aldridge, J., van der Molen, J., and Forster, R.: Wider ecological implications of Macroalgae culitvation, The Crown Estate, 95 pp, ISBN: 978-1-906410-38-4, 2012.

Allen, J. I. and Clarke, K. R.: Effects of demersal trawling on ecosystem functioning in the North Sea: a modelling study, Mar. Ecol.-Prog. Ser., 336, 63-75, 2007.

Anderson, T. R.: Plankton functional type modelling: running before we can walk?, J. Plankton Res., 27(11), 1073-1081, 2005.

Balmaseda M. A., Mogensen, K., and Weaver, A.: Evaluation of the ECMWF Ocean Reanalysis ORAS4, Q. J. Roy. Meteor. Soc., 139(674), 1132-1161, 2013.

Baretta, J. W., Ebenhöh, W., and Ruardij, P.: The European Regional Seas Ecosystem Model, a complex marine ecosystem model, Neth. J. Sea Res., 33(3), 233-246, 1995.

Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J., and Follows, M. J.: Patterns of diversity in marine phytoplankton, Science, 327(5972), 1509-1511, 2010.

Blackford, J. C., Allen, J. I., and Gilbert, F. J.: Ecosystem dynamics at six contrasting sites: a generic modelling study, J. Marine Syst., 52(1), 191-215, 2004.

Blauw, A. N., Los, F. J., Huisman, J,, and Peperzak, L.: Nuisance foam events and *Phaeocystis* globosa blooms in Dutch coastal waters analyzed with fuzzy logic, J. Marine Syst., 83, 115-126, 2010.

Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., Mirouze, I., Peterson, K. A., Sellar, A. and Storkey, D.: Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts, Geoscientific Model Development, 7(6), 2613-2638, 2014.

Borja, A., Elliott, M., Andersen, J. H., Cardoso, A. C., Carstensen, J., Ferreira, J. G., Heiskanen, A. S., Marques, J. C., Neto, J. M., Teixeira, H., and Uusitalo, L.: Good Environmental Status of marine ecosystems: What is it and how do we know when we have attained it?, Mar. Pollut. Bull., 76(1), 16-27, 2013.

Boyes, S. J. and Elliott, M.: Marine legislation - The ultimate "horrendogram": International law, European directives and national implementation, Mar. Pollut. Bull., 86(1), 39-47, 2014.

Brandsma, J., Martínez, J.M., Slagter, H.A., Evans, C., and Brussaard, C.P.: Microbial biogeography of the North Sea during summer, Biogeochemistry, 113, 119-136, 2013.

Brewin, R. J., Sathyendranath, S., Hirata, T., Lavender, S. J., Barciela, R. M. and Hardman-Mountford, N. J.: A three-component model of phytoplankton size class for the Atlantic Ocean, Ecol. Model., 221(11), 1472-1483, 2010.

Brewin, R. J., Hardman-Mountford, N. J., Lavender, S. J., Raitsos, D. E., Hirata, T., Uitz, J., Devred, E., Bricaud, A., Ciotti, A., and Gentili, B.: An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing, Remote Sens. Environ., 115(2), 325-339, 2011.

Brito, A. C., Sá, C., Brotas, V., Brewin, R. J., Silva, T., Vitorino, J., Platt, T., and Sathyendranath, S.: Effect of phytoplankton size classes on bio-optical properties of phytoplankton in the Western Iberian coast: Application of models, Remote Sens. Environ., 156, 537-550, 2014.

Brotas, V., Brewin, R. J., Sá, C., Brito, A. C., Silva, A., Mendes, C. R., Diniz, T., Kaufmann, M., Tarran, G., Groom, S. B., and Platt, T.: Deriving phytoplankton size classes from satellite data: Validation along a trophic gradient in the eastern Atlantic Ocean, Remote Sens. Environ., 134, 66-77, 2013.

Bruggeman, J.: Succession in plankton communities: A trait-based perspective, Ph.D. thesis, Vrije Universiteit Amsterdam, 2009.

5   Bruggeman, J. and Bolding, K.: A general framework for aquatic biogeochemical models, Environ. Modell. Softw., 61, 249-265, 2014.

Burchard, H. and Bolding, K.: GETM: A General Estuarine Transport Model; Scientific Documentation, Tech. Rep. EUR 20253 EN, European Commission, 2002.

Butenschön, M., Clark, J., Aldridge, J. N., Allen, J. I., Artioli, Y., Blackford, J., Bruggeman, J., Cazenave, P., Ciavatta, S., Kay, S., Lessin, G., van Leeuwen, S., van der Molen, J., de Mora, L., Polimene, L., Sailley, S., Stephens, N., and Torres, R.: ERSEM 15.06: a generic model
10   for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels, Geoscientific Model Development, 8, 7063-7187, 2015. 9, 1293-1339, 2016.

Campbell, J. W.: The lognormal distribution as a model for bio-optical variability in the sea, J. Geophys. Res.-Oceans, 100(C7), 13237-13254, 1995.

Capuzzo, E., Stephens, D., Silva, T., Barry, J., and Forster, R. M.: Decrease in water clarity of the southern and central North Sea during the
15   20th century, Global Change Biol., 21(6), 2206-2214,, 2015.

Chassot, E., Mélin, F., Le Pape, O., and Gascuel, D.: Bottom-up control regulates fisheries production at the scale of eco-regions in European seas, Mar. Ecol.-Prog. Ser., 343, 45-55, 2007.

Chavez, F. P., Messié, M., and Pennington, J. T.: Marine Primary Production in Relation to Climate Variability and Change, Annu. Rev. Mar. Sci., 3, 227-260, 2011.

20   Ciavatta, S., Torres, R., Saux-Picart, S., and Allen, J. I.: Can ocean color assimilation improve biogeochemical hindcasts in shelf seas?, J. Geophys. Res.-Oceans, 116(C12), 2011.

Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall'Olmo, G., Polimene, L., and Allen, J. I.: Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling, Prog. Oceanogr., 127, 74-95, 2014.

Ciavatta, S., Kay, S., Saux-Picart, S., Butenschön, M., and Allen, J.I.: Decadal reanalysis of biogeochemical indicators and fluxes in the
25   North West European shelf-sea ecosystem, J. Geophys. Res.-Oceans, 121, 1824-1845, 2016.

Courant, R., Friedrichs, K., and Lewy, H.: Über die partiellen Differenzengleichungen der mathematischen Physik, Math. Ann., 100(1), 32-74, 1928.

Daan, N., Gislason, H., Pope, J. G., and Rice, J. C.: Changes in the North Sea fish community: evidence of indirect effects of fishing?, ICES J. Mar. Sci., 62(2), 177-188, 2005.

30   De Mora, L., Butenschön, M., and Allen, J. I.: How should sparse marine in situ measurements be compared to a continuous model: an example, Geoscientific Model Development, 6(2), 533-548, 2013.

De Mora, L., Butenschön, M., and Allen, J. I.: The assessment of a global marine ecosystem model on the basis of emergent properties and ecosystem function: a case study with ERSEM, Geoscientific Model Development, 9(1), 59-76, 2016.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P.,
35   Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553-597, 2011.

Defra: Making the most of our evidence?: a strategy for Defra and its network, Department for Environment, Food and Rural Affairs, 2014.

Devred, E., Sathyendranath, S., Stuart, V., and Platt, T.: A three component classification of phytoplankton absorption spectra: Application to ocean-color data, Remote Sens. Environ., 115(9), 2255-2266, 2011.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W: The operational sea surface temperature and sea ice analysis (OSTIA) system, Remote Sens. Environ., 116, 140-158, 2012.

Donnelly, C., Andersson, J. C., and Arheimer, B.: Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe, Hydrolog. Sci. J., 61(2), 255-273, 2015.

Dupont, N. and Aksnes, D. L.: Centennial changes in water clarity of the Baltic Sea and the North Sea, Estuar. Coast. Shelf S., 131, 282-289, 2013.

ECMWF: European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year Re-Analysis (ERA-40) model data, NCAS British Atmospheric Data Centre, http://catalogue.ceda.ac.uk/uuid/775634f7e339b5262067e28a5d7b679d, 2006a.

ECMWF: Assimilated Data from the European Centre for Medium-Range Weather Forecasts (ECMWF) operational analysis program, NCAS British Atmospheric Data Centre, http://catalogue.ceda.ac.uk/uuidf6ce34fc/c462480467660a36d9b10a71, 2006b.

Edwards, K. P., Barciela, R., and Butenschön, M.: Validation of the NEMO-ERSEM operational ecosystem model for the North West European Continental Shelf, Ocean Sci., 8, 983-1000, 2012.

Engel, A.: The role of transparent exopolymer particles (TEP) in the increase in apparent particle stickiness (α) during the decline of a diatom bloom, J. Plankton Res., 22, 485-497, 2000.

EU: Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, European Parliament and Council of the European Union, Off. J. Eur. Communities, 327, 1-72, 2000.

EU: Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive), European Parliament and Council of the European Union, Off. J. Eur. Union, 164, 19-40, 2008.

EU: Regulation (EU) No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy, amending council Regulations (EC) No 1954/2003 and (EC) No 1224/2009 and repealing Council regulations (EC) No 2371/2002 and (EC) No 639/2004 and Council Decision 2004/585/EC, Off. J. Eur. Union, 354, 22-61, 2013.

Finkel, Z. V., Beardall, J., Flynn, K. J., Quigg, A., Rees, T. A. V., and Raven, J.A.: Phytoplankton in a changing world: cell size and elemental stoichiometry, J. Plankton Res., 32(1), 119-137, 2010.

Follows, M. J., Dutkiewicz, S., Grant, S., and Chisholm, S. W.: Emergent biogeography of microbial communities in a model ocean, Science, 315(5820), 1843-1846, 2007.

Friedrichs, M. A., Dusenberry, J. A., Anderson, L. A., Armstrong, R. A., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., Hood, R., and McGillicuddy, D. J.: Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups, J. Geophys. Res.-Oceans, 112(C8), 2007.

Fulton, E. A. and Link, J. S.: Modeling Approaches for Marine Ecosystem-based Management, in: The Sea, Volume 16: Marine Ecosystem-Based Management, editors: Fogarty, M. J. and McCarthy, J. J., Harvard University Press, 2014.

Fulton, E. A., Link, J. S., Kaplan, I. C., SavinaRolland, M., Johnson, P., Ainsworth, C., Horne, P., Gorton, R., Gamble, R. J., Smith, A. D., and Smith, D. C.: Lessons in modelling and management of marine ecosystems: the Atlantis experience, Fish Fish., 12(2), 171-188, 2007.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Zweng, M. M., Baranova, O. K., and Johnson, D. R.: World Ocean Atlas 2009, Volume 4: Nutrients (phosphate, nitrate, silicate), edited by Levitus, S., NOAA Atlas NESDIS 71, U.S. Government Printing Office, Washington, D.C., 398 pp., 2010.

Gårdmark, A., Lindegren, M., Neuenfeldt, S., Blenckner, T., Heikinheimo, O., Müller-Karulis, B., Niiranen, S., Tomczak, M. T., Aro, E., Wikström, A., and Möllmann, C.: Biological ensemble modeling to evaluate potential futures of living marine resources, Ecol. Appl., 23(4), 742-754, 2013.

Garmendia, M., Borja, A., Franco, J., and Revilla, M.: Phytoplankton composition indicators for the assessment of eutrophication in marine waters: Present state and challenges within the European directives, Mar. Pollut. Bull., 66, 7-16, 2013.

Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D. and Lehodey, P.: Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments, Journal of Operational Oceanography, 8(sup1), s168-s187, 2015.

Geider, R. J., H. L. MacIntyre, and T. M. Kana: Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature, Mar. Ecol.-Prog. Ser., 148, 187-200, 1997.

Gohin, F., Druon, J. N., and Lampert, L.: A five channel chlorophyll concentration algorithm applied to SeaWiFS data processed by SeaDAS in coastal waters, Int. J. Remote S., 23(8), 1639-1661, 2002.

Gohin, F., Loyer, S., Lunven, M., Labry, C., Froidefond, J. M., Delmas, D., Huret, M., and Herbland, A.: Satellite-derived parameters for biological modelling in coastal waters: Illustration over the eastern continental shelf of the Bay of Biscay, Remote Sens. Environ., 95(1), 29-46, 2005.

Gohin, F., Saulquin, B., Oger-Jeanneret, H., Lozac'h, L., Lampert, L., Lefebvre, A., Riou, P., and Bruchon, F.: Towards a better assessment of the ecological status of coastal waters using satellite-derived chlorophyll-a concentrations, Remote Sens. Environ., 112(8), 3329-3340, 2008.

Greenwood, N., Parker, E. R., Fernand, L., Sivyer, D. B., Weston, K., Painting, S. J., Kröger, S., Forster, R. M., Lees, H. E., Mills, D. K., and Laane, R. W. P. M.: Detection of low bottom water oxygen concentrations in the North Sea; implications for monitoring and assessment of ecosystem health, Biogeosciences, 7(4), 1357-1373, 2010.

Gregg, W. W. and Casey, N. W.: Modeling coccolithophores in the global oceans, Deep-Sea Res. Pt II, 54(5), 447-477, 2007.

Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T. J., and Barlow, R.G.: An absorption model to determine phytoplankton size classes from satellite ocean colour, Remote Sens. Environ., 112(6), 3153-3159, 2008.

Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M. and Yamanaka, Y.: Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types, Biogeosciences, 8(2), 311-327, 2011.

Hirata, T., Saux-Picart, S., Hashioka, T., Aita-Noguchi, M., Sumata, H., Shigemitsu, M., Allen, J. I., and Yamanaka, Y.: A comparison between phytoplankton community structures derived from a global 3D ecosystem model and satellite observation, J. Marine Syst., 109, 129-137, 2013.

Holt, J. T. and James, I. D.: An s coordinate density evolving model of the northwest European continental shelf: 1. Model description and density structure, J. Geophys. Res.-Oceans, 106(C7), 14015-14034, 2001.

Holt, J. T., Allen, J. I., Proctor, R., and Gilbert, F.: Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 1. Model overview and assessment of the hydrodynamics, J. Marine Syst., 57(1), 167-188, 2005.

Holt, J., Allen, J. I., Anderson, T. R., Brewin, R., Butenschön, M., Harle, J., Huse, G., Lehodey, P., Lindemann, C., Memery, L. and Salihoglu, B.: Challenges in integrative approaches to modelling the marine ecosystems of the North Atlantic: Physics to fish and coasts to ocean, Prog. Oceanogr., 129, 285-313, 2014.

Hyder, K., Rossberg, A. G., Allen, J. I., Austen, M. C., Barciela, R. M., Bannister, H. J., Blackwell, P. G., Blanchard, J. L., Burrows, M. T., Defriez, E., and Dorrington, T.: Making modelling count-increasing the contribution of shelf-seas community and ecosystem models to policy development and management, Mar. Policy, 61, 291-302, 2015.

Jennings, S. and Collingridge, K.: Predicting consumer biomass, size-structure, production, catch potential, responses to fishing and associated uncertainties in the world's marine ecosystems., PloS one, 10(7), e0133794, 2015.

Jennings, S., Greenstreet, S., Hill, L., Piet, G., Pinnegar, J., and Warr, K.J.: Long-term trends in the trophic structure of the North Sea fish community: evidence from stable-isotope analysis, size-spectra and community metrics, Mar. Biol., 141(6), 1085-1097, 2002.

Jupp, T. E., Lowe, R., Coelho, C. A., and Stephenson, D. B.: On the visualization, verification and recalibration of ternary probabilistic forecasts, Philos. T. Roy. Soc. A, 370(1962), 1100-1120, 2012.

Kurekin, A.A., Miller, P.I., and ~~Van~~ van der Woerd, H.J.: Satellite discrimination of Karenia mikimotoi and *Phaeocystis* harmful algal blooms in European coastal waters: Merged classification of ocean colour data, Harmful Algae, 31, 163-176, 2014.

Kwiatkowski, L., Yool, A., Allen, J. I., Anderson, T. R., Barciela, R., Buitenhuis, E. T., Butenschön, M., Enright, C., Halloran, P. R., Le Quéré, C., and ~~De~~ de Mora, L.: iMarNet: an ocean biogeochemistry model intercomparison project within a common physical ocean modelling framework, Biogeosciences, 11(24), 7291-7304, 2014.

Le Quéré, C., Harrison, S. P., Prentice, I. C., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., and Klaas, C.: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, Global Change Biol., 11(11), 2016-2040, 2005.

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

Lewis, K., Allen, J. I., Richardson, A. J., and Holt, J. T.: Error quantification of a high resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 3, validation with Continuous Plankton Recorder data, J. Marine Syst., 63(3), 209-224, 2006.

Lewis, K. and Allen, J. I.: Validation of a hydrodynamic-ecosystem model simulation with time-series data collected in the western English Channel, J. Marine Syst., 77(3), 296-311, 2009.

Mackinson, S.: Combined analyses reveal environmentally driven changes in the North Sea ecosystem and raise questions regarding what makes an ecosystem model's performance credible? , Can. J. Fish. Aquat. Sci., 71(1), 31-46, 2014.

MacLachlan, C., Arribas, A., Peterson, D., Maidens, A., Fereday,D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., and Xavier, P.: Global Seasonal Forecast System 5 (GloSea5): a high resolution seasonal forecast system, Q. J. Roy. Meteor. Soc., 141(689), 1072-1084, 2014.

Madec, G.: NEMO ocean engine, Note du Pole de modélisation, Insititut Pierre-Simon Laplace (IPSL), France, No 27, ISSN No 1288-1619, 2008.

Mogensen, K., Balmaseda, M., and Weaver, A.: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System4, ECMWF Technical Memorandum 668, 59 pp., 2012.

Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M-H., Devred, E., and Bouman, H.: Remote sensing of phytoplankton functional types, Remote Sens. Environ., 112, 3366-3375, 2008.

Núñez-Riboni, I. and Akimova, A.: Monthly maps of optimally interpolated in situ hydrography in the North Sea from 1948 to 2013, J. Marine Syst., 151, 15-34, 2015.

5 O'Dea, E. J., Arnold, A .K., Edwards, K. P., Furner, R., Hyder, P., Martin, M. J., Siddorn, J. R., Storkey, D., While, J., Holt, J. T. and Liu, H.: An operational ocean forecast system incorporating NEMO and SST data assimilation for the tidally driven European North-West shelf, Journal of Operational Oceanography, 5(1), 3-17, 2012.

O'Dea, E. J., et al.: Siddorn, J., Furner, R., While, J., Sykes, P., King, R., Holt, J., and Wakelin, S.: The CO5 configuration of the 7km Atlantic Margin Model: Large scale biases and sensitivity of CO5 to forcing, physics options and vertical resolution, to be submitted to
10 Geoscientific Model Development, in preparation.

Payne, M. R., Barange, M., Cheung, W. W., MacKenzie, B. R., Batchelder, H. P., Cormon, X., Eddy, T. D., Fernandes, J. A., Hollowed, A. B., Jones, M. C., and Link, J. S.: Uncertainties in projecting climate-change impacts in marine ecosystems, ICES J. Mar. Sci., fsv231, 2015.

Peloquin, J., Swan, C., Gruber, N., Vogt, M., Claustre, H., Ras, J., Uitz, J., Barlow, R., Behrenfeld, M., Bidigare, R. and Dierssen, H.: The
15 MAREDAT global database of high performance liquid chromatography marine pigment measurements, Earth System Science Data, 5(1), 109-123, 2013.

Peperzak, L., Colijn, F., Gieskes, W. W. C., and Peeters, J. C. H.: Development of the diatom-*Phaeocystis* spring bloom in the Dutch coastal zone of the North Sea: the silicon depletion versus the daily irradiance threshold hypothesis, J. Plankton Res., 20, 517-537, 1998.

Petersen, W., Wehde, H., Krasemann, H., Colijn, F., and Schroeder, F.: FerryBox and MERIS - Assessment of coastal and shelf sea ecosystems
20 by combining in situ and remotely sensed data, Estuar. Coast. Shelf. S., 77(2), 296-307, 2008.

Piroddi, C., Teixeira, H., Lynam, C. P., Smith, C., Alvarez, M. C., Mazik, K., Andonegi, E., Churilova, T., Tedesco, L., Chifflet, M., and Chust, G.: Using ecological models to assess ecosystem status in support of the European Marine Strategy Framework Directive, Ecol. Indic., 58, 175-191, 2015.

Plagányi, É. E., Punt, A. E., Hillary, R., Morello, E. B., Thébaud, O., Hutton, T., Pillans, R. D., Thorson, J. T., Fulton, E. A., Smith, A. D.
25 and Smith, F.: Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity, Fish Fish., 15(1), 1-22, 2014.

Richardson, A. J., Walne, A. W., John, A. W. G., Jonas, T. D., Lindley, J. A., Sims, D. W., Stevens, D., and Witt, M.: Using continuous plankton recorder data, Prog. Oceanogr., 68(1), 27-74, 2006.

Rombouts, I., Beaugrand, G., Fizzala, X., Gaill, F., Greenstreet, S. P. R., Lamare, S., Le Loc'h, F., McQuatters-Gollop, A., Mialet, B., Niquil,
30 N., and Percelay, J.: Food web indicators under the Marine Strategy Framework Directive: From complexity to simplicity?, Ecol. Indic., 29, 246-254, 2013.

Roselli, L. and Basset, A.: Decoding size distribution patterns in marine and transitional water phytoplankton: from community to species level, PloS one, 10(5), e0127193, 2015.

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

Ruardij, P., van Haren, H., and Ridderinkhof, H.: The impact of thermal stratification on phytoplankton and nutrient dynamics in shelf seas, J. Sea Res., 38, 311-331, 1997.

Ruardij, P., Veldhuis, M. J. W., and Brussaard, C. P. D.: Modeling the bloom dynamics of the polymorphic phytoplankter *Phaeocystis* globosa: impact of grazers and viruses, Harmful Algae, 4, 941-963, 2005.

Ryther, J. H.: Photosynthesis and fish production in the sea. The production of organic matter and its conversion to higher forms of life vary throughout the world ocean, Science, 166, 72-76, 1969.

Schlüter, L., Henriksen, P., Nielsen, T. G., and Jakobsen, H. H.: Phytoplankton composition and biomass across the southern Indian Ocean, Deep-Sea Res. Pt I, 58(5), 546-556, 2011.

Schlüter, L., Møhlenberg, F., and Kaas, H.: Temporal and spatial variability of phytoplankton monitored by a combination of monitoring buoys, pigment analysis and fast screening microscopy in the Fehmarn Belt Estuary, Environ. Monit. Assess., 186(8), 5167-5184, 2014.

Shephard, S., Greenstreet, S. P., Piet, G. J., Rindorf, A. and Dickey-Collas, M.: Surveillance indicators and their use in implementation of the Marine Strategy Framework Directive, ICES J. Mar. Sci., fsv131, 2015.

Sherrard, N. J., Nimmo, M., and Llewellyn, C. A.: Combining HPLC pigment markers and ecological similarity indices to assess phytoplankton community structure: an environmental tool for eutrophication?, Sci. Total Environ., 361(1), 97-110, 2006.

Shutler, J. D., Warren, M. A., Miller, P. I., Barciela, R., Mahdon, R., Land, P. E., Edwards, K., Wither, A., Jonas, P., Murdoch, N., and Roast, S. D.: Operational monitoring and forecasting of bathing water quality through exploiting satellite Earth observation and models: The AlgaRisk demonstration service, Computers & Geosciences, 77, 87-96, 2015.

Siddorn, J. R., and Furner, R.: An analytical stretching function that combines the best attributes of geopotential and terrain-following vertical coordinates, Ocean Model., 66, 1-13, 2013.

Sinha, B., Buitenhuis, E.T., Le Quéré, C., and Anderson, T.R.: Comparison of the emergent behavior of a complex ecosystem model in two ocean general circulation models, Prog. Oceanogr., 84, 204-224, 2010.

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

Smaal, A. C., Schellekens, T., van Stralen, M. R., and Kromkamp, J. C.: Decrease of the carrying capacity of the Oosterschelde estuary (SW Delta, NL) for bivalve filter feeders due to overgrazing?, Aquaculture, 404, 28-34, 2013.

Stewart, I. J. and Martell, S. J.: Reconciling stock assessment paradigms to better inform fisheries management, ICES J. Mar. Sci., fsv061, 2015.

Stips, A., Bolding, K., Pohlmann, T., and Burchard, H.: Simulating the temporal and spatial dynamics of the North Sea using the new model GETM (General Estuarine Transport Model), Ocean Dynam., 54(2), 266-283, 2004.

Storkey, D., Blockley, E. W., Furner, R., Guiavarc'h, C., Lea, D., Martin, M. J., Barciela, R. M., Hines, A., Hyder, P. and Siddorn, J. R.: Forecasting the ocean state using NEMO: The new FOAM system, Journal of Operational Oceanography, 3(1), 3-15, 2010.

Sutherland, W. J., Armstrong-Brown, S., Armsworth, P. R., Tom, B., Brickland, J., Campbell, C. D., Chamberlain, D. E., Cooke, A. I., Dulvy, N. K., Dusic, N. R., and Fitton, M.: The identification of 100 ecological questions of high policy relevance in the UK, J. Appl. Ecol., 43(4), 617-627, 2006.

Sykes, P. A. and Barciela, R. M.: Assessment and development of a sediment model within an operational system, J. Geophys. Res.-Oceans, 117(C4), 2012.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res.-Atmos., 106(D7), 7183-7192, 2001.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philos. T. Roy. Soc. A, 365(1857), 2053-2075, 2007.

Thorpe, R. B., Le Quesne, W. J., Luxford, F., Collie, J. S., and Jennings, S.: Evaluation and management implications of uncertainty in ~~amultispecies~~ a multispecies size-structured model of population and community responses to fishing, ~~Journal~~Methods Ecol. Evol., 6(1), 49-58, 2015.

Thyssen, M., Alvain, S., Lefèbvre, A., Dessailly, D., Rijkeboer, M., Guiselin, N., Creach, V., and Artigas, L. F.: High-resolution analysis of a North Sea phytoplankton community structure based on in situ flow cytometry observations and potential implication for remote sensing, Biogeosciences, 12(13), 4051-4066, 2015.

Uitz, J., Claustre, H., Morel, A., and Hooker, S. B.: Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, J. Geophys. Res.-Oceans, 111(C8), 2006.

Uitz, J., Huot, Y., Bruyant, F., Babin, M., and Claustre, H.: Relating phytoplankton photophysiological properties to community structure on large scales, Limnol. Oceanogr., 53(2), 614-630, 2008.

Van der Molen, J., Aldridge, J. N., Coughlan, C., Parker, E. R., Stephens, D., and Ruardij, P.: Modelling marine ecosystem response to climate change and trawling in the North Sea, Biogeochemistry, 113, 213-236, 2013.

Van der Molen, J., Smith, H. C. M., Lepper, P., Limpenny, S., and Rees, J.: Predicting the large-scale consequences of offshore wind turbine array development on a North Sea ecosystem, Cont. Shelf Res., 85, 60-72, 2014.

Van der Molen, J., van Beek, J., Augustine, S., Vansteenbrugge, L., van Walraven, L., Langenberg, V., van der Veer, H.W., Hostens, K., Pitois, S., and Robbens, J.: Modelling survival and connectivity of *Mnemiopsis leidyi* in the south-western North Sea and Scheldt estuaries, Ocean Sci., 11, 405-424, 2015.

Van der Molen, J., Ruardij, P., and Greenwood, N.: Potential environmental impact of tidal energy extraction in the Pentland Firth at large spatial scales: results of a biogeochemical model, Biogeosciences, 13(8), 2593-2609, ~~2016a.~~ 2016.

Van der Molen, J., Ruardij, P., and Greenwood, N.: A 3D SPM model for biogeochemical modelling, with application to the northwest European continental shelf, ~~Submitted to~~ J. Sea Res., ~~2016b.~~ in press, 2017.

Van Leeuwen, S. M., ~~Van~~ van der Molen, J., Ruardij, P., Fernand, L., and Jickells, T.: Modelling the contribution of deep chlorophyll maxima to annual primary production in the North Sea, Biogeochemistry, 113(1), 137-152, 2013.

Van Leeuwen, S., Tett, P., Mills, D., and ~~Van~~ van der Molen, J.: Stratified and non-stratified areas in the North Sea: Long-term variability and biological and policy implications, J. Geophys. Res.-Oceans, 120(7), 4670-4686, 2015.

Van Leeuwen, S. M., Le Quesne, W. F., and Parker, E.R.: Potential future fisheries yields in shelf waters: a model study of the effects of climate change and ocean acidification, Biogeosciences, 13(2), 441-454, 2016.

Veldhuis, M. J. W. and Admiraal, W.: Influence of phosphate depletion on the growth and colony formation of Phaeocystis pouchetii, Mar. Biol., 95, 47-54, 1987.

Vichi, M., Oddo, P., Zavatarelli, M., Coluccelli, A., Coppini, G., Celio, M., Umani, S. F., and Pinardi, N.: Calibration and validation of a one-dimensional complex marine biogeochemical flux model in different areas of the northern Adriatic Sea, Ann. Geophys., 21(1), 413-436, 2003.

Vichi, M., Ruardij, P., and Baretta, J.W.: Link or sink: a modelling interpretation of the open Baltic biogeochemistry, Biogeosciences, 1, 79-100, 2004.

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

Wakelin, S. L., Artioli, Y., Butenschön, M., Allen, J. I., and Holt, J. T.: Modelling the combined impacts of climate change and direct anthropogenic drivers on the ecosystem of the northwest European continental shelf, J. Marine Syst., 152, 51-63, 2015a.

Wakelin, S., While, J., King, R., O'Dea, E., Holt, J., Furner, R., Siddorn, J., Martin, M., McEwan, R., Blockley, E., and Tinker, J.: Quality Information Document: North West European Shelf Reanalysis - NORTHWESTSHELF_REANALYSIS_PHYS_004_009 and NORTHWESTSHELF_REANALYSIS_BIO_004_011, EU Copernicus Marine Environment Monitoring Service, http://marine.copernicus.eu/documents/QUID/CMEMS-NWS-QUID-004-009-011.pdf, 2015b.

5   Ward, B. A., Schartau, M., Oschlies, A., Martin, A. P., Follows, M. J., and Anderson, T. R.: When is a biogeochemical model too complex? Objective model reduction and selection for North Atlantic time-series sites, Prog. Oceanogr., 116, 49-65, 2013.

Waters, J., Lea, D. J., Martin, M. J., Mirouze, I., Weaver, A., and While, J.: Implementing a variational data assimilation system in an operational 1/4 degree global ocean model, Q. J. Roy. Meteor. Soc., 141(687), 333-349, 2015.

Wollschläger, J., Wiltshire, K. H., Petersen, W., and Metfies, K.: Analysis of phytoplankton distribution and community structure in the
10   German Bight with respect to the different size classes, J. Sea Res., 99, 83-96, 2015.

Xiao, Y. and Friedrichs, M. A. M.: Using biogeochemical data assimilation to assess the relative skill of multiple ecosystem models in the Mid-Atlantic Bight: effects of increasing the complexity of the planktonic food web, Biogeosciences, 11(11), 3015-3030, 2014.
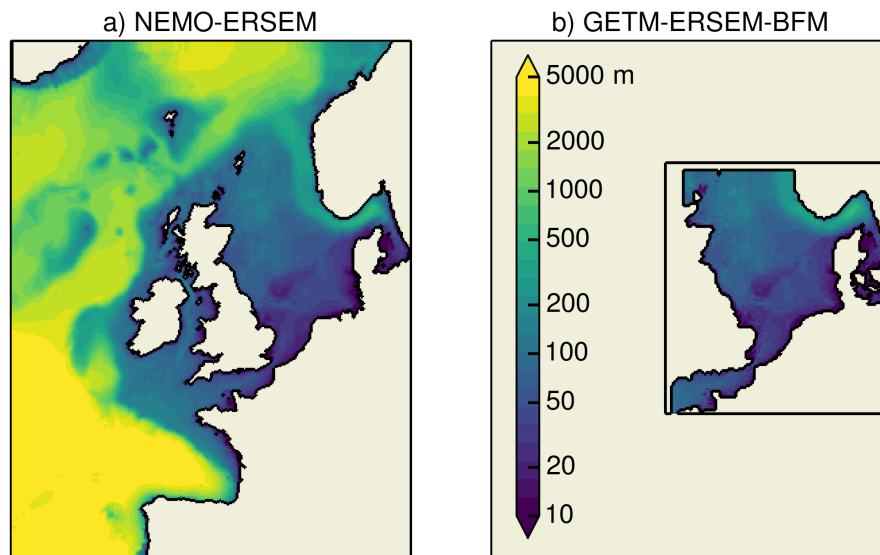
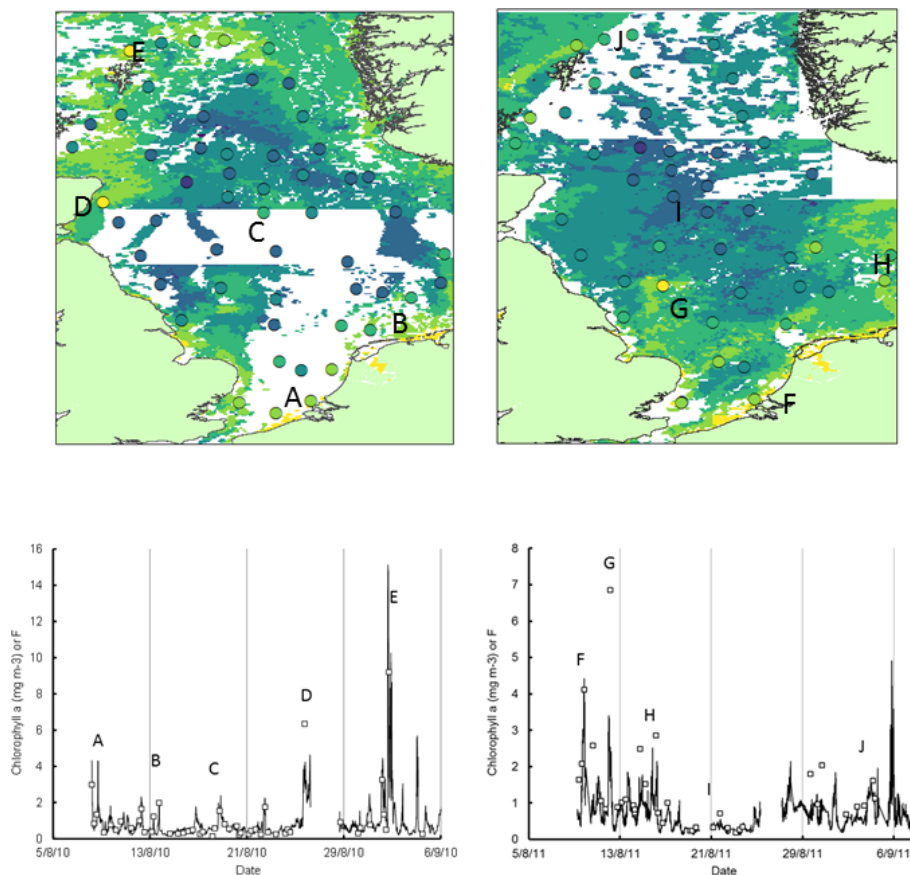**Figure 1.** Maps of the model domain and bathymetry for a) NEMO-ERSEM and b) GETM-ERSEM-BFM.

Figure 2. Maps of satellite-derived surface phytoplankton chlorophyll distribution during the summer IBTS cruises of a) 2010 and b) 2011, overlaid with the in situ observations in circles. White areas are where no satellite data were available. Time series of phytoplankton chlorophyll along the IBTS cruise track in c) 2010 and d) 2011 as assessed by continuous measurements of chlorophyll fluorescence (solid black line), and sampling of surface water for quantification of chl-a (open squares). Specific events along the tracks are referenced with a letter.
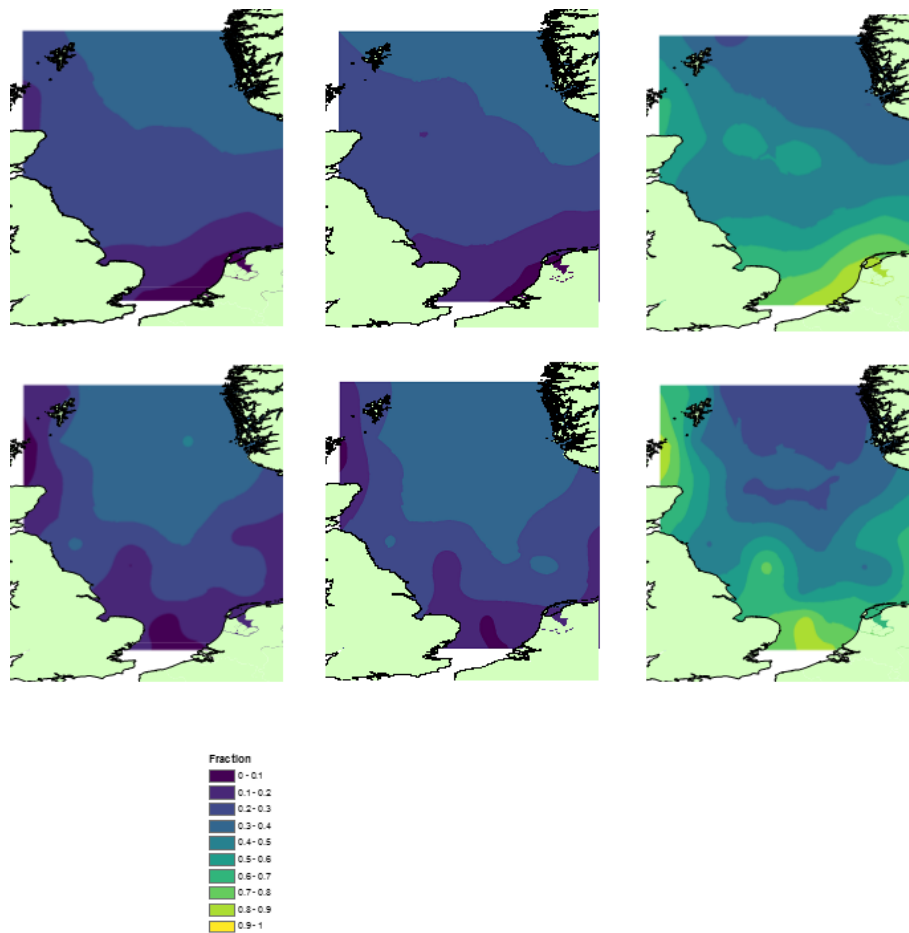
Figure 3. Maps of percentage surface PFT distribution during the summer IBTS cruises of 2010 (upper maps, a-c) and 2011 (lower maps, d-f) for pico-phytoplankton (a, d), nano-phytoplankton (b, e) and micro-phytoplankton (c, f).
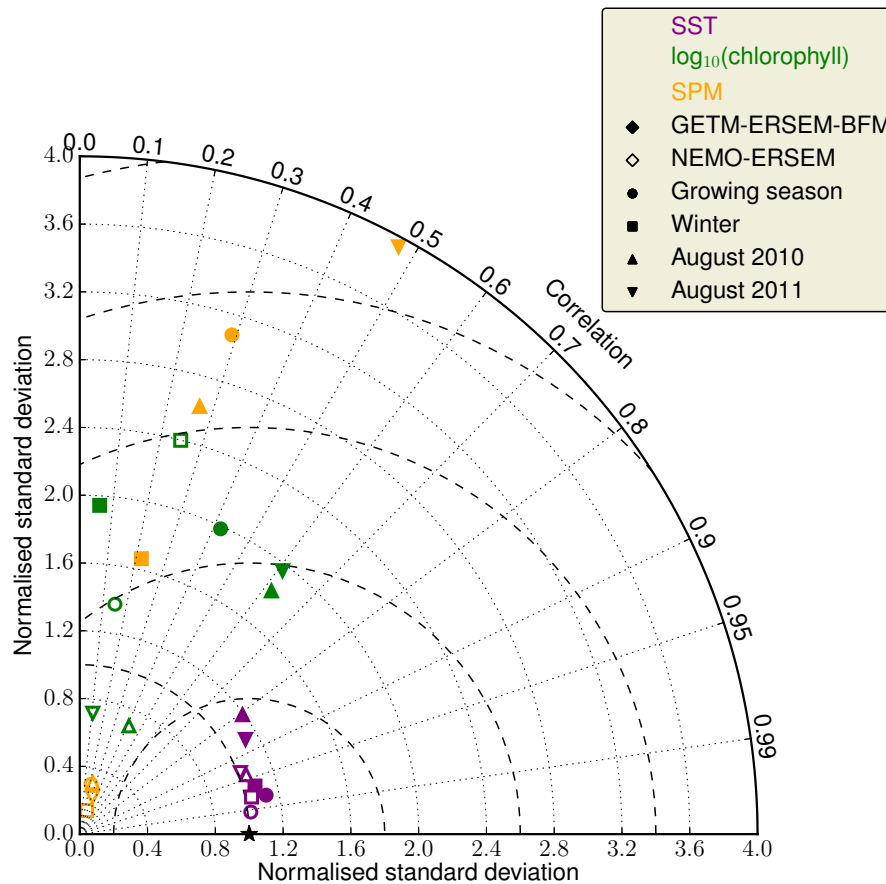
Figure 4. Taylor plot of SST (purple), $\log_{10}$(chlorophyll) (green), and SPM (orange) for the growing season (circles), winter (squares), August 2010 (upwards triangles), and August 2011 (downwards triangles). Filled symbols are GETM-ERSEM-BFM, unfilled symbols are NEMO-ERSEM. Validation has been performed in observation space, validating the models against OSTIA for SST, and OC5 ocean colour data for $\log_{10}$(chlorophyll) and SPM. A perfect model would plot at 1.0 on the x-axis, marked with a black star.
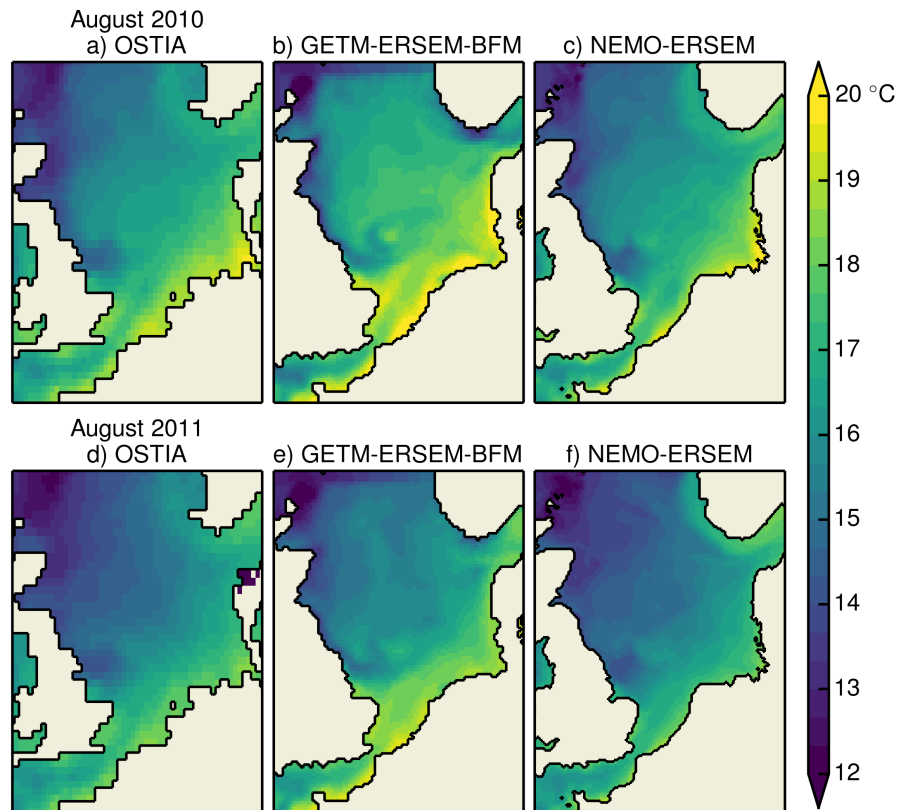
Figure 5. ~~Monthly~~ Maps of monthly mean SST for August 2010 (a-c) and August 2011 (d-f): observational data (a, d), GETM-ERSEM-BFM (b, e), and NEMO-ERSEM (c, f).
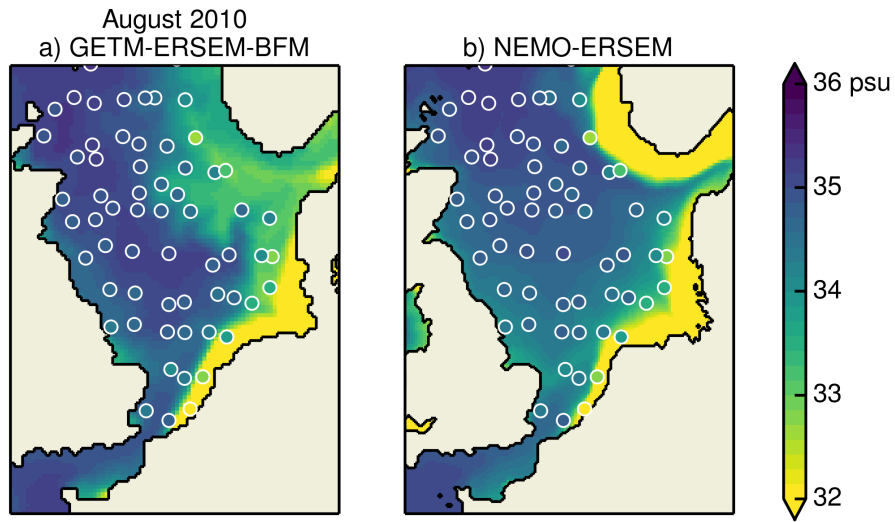
Figure 6. Maps of monthly mean SSS for August 2010 from a) GETM-ERSEM-BFM and b) NEMO-ERSEM. The in situ IBTS observations from August 2010 are overlaid in circles.
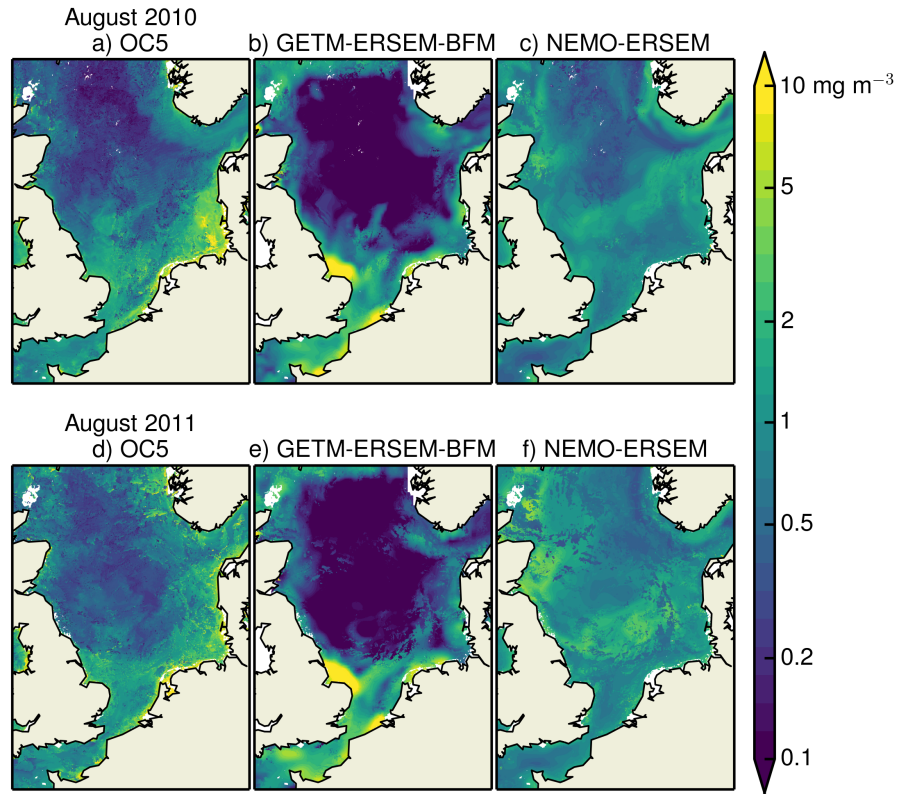
Figure 7. Composites of sea surface chlorophyll at ocean colour observation points for August 2010 (a-c) and August 2011 (d-f): satellite observations (a, d), GETM-ERSEM-BFM (b, e), and NEMO-ERSEM (c, f).
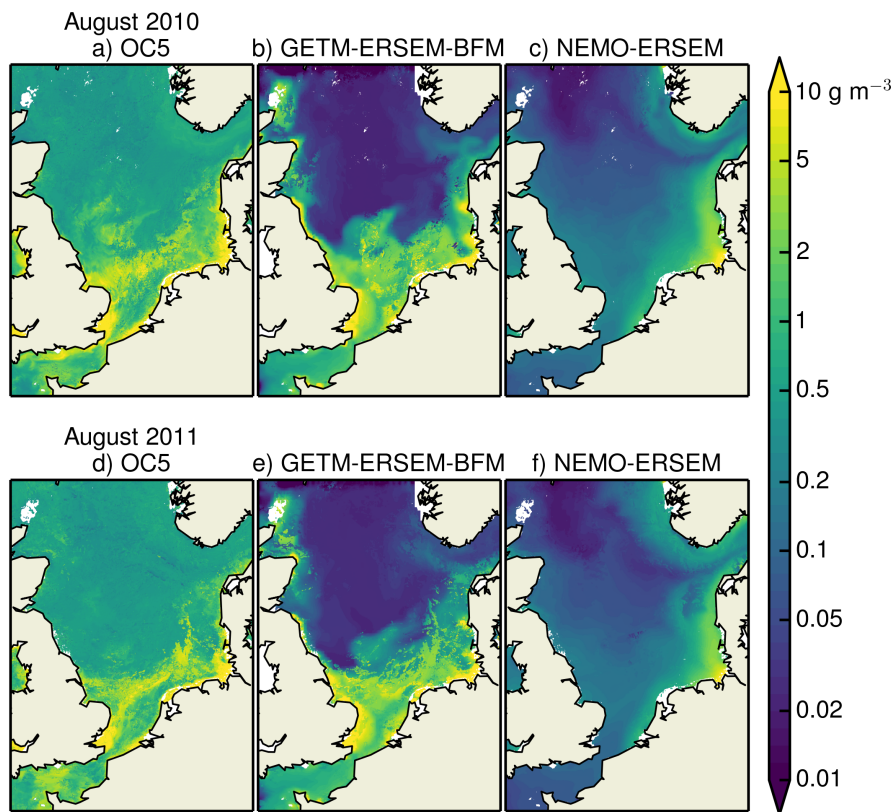
Figure 8. Composites of sea surface SPM at ocean colour observation points for August 2010 (a-c) and August 2011 (d-f): satellite observations (a, d), GETM-ERSEM-BFM (b, e), and NEMO-ERSEM (c, f).
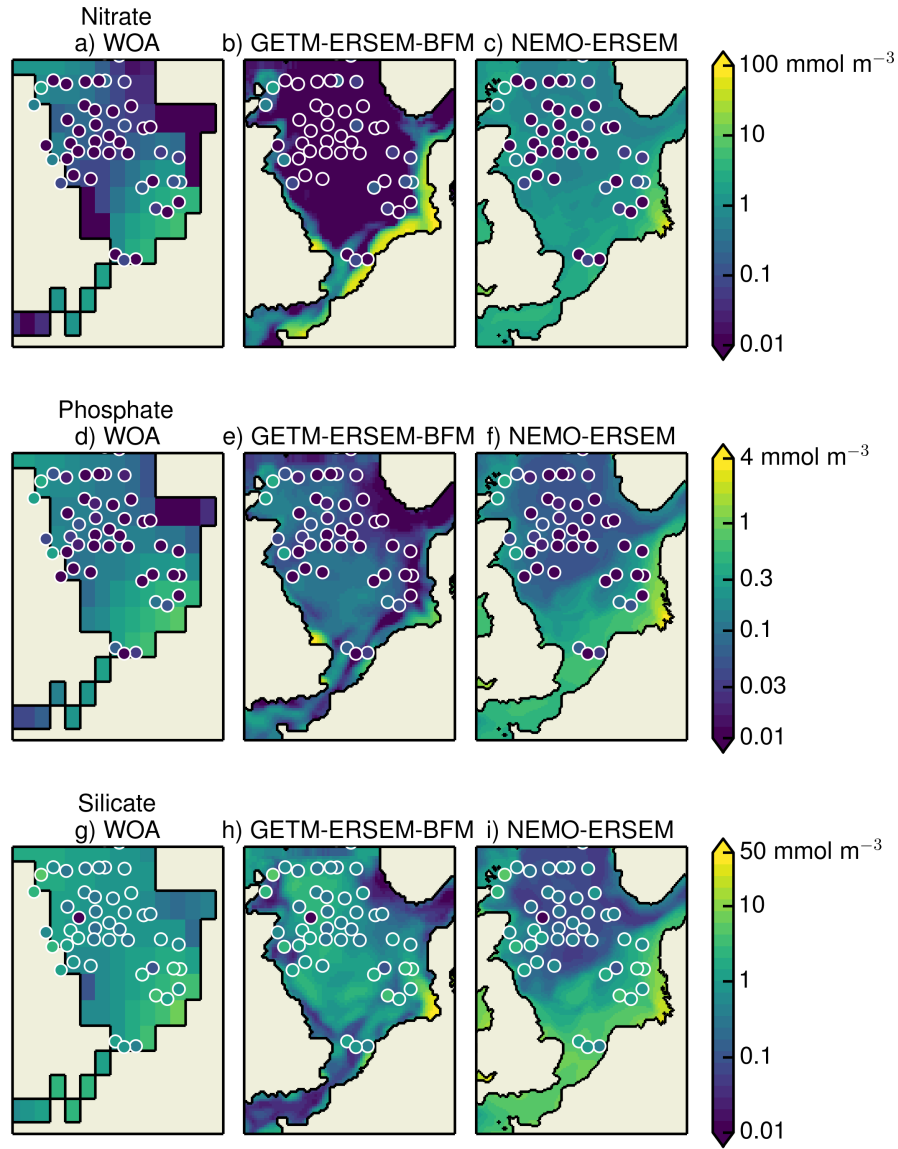
Figure 9. ~~Monthly~~ <ins>Maps of monthly</ins> mean sea surface nitrate (a-c), phosphate (d-f), and silicate (g-i) for August 2010: World Ocean Atlas climatology (a, d, g), ~~GETM-ESREM-BFM~~ <ins>GETM-ERSEM-BFM</ins> (b, e, h), and NEMO-ERSEM (c, f, i). <ins>The in situ IBTS surface observations from August 2010 are overlaid in circles.</ins>
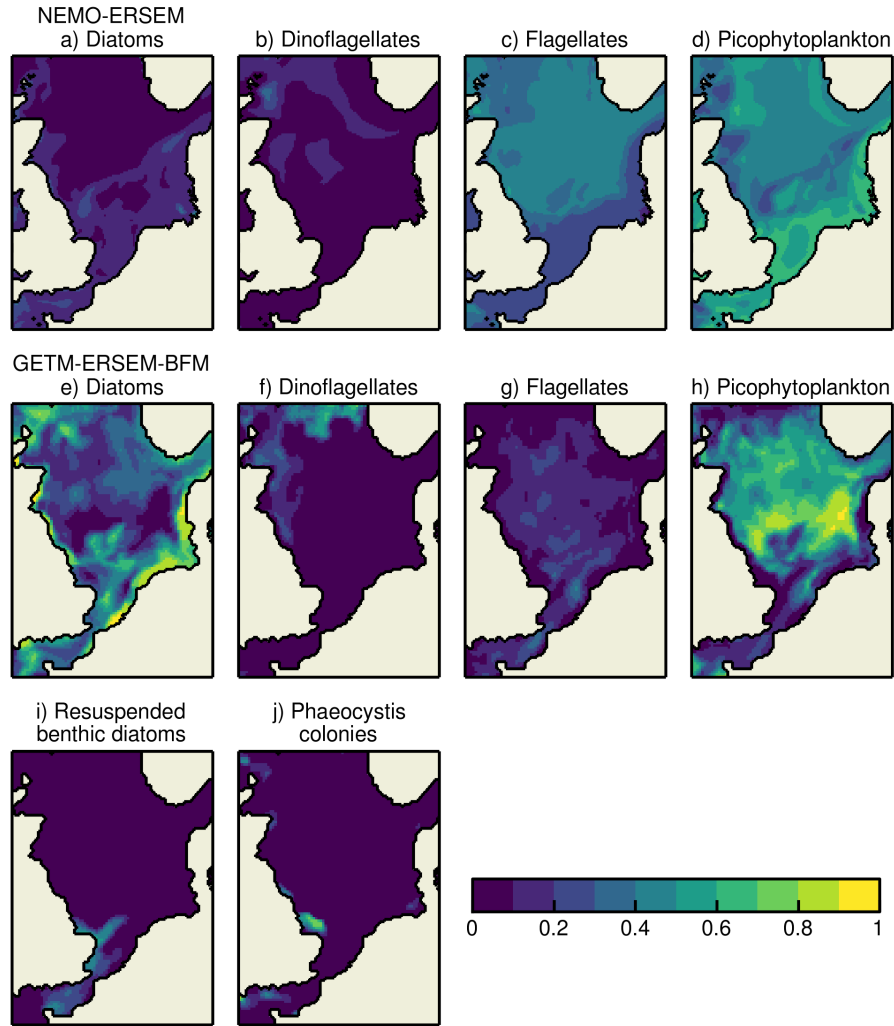
Figure 10. ~~Monthly~~ Maps of monthly mean sea surface PFT fractions for August 2010 from NEMO-ERSEM (a-d) and GETM-ERSEM-BFM (e-j). PFT fractions have been calculated as the proportion of the total sea surface chlorophyll concentration.
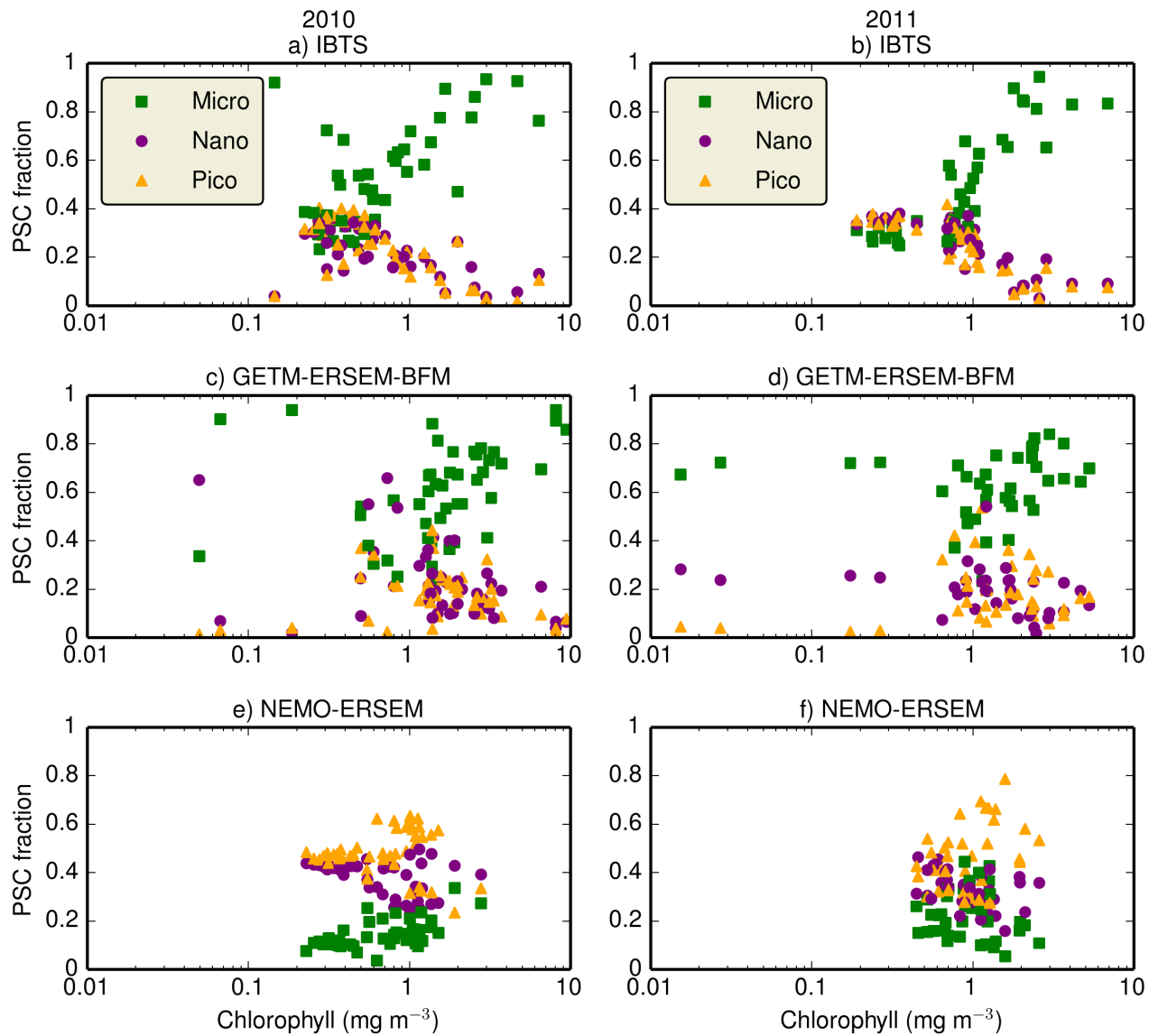
40

Figure 11. ~~PSC~~ Phytoplankton size class (PSC) distribution as a function of chlorophyll concentration, plotted at each IBTS observation location in 2010 (a, c, e) and 2011 (b, d, f) from the IBTS observations (a, b), GETM-ERSEM-BFM (c, d), and NEMO-ERSEM (e, f).
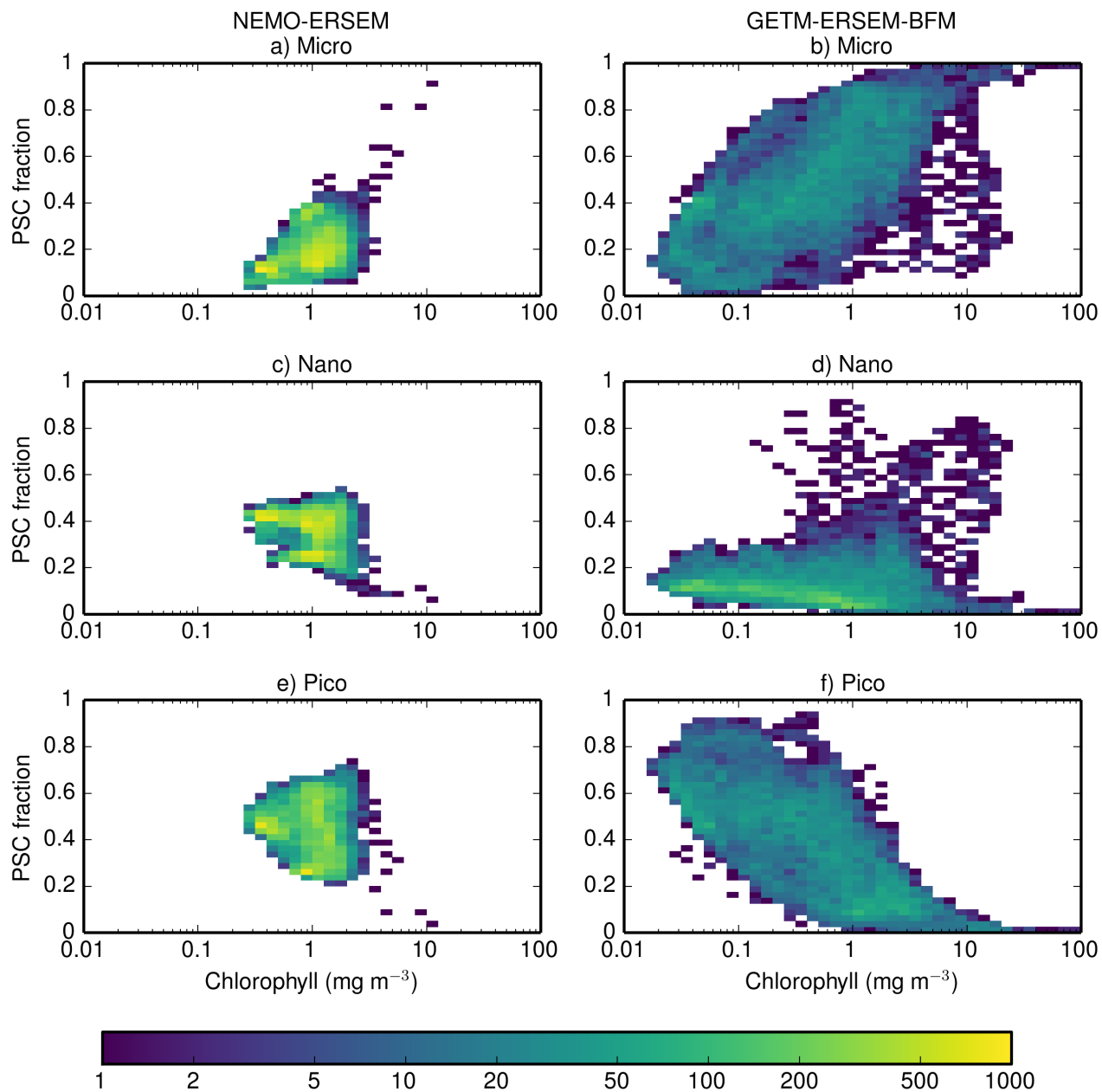
Figure 12. 2D histograms of the distribution of each phytoplankton size class (PSC) with chlorophyll, from NEMO-ERSEM (a, c, e) and GETM-ERSEM-BFM (b, d, f). The PSC fractions have been calculated at each surface model grid point in the monthly means for August 2010 and August 2011.
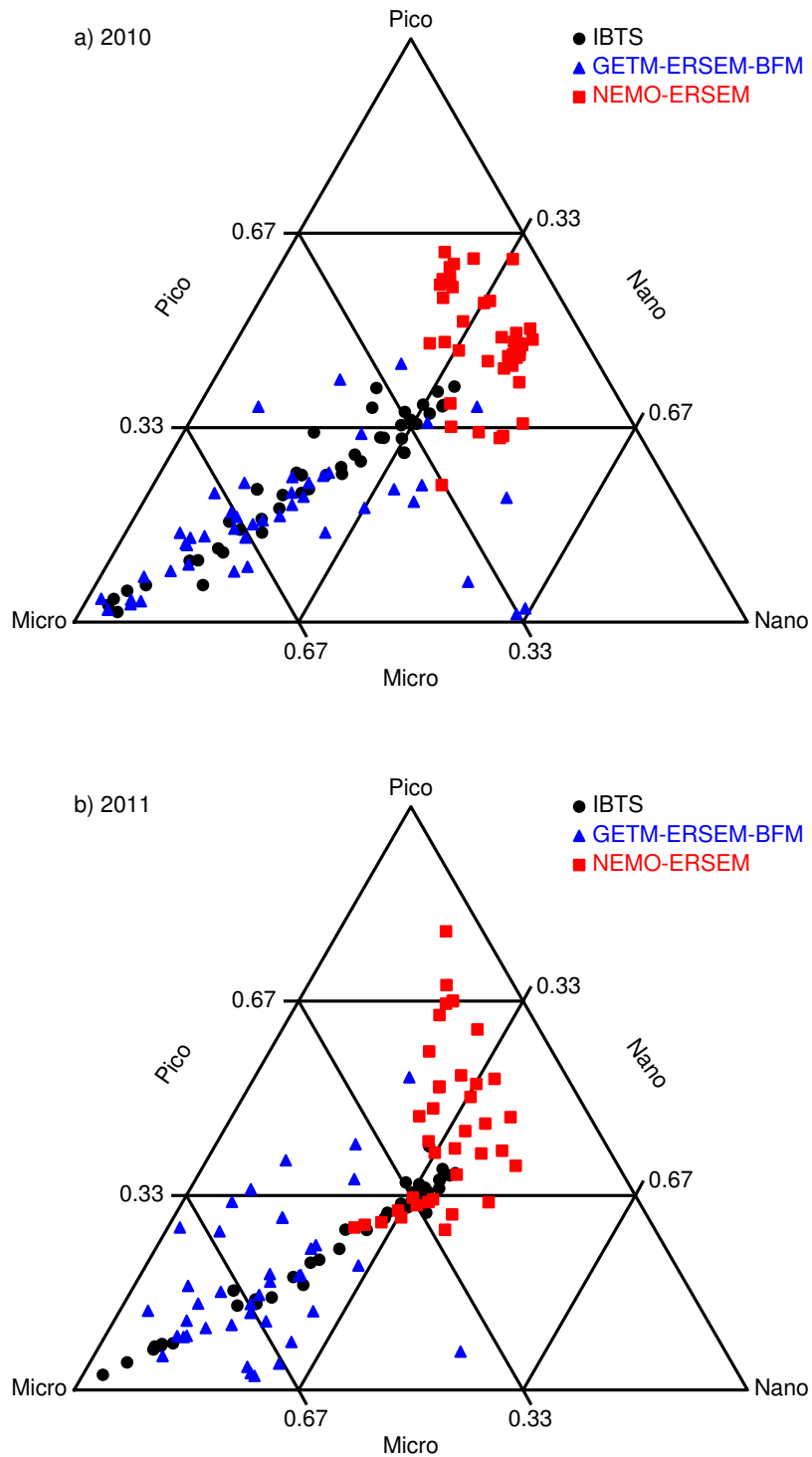
Figure 13. Ternary plots of phytoplankton community structure , showing the phytoplankton size class distribution at IBTS observation locations in a) 2010 and b) 2011, from the IBTS observations, GETM-ERSEM-BFM, and NEMO-ERSEM.

Figure 14. Maps of modelled surface phytoplankton community structure for August 2010 (a, b) and August 2011 (c, d) from NEMO-ERSEM (a, c) and GETM-ERSEM-BFM (b, d). The IBTS observations, sampled in August and early September of each year, are overlaid in circles.

Table 1. ~~General-level~~ Summary of general-level model characteristics.
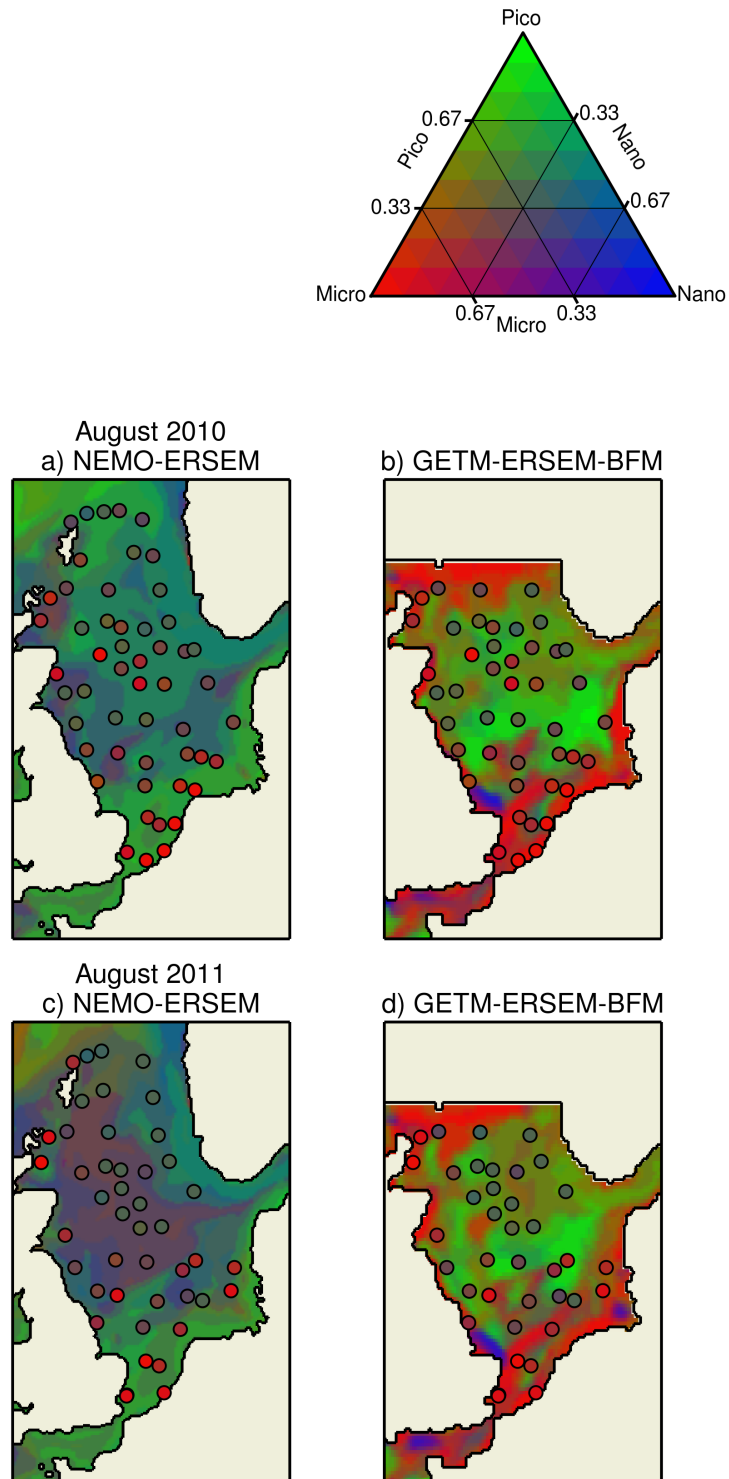
	NEMO-ERSEM	GETM-ERSEM-BFM
Hydrodynamics	NEMO	GETM
Biogeochemistry	~~PML-ERSEM~~ ERSEM	ERSEM-BFM
Domain	North-west European Shelf	North Sea
Horizontal resolution	~7 km	~10 km
Vertical resolution	50 levels, terrain-following with constant 1m top box	25 levels, terrain following General Vertical Coordinates
Tidal boundary	Elevation and currents from Met Office global model, Flather radiation condition	Elevations and currents from shelf model, Flather radiation condition
Temperature and salinity boundary	Met Office global model (GloSea5 reanalysis)	ECMWF global model
Nutrients boundary	World Ocean Atlas climatology	World Ocean Atlas climatology
Meteorological forcing	ECMWF ERA-Interim reanalysis: surface temperature, 2m wind, air pressure, heat fluxes, precipitation. 3-hourly	ECMWF ERA-40 and operational hindcast: surface temperature, 10m wind, air pressure, humidity, cloud cover. 6-hourly
Atmospheric nutrient deposition	Not included	Not included
River inputs	Freshwater flow: E-Hype; nutrients: climatology; sediments: daily climatology of satellite SPM at river points	Cefas data base, interpolated daily values of runoff and nutrients based on various observational sources
SPM concentrations	Modelled, two size classes, full transport, resuspension, aggregation and disaggregation	Modelled, 1 size class with concentration-dependent settling, full transport, resuspension by waves and currents
Nutrients	N, P, Si, C, O (Fe available but not used)	N, P, Si, C, O reduction equivalents
Pelagic autotrophic types	Diatoms, flagellates, dinoflagellates, picophytoplankton	Diatoms, flagellates, dinoflagellates, picophytoplankton, *Phaeocystis* colonies, resuspended benthic diatoms, pelagic nitrifiers
Zooplankton functional types	Mesozooplankton, microzooplankton, heterotrophic nanoflagellates	Filterfeeder larvae, mesozooplankton, omnivorous mesozooplankton, microzooplankton, heterotrophic nanoflagellates
Pelagic bacteria	Pelagic bacteria	Pelagic bacteria
Pelagic detritus	Labile dissolved organic matter, semi-labile dissolved organic matter, small particulate organic matter, medium particulate organic matter, large particulate organic matter	Labile organic carbon, TEP, particulate organic carbon (POC). Degradability of POC depends on nutrient:C quota. Vertical exchange of POC coupled to SPM transport
Type of benthic model	3-layer model: oxic layer, denitrification layer, anoxic layer	3-layer model: oxic layer, denitrification layer, anoxic layer
Seabed characterisation	Distribution of the two modelled SPM size classes, dependent on model dynamics	Porosity interpolated from North Sea Benthos Survey grain size data
Benthic autotrophic types	Not included	Benthic diatoms, benthic nitrifying bacteria, nitrifying archaea
Benthic macrofauna	Deposit feeders, suspension feeders, meiobenthos	Epibenthos, Deposit Feeders, Filter Feeders, Meiobenthos, Benthic predators
Benthic bacteria	Aerobic bacteria, anaerobic bacteria	Aerobic bacteria, anaerobic bacteria
Benthic detritus	Dissolved organic matter, particulate organic matter, buried organic matter	Labile organic carbon, particulate organic carbon
CO_2 method	Available but not used	Benthic and pelagic CO_2, pH, alkalinity
Pelagic nutrient regeneration	Nitrification depends on dynamics of nitrifying bacteria	Nitrification depends on dynamics of nitrifying archaea and bacteria
Benthic nutrient regeneration	Modelling of fluxes based on estimation of nutrient gradients on basis of processes and concentrations in the 3 benthic layers	Modelling of fluxes based on estimation of nutrient gradients on basis of processes and concentrations in the 3 benthic layers for phosphate, ammonium, nitrate, reduction equivalents, silicate, DIC, alkalinity. Dynamic determination of nitrification rate from benthic nitrifier model
Spinup period	Previous hindcast of Edwards et al. (2012), run for 2007 from previously spun-up fields	1991-2001
Production run	2003-2012 (also run for 1983-1989 and 1989-2003, but sections not continuous)	2002-2011
Data assimilation	Satellite and in situ SST; 3D-Var	Not included
References	Blackford et al. (2004); Edwards et al. (2012)	Baretta et al. (1995); Ruardij and van Raaphorst (1995); van der Molen et al. (2016)

Table 2. Parameters of the four coinciding phytoplankton functional types related to inter-species competition.

	NEMO-ERSEM					GETM-ERSEM-BFM				
	Method/ parameter	Diatoms (P1)	Flagellates (P2)	Pico-phytoplankton (P3)	Dinoflagellates (P4)	Method/ parameter	Diatoms (P1)	Flagellates (P2)	Pico-phytoplankton (P3)	Dinoflagellates (P4)
Doubling temperature	Q_{10}	2	2	2	2	Q_{10}	2	2	2	2
Maximum productivity at 10 °C	sum	2.5	2.7	3.3	1.5	sum	3.0	4.875	5.6	1.75
Respiration rate at 10 °C	srs	0.05	0.05	0.05	0.05	srs	0.125	0.1	0.1	0.125
Fraction of PP excreted as PLOC/PDET	pu_ae	0.05	0.2	0.2	0.05	pu_ae	0.05	0.1	0.1	0.05
Activity respiration rate	pu_ra	0.1	0.25	0.25	0.25	pu_ra	0.1	0.1	0.2	0.1
Half-value of SiO_4-limitation	chP1sX	0.3	-	-	-	P_chPs	0.3	-	-	-
~~Miniumum~~ Minimum quota N	~~qnlP?cX~~ qnlP1cX … qnlP4cX	0.00687	0.00687	0.00687	0.00687	p_qnlc	0.00687	0.00687	0.00687	0.00687
Minimum quota P	~~qplP?cX~~ qplP1cX … qplP4cX	0.0004288	0.0004288	0.0004288	0.0004288	p_qplc	0.0003931	0.0003931	0.0003931	0.0003931
Minimum quota Si	-	-	-	-	-	-	0.09	-	-	-
Multiplication factor for critical N:C ratio	~~xpcP?nX~~ xpcP1nX … xpcP4nX	1	1	1	1	Not included	-	-	-	-
Multiplication factor for critical P:C ratio	~~xpcP?pX~~ xpcP1pX … xpcP4pX	1	1	1	1	Not included	-	-	-	-
Multiplication factor for maximum quotum nitrate uptake	~~xqnP?X~~ xqnP1X … xqnP4X	2	2	2	2	p_xqn	2	2	2	2
Multiplication factor for maximum quotum phosphate uptake	~~xqpP?X~~ xqpP1X … xqpP4X	2	2	2	2	p_xqp	2	2	2	2
Multiplication factor for maximum quotum silicate uptake	-	-	-	-	-	p_xqs	1.5	-	-	-
Affinity for N(3) [nitrate]	~~quP?n3X~~ quP1n3X … quP4n3X	0.0025	0.0025	0.0025	0.0025	p_qun	0.15	0.215	1.29	0.0084
Affinity for N(4) [ammonium]	~~quP?n4X~~ quP1n4X … quP4n4X	0.01	0.01	0.02	0.01	Grouped with N(3)	-	-	-	-
Affinity for P	~~qurP?pX~~ qurP1pX … qurP4pX	0.0025	0.0025	0.0025	0.0025	p_qup	0.15	0.215	1.29	0.0084
Affinity for Si	qsP1cx	0.03	-	-	-	p_qus	0.1	0	0	0
Nutrient stress threshold for sinking	~~esNIP?X~~ esNIP1X … esNIP4X	0.7	0.75	0.75	0.75	Different method: based on TEP production	-	-	-	-
Sinking by formation of macro-aggregates	-	-	-	-	-	-	Threshold process: in presence of sufficient TEP and diatoms	Sticking to macro-aggregates	Sticking to macro-aggregates	Sticking to macro-aggregates
Lysis rate	~~sdoP?X~~ sdoP1X … sdoP4X	0.05	0.05	0.05	0.05 + 0.2	Max. by nutrient stress, p_sdmo	No lysis	0.025	0.15	0.001 + 0.0
Stress excretion	-	-	-	-	-	-	Extraction of TEP (carbohydrates)	-	-	-
Light susceptibility	P-I curve, alpha	2.98	2.98	2.98	2.98	P-I curve, Ke	35	70	124	116
Maximum Chl to C ratio	~~phimP?X~~ phimP1X … phimP4X	0.035	0.035	0.035	0.035	p_qchlc	0.02	0.035	0.035	0.035
Minimum Chl to C ratio	~~phiP?HX~~ phiP1HX …	0.025	0.025	0.025	0.025	p_qlPlc	0.0025	0.0035	0.0025	0.00225

Table 3. Statistical comparison of log₁₀(chlorophyll) ~~statistics~~ and phytoplankton size classes against IBTS observations.

Year	Model	Variable	Bias	RMSE	Correlation	No. observa
2010	~~GETM-ERSEM-BFM~~ ~~NEMO-ERSEM~~ GETM-ERSEM-BFM	$\log_{10}$(chlorophyll) ($\log_{10}$(mg m^{-3}))	0.369	0.597	0.334	
		Micro (fraction)	0.069	0.285	0.010	
		Nano (fraction)	-0.006	0.170	0.116	
		Pico (fraction)	-0.062	0.182	-0.265	46
	NEMO-ERSEM	$\log_{10}$(chlorophyll) ($\log_{10}$(mg m^{-3}))	0.017	0.339	0.446	
		Micro (fraction)	-0.385	0.434	0.160	
		Nano (fraction)	0.146	0.165	0.604	
		Pico (fraction)	0.240	0.292	-0.369	
2011	~~GETM-ERSEM-BFM~~ ~~NEMO-ERSEM~~ GETM-ERSEM-BFM	$\log_{10}$(chlorophyll) ($\log_{10}$(mg m^{-3}))	0.178	0.543	0.343	
		Micro (fraction)	0.138	0.261	0.265	
		Nano (fraction)	-0.076	0.134	0.387	
		Pico (fraction)	-0.062	0.178	-0.061	39
	NEMO-ERSEM	$\log_{10}$(chlorophyll) ($\log_{10}$(mg m^{-3}))	0.067	0.389	0.157	
		Micro (fraction)	-0.270	0.382	-0.416	
		Nano (fraction)	0.063	0.121	0.359	
		Pico (fraction)	0.207	0.299	-0.530	
2010+2011	~~GETM-ERSEM-BFM~~ ~~NEMO-ERSEM~~ GETM-ERSEM-BFM	$\log_{10}$(chlorophyll) ($\log_{10}$(mg m^{-3}))	0.282	0.573	0.320	
		Micro (fraction)	0.101	0.274	0.097	
		Nano (fraction)	-0.039	0.154	0.178	
		Pico (fraction)	-0.062	0.180	-0.163	85
	NEMO-ERSEM	$\log_{10}$(chlorophyll) ($\log_{10}$(mg m^{-3}))	0.040	0.363	0.358	
		Micro (fraction)	-0.332	0.411	-0.210	
		Nano (fraction)	0.108	0.146	0.401	
		Pico (fraction)	0.225	0.295	-0.452	