

Interactive comment on “Observing and modelling phytoplankton community structure in the North Sea: can ERSEM-type models simulate biodiversity?” by David A. Ford et al.

David A. Ford et al.

david.ford@metoffice.gov.uk

Received and published: 18 October 2016

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 will work to expand our discussion of these points, going into more details. Furthermore, we will include some validation against in situ temperature, salinity and nutrient

C1

data. However, we also stress that a detailed comparison of these processes was not our aim, and we will make 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 will also expand our discussion to propose targeted experiments which would lead on from our study. We will also revise the title to remove the reference to “biodiversity”, add a figure showing the model domains and cruise tracks, replot other figures to be as clear, consistent and useful as possible, discuss and compare results from other studies in the literature, and address 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.”

This has also been raised by the other reviewers, and we will replot it with more contrasting colours.

“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 will state this more clearly. Nonetheless, where differences in results are discussed, we will endeavour to do so in more detail. Figure 4 was

C2

included to demonstrate inter-annual and spatial variability, as well as for validation, but we can remove it if the editor wishes.

"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 will reword 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 will add some discussion regarding these points.

"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 will include some discussion regarding other results in the literature.

"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 will reword the conclusion to come across as more pragmatic.

Detailed response to Reviewer 2

C3

"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 will do so accordingly. We will also replace "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)."

We will include a discussion of previous studies in the literature. 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 will check the document and reword accordingly where appropriate.

"3) 'Observations' section is incomplete: In this section, only the procedures relating

C4

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 will move 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 et 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 will expand the description of the models as requested, and include a figure comparing the domains. See also our response to the related comment by Reviewer 3 below.

"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 will replot Fig. 11 using a more contrasting colour scheme, and investigate the value of adding a histogram plot.

C5

"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 will investigate this point further.

"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 will add more discussion to this point, and investigations suggest it may be related in part to inter-annual variability in wind speeds and therefore mixed layer depths, which alters nutrient concentrations and temperatures, with knock-on effects on the ecosystem dynamics. Although as noted in our response to Reviewer 1, this is difficult to assess thoroughly given 1) that we have only one month in two years of data, 2) that the preceding months (for which we do not have PFT observations) are likely to play a role, and 3) the many differences in model configurations. Moreover, a thorough assessment was not our intention – that is left for future studies, and we will make this clearer.

C6

“Technical Corrections”

We will address these. The question marks in Table 2 are not errors, but we accept their meaning is unclear and will modify the table accordingly. The comparison against IBTS chlorophyll was included in section 4.3 rather than 4.2 as we felt it fitted better with the rest of the validation against IBTS data.

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 will make this clearer in the manuscript. That said, we will expand our discussion of the differences in results, and justify more clearly our conclusions.

“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

C7

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 agree, and will note these points in the manuscript. In particular, we will discuss two mechanisms that may be at work here (but which can't be separated): 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 uniform across PFTs in NEMO-ERSEM). 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 will include 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 parametrizations 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,

C8

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 parametrizations.”

We will expand the discussion and conclusions accordingly to try and address these points, and also propose further work which would answer these questions more fully. See also our comments above.

“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 will revise the title accordingly.

“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 will expand the validation to include some assessment against in situ observations of temperature, salinity and nutrients.

“I find the figure and table captions to be too brief. I would prefer having more informa-

C9

tion 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 will expand 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 will address these points as far as is possible.

“L13-L16. “A comparison of the ...is key to capturing the observed biodiversity”. While I agree with this statement (after replacing “biodiversity” with “Phytoplankton 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 will both reword this sentence and expand the discussion section.

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

We will do this.

“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 will clarify these points.

“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.”

C10

We will address this.

“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. Do 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 will mention this in the manuscript, although 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 Fig. 11 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 will replot these figures and reword the text accordingly.

C11

“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 will include 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 will reword 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 will expand the discussion and justify further our reasoning.

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

C12

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

We feel these would be best kept where they are, but will improve 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 will investigate this possibility.

"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 will experiment with this, to see whether or not this gives clear results.

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

We will expand on this in the manuscript, but getting the details of the stratification, nutrient concentrations and therefore phytoplankton concentrations is dependent on having successfully simulated processes in previous months, more so than at other times of year, as well as the processes seen during August. Also, being the height of summer, there are no strong temporal gradients driving the response of the system, like in spring and autumn, so the internal biogeochemical dynamics can play out most freely (both in reality and in the model).

"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

C13

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 will discuss this further in the manuscript.

"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 will expand the discussion of this point.

"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 will try and tighten this section up accordingly.

"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 will change 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 will make these consistent.

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

C14

We will add this. The North Sea domain used for validation is the same as the GETM-ERSEM-BFM model domain, and we will make this clear in the manuscript.

“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 markers (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 will replot this figure to be easier to read and consistent with other figures. c) and d) are intended to compare the different in situ sampling methods of chlorophyll, as well as variability, and we will discuss them further.

“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 will replot this to be consistent with the other figures.

“Figure 3: The “closest to the reference data” point of a Taylor plot is on the x-axis at a value 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 will add this.

“Figure 5: This figure shows the satellite Chlorophyll and both models, but it may benefit

C15

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 decided against including the in situ data as we felt that would duplicate the information in Fig. 1, and also they do not coincide in time (the satellite data being monthly composites, and the IBTS data discrete data with each data point collected at a different time). Hence it would risk sending the wrong message about the validity of the satellite data. However, we can add it if the editor wishes. We will also modify the colour scale, as with Fig. 1.

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

We debated whether the scale should be linear or logarithmic, and settled on linear, but we will revisit that debate.

“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 will investigate these different options and see what seems to work best.

“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

C16

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 overlaid 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 will replot these to be short and wide, and look into including the three population fits either overlaid or as a separate subplot. 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 will look into including additional subplots using all the model grid points, and applying the three population model to these. 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. We will mention this in the text.

"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?"

Yes, we will do 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 will change this to be discrete, and also modify the colour scale, as has been raised by all reviewers.

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

C17

We will add 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 will modify the headers, and make the parameter names clearer as raised by Reviewer 2.

"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 will add these statistics.

"Technical Corrections"

We will address all these.

Detailed response to M. Butenschön

"However, with respect to the speculation on the reason of differences in community 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 community 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 will expand our discussion to further consider the roles of different drivers, and will take the work of Sinha et al. (2010) into account as part of this discussion. We will also make clear that further experiments are likely required to come to firm conclusions on this issue, and will outline potential future work.

C18

“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 community 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 will include further discussion about recent developments which have been made to PML-ERSEM, particularly within the SSB framework, and the results that have been obtained from this work. We will also outline 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.

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2016-304, 2016.