Response to the comments of reviewer 3

General comments:

The paper by Losa et al. (2019) describes marine ecosystem model development in order to better represent the marine phytoplankton community in the Southern Ocean. This is a worth while end eavour, and, if done correctly, could lead to major improvements in our understanding of marine ecosystems, and global biogeochemical cycling in the high latitudes. However, unfortunately, the current manuscript has multiple severe shortcomings, that - in my view - preclude a publication in Biogeosciences in its current form. In essence, I have strong reservations about (1) the lack of a scientific purpose of the paper presented, (2) the modelling work itself, which is not following standard protocols in the field with regard to model set-up, testing and quantitative validation, and (3) the interpretation of the results as a result of point (2).

Since reviewer 1 and 2 have done an excellent job in pointing out specific shortcomings of the current work already, I would like to highlight my concerns with regard to these three general issues

We thank the reviewer for the comments on the manuscript. Our replies are presented in blue, labeled "R:" and follow each reviewer comment. (For few additional figures we refer the reviewer to our responses to reviewer 2)

1. No scientific hypothesis pursued

As becomes clear from the abstract, and later throughout the entire manuscript, no concrete scientific question is pursued by the paper. Hence, this paper is not suitable for publication in Biogeosciences in its current state. Rather, the current development work should be published in journals such as "Geoscientific Model Development," or "Ecological Modelling" instead. For a successful submission to any of these journals, however, a proper documentation of the model, a documentation of its sensitivity to parameters chosen and assumptions made, and thorough model evaluation and validation would be required.

R: The purpose of this study is to understand some of the traits that leads to observed distribution of the key phytoplankton in the Southern Ocean: the silicifying diatoms, calcifying coccolithophores, and colony-forming *Phaeocystis*. We provide a synthesis of observations, and then use a model to explore what traits need to be incorporated to allow these observed distributions. In the manuscript we provide details on model development and on the evaluation of the simulations with in situ und satellite information. The use of as much as possible observational information to constrain the model allowed to define the required differences in the PFT traits, therefore, to test the hypothesis on the factors controlling the biogeography of key Southern Ocean phytoplankton groups.

In the manuscript we write:

"Observational information from in situ and satellite measurements was used to help to define differences in the PFT traits, to constrain the model, as well as to quantitatively evaluate the model performance to overall find a representation of the phytoplankton community in the Southern Ocean that is close to observations. We used the optimized Darwin model to test three hypotheses on the factors controlling the biogeography of Southern Ocean phytoplankton groups:"

In particular, the following three hypotheses have been tested which are also supported by literature:
- Size diversity of the diatoms (Queguiner, 2013; Tréguer et al., 2018) leads to the distribution of small diatoms ("slightly silicified and fast growing") at lower latitudes and large diatoms ("strongly silicified and slowly growing") at higher latitudes in the Southern Ocean.
– Distribution of coccolithophores in the Great Calcite Belt is not necessarily controlled by temperature (Smith et al., 2017) but determined by the ability of this PFT to escape grazing because of their exoskeleton (Nejstgaard et al., 1997; Huskin et al., 2000; Monteiro et al., 2016), and to grow under nutrient depleted conditions (especially phosphate and iron) (Paasche, 2001; Iglesias-Rodríguez et al., 2002). These characteristics of coccolithophores would make them more competitive among other phytoplankton of larger or similar size, small diatoms and *Phaeocystis* solitary cells.
– Phaeocystis sp. exists in two life stages, solitary cells and colonies, depending on iron availability (Bender et al., 2018). This additional difference in the traits of distinct haptophytes , coccolithophores and *Phaeocystis*, allows them to co-exist.

These hypotheses were formulated in Section 3.1 and 3.3. In the revised version, we state them in the introduction to make it clearer for the reader. We have also revised the manuscript following the comments of all three reviewers.

2. Model set-up characterized by substantial flaws that preclude a conclusive assessment of the main model dynamics and conclusions of this paper. It is evident in every section that a non-modeller with very little background on marine ecosystem model development, testing, and evaluation has conducted the current analysis. This becomes evident from an under-referenced introduction section, which is limited in scope and does not point out current gaps in marine ecosystem development, a poor, incomplete and flawed methods section that reveals major critical issues with the model set-up, spin-up, the initialisation and documentation of the runs conducted, and major critical issues with model stability and tuning.
R: The manuscript was revised accordantly. In the introduction we show that current biogeochemical models focusing on the SO phytoplankton community are either regional (e.g. Ross Sea, Crozet Islands), struggle to represent more different PFTs or are not well validated against available in situ/satellite datasets. The difficulties primarily originate from uncertain parameters employed in the parameterizations of e.g. phytoplankton growths and grazing that on the other hand define the differences in the phytoplankton traits.

We added the following paragraphs to the introduction (list of references can be found in the end of this document):

"Despite the recognized importance of the PFTs, global biogeochemical models struggle to represent the Southern Ocean phytoplankton community accurately. The difficulties primarily originate from uncertain parameters employed in the parametrizations of, e.g., phytoplankton growth and grazing (Anderson, 2005), that define the differences in the phytoplankton traits. On the other hand, the available observational information is still limited in the Southern Ocean to allow to properly constrain the models.

One of the most investigated regions in the Southern Ocean is the Ross Sea, where many in situ observations on diatoms and *Phaeocystis* have been collected and inspired regional coupled ocean-sea ice-ecosystem modeling activities (Arrigo et al., 2003; Worthen and Arrigo, 2003; Kaufman et al., 2017). Several studies that include *Phaeocystis* in the list of simulated PFTs in the frame of global coupled ocean-biogeochemical models have focused on the Southern Ocean (Lancelot et al., 2009; Wang and Moore, 2011; Le Quéré et al., 2016). These studies specified differences in (photo-)physiological parameters between diatoms and *Phaeocystis*, considering *Phaeocystis* in colony form. In a regional study (Popova et al. 2007, Crozet Islands) within the Southern Ocean, *Phaeocystis* was represented by two different life-stages: colonies and solitary cells. This approach was also successfully used by Kaufman et al. (2017) to examine the influence of climatic changes on the Ross Sea phytoplankton.

Nevertheless, an in-depth evaluation of the model simulations of diatoms and *Phaeocystis* with PFT observations either has not been done (e.g. Lancelot et al. 2009) or has been only performed based on a sparse in situ dataset (Wang and Moore, 2011). A more complete evaluation of these PFTs was presented by Le Quéré et al. (2016) by comparing the dominance of the PFTs to satellite-based dominance retrievals, and to a global dataset of in situ-based integrated PFT biomass within upper 200 m of Alvain et al. (2008) and (Buitenhuis et al., 2013), respectively. In general, as compared to the satellite retrievals, the dominance of diatoms and *Phaeocystis* has been overestimated by Le Quéré et al. (2016), while dominance of coccolithophores was underestimated.

Coccolithophore biogeography has recently been investigated globally by Monteiro et al. (2016), Krumhardt et al. (2017) and Krumhardt et al. (2019), and particularly for the Southern Ocean by Nissen et al. (2018). With respect to specific coccolithophore traits, the study by Krumhardt et al. (2017), Monteiro et al. (2016), as well as earlier studies by Paasche, (2001) and Iglesias-Rodríguez et al. (2002), emphasized the high nutrient affinity of the coccolithophores and high grazing protection of this PFT (Monteiro et al., 2016). Nissen et al. (2018) reported on higher grazing pressure on coccolithophores than on diatoms. Krumhardt et al. (2019) used lower grazing pressure on coccolithophores than on diatoms and related the distribution of coccolithophores to a specific temperature function in dependence to its growth rate. However, none of these studies included *Phaeocystis* in their model simulations."

We also added more details here and in the manuscript concerning the methods to show that our work is consistent. We provide detailed information about model spin up as well as detailed description of how the used MITgcm configuration was span up since 1979 (in original version we referred the reader to the study by Tailor et al., 2013, which is a correct reference as well as the reference to the study by Losch et al. 2010). To initialise (in 1992) the biogeochemical model variables, we used the results of the study by Taylor et al. (2013) that used a similar MITgcm configuration coupled with the Regulated Ecosystem Model (REcoM, Schartau et al. 2007) to examine the mechanisms behind the phytoplankton bloom in the Antarctic seasonal ice zone. Since their REcoM-MITgcm simulations were validated for the SO and the variables involved in the cycling of N, C, Fe, Si (including inorganic and organic particular and dissolved pools) and chlorophyll-a (decoupled from carbon) are presented in both Darwin-MITgcm and REcoM-MITgcm models, we use correspondent REcoM-based model states as initial conditions for these variables. The model variables describing the P cycle were initialised given N-based variables and the Redfield N:P ratio. The REcoM-based phytoplankton and zooplankton biomasses from Taylor et al. (2013) were distributed equally between six and two Darwin PFTs and zooplankton groups, respectively.

As in previous studies using the ecosystem model (e.g. Dutkiewicz et al 2015; Clayton et al 2017) the plankton establish a repeating seasonal cycle after about 3 years such that we can assume a "quasi-steady state" by year 2002. Surface nutrients are also in quasi-steady state. Longer term drift in deep nutrient distributions do not significantly change the results for the rest of the period that we consider here. It is not computationally possible to reach a totally adjusted system, and the ecological questions we address in this paper do not require such adjustments."

W added the following paragraphs to the manuscript:

"The biogeochemical model is coupled to a global configuration of the Massachusetts Institute of Technology general circulation model (MITgcm, 2012) on a cubed-sphere grid (Adcroft et al., 2004) with a mean horizontal grid spacing of 18 km and 50 vertical levels with the resolution ranging from 10 m near the surface to 450 m deep ocean (Menemenlis et. al., 2005, Losch et al. 2010). The simulation includes a dynamic sea-ice model with a viscous-plastic rheology and a zero-layer thermodynamic submodel (Losch et al., 2010). Penetrating light is attenuated within sea ice with an exponential law (Taylor et al. 2013, Appendix A2).

Initial conditions of the physical model were obtained from a short spin-up simulation initialised in January 1979 from rest and from temperature and salinity fields derived from the Polar Science Center Hydrographic Climatology (PHC) 3.0 (Steele et al., 2001). In the spin-up phase, the model forced until the end of 1991 by 6-hourly atmospheric surface fields derived from the European Centre for Medium-Range Weather Forecasts (ECMWF) 40 year re-analysis (ERA-40) (Uppala et al., 2005). More details can be found in Losch et al. 2010 (section 3). Starting on January 1st, 1992, the model with biogeochemistry is forced until 2012 by 3-hourly atmospheric surface fields of the Japanese 55-year reanalysis (JRA55, Kobayashi et al. 2015). Initially, the model time step had to be decreased to 10 min because of the higher forcing frequency, this constraint was slowly

relaxed to 20 min by January 1st 1996. The change in forcing also required an adjustment of some the sea-ice model parameters. The albedos for dry ice, wet ice, dry snow, and wet snow were set to 0.75, 0.71, 0.87, and 0.81, respectively; the simulation does not use the replacement pressure method (Kimmritz et al., 2017). After spinning up the biogeochemistry for six years, during which also the physical simulation adjusts to the new forcing, the years 1999 – 2012 are integrated and the period of Aug 2002 – Apr 2012 is used for analysis."

The paper is further characterised by a complete lack of a suitable model evaluation that would show that the model decisions taken (e.g. two diatoms, Phaeocystis parameterisation, etc.) are sound and robust (how do your nutrient patterns look like compared to observations, what is your NPP and export, what are your zonal and vertical standing stocks of carbon and chlorophyll-a for each of the biological tracers, what is the zonal and vertical biomass of your zooplankton functional types, how sensitive is your model to each of the parameters of your new tracers, are these values and dynamics realistic?), a profoundly disorganised documentation of model equations and parameters (no units in tables, wrongly declared equations, random selection of topics presented), followed by an erratic presentation of random model results, and finally, as a consequence of the above, a questionable interpretation of the model results that documents a severe lack of understanding of the model behaviour. In my view, the senior model developers that are co-authors on this paper would need to guide and advise the lead author(s) on what is standard practice in our field.

R: We disagree with the reviewer's assessment. We have evaluated the model against the best available data. We now also include evaluation against MAREDAT. However, this data is sparse in space and time, and especially in vertical. Thus, we would be unable to perform all the metric that the reviewer suggests. However, we performed for our model even more specific evaluation. Our interests are ecological, thus we consider specifically community structure. The data we use are:

The evaluation of the PFTs was performed with the best available information:

- satellite PFT-Chla (SynSenPFT, Losa et al. 2017) and phytoplankton dominance retrievals (PHYSAT, Alvain et al. 2008) in terms of spatial/temporal distribution;
- quantitatively using *in situ* HPLC-based PFT-Chla (large dataset from August 2002 – April 2012, Soppa et al. 2017), diatom vs. coccolithophores dominance based on cell counts sampled in the Great Calcite Belt (during January - February 2011 and February – March 2012, Smith et al. (2017). The timing of the satellite and in situ observations was considered when choosing the time period to be shown in the results. In the revised version we also include a statistical analysis of matchups between model PFT (carbon) biomass and MAREDAT observational data (as climatological monthly composites) for diatoms, *Phaeocystis* and coccolithophores. However, these data are rather sparse (not evenly distributed over the Southern Ocean).

As about nutrients, in the original manuscript we indeed show the nutrients distribution, and also report on (visual) agreement with the Garcia et al. (2014). In revised version of the manuscript we provide quantitative estimates of the nutrient assessment (see also our reply to reviewer's specific comments).

However, prior to provide any NPP or export information and any other standing stocks, one has to evaluate model performance with respect to observed PFTs, since these estimates mentioned by the reviewer as well as "…bulk properties such as total chlorophyll" depends on the PFT distribution itself (Anderson, 2005). Thus, new primary and export productions are beyond the scope of the current study.

The total chlorophyll-a concentration, nevertheless, agreed with OC-CCI total Chla product with a correlation coefficient r = 0.67 and mean absolute error MAE = 0.21 mgChla m^{-3}.

Zonally averaged zooplankton biomass showing the range from 0 - 20 (mgC m^{-3}) was presented in Figure 4 and Figure S8 (Supplementary Material). The agreement with MAREDAT data (Moriarty and O'Brien, 2013) shown in Dutkiewicz et al. (2015) is now explicitly introduced in the revised version of the manuscript and supplementary material.

In the method section we provided all the modification made with respect to original Darwin-2015 module (Dutkiewicz et al. 2015) (some of the listed modifications the reviewer proposed to remove). The units of the biogeochemical parameters were given in the text where they were introduced and in the revised manuscript they were also added in the tables. We also added few more parameters in Table 1 as suggested by the reviewer 2. Sensitivity experiments are provided in the Supplementary Material and we thank the reviewer for finding the typo in the name of the function "Holling III" used to parameterize grazing.

We now explain the logic underpinning the presentation of our results. Figures 9 – 11 depicting selected ("random") model snapshots against *in situ* observations were presented as an example of the modeled and in situ HPLC data agreement, with the link to video materials showing 2 weekly PFT spatial distributions and available *in situ* observations over the 10 year period (August 2002 – April 2012). The statistical analysis of the model and data matchups between PHAEO PFT-Chla and HPLC derived PFT-Chla were provided in Tables 3 to 5 (and Tables S7-S9 in the Supplementary Material). To avoid repeating information presented in the tables 3-5, we do not include the three figures corresponding to the model/*in situ* comparison.

An example that will illustrate why I have strong doubts about the methodology used is found in the description of the model set-up described in the Methods section. Here, the authors mention, that

- they use a spin-up period of 6 years for the physical model, which is far too short to equilibrate even the surface layer of the Southern Ocean,- then initialise biomasses and nutrient fields with results from a model coupled to another, completely different biogeochemical model (Recom-MITgcm) without carefully validating these fields,

R: We hope that with the details on the model spin up provided above and in the revised manuscript we prove that the model spin up was properly performed (w.r.t REcoM-MITgcm results of Taylor et al. (2013) validated for the Southern Ocean). The reviewer may be more used to low resolution biogeochemistry models that are, for instance used to study carbon cycling. Such

models are indeed spun up for long periods (often thousands of years) such that physics and biogeochemistry reach steady state. Such long spin ups are not possible with higher resolution models. Often initialization choices are important such the physics transitions smoothly. Similarly, when the questions are more of ecological (e.g. community structure) and not biogeochemical (air-sea fluxes of carbon), long spin ups needed to adjust the deep ocean are not required and likely not helpful since it is aimed to maintain an intermediate depth nutrient profile similar to the real ocean. Short time spin-ups (10 years) are what have been done in many of the recent Darwin model studies (Follows et al, 2007; Barton et al, 2010; Monteiro et al, 2016; Dutkiewicz et al. 2015; Clayton et al. 2017). Our tests running longer periods show that the ecological results are not significantly affected.

Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J., Follows M. J.: Patterns of diversity in marine phytoplankton, Science, 327 (5972), 1509-1511, doi: 10.1126/science.1184961, 2010

- and finally run the coupled model for 13 years only, without spinning up the biogeochemical module first (the authors point out that coccolithophores die out by the end of the simulation, which shows that the model isn't in a stable equilibrium yet for that target region),

R: The model experiment PHAEO (the main results we present in the manuscript) was in the quasi-steady state by the time of August 2002 – April 2012, which was used in our study (after physical model spin up since 1978 and biogeochemical module since 1992 over 10 years). Seasonal cycles of nutrients and PFT biomass are clearly present with some interannual variability but no significant drift. (Figure R3.3 and Figure R1.1) In addition, we only show results after this spin-up period.

However, for experiment REF as well as for other sensitivity experiments (overviewed in the Supplementary Material) there were no sufficient differences between the traits assumed for coccolithophores and "other large" (or Phaeo) phytoplankton. As a result, it took longer for the model to get in a quasi-steady state and finally just one of "similar" PFTs survived (taking over for another PFTs). Thus, in the experiment REF coccolithophores do not survive and Phaeo-analogue indeed represents haptophytes in general. In original Darwin-2015 model (Dutkiewicz et al. 2015) "other large" phytoplankton did not survive either. In this respect, we do acknowledge, that the REF simulations we showed for the year 2003/2004 still included slow drift towards dying coccolithophores (but nutrients and other PFTs did not drift).

We apologize as in the original version of this paper we presented some of the results in a confusing way. We realize that it was confusing to show REF output during this drift, though in some way we did want to make the point that the "generic other" needed to have significant different traits to enable both haptophytes to co-exist. This was the reason to add the bimorphic *Phaecystsis* traits. We wanted also to emphasize the problem of the instability of a complicated ecosystem model as a non-linear system in case of many variables and uncertain parameters.

To avoid any confusions, in the revised version of the manuscript we edited the sentence in L236 (original version) replacing "to the end of the considered period of time" by "after reaching a quasi-steady state" and only show results from both simulations for 2006-2012 when comparing PHAEO and REF.

- and present results from random months within the first and last few years of the simulation (across all figures we see patterns from: January, June, July, August, December 2003 and January, February, March 2004, February 2008, March 2012), where the biogeochemical model is not even remotely in equilibrium.

In the experiment PHAEO (the main experiment we present, evaluate and use for testing our hypothesis), the model is in quasi-steady state by the year 2002 and we no longer show results from 2003 and 2004 for the experiment REF as explained above. As to the "random" months: these were chosen to match with the HPLC observations. As this was not clear also for the other reviewers we now show results from climatological months (2006-2012) instead. The specific model snapshots (video supplement) and matchup statistics (Tables 3 to 6) are however still presented and movies show the interannual variability and the visual matchups from year to year.

Needless to say that common practice in the field is to carefully spin up the physics for multiple decades, then couple to the biogeochemical model, initialise that model from observations (available as gridded and extrapolated products in standard netcdfformat, so no excuses here), spin-up the coupled model for another ten years or sountil the biogeochemistry does not drift anymore, and then run the model with varying forcing and finally quantitatively (!!! -> Taylor diagrams, other model evaluation metrics, observational constraints, etc.) analyse the 5- or 10-year averages (or what-ever is appropriate to filter out inter-annual and multi-decadal variability in the target region) of the last few decades of the simulation, provided that the point of the study is to present average biogeographic patterns (well, if that was the point of the study, other approaches may hold for other scientific questions). And needless to point out that many of the required observations for the model are actually carefully processed by and available within the team of senior co-authors. I have hardly ever seen such a botch job.

R: As we explained above we did carefully spin up the physical model (accordingly to the application). As about the biogeochemical model initialization we can add the following: in the common practice, any initialization is worth to be done from a realistic and physically-consistent state of the system. In an ideal case, it could be a state estimation obtained via observational data assimilation into the model. Indeed, initialization just by pure observations can result in so-called "initialization shock". From the other side, normally biogeochemical ocean models forget fast their state in the upper ocean. Plankton biomass and detrital matter relatively quickly reach quasi-steady state, and the model results are relatively insensitive to the initial conditions. For phytoplankton, zooplankton and detritus the biogeochemical models could be initialized by standard profiles (Losa et al. 2006). For nutrients and proper carbon cycling simulations, it is more crucial to use nutrient initial states as close to observations as possible (but consistent with the used physics). With respect to nutrients state in the interior ocean, therefore, the utilization of the Taylor et al. (2013) validated run (which was initialized with the state accounting for

the World Ocean information) is a consistent way to initialize our model with a similar physical and identical grid configuration. The system is more inertial and has longer memory in the deep ocean.

In the revised version of the manuscript we provide additional information on the nutrient evaluation. For phosphate, that in our case was initialized given nitrate solution of Taylor et al. (2013) and Redfield N:P ratio, our climatology agrees well with WOA18 given the correlation coefficient 0.97 and normalized standard deviation 1.13 (one can consider these statistical criteria as inputs for Taylor diagram). We added the following information in the revised manuscript:

"In general, the simulated surface nutrient climatology agrees well with the World Ocean Atlas (Garcia et al., 2014) with correlation coefficient of 0.90 and 0.97 and normalised standard deviation of 1.27 and 1.13 for silicon and phosphate, respectively."

With respect to the observational data processed and produced in the Phytooptics group in Alfred-Wegener-Institute, for the Southern Ocean the satellite retrievals possess a number of limitations also discussed in the original version of this manuscript (see lines 392 - 403):

"In situ measurements in the Southern Ocean are sparse in space and time and only provide a fraction of the information obtained by the model. Satellite observations cover larger areas frequently but only cloud-free scenes which leads to a temporal bias in the often cloud-covered Southern Ocean. In addition, they are limited to only observe the first optical depth, which often limits the detection of the chlorophyll maximum. The development of algorithms for deriving PFT information requires a large in situ dataset with homogeneous temporal and spatial distribution. Nevertheless, scientific cruises in the Southern Ocean are often carried out close to the continents/ice shelf or in regions with high phytoplankton concentration (Figure 1). The diagnostic pigment analysis used to estimate PFTs from HPLC pigments assumes that different PFTs have different marker pigments, but it is known that they can also have pigments in common (Hirata et al. 2011). This ambiguity leads to uncertainties in the in situ database which is, on the one hand, needed as fundamental input for the algorithms of PFT retrievals and, on the other hand, used for direct comparison with model output here. Concerning spectral based methods applied to either in situ or satellite data, it is difficult to distinguish the specific absorption spectra of PFTs (e.g., coccolithophores and Phaeocystis). These and more limitations are well discussed by Sathyendranath (2014) and Bracher et al. (2017)."

One purpose of the study is to get the model set up that would allow (in future) to complement the observations either as additional information, or by simulating (if possible) the initial measurements and the retrieval algorithm itself (Dutkiewicz et al., 2017, 2018) and by assimilating directly the measurements.

Since the modifications of the biological module are documented even more poorly, there is no quantitative model evaluation, and neither is there a comprehensive documentation of carefully designed sensitivity tests that would allow us to understand the model sensitivity to the new parameterisation, it is impossible to critically evaluate the biological results of this work.

R: We provided all the modification made with respect to original Darwin-2015 (Dutkiewicz et al. 2015) module and equations in section 2.1.1. In the revised version we added few more parameters in Table 1 as suggested by reviewer 2. The sensitivity experiments have been added to the Supplementary Material.

Furthermore, the reported model instability with a high sensitivity of model results to parameter choice, as well as the disappearance of major functional groups (coccolithophores) throughout the simulation shows that this model configuration has major stability issues and a huge drift in the biological compartments with likely substantial consequences for productivity, nutrient distribution and biogeochemical cycling in this basin, and is thus unsuitable for publication at this point in time, as it clearly needs to be further tested.

R: As explained above, the model configuration used in the evaluated experiment (PHAEO) does not reveal any "huge drift" and we no longer show REF output prior to 2005.

3. Flawed analysis and interpretation of model results

Due to the above issues, any interpretation of the findings in this paper will obviously suffer from major uncertainties due to a lack of conclusive sensitivity tests and a thorough validation of the approach, and thus must remain entirely speculative. In its current form, it is impossible to say whether (1) the implementation of two diatoms is meaningful and leads to a gain in model performance (most models suffer from a nover-estimation of diatom biomass in this ocean basin), (2) the Phaeocystis module is correctly implemented, (3) the modifications of coccolithophore physiology are justified (since the author claims that they die out anyway), and (4) if, in fact, the model produces reasonable biogeography, primary production, depth patterns of biomass, and carbon standing stocks in this basin that would point at an actually improvement of the model compared to previous versions.

R: (1) We now straighten out the presentation and discussion of our results on consequences of including small diatoms, for instance by explicitly showing and discussing the diatom phenological indices in line with Chla distribution of small and large diatoms. The adjusted model reveals spatial distribution of small diatoms at lower latitudes and large diatoms at the higher latitudes of the Southern Ocean (as it was also shown in figure 3 of the original manuscript). "To capture the regional timing of diatom blooms obtained from satellite, it was required including both a lightly silicified diatom type and a larger and heavy silicified type in the model." Thus, our results support the hypothesis that introducing two size classes of diatoms in biogeochemical models is a prerequisite to simulate the observed diatom phenology and PFT distribution in general.

(2) What is the correct implementation of the "Phaeocystis module"? To consider *Phaeocystis* in two life stages was initially proposed by Popova et al. (2006), later used by Kaufman et al. (2017). We use similar approach, but with different implementation: these two *Phaeocystis* life stages are considered only as a function of iron availability as shown in the study by Bender et al. 2018; just one tracer is considered. We are motivated by this recent study by Bender et al. 2018 who reported

on the role of iron "as a trigger" for colony formation and now state in the introduction that we test the hypothesis that the transition in the Phaeocystis life cycle is determined by iron availability. Becquevort et al. (2007) and Hassler & Schoemann (2009) also showed that Fe addition had an effect on the morphotype dominance (colonies vs. solitary cells) of *Phaeocystis antarctica* with proportionally more solitary cells under low Fe conditions.)

In the manuscript we state:

"The implementation of two life stages of *Phaeocystis* to simulate both solitary and colonial forms of this PFT (with switching between forms being driven by iron availability) improved the co-existence of coccolithophores and *Phaeocystis* north of the Polar Front."

We also write that:

"In this study, we use very simplistic approach to parameterize life cycle transition of *Phaeocystis* given just one model tracer. In our model this transition is triggered only by iron variability (as reported by Bender et al. 2018), but not by light availability (as previously reported by Pererzak, 1993). Since we reported on our first trial, it is worth keeping in mind that the model is expected to be sensitive to the differences we specify for the mortality and grazing rates and iron uptake for colonial and single cell stage. A careful model calibration of these parameters could further improve the model performance."

Hassler, C. S. and Schoemann, V.: Bioavailability of organically bound Fe to model phytoplankton of the Southern Ocean, Biogeosciences, 6, 2281–2296, https://doi.org/10.5194/bg-6-2281-2009, 2009.

Becquevort, S., Lancelot, C., and Schoemann, V.: The role of iron in the bacterial degradation of organic matter derived from *Phaeocystis Antarctica*, Biogeochemistry, 83, 119–135, doi:10.1007/s10533-007-9079-1, 2007.

(3) Coccolithophores do exist in the final (PHAEO) model set up presented and evaluated. The modifications of coccolithophore physiology are justified by validation with an extensive synthesis of datasets (considered in this study), our previous experience and sensitivity tests. These modifications are consistent with and backed up by Nejstgaard et al. (1997), Huskin et al. (2000), Paasche (2001), Iglesias-Rodríguez et al. (2002), Losa et al. (2006), Monteiro et al. (2016), Krumhardt et al. (2017).

(4) Reasonable biogeography of the Southern Ocean PFTs has been validated with available in situ and satellite datasets. Since our interest is primarily ecological, precise estimates of carbon standing stocks are beyond this study.

In figure 2, model results are presented for the month of July, i.e. austral winter, where biomass in this basin is clearly very low, as it is dark. Yet, much ado is made about "dominance patterns" of specific PFTs during that time. However, the quantification of dominance on very low background biomass values is meaningless, as we're looking at percentages of zeros, essentially.

R: Chlorophyll concentrations north of the Subantarctic Front are not necessarily small in austral winter (see supplementary videos) and that is the reason for showing the results in Figure 2 for July. A similar figure panel was also presented in Dutkiewicz et al. (2015). Note that in the revised version of the manuscript Figure 2 shows the climatological values for all three datasets: PHYSAT, Dutkiewicz et al. (2015) and this study.

Furthermore, plankton biogeography in winter is very likely strongly linked to sea ice dynamics and factors not represented in current (global)models (resting spores, etc.) – whereas the authors even discuss spurious dominance patterns for areas clearly covered by ice, e.g. in the Dutkiewicz et al. (2015) set-up (which, apparently, isn't coupled to an ice model in its original configuration).

R: The authors referred to the patterns not covered by ice, we made that clearer in the text. Even though the original Darwin-2015 was not coupled to any ice model, it did have an ice mask. However, not the entire Southern Ocean is ice covered in the winter and it is informative to look at the ice-free water standing stocks at lower latitudes.

In addition, dominance patterns are compared for a random year of the simulation (2003 and 2004), where biogeography is still reported to contain major drifts (as coccolithophores are reported to die out). This makes the claimed "improvement" of the model quality interms of phytoplankton biogeography highly questionable.

R: Now we present the climatological monthly mean PFT dominance for the period between 2006 and 2012 (see Figure R2.6 and R2.7 in our responses to reviewer 2), which confirms that the dominance we showed for the year 2003/2004 in the former version of the manuscript were typical.

Furthermore, since the MIT-gcm set-up is likely global in scale (we do not know, as this is not described), and since the Dutkiewicz et al. (2015) set-up was likely tuned at the global scale, a comparison with a regionally tuned model is just simply unfitting – a global model will never do as well as a model tuned for a specific region, and it doesn't have to. We do not know, how the current set-up was tuned (I assume the MITgcm is still run in its global set-up, and the rest of the ocean is just not shown on the maps), and how it performs for the rest of the ocean.

R: The current coupled Darwin-MITgcm also runs globally. We agree though that the Darwin-MITgcm model configuration used in Dutkiewicz et al. (2015) was developed for global applications, and make this statement in the revised manuscript (lines 118-119). The problems we highlight and use as motivation for the present study are independent of the differences between Darwin-2015 and current configurations: 1) too early diatom bloom north of the Subantarctic Front, which results in diatom dominance north of the Subantarctic Front; 2) the survival only of one of two similar (w.r.t size class and traits) phytoplankton types ("other large" did not survive in Dutkiewicz et al.,2015).

As a marine ecosystem modeller, I am all in favour of seeing further modelling work published that would illuminate the drivers of phytoplankton biogeography, and the respective role of competition, predation and environmental niche dynamics

in shaping phytoplankton communities and associated ecosystem services. However, unfortunately, due to the poor quality of the submitted manuscript, my recommendation for this manuscript is forced to be: reject, revise and resubmit. Note to the author: The posted videos are in no way helpful for a quantitative (!) evaluation of the paper.

R: The videos were supported by three tables (Tables 3 – 5) in the main text of the manuscript and four tables (Tables S7 – S9) and three figures (Figures S9 – S11) in the supplement presenting the quantitative assessment of the model simulations against in situ observations over the period of time August 2002 – April 2012, which the reviewer seems to have misinterpreted.

Specific comments:

Abstract:

Lines 8-14: Revise. Vague and unsubstantiated claims. Be quantitative. Give numbers.

R: We did not feel that details on quantitative assessment belonged in the abstract. The abstract has been re-written to be tighter, and to focus on the novel traits added.

What scientific questions would you like to address with your model?

R: We added the overall aim of this study in the Abstract and the three hypotheses we investigate in the Introduction (L80-90).

We have rewritten the abstract as following:

"Phytoplankton in the Southern Ocean support important ecosystems and play a key role in the earth's carbon cycle, hence affecting climate. However, current global biogeochemical models struggle to reproduce the dynamics and co-existence of key phytoplankton functional types (PFTs) in this Ocean. Here we explore the traits important to allow three key PFTs (diatoms, coccolithophores and *Phaeocystis*) to have distributions, dominance and composition consistent with observations. In this study we use the Darwin biogeochemical/ecosystem model coupled to the Massachusetts Institute of Technology (MIT) general circulation model (Darwin-MITgcm). We evaluated our model against an extensive synthesis of observations, including in situ microscopy and high-performance liquid chromatography (HPLC), and satellite derived phytoplankton dominance, PFT chlorophyll-a (Chla), and phenology metrics. To capture the regional timing of diatom blooms obtained from satellite required including both a lightly silicified diatom type and a larger and heavy silicified type in the model. To obtain the anticipated distribution of coccolithophores, including the Great Calcite Belt, required accounting for a high affinity for nutrients and an ability to escape grazing control of this PFT. The implementation of two life stages of *Phaeocystis* to simulate both solitary and colonial forms of this PFT (with switching between forms being driven by iron availability) improved the co-existence of coccolithophores and *Phaeocystis* north of the Polar Front. The dual life-stages of *Phaeocystis* allowed it to

compete both with other phytoplankton of larger size and/or similar sizes. The evaluation of simulated PFTs showed significant agreement to a large set of matchups with in situ PFT Chl-a data derived from pigment concentrations. Satellite data provided important qualitative comparisons of PFT phenology and PFT dominance. With these newly added traits the model produced the observed >50% coccolithophore contribution to the biomass of biomineralizing PFTs in the Great Calcite Belt. The model together with the large synthesis of observations provides a clearer picture of the Southern Ocean phytoplankton community structure, and new appreciation of the traits that are likely important in setting this structure."

Intro:

General: Does not identify major challenges in the field addressed by this work. Does not introduce Southern Ocean community structure and function. Poorly referenced. Lack of original references.

R: The introduction has been revised and improved substantially considering all the comments from the reviewers. We have further considered in the introduction the challenges that the model community is facing despite recent advances and that this has motivated us to adjust a coupled model to predict more realistically the dominating phytoplankton groups in the Southern Ocean.

Fairly irrelevant discussion of the multiple meanings of PFT, and the criteria in Le Quéré et al. (2005).

R: We considered this issue important because of existing mismatch between grouping and "dimension of diversity" used in different observational techniques and models. But we agree with the reviewers that we indeed added to much information on how PFTs are defined. In the revised version of the manuscript we removed the discussion on the definition PFTs.

Lines 16-23: Poorly referenced. Give evidence for each of your claims.

R: The paragraph was revised and references added to it.

"The Southern Ocean is one of the most important regions in regulating climate via the uptake of about 40% of the global oceanic anthropogenic CO2 (DeVries, 2014) and at the same time, is a region with the dynamics evidently altered by past and present climate change (Stocker et al., 2013). The climatic changes in the Southern Ocean environmental conditions affect the spatial distribution of phytoplankton (Deppeler and Davidson, 2017). The phenology and dominance of different phytoplankton functional types (PFTs) sustaining the marine food web affect the diversity of higher trophic levels (Edwards and Richardson, 2004). Playing distinct roles in biogeochemical cycling, PFTs may determine how and on which spatial and temporal scales the ocean mediates climate (Wilson et al., 2018)."

Line 28-30: I disagree. Not the main planktonic calcifiers, if we trust modern estimates. The main PHYTOplanktonic calcifiers. See Buitenhuis et al. (2013) and Buitenhuis etal. (2019).

R: Corrected to “phytoplankton calcifiers”.

Line 32: References for Phaeocystis contribution to biogeochemical cycles missing.

R: References added (Arrigo et al., 2000; DiTullio et al., 2000; Wang and Moore, 2011).

Lines 32-36: Same thing. Original references for the importance of named groups missing.

R: In the revised version of the manuscript we removed the discussion on the definition PFTs.

Lines 49-50: "..different algorithms...use various approaches..." Be precise. What is relevant in this context. Focus on the issue with carbon to chlorophyll conversions. Your model calculates biomass in carbon units and derived chlorophyll-a biomass. You validate against chlorophyll-based algorithms (e.g. PHYSAT). What are the challenges? How can DARWIN help?

R: In the model carbon and Chla are decoupled in accordance to Geider (1998). Besides, PHYSAT is not a chlorophyll-based algorithm, but spectral radiance based method. Darwin helps due to considering different aspects/several dimensions of phytoplankton diversity (Dutkiewicz et al., BSD). Nevertheless, we removed this paragraph in the revised version of the manuscript.

Line 54-55: Do not use multiple names for the same thing. One term – one meaning.

R: In this respect, we have to admit this a general situation with the terms “phytoplankton functional types” and “phytoplankton group”. Nevertheless, we removed this paragraph in the revised version of the manuscript.

Line 56: No. Le Quéré and Follows were not the only researchers to initiate efforts in PFT modelling. They, as many others, contributed important material and thoughts to an ongoing discussion and effort. Have you read Hood et al. (2006)? Anderson et al. (2005)?

R: The introduction as revised accordantly.

Line 61: NOBM is not the only model with 4 PFTs. In fact many others do. BFM, GFDL, BEC (see e.g. Krumhardt et al. (2019)), etc.

R: The introduction as revised accordantly.

Line 63: No. DARWIN in its 2007 configuration does not contain all PFTs proposed by Le Quéré et al. (2005).

R: Agree, we referred here to Darwin (2010, 2015). Again, the introduction as revised accordantly and this sentence was removed.

Line 75 ff: What is the point of your paper? You state no scientific purpose. Model development should be published elsewhere.

R: Please, check our reply to your comment 2.

Methods:

General: Poor and disorganised. Documents strong lack of understanding of model structure and functioning.

R: For the model description and detailed equations and parameterizations the reader is referred to the study by Dutkiewicz et al (2015). In the current study, we introduce the modifications and changes carried out for our study (relative to their configuration) and list the parameterizations and parameters specifying the differences in the PFT traits. In the revised version we include few more parameters as suggested by reviewer 2 (the maximum quantum yield of carbon fixation; the spectrally averaged phytoplankton-specific light absorption and mortality rate). We also edited section 2.1.1 explaining how we introduce and parameterized the two different life stages for the *Phaeocystis* (L157-168).

Lines 85-876: Use consistent nomenclature to designate the PFTs included in the model. Your statement here lists other groups than those, e.g. represented in Figure 2.

R: Indeed, the grouping is done differently in the observations (PHYSAT, PhytoDOAS, HPLC-Chla), original Darwin-2015 model and current version. It is why we introduced Table 2 and, in the introduction, put lots of attention on the differences in the nomenclature used in modeling and satellite and in situ observation.

Lines 90-100: You must show all parameters and all equations for novel tracers. Need to document how diatoms differ, justify coccolithophore modifications, give equations for full Phaeocystis module. Justify each and every parameter, back up with literature values. You also need to show the results of your sensitivity analysis. Any deviation from the conventional PFT equation structure, i.e. the inclusion of life stages needs to be carefully motivated and evaluated. Get Phaeocystis data for the Southern Ocean. Evaluate fraction of biomass in colonial versus single-cell stage. Yes, there is data.

R: All parameters are shown in Table 1. We do not change equations *per se* except for the modification that are listed in 2.1.1, Equations 10 and 11. The detailed literature parameter values for photophysiological parameters are presented in section S1 (Supplementary Material). In the Supplementary Material we also provide an analysis of a number of parameters (traits) determined by specific sensitivity tests.

We did not perform any sensitivity analysis to the parameters mortality rate and grazing pressure for *Phaeocystis* solitary cell. But we expect that the model is sensitive to these parameters. We introduce a discussion on this issue in the revised version (L167-168):

"We have not performed any sensitivity experiments with respect to the new parameters. However, we expect the model to be sensitive to their specification since it will also determine the competition between *Phaeocystis* and small diatoms."

We also state this in the new subsection "Perspectives and limitations of the study" and additionally write (L532-533): "A careful model calibration of these parameters could further improve the model performance."

The inclusion of *Phaeocystis* life stages from our side was motivated by necessity to introduce additional differences in the haptophytes traits. While in general it was introduced by previous studies by Popova et al. (2006), Kaufman et al. (2016). It would be interesting later on to evaluate the simulated fractions of *Phaeocystis* in colonial *vs.* single-cells with available observational data. However, at this point we did not trace the particular stage of modeled *Phaeocystis* and leave this topic for future publications.

Line 100: Why did you choose to replace other large phytoplankton with Phaeocystis, and not nitrogen fixers? Nitrogen fixers have been shown to only play a very minor role in the Southern Ocean due to their temperature limitation. On the contrary, other large phytoplankton species, such as dinoflagellates, are regularly observed in this ocean basin. This decision seems questionable.

R: We were motivated by the following reasons:

1. "Other large" did not survive in the original Darwin 2015 version.
2. Distributions of N-fixer as well as prokaryotic picophytoplankton determine the extent and abundance of small phytoplankton and coccolithophores north of the Subantarcic and Suptropical Fronts. Additionally, we wanted to maintain a reasonably good performance of the model globally. Thus, we kept N-fixer.
3. We cannot strictly state that the *Phaeocystis*-analogue considered is pure *Phaeocystis* sp., it could be also other misrepresented nano-PFTs.

We clarify in the text (L.114-120):

" …Thus, in the modified Darwin version the following six PFTs are considered: large and small diatoms, *Phaeocystis* and coccolithophores, *Proclorococcus*-like and N-fixers. Although later two PFTs only play a very minor role in the Southern Ocean, their distributions determine the extent and abundance of small phytoplankton and coccolithophores north of the Subantarcic and Suptropical Fronts. Hence, we keep N-fixer and *Proclorococcus*-like prokarytes in our version (it would also allow to maintain a reasonably good performance of the model globally). *Phaeocystis* are considered as adjusted (with respect to the traits) "other large", since "other large" did not survive in the original (Dutkiewicz et al., 2015) version that was developed for the global ocean. However, we cannot strictly state that the *Phaeocystis*-analogue considered is pure *Phaeocystis sp.*, it could be also other misrepresented nano-PFTs."

Line 105ff: The light module is completely irrelevant in this context. Move to appendix or omit.

R: Moved to the Supplementary Material.

Line 109; Same holds for kCDOM. You don't ever discuss this. Omit.

R: This light module, as well as correspondent kCDOM, was introduced as differences from the original Dutkiewicz et al. (2015) Darwin configuration.

Table 1: No units given. No references included.

R: The units were provided in the text in the revised version of the manuscript and also in the table.

What is mfunc? It does not appear in equations (3) – (6).

R: "mfunc" was/is determined in line 115. We explain it now in the table footnote "as a biomineralizing function".

Use consistent abbreviations and names for all PFTs throughout entire paper.

R: Sometimes it was not possible it is why Table 2 was introduced.

Line 118: I am 100% positive it was not Geider et al. (1998) who invented the growth functions. This parameterisation is quite similar to what Riley developed in 1946 (!). Please check your referencing.

R: With the reference to Geider et al. (1998) we refer to the set of the equations (3) – (6) accounting for decoupling between carbon and Chla.

Lines 119 – 123: Not all parameters defined, all functions named must be given as equations. Incomplete set of equations. For example, what is f(k_sat)?

R: For the complete set of the equations the reader is referred to the study by Dutkiewicz et al. (2015). The equations (3) - (11) were provided to emphasize where the main differences between considered PFT traits were specified.

Formulation for Phaeocystis is clearly not as given in (3) – (6).

R: Indeed, it is as presented in equations (3) – (6) plus equations (10) and (11).

Line: 130: This is not a Holling II function.

R: It is Holling III function

Also, you state that DARWIN has two zooplankton types on line 84. Where are the zooplankton traits? Need to report.

R: The palatability parameters r_{ij} also determine the zooplankton traits, for rest the reader is referred to Dutkiewicz et al. (2015).

In general, the role of zooplankton grazing pressure on plankton biogeography is not discussed in this manuscript. Since zooplankton usually play a vital role in determining the relative biomass fractions and dominance patterns in these models, the role of top-down control must be addressed else in the manuscript.

R: This crucial issue was discussed in part 3.3 (L256-262, in the revised version in L 366-376), where the reviewer "stopped detailed review".

"The discussed distribution of coccolithophores have been obtained under the assumption of lower palatability function (leading to lower grazing pressure) in comparison with what is assumed for other PFTs. This contradicts the study by Nissen et al. (2018), who reported on an increased (relative to diatoms) grazing of coccolithophores as a factor controlling the coccolithophore biogeography in the Southern Ocean. Our assumptions on low palatability factor of coccolithophores are, nevertheless, backed up by studies by Nejstgaard et al. (1997), Huskin et al. (2000), Losa et al. (2006) and Monteiro et al. (2016). In the study by Losa et al. (2006) on optimised biogeochemical parameters, it was shown that coccolithophore blooms are associated with low grazing pressure. Based on laboratory experiments, Nejstgaard et al. (1997) and Huskin et al. (2000) concluded that coccolithophores (due to its "stony" structure) do not influence the microzooplankton growth. While the exact mechanisms of how this PFT uses the coccolith to protect itself against grazing is not fully understood (Monteiro et al., 2016), the ability of coccolithophores to escape grazing control has "relatively well-supported evidence" (see Monteiro et al. 2016 for review)."

And later (L274-277, in the revised version in L409-412):

"…in the region between the Subtropical and Subantarctic Fronts the occurrence of coccolithophores is more evidently linked to low grazing pressure on this PFT due to its much lower palatability for zooplankton in comparison with small diatoms or *Phaeocystis* presented by single solitary cells."

Line 140ffff: Where does this parameterisation come from. Why did you not choose to follow e.g. Schoemann et al. (2005), the most comprehensive review on Phaeocystis dynamics.

R: Agree that the study by Schoemann et al. (2005) nicely reviews the *Phaeocystis* distribution, also suggesting a specific temperature limitation function for *Phaeocystis*.

We, nevertheless, decided to follow Popova et al. (2007) and Kaufman et al. (2017) who proposed to consider *Phaeocystis* in two different life stages and recent study by Bender et al. (2018).

What temperature function did you choose, and why?

As in Dutkiewicz et al. (2015), the original study we are based on:

$$\gamma_j^T = \tau_T e^{\left(A_T\left(\frac{1}{T+273.15}-\frac{1}{T_O}\right)\right)},$$

given the coefficient $\tau_T = 0.8$ normalized the maximum value (unitless), the temperature coefficient $A_T = -4000K$, and the optimal temperature $T_o = 293.15K$.

We have added it the manuscript.

Do you use one or two tracers for Phaeocystis (Phaeo versus Phaeo_cell)? Are there two combined?

R: We use one tracer.

Lines 144 – 145 (original version): "Note, that in the model *Phaeocystis* is still the same variable/array, but the assumed morphology and, therefore, physiology is different for the different life stages of *Phaeocystis*".

Do you get realistic fractions of biomass in colonial stage?

R: We did not trace it explicitly for the outputs, since the priority was to get *Phaeocystis* and coccolithophores co-existing. However, we agree that this would be interesting and we will do this in future studies.

Line 144-145: Rewrite. Utterly unclear.

R: We revised as following: "Note, that in the model *Phaeocystis*, independently on the life stage – colonial phase or solitary cells – is considered as one tracer. However, the assumed morphology and, therefore, physiology (mortality rate, grazing pressure, k_{satFe} ,sinking rate) differ as described above" (L165-168).

2.1.2. Physics: 6 years is not a spin-up. What are you spinning up?

R: Please see our response to general comments above. Section 2.1.2 has been revised accordantly

Are you using a global model? Documentation of model set-up is insufficient – do not just say it was "similar" to something else.

R: We are using global model. Please see our response to general comments above. Section 2.1.2 has been revised accordantly.

And, by the way, it was not similar to the model used in Taylor et al. (2013), since these authors used a completely different coupled GCM-biogeochemical model.

R: The study by Taylor et al. (2013) used the same MITgcm sea ice – ocean model configuration (except for external forcing and some of the sea ice model parameters). The similarity in the configuration also includes the initialization (1992) of physical state based on the model spin up from the year 1979. Please see our response to general comments above.

What is the temporal resolution of your simulation?

R: The time step of model integration is 20 min (L181-183):

"Initially, the model time step had to be decreased to 10 min because of the higher forcing frequency. This constraint was slowly relaxed to 20 min by January 1st, 1996."

The model output is stored as two-weekly model snapshots. We clarify this in the manuscript (L227-228): "Two-weekly PHAEO model snapshots from August 2002 to April 2012"

Why do I care about light penetration, when you haven't even described the ice module?
Which ice-model do you use? Is it dynamic?
R: We have to care about the light penetration since it is one of the factors controlling the phytoplankton dynamics (particularly, in the marginal ice zone). In the revised version we explicitly provide information about the dynamical sea-ice model used. Please see our response to general comments above.

2.1.3 Biogeochemical tracer initialization

Botch job. Really. Initialise from observations, not from unvalidated model runs con-ducted with a completely different model. These observations are generated in your group.
R: Initialization with the satellite observation: the satellite data provide only surface information, interpolation to the model grid will introduce representation errors even for the surface layer since the satellite values are provided for the first optical depth which varies in accordance to the variation of the optical constituents. Satellite Chla is available only for haptophytes (problem with distinguishing between Phaeo and coccolithophores), diatoms, and prokaryotes.

Spin up the biogeochemistry. Evaluate (quantitatively) your NPP, export, nutrient fields, total chlorophyll-a before you even start thinking about a reference run. Spinup your biogeochemistry module.
R: The biogeochemistry had been spun up over the period 1992 – 1998 and, further, 1998 – 2002. Evaluation of nutrients fields had been performed, the nutrients were validated and the results are presented (above and in the revised version). As about evaluation of NPP, export production, we would state opposite: before estimating NPP, export production or any other standing stock, one has to validate PFTs (Anderson, 2005).
(Nevertheless, the model climatological TChla agreed with OC-CCI TChla given correlation coefficient 0.67, mean absolute error 0.21 mgChla/m^3, rmse = 0.4 mgChla/m^3 (0.33 in log10 scales) and normalized standard deviation 0.69)

Initialisation of biomasses from Recom is absolute nonsense. Below you describe your satellite validation data. Use that as an initialisation.

R: Please see our responses to the two comments above and also to the general comments. Taylor et al. (2013, which showed its evaluation) provided Chla 3D distribution consistent with the physical model dynamics and grid. Moreover, the model "forgets" quite fast the initial Chla and biomass conditions (distributed mostly in upper ocean).

Do not point to Taylor et al. (2013), as this leads me to Losch et al. (2010), and they use a different model. Use new World Ocean Atlas nutrients, for instance.

R: In the study by Losch et al. (2010) the same physical model configuration was used to provide initial conditions for 1992 after spinning up the model from 1979).

Equilibrate your model. If major Southern Ocean players show a strong in their biomass (e.g. your coccolithophores die out, as you write), then there is a major problem in your model.

R: After biogeochemical spin up via model integration from 1992 to 1998 and further, the model reaches for the experiment REF steady state in 2004 – 2005 and in 2002 for the experiment PHAEO. Please see our response to the general comments above.

2.2.1 In situ observation

Cryptic and poorly written. List all data sets used.

R: We have edited and extended the description of section 2.2.1 (see below)

Which units does the data have, which spatio-temporal resolution, are these surface data, or are they depth-resolved. Pointing at another study is insufficient.

R: The details are included in the revised version as written below.

How do you treat, bin, grid, quality control this data to be useful for model validation?

Do you convert any of this to biomass, in order to evaluate biomasses?

R: The surface HPLC measurements used for PFT Chla retrievals do not require any conversion to be directly compared with model PFT Chla (PFT Chla is also a prognostic model variable due to Chla decoupling following Geider et al., 1998).

What about the vertical pattern?

R: The in situ dataset are taken at the surface, within the first 12 m.

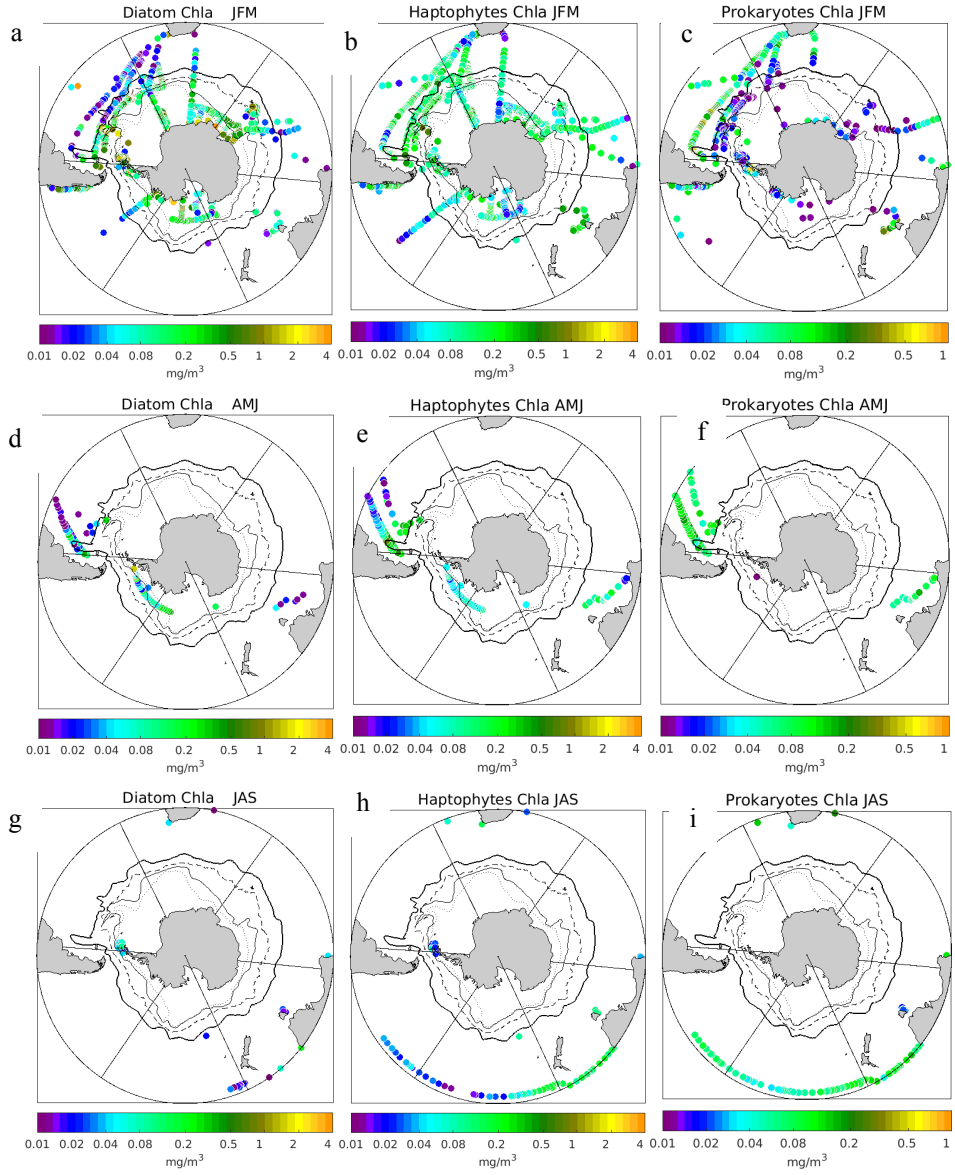
What about the comprehensiveness of this data – how many months, years, seasons, latitude bands, depth levels does your data cover (all this needs to be described in the main text).

R: We added information into the section and edited this the text as following:

"A quantitative assessment of the model has been carried out using in situ observation from a large global and quality controlled dataset of in situ chlorophyll-a concentrations (mg m^{-3}) of diatoms, haptophytes and prokaryotes derived from high precision liquid chromatography (HPLC) phytoplankton pigments (Soppa et al. 2017, https://doi.pangaea.de/10.1594/PANGAEA.875879). The dataset is composed **surface (first 12 m)** measurements collected by different expeditions in the Southern Ocean (south of 30°S, see Figure 1) over the time period August 2002 – April 2012, sampled mostly during austral spring and summer months (see supplemental video materials). The phytoplankton groups were derived using the Diagnostic Pigment Analysis (DPA) following Vidussi et al. (2001) and Uitz et al. (2006) and modified as in Hirata et al. (2011) and Brewin et al. (2015) and adapted to a much larger data set. Briefly, different PFTs have different and specific pigments (marker pigments, e.g. fucoxanthin – diatoms) that allow distinguishing the PFTs. The biomass of a specific PFT can be quantified by determining the contribution of the corresponding diagnostic pigment to total phytoplankton biomass (represented by the weighted sum of the diagnostic pigments). It is worth mentioning that DPA allows also to retrieve other phytoplankton groups – like dinoflagellates, cryptophytes and green algae – however, they were not included in the referred data set originally generated for the evaluation of satellite retrievals of diatoms, coccolithophores (haptophytes) and prokaryotes. For more details on the method and data quality control of this in-situ data set, we refer the reader to the study by Losa et al. (2017, Supplementary Material, Sections 1 and 3).

Figure 1 shows the locations of the available *in situ* data in the Southern Ocean. As we can see there and in Table 2" (table 2 is a newly introduced table), "this large dataset gives us the possibility for a quantitative validation of our model results. Two weekly PHAEO model snapshots from August 2002 to April 2012 have been collocated against in situ HPLC-based Chla observations, if available, within a time window +-1 week. We compare the simulated Chla of diatoms (large + small), haptophytes (coccolithophores + Phaeocystis) and prokaryotic pico-phytoplankton against HPLC-derived Chla for diatoms, haptophytes and prokaryotes. The matchup statistics is presented for several biogeochemical provinces (Longhurst, 1998) distributed over the Southern Ocean (Figure 1): Austral Polar Province (APLR), Antarctic Province (ANTA), Subantarctic Water Ring Province (SANT), South Subtropical Convergence Province (SSTC), Humbold Current Coastal Province (CHIL), Southwest Atlantic Shelves Province (FKLD), Eastern Africa Coastal Province (EAFR), Australia-Indonesia Coastal Province (AUSW), East Australian Coastal Province (AUSE). In the Supplementary Material (Figure S12) we also present the distribution of the HPLC-derived Chla dataset (Soppa et al., 2017) as seasonal climatological PFT composites."

Figure R3.1 (Figure S12 in the Supplementary Material) depicts distribution of the in situ HPLC based Chla (HPLC-Chla) for diatoms, haptophytes and prokaryotes (Soppa et al., 2017) available over the period of time August 2002 – April 2012 and composed for different seasons.


a
Diatom Chla JFM
b
Haptophytes Chla JFM
c
Prokaryotes Chla JFM
0.01 0.02 0.04 0.08 0.2 0.5 1 2 4
mg/m3
d
Diatom Chla AMJ
e
Haptophytes Chla AMJ
f
Prokaryotes Chla AMJ
g
Diatom Chla JAS
h
Haptophytes Chla JAS
i
Prokaryotes Chla JAS

775

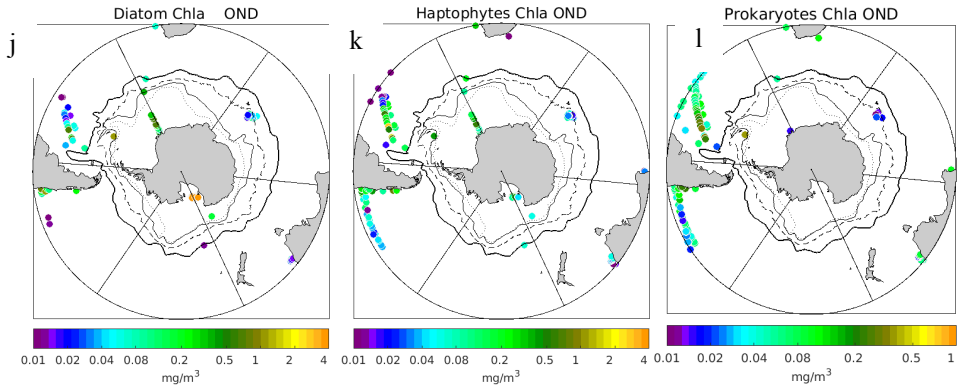


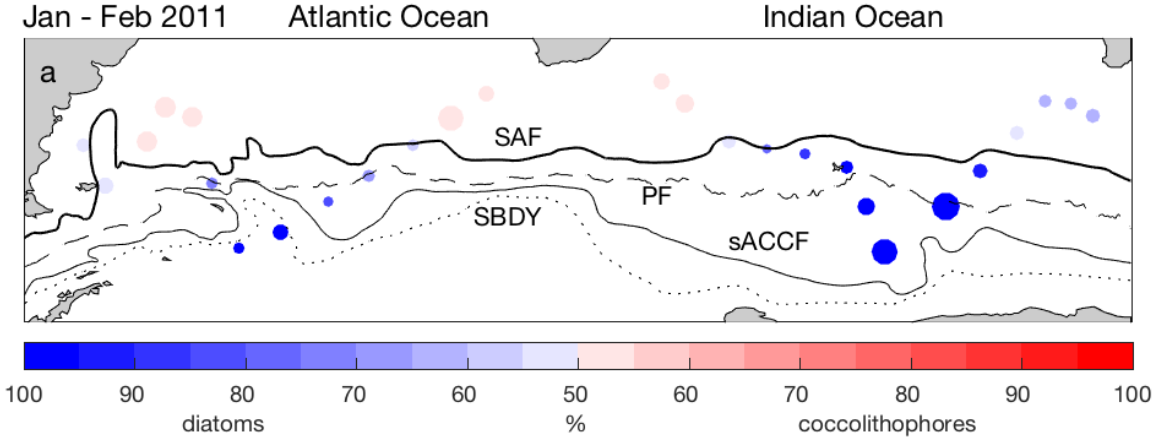
Figure R3.1. Distribution of seasonal composites of HLPC-Chla (Soppa et al.2017) for diatoms, hyptophytes and prokaryotes. Black counters represent Southern Ocean fronts (as white contours in Figure 1 of the manuscript).

780

What do the "measurements" by Smith et al. 2017 comprise.

R: We write "Predicted PFTs were additionally compared to diatom and coccolithophores measurements (as cell counts) reported by Smith et al. (2017). These data were obtained by scanning electron microscopy in the North Atlantic and Indian Ocean sections of the Southern Ocean during the time period of January – February 2011 and February – March 2012."

785 Additionally, for qualitative assessment of the agreement of the simulated diatom and coccolithophore distributions to these data, we provide now estimates of the diatom *vs.* coccolithophores dominance (Figure R3.2) to compare to the similar presentation of estimates by Smith et al. (2017).



790

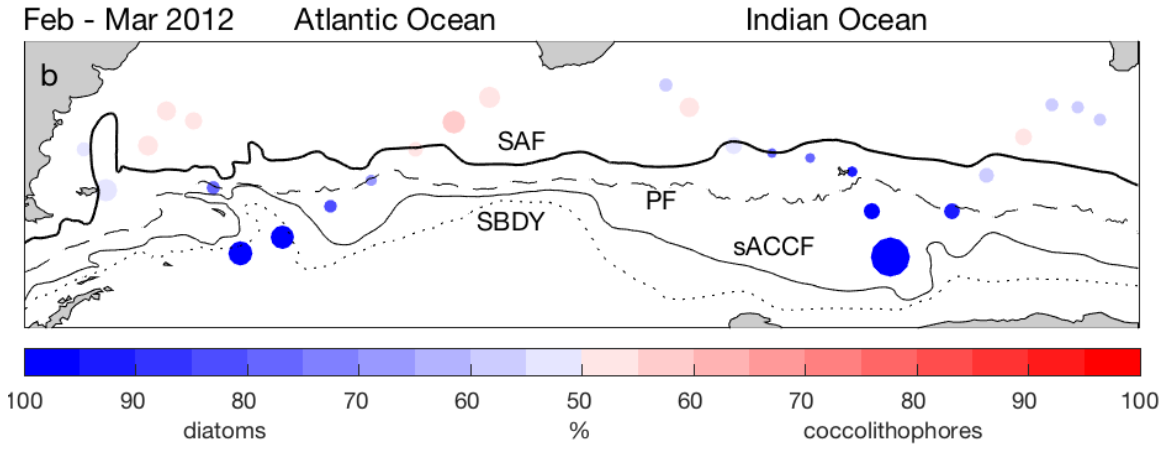


Figure R3.2: PHAEO Diatom *vs.* coccolithophores dominance averaged over January-February 2011 (a) February - March 2012 (b). The size of the circles is relative to phytoplankton carbon content (mmolC/m^3). The largest size of the circle corresponds to 3.12 (mmolC/ m^3). (Figure 5 in the revised version of the manuscript)

Line 173: Table 2, in fact, does not contain any useful information.

R: Table 2 contains the information about the evaluated phytoplankton groups as classified in the model and from observations, which based on different phytoplankton grouping.

Figure 1: What kind of observations do you show? Cite the appropriate reference.

R: Figure 1 depicts the locations of HPLC measurements collected over the period of time August 2002 – April 2012 (Soppa et al. 2017). The figure caption has been extended accordantly in the revised version of the manuscript.

Why do you show the Longhurst provinces? These are of no relevance in the rest of the paper. They only confuse the reader. Else aggregate and evaluate your model based on these provinces throughout the entire manuscript.

R: The Longhurst provinces are depicted in this figure, since quantitative assessment of the model simulated PFT Chla against in situ HPLC based estimates were performed for each of these provinces (Tables 3 – 5 and Tables S7 – S9).

2.2.2 Remote sensing

Why would you use the old 2008 PHYSAT product? There is an updated algorithm using a neural network approach. This should be far better than this outdated version.

R: We use the PHYSAT version that is freely available. We tried to contact Severine Alvain several times but we received no response and therefore could not obtain the neural network results.

What type of "abundance" are you referring to, and how can this be compared to the model output?

R: Abundance-based approach is the term used by the ocean colour community to classify algorithms derived to estimate either PFTs or phytoplankton size classes based on observed relationships between some measure of abundance of phytoplankton and their type or size structure (IOCCG 2014). The one we mention in the paper is the OC-PFT of Hirata et al. (2011) and it gives the information of the Chla in mg/m3 of each PFT. Note that we did not compare the OC-PFT product directly to our simulations but we used the SynSenPFT product which combines the information of OC-PFT and PhytoDOAS (spectral-based algorithm).

This part of the manuscript was revised to make it clearer for the reader:

"Model results are compared to phytoplankton dominating groups from the climatological monthly mean satellite derived product PHYSAT (1998-2006, Alvain et al., 2008). PHYSAT is based on the analysis of normalized water-leaving radiance anomalies, computed after removing the impact of chlorophyll-a variations. Specific water-leaving radiance spectra anomalies (in terms of spectral shapes and amplitudes) have been empirically associated to the presence of dominant phytoplankton groups, based on in situ diagnostic pigment observations. This product is based on the multispectral Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) information and available in http://log.cnrs.fr/Physat-2?lang=fr.

We also evaluated the model simulations (mg m^{-3}) against the satellite PFT Chla (mg m^{-3}) product SynSenPFT (Losa et al. 2017, https://doi.org/10.1594/PANGAEA.875873). The SynSenPFT product combines the information of two satellite PFT Chla products: one retrieved with the differential optical absorption spectroscopy method (PhytoDOAS, Bracher et al. 2009; Sadeghi et al. 2012) applied to hyperspectral information from the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY, Bracher et al. 2017; https://doi.org/10.1594/ PANGAEA.870486) and the OC-PFT abundance-based approach (Hirata et al. 2011 and refined in Losa et al. 2017) applied to multi-spectral satellite total Chla data from the Ocean Colour Climate Change Initiative (OC-CCI). While the PhytoDOAS products from the SCIAMACHY sensor are only available at 0.5° spatial resolution and monthly means, OC-PFT applied to OC-CCI Chla products can be obtained daily and at 4 km resolution.

PhytoDOAS and PHYSAT satellite products are derived based on phytoplankton absorption properties captured by the satellite sensors and distinguished by the retrieval algorithms either as a particular PFT optical imprint ("finger print") in case of available hyperspectral information (in PhytoDOAS) or as anomalies in a multispectral signal (in PHYSAT). Thus, the PhytoDOAS allows to retrieve quantitatively major PFTs (coccolithophores, diatoms, cyanobacteria), while PHYSAT provides information about five dominant phytoplankton groups: prokaryotes (presented by *Prochloroccocus* and *Synechococcus*-like SCL), diatoms, haptophytes in general and *Phaeocystis* in particular.

We compare model climatology of Southern Ocean PFT dominance (averaged over the years 2006 – 2012) to the PHYSAT PFT dominance (dominance of the modeled PFT is defined if its Chla fraction is more than 55% of the total Chla). In line with the evaluation against the PHYSAT PFT dominance, the simulated PFT dominance are compared to similar estimates obtained in the study by Dutkiewicz et al. (2015). Two SynSenPFT products (at 4 km and daily) – diatoms Chla that combines diatoms Chla from PhytoDOAS and OC-PFT, and coccolithophores Chla that combines coccolithophores Chla from PhytoDOAS with haptophytes Chla from OC-PFT – are used in addition to the in situ based diatom vs. coccolithophores dominance by Smith et al. (2017). Hence, we only use the same areas and time period as in their study for comparisons to the SynSenPFT results. Here as well the comparison is qualitative as the SynSenPFT products are mostly based on OC-PFT in our study region and the global relationships between Chla and the fraction of PFTs from the OC-PFT algorithm might differ in the Southern Ocean, as shown by Soppa et al. (2014) for diatoms."

You mention this data is high resolution – did you regrid the data to fit the model grid?

R: If the reviewer refers to SynSenPFT product, the data were gridded only for the visualization since, so far, being used for qualitative assessment because of product limitations as discussed in lines 392 – 403.

Results & Discussion

General: This section is characterised by a complete chaos in terms of information presented, spatio-temporal scales shown. None of the figures are in any way useful to evaluate the quality of the model development you presented in your methods section.

R: Below we provide our motivation/explain why our results are presented in the way that the reviewer criticizes and what has been changed in the revised version.

All you show is colourful surface ocean maps, and most of these maps are useless. Redo entire section from scratch. Does not contain any discussion, as there is no quantitative evaluation and interpretation of the work.

R: The colourful figures (as well as video supplementary) were presented for qualitative assessment or to support the discussion on the hypothesis. Quantitative evaluation was also presented in several tables and figures. We now extend the discussion on quantitative assessment previously provided (Tables 3-5, S7-S9) and additionally performed. As a summary on the model evaluation performed is presented below.

The evaluation of the coupled model skill with respect to predicted PFT Chla was performed given *in situ* HPLC-based Chla retrievals for diatoms, haptophytes and prokaryotes (2166, 2388 and 1425 matchups, respectively) over the time period of August 2002 – April 2012 (Soppa et al. 2017). Quantitative assessment of the agreement between model and data were/are provided for several biogeochemical provinces (Tables 3 – 5). Three additional tables and three figures depicting probability

density of the predicted and observed PFT-Chla and their differences were provided in the Supplementary Material (Tables S7-S9, Figures S9 - S11).

Qualitative assessment for simulated PFTs was possible for:

1) simulated Southern Ocean PFT dominance when comparing to the PHYSAT PFT dominance climatological data product (Alvain et al. 2018);

2) diatoms vs. coccolithophores dominance in the Great Calcite Belt (compared with in situ cell counts by Smith et al. 2017) for January – February 2011 and February – March 2012;

3) satellite data based PhytoDOAS coccolithophore fit (Losa et al. 2018) and SynSenPFT Chla results for diatoms and haptophytes(coccolithophores) over the same area and time period as shown in Smith et al. (2017);

4) Southern Ocean Diatom phenological indexes (we additionally present here for a typical year 2007/2008) compared to Soppa et al. (2016).

"We chose this particular year because: 1) with the two-weekly model output the phenological indices can be more precisely calculated than based on the two-weekly or monthly mean climatology; 2) it is a typical year over the period 2006 – 2012 with respect to the simulated PFT distribution (after model reached the quasi-steady state) and climate oscillations (Soppa et al., 2016)."

The spatial distributions of the PHAEO simulated March 2004 nutrients were shown in Figure 5 (compared, agreed visually, with World Ocean Atlas (WOA) climatological March, Garcia et al. 2014). These nicely spatially resolved nutrients distributions along with PFTs distributions were shown to support the discussion on the factor controlling the Southern Ocean Biogeography. For this purpose, we did want to see fine structures of these distributions, thus we do not performed larger temporal scale (over several years) averaging. We now provide quantitative assessment of the model – data nutrients agreement in terms of correlation and normalized standard deviation based on model climatological means (averaged over the years 2006 – 2012) and WOA data. Temporal evolution of the simulated surface phosphate, silica and iron averaged over several biogeochemical provinces are presented in figure R3.5.

Simulated zooplankton biomass was shown in Figure 4 and Figure S8 to be compared with MAREDAT (Moriarty and O'Brien, 2013) data also depicted in Dutkiewicz et al. (2015, their Figure 12, panel i). We now explicitly add quantitative assessment of the agreement between the model and MAREDAT observations with statistical analysis of model zooplankton climatology matchups with climatological monthly composites of the MAREDAT micro- and mesozooplankton (Figures R2.2, responses to reviewer 2; in the Supplementary Material, Figure S14).

PIC:POC we only discuss in a context of a qualitative agreement with previously reported studies (Balch et al., 2005, 2016) and only in the context of importance of distinguishing among haptophytes (Figures R3.7).

Line 205: You do not discuss phenology in this section.

R: The term "phenology" is used here as PFT Chla dynamics in general, but the reviewer is right and we removed the term phenology and now call this section as "Diversity within diatoms".

Lines 207: The satellite estimate you are comparing your results against is not "the truth", merely another algorithm. Be sure to correct.

R: We have never stated that the satellite estimate is "the truth" but an additional information with possible limitations and uncertainties (see lines 392 – 403 of the original version).

Why, on Earth, would you compare winter values for the Southern Ocean? It's dark in winter, and background biomass is likely very challenging to model, as it will depend on features not included in the current generation of ecosystem models, such as resting spores, overwintering strategies, ice associations, etc. Hence, your entire dominance analysis is severely flawed, see detailed comment above. Why don't you evaluate e.g. a 5-year seasonal average over December – February period (most commonly used)?

R: The months shown were chosen in consistency to the PFT dominance presented in original Dutkiewicz et al. (2015). In the revised version we show dominance plots in climatological context (given these climatological maps, it is seen, however, that the year 2003/2004 shown in the original version is representative). In the supplementary material (Figures S15-S18), we now show climatological monthly mean of the PFT dominance as retrieved PHYSAT (1998-2006) and obtained in Darwin-2015 (Dutkiewicz et al. 2015), REF (2006-2012) and PHAEO (2006-2012). Please see these figures included in our responses to reviewer 2 (R2.4 – R2.7). In the manuscript we show monthly climatology for December, January and February (Figure R3.3). However, we would like to show PFT dominance for July since it appeared a kind of diagnostic month indicating whether the model has deficiencies in reproducing observed phytoplankton dynamics. We agree that current biogeochemical models still underrepresent some processes or phenomena. One of these phenomena is the size diversity within diatoms. After evaluation of our adjusted model against the extensive synthesis of observations (including *in situ* SEM and HPLC, and satellite derived phytoplankton dominance, PFT-Chla, and phenology metrics) we confirm that including both a lightly silicified diatom type and a larger and heavy silicified type is required to capture the regional timing of observed diatom blooms and, therefore, seasonality of the observed phytoplankton composition.

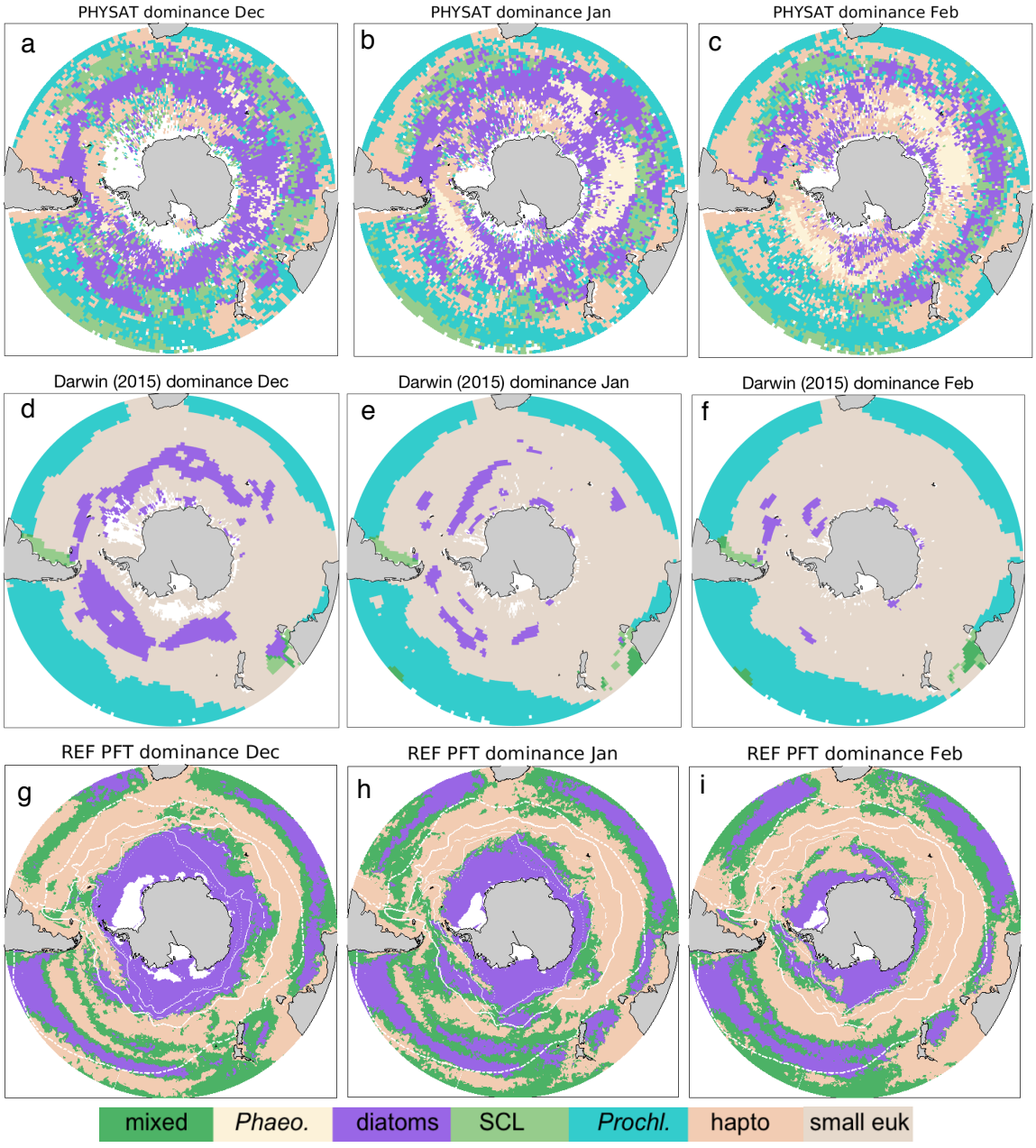


Figure R3.3. Climatology of surface PFT dominance retrieved by PHYSAT algorithm (1998-2006, upper), simulated with the Darwin-MITgcm version of Dutkiewicz et al. (2015) (middle) and the current model set up REF (bottom) (2006 - 2012). "SCL" represents Synechococcus-like prokaryotic phytoplankton (not considered in the current model version). Simulated haptophytes include coccolithophores and *Phaeocystis*. Model PFT is considered dominant if its Chla fraction of total Chla is more than 55%. The model output (REF) is masked by the area with sea ice concentration > 75% during respective month. Darwin-15 is masked by PHYSAT missing values. White contours denote the Southern Ocean fronts (Orsi et al., 1995; Orsi and Harris, 2001) as in Figure 1

Lines 219: This comparison appears flawed. As far as I know, Dutkiewicz et al. (2015) was tuned for a global fit.

R: The current version also runs globally, moreover, the problem discussed is independent from that whether the model is tuned globally or regionally (see our response above).

Figure 2: Names in legend do not match those given for your PFTs elsewhere in the manuscript. Do not show results for random months and years.

R: The names in legend are given as in Alvain et al. (2008). The names "diatom" and "hapto" match our diatom and haptophytes. We extend the figure capture to better clarify the correspondence of legend names to Darwin-2015 and the current Darwin version (please also see Table 2).

"Climatology of surface PFT dominance retrieved by PHYSAT algorithm (Alvain et al. 2008) (left), simulated with the Darwin-MITgcm version of Dutkiewicz et al. (2015) (middle) and current model set up REF (right) for 2006 – 2012. "SCL" represents *Synechococcus*-like prokaryotic phytoplankton (retrieved by PHYSAT, but not considered in current model version); "*Procl"* denotes *Prochlorococcus*, "small euk" is other small eukaryotes. Simulated haptophytes (hapto) include coccolithophores and *Phaeocystis*. Model PFT is considered dominant if its Chla fraction of total Chla is more than 55%."

Lines 213-2014: This must be included in the main text, not the supplementary. Along with a thorough discussion of the parameter choices in the methods section.

R: We opted to have the detailed protocol of the experiments as supplement since it would make the manuscript extremely long. Moreover, today it is an established practice to use supplementary (including videos) to provide additional information along with the details the study is focusing on in the main manuscript. It is common practice that the supplementary materials are read as well.

Lines 220 – 223: This, if true, would deserve a paper on its own. Unfortunately, you do neither quantify nor show phonological patterns. All we see in figure 2 is dominance plots for two selected months in random years.

R: We indeed have performed the analysis and we planned a separate paper focusing on the analysis of phenological indices. As mentioned above, we now compare dominance monthly mean climatology instead of PFT dominance for specific years; and show diatom phenological indices estimated for the year 2007/2008.

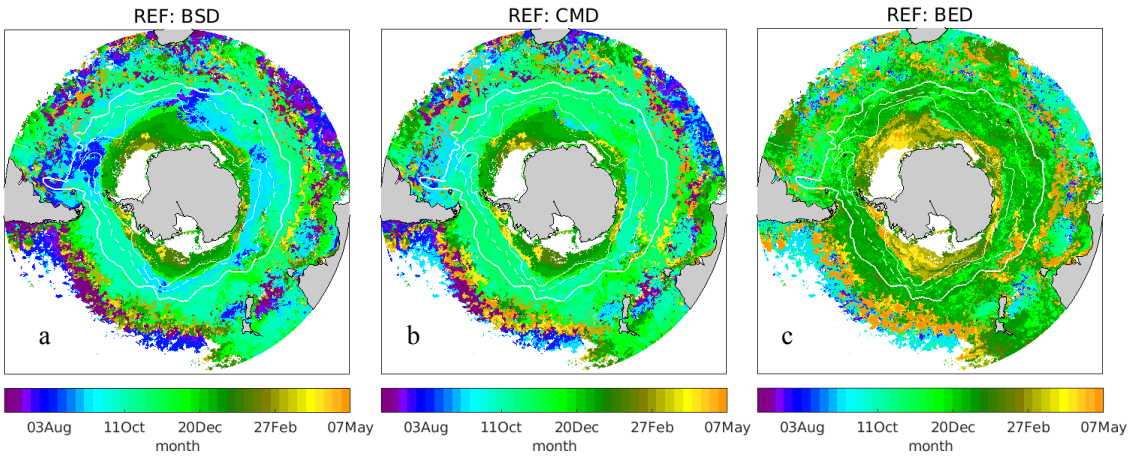


Figure R3.4: REF diatom Chla phenological indices: bloom start date (BSD, a), chlorophyll maximum date (CMD, b), bloom end date (BED, c) (Figure 3(a,b,c) **in ther revised version of the manuscript).**

Section 2 has been extended by additional Subsection 2.3 "Diatom phenological indices" with the following text:
"Following Soppa et al. (2016) we evaluate the diatom phenology by calculating phenological indices based on a threshold method proposed and initially applied for assessing the TChla phenology by Siegel et al. (2002). In particular, we use the following indices: the Chla maximum date, the bloom start date, and the bloom end date. These indices are calculated based on the REF Chl simulations for diatoms (including small and large) over the year 2007/2008…"

Line 225: Claimed "augmentation" hasn't been shown. What is augmented? NPP?Export? Nutrient fields? Silicification and calcification rates? Opal export? Relative biomass fractions? None of this has been shown so far.
R: Augmentation is mentioned here with respect to model set up. The model was augmented/extended by considering two size classes of diatoms and two life stages of *Phaeocystis*.

Line 228: "in agreement with" How is anything shown so far in agreement with anything. No quantitative evaluation has been provided.
R: The discussed is the spatial distribution and dominance of co-existing small diatom and coccolithophores with results reported by Signorini et al. (2006), Balch et al. (2016) and Smith et al. (2017). Qualitative comparison is still an assessment. We, nevertheless, remove this sentence.

Line 230: "..results are supported.." How, and in which way? Quantitative comparison with data is missing (model evaluation).
R: Direct quantitative comparison with the PhytoDOAS PFT retrieved fit factor (Losa et al. 2018) is not possible simply due to different units. However, qualitative assessment of the coccolithophores distribution in the Great Calcite Belt, with respect

to its co-occurrence with diatoms north of the Subantarctic Front and to the position of the northern edge of the Great Calcite Belt, is still possible.

Line 233-236: "However,...." This shows that your model is not stable, not in any kind of sensible equilibrium, with likely large consequences for all tracers associated with global biogeochemical cycles, and thus not publishable yet. Absolute game stopper.

R: We carefully checked the model solution in different regions with respect to reaching a quasi-steady state (see figure R3.5 below). The silica is still drifting slightly, but does not affect the results. The issue we stressed here is about the biochemical

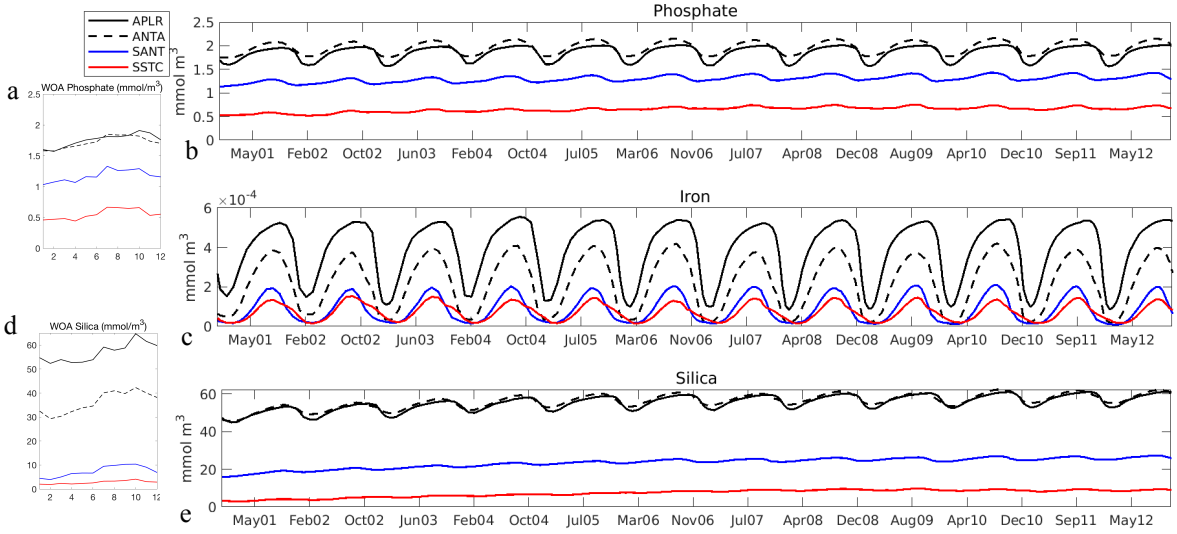


Figure R3.5: Temporal evolution of monthly mean nutrient concentrations averaged over the APLR, ANTA, SANT and SSTC biogeochemical provinces (Longhurst, 1998).

And by the way, in fig 8 you show us coccolithophore biomass for March 2012, which is "towards the end of your simulation", and thus it looks like coccolithophores are abundantly populating the low latitude Southern Ocean, with a substantially overestimated chlorophyll-a biomass relative to the SynSenPFT estimate.

R: Figure 8 depicts results of experiment PHAEO (the experiment where both Phaeocystis and coccolithophores co-exist).

3.3. "To cope with the aforementioned chaoticity of the system...."

I stop my detailed review of this section here. All the remainder is speculation. If the model couldn't be tuned to reliably reproduce the biogeography of major players, then this paper shouldn't have been written, but the model should have been developed further.

R: Please see our response to the "first general issue".

Figure 3: Why do you show another month? Evaluate model results on same spatio-temporal scales throughout entire manuscript.

R: We now introduce a new table (Table 2). This table presents the datasets used to constrain the model, including information on spatial and temporal representation of these data and corresponding model output (please Table R2 in our responses to reviewer 2).

The purpose of showing Figure 3 was to, in particular, illustrate the distribution of small *vs.* large diatoms depicting also fine spatial scales (which would be vanished if presented as climatological mean). The January was chosen as a month showing more prominently (compare to the months) diatoms dominance in the Southern Ocean.

In the revised version we replace this figure with a figure depicting diatom Chla phenological indices (for the year 2007/2008, Figure R3.4) in line with the distribution of large *vs.* small diatoms for few months of this typical year (to explore this distribution within different bloom periods, Figure 3d-f: large diatom, Figure 3g-i: small diatom).

Compare total model chlorophyll-a estimate to total satellite chlorophyll-a. Compare group-specific chlorophyll-a to your different SynSenPFT etc. algorithms. Quantitatively.

R: The quantitative evaluation with SynSenPFT was not performed because of the limitations of the SynSenPFT product for the Southern Ocean (as discussed in lines 298 – 304). The model climatological TChla agreed with OC-CCI TChla given correlation coefficient 0.67, mean absolute error 0.21 mgChla/m^3, rmse = 0.4 mgChla/m^3 (0.33 in log10 scales) and normalized standard deviation 0.69. However our prior goal is the PFT evaluation (Anderson, 2005).

As about comparison to SynSenPFT, we clarify above and in the revised version of the manuscript:

"Two SynSenPFT products (at 4 km and daily) – diatoms Chla that combines diatoms Chla from PhytoDOAS and OC-PFT, and coccolithophores Chla that combines coccolithophores Chla from PhytoDOAS with haptophytes Chla from OC-PFT – are used in addition to the in situ based diatom *vs.* coccolithophores dominance by Smith et al. (2017). Hence, we only use the same areas and time period as in their study for comparisons to the SynSenPFT results. Here as well **the comparison is qualitative** as the SynSenPFT products are mostly based on OC-PFT in our study region and the global relationships between Chla and the fraction of PFTs from the OC-PFT algorithm might differ in the Southern Ocean, as shown by Soppa et al. (2014) for diatoms."

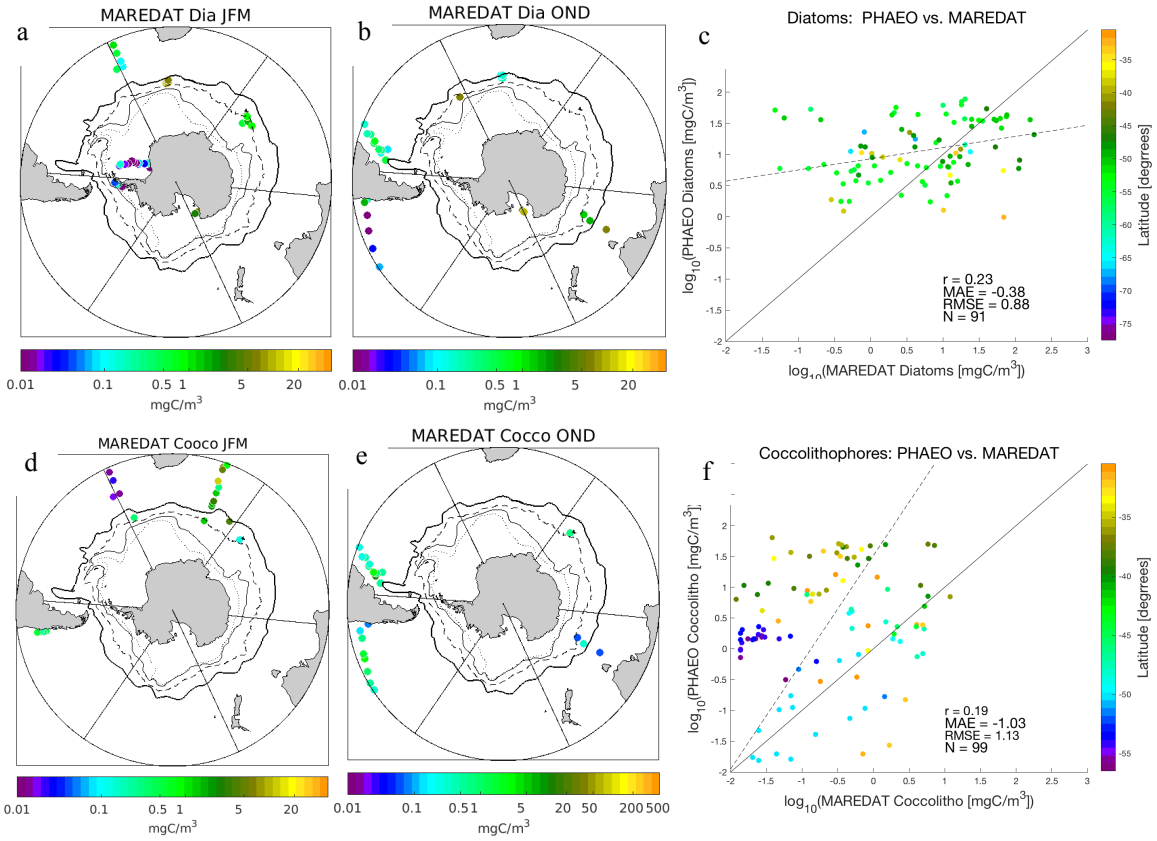
Figure 4: Label your axes.

R: We label the axes of the revised version of the figure.

What is the total biomass level in each of these plots?

R: We do not provide the biomass level yet (since no total carbon biomass data available), instead we now show a more precise evaluation of predicted carbon biomass of diatom, coccolithophores and *Phaeosystis* against available MAREDAT datasets.

"In addition, simulations are also compared to the global MAREDAT *in situ* datasets of diatoms (Leblanc et al. 2012, https://doi.org/10.1594/PANGAEA.777384), coccolithophores (O'Brien et al. 2013, https://doi.org/10.1594/PANGAEA.785092), Phaeocystis spp. (Vogt et al. 2013, https://doi.org/10.1594/PANGAEA.779101) …These datasets are based on a data collection spanning between 55 to 75 years and are provided as climatological monthly composites."



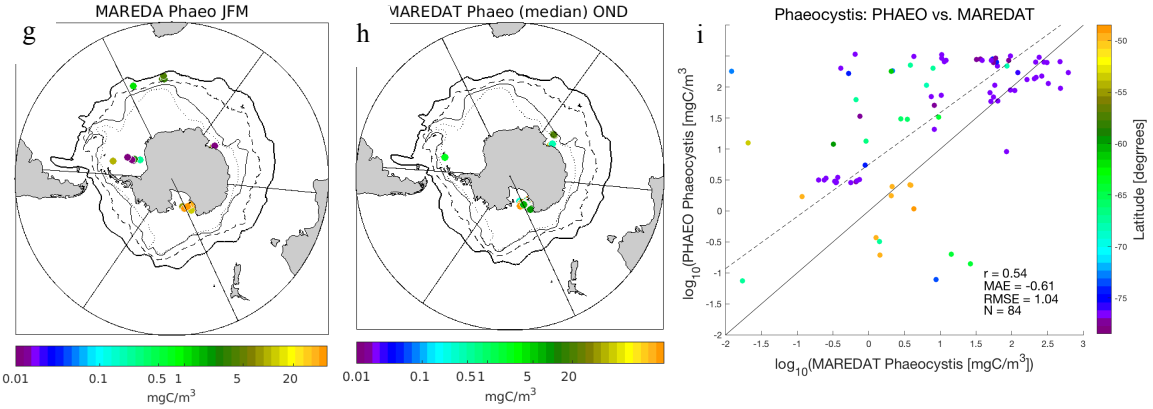


Figure R3.6: Climatological sesonal composites of the MAREDAT surface phytoplankton biomass for diatoms (a: for January – March, b: for October - December), coccolithophores (d: for January – March, e: for October - December) and Phaeocystis (g: for January – March, h: for October - December); scaterplot of the model vs. MAREDAT matchups based on all surface climatological monthly means: c) for diatoms; f) for coccolithophores; i) for Phaeocystis. (PHAEO model climatology is based on the years 2006 – 2012). Statistics are presented for logtransformed concentrations. Black counters represent Southern Ocean fronts (as white contours in Figure 1 of the manuscript).

However, the coverage of the Southern Ocean PFT biomass is, indeed, very limited (see figure R3.6, same as S13).

"Figure S13 shows the distribution of MAREDAT seasonal (summer and spring) composites of diatom (panels a and b), coccolithophores (panels d and e) and *Phaeocystis* (panels g and h) biomass data vs. PHAEO monthly climatology matchups to MAREDAT monthly climatology for diatoms (panel c), coccolithophores (f) and *Phaeocystis* (i). Because of the poor data coverage and large discrepancies in representation temporal scales, differences between the model and data (due to the representation error) are expected to be large. As a result, correlation between model and data PFT biomass is weak but significant (0.23, 0.19 and 0.54 for diatoms, coccolithophores and *Phaeocystis*, respectively). In general, the model overestimates PFT-carbon biomass in comparison with the data. At the end, showing the quantitative estimates of the data and model agreements, we still give a qualitative assessment. Moreover, MAREDAT measurements are not always collocated for different PFTs, thus, it is not always possible to draw any conclusions on the phytoplankton compositions. However, one, can notice, that diatoms, coccolithophores and *Phaeocysts* do co-exist in the areas along the subantarctic and polar fronts."

There seem to be far too many nitrogen fixers at 40 South.

R: We agree with the reviewer, however unfortunately, for this area the observational information is rather sparse.

Same for coccolithophore contribution in this area. Are you under- or overestimating total biomass here?

R: We cannot *per se* conclude whether we over- or underestimate total carbon biomass (since there is no data available), but biomass of diatom, *Phaeocystis* and coccolithophores. In respect to MAREDAT coccolithophores data base (however, over

the Southern Ocean there are not many observations, see Fig. R3.6 d,e), our climatological mean biomass (mgC/m^3) are higher (Fig. R.3.6 f), especially in the Atlantic Ocean Section north of the South Subtropical Convergence Province (SSTC), where the seemingly simulated Great Calcite Belt is shifted northward. So we do have overestimation for coccolithophores there. However, it is worth mentioning that the MAREDAT data error because of the conversion from cell counts to carbon biomass is several 100%.

Nevertheless, as we write in the revised version of the manuscript (L369-376):

"For a more precise evaluation of the PHAEO results with the study by Smith et al. (2017), we show diatom *vs.* coccolithophores dominance collocated in space and time with observations of Smith et al. (2017) (Figure 5). Even though our estimates have been obtained based on phytoplankton biomass (mmol C m^{-3}), but not on cell counts as in Smith et al. (2017), our results agree well to their higher concentrations and dominance of diatoms in the SBDY and SACCF, while north of the Polar Front coccolithophores become more abundant (better seen in Fig. 9). As compared with Smith et al. 2017 (their figure 2), in the Atlantic section, the dominance of simulated coccolithophores (55%) is shifted northward of the Subantarctic Front leading to underestimation of the coccolothophore dominance along the polar front and south of SAF and overestimation north of SAF."

Why have you chosen to present a zonal view of the community structure?

R: As seen, the distribution of simulated and observed PFTs reveal prominent zonal features (one can also notice zonally determined Longhurst provinces over the Southern Ocean, defined Southern Ocean Fronts, zonal gradients in hydrography). The authors find, that is a nice representative way of showing latitudinal changes in the PFT composition.

In figure 1 you show Longhurst biomes. Can you evaluate the depth pattern?

R: Figure 1 shows *in situ* surface HPLC-Chla (Soppa et al., 2017) distributed over the Longhurst's biogeochemical provinces. For these provinces the quantitative assessment of the agreement between model and HPLC-based Chla are presented. Is there any reason to evaluate the depth pattern/topography in this context or context of the manuscript? If the reviewer refers to an evaluation of PFT-Chla vertical distribution, we have to admit that it is not currently possible since the dataset (Soppa et al., 2017) contains surface information.

What is the link between zooplankton biomass dynamics and the relative fractional contribution of individual groups to total biomass? Furthermore, Phaeocystis is not usually known to dominate biomass between 60-50S.

R: In experiment PHAEO (right panels of Figure 4) *Phaeocystis* is not the dominating biomass between 60-50S.

Figure 5: Why, again, do you show another month? Compare total chlorophyll-a to observations. Compare the groups to observational estimates from space, or in situ data.

R: Figure 5 (6 in the revised version) depicts PFT and nutrients distribution for a typical summer month February and typical year 2008 to back up the discussion on the co-existence and distribution of the Southern Ocean key phytoplankton groups and drivers of the PFT biogeography. The particular month (but not climatological one) was chosen also to show the resolved spatial scales of the discussed distribution of limiting nutrients and PFTs.

As we clarify it in the text (L394-416):

"Figure 6 depicts the Chla spatial distribution for diatoms (a), *Phaeocystis* (b) and coccolithophores (c) for February 2008 from PHAEO. We present this particular summer month of a typical year to clearly show the patterns of the depicted distribution, which could not be very obviously seen on seasonal or climatological mean maps. One can notice co-existence of simulated PHAEO diatoms and *Phaeocystis* south of the Polar Front and the co-occurrence of diatoms and coccolithophores in the Subantarctic Zone north of the Subantarctic Front. This agrees to (Smith et al., 2017) and is supported by the PhytoDOAS PFT retrievals from SCIAMACHY hyper-spectral information within the same time frame and region in Losa et al. (2018) and Smith et al. (2017).

Figure 6 presents the spatial distribution of silicon (d), dissolved iron (f) and phosphate (g) in February 2008 from PHAEO. … The spatial distribution of silicon, dissolved iron and phosphate is discussed in line with the simulated PFT Chla biogeography. The regions with high iron concentrations (in the Ross Sea, along the Western Antarctic Peninsula, around the Falkland, South Georgia and South Sandwich, Crozet and Kerquen Islands) indicate the area of Phaeocystis potential existence in colonial form. Thus Figure 6 shows that the simulated abundance of coccolithophores north of the Subtropical Front (STF) – where phosphate occurs in very low concentrations – is explained by the introduced high affinity of this PFT to phosphate (small half-saturation rate in $\gamma_{\eta\_}$ function) allowing coccolithophores to grow in nutrient depleted conditions. However, in the region between the Subtropical and Subantarctic Fronts the occurrence of coccolithophores is more evidently linked to low grazing pressure on this PFT due to its much lower palatibility for zooplankton in comparison with small diatoms or Phaeocystis presented by single solitary cells. As in the study by Smith et al. (2017) reported biogeography of observed coccolithophores in the Great Calcite Belt, our simulated coccolithophore Chla is distributed in the silica-depleted area, where small diatom cells, even if they could still compete for other nutrients, have higher palatability for grazers. Coccolithophores do not compete with small diatoms on silica resources and might survive due to its lower palatability factor. It could also be that in this area silica limited diatoms slowly grow allowing coccolithophores for earlier access to other (not used yet by diatoms) macronutrients and iron."

Showing total Chla in this respect would not help even though it agrees with OC-CCI TChla (as already mentioned above) with a correlation coefficient r = 0.67 and mean absolute error MAE = 0.21 mgChla m^{-3}. The comparison of model PFT Chla with in situ HPLC-Chla (Soppa et al., 2017) over the time period of August 2002 – April 2012 are shown in three supplementary video materials and tables 3-5, S7-S9.

Do not just show model nutrients. Compare quantitatively with observational estimates, e.g. from World Ocean Atlas.

R: Figure 5 (6 in the revised version) is used to support the discussion on the drivers of nano-phytoplankton distribution in the Great Calcite Belt. A quantitative assessment of climatological mean surface nutrients against the World Ocean Atlas is presented below (as correlation coefficient r and normalized standard deviation). We add this information into the text of the manuscript:

"In general, the simulated surface nutrient climatology agrees well with the World Ocean Atlas (Garcia et al., 2014) with correlation coefficient of 0.90, 0.92 and 0.97 and normalised standard deviation of 1.27, 0.67 and 1.13 for silicon, nitrate and phosphate, respectively."

Figure R3.5 (panels a, b and d, e) depicts WOA and modelled seasonal evolution of phosphate and silica averaged over the APLR, ANTA, SANT and SSTC biogeochemical provinces (Longhurst, 1998).

Figure 6: Why would you only show modeled PIC? Compare to observational PIC from space.

R: Figure 6 (7 in the revised version) was presented to emphasize the importance of distinguishing among haptophytes before estimating any carbon standing stocks (as it was suggested by the reviewer). Comparing with PIC from space model would show underestimation of PIC south of the Polar Front in the line also with the qualitative comparison with PhytoDOAS. We think that the quantitative estimates of the differences will not add anything to the conclusion. However, the results are also in line with the study by Holligan et al. (2010) concluded that current satellite algorithms may significantly overestimate PIC in cold waters of the Southern Ocean.

Holligan, P.M. Charalampopoulou, A., Hutson, R.: Seasonal distributions of the coccolithophore, *Emiliania huxleyi*, and of particulate inorganic carbon in surface waters of the Scotia Sea, *Journal of Marine Systems*, 82 (4), 195 – 205, doi: 10.1016/j.jmarsys.2010.05.007, 2010.

Where is the "Great Calcite Belt"? Do you represent it well? And if your coccolithophores die out throughout your simulation, how does this affect PIC patterns?

R: Coccolithophores do not die in experiment PHAEO and do not have any strong drift. The Great Calcite Belt can be seen from figure R.3.7 depicting PIC:POC ratio for experiment PHAEO (here for March 2008, in the revised version for February 2008).

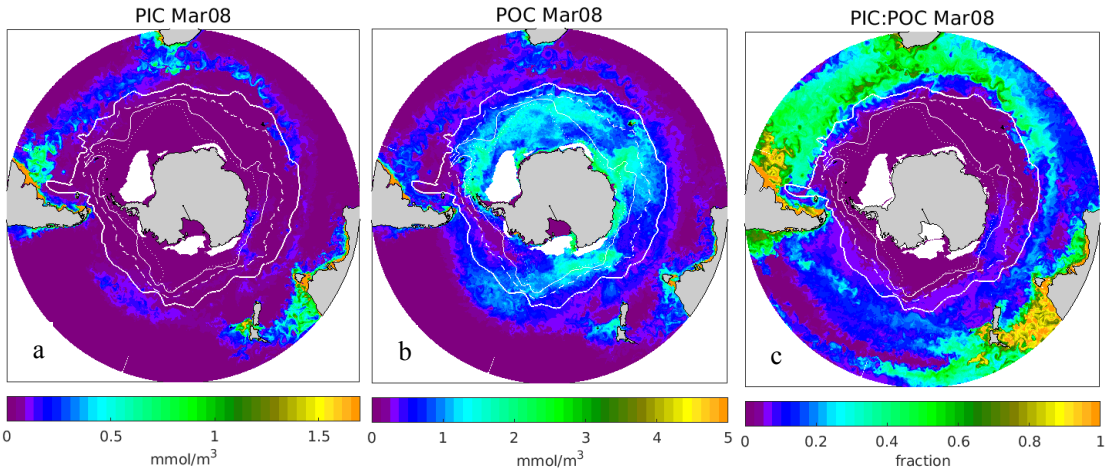


Figure R3.7: Spatial distribution of PIC, POC the ratio of the model surface particulate inorganic carbon to particulate organic carbon (PIC:POC) for experiment PHAEO in March 2008.

2004 does not seem to represent a "typical" year, as there is no typical year in a model with a strong drift.

R: Indeed, in experiment REF for the year 2004 PIC/POC was slightly affected by still slow drift shifting towards *Phaeosystis* (we no longer show 2003/2004 results for REF). However, it is worth mentioning that for the year 2004, as well as for the following years, REF shows unrealistically low PIC:POC ratio since this model configuration does not distinguish among haptophytes.

The results for typical year 2008 (long after the model solution reached quasi-steady state) for experiment REF are presented in figures R3.4. Please mind, that there was/is no strong drift in the experiment PHAEO that was/is finally evaluated.

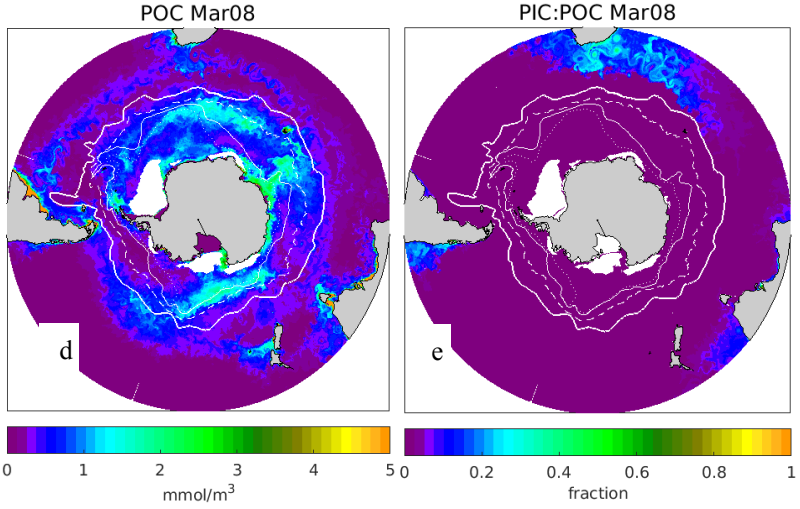


Figure R3.8: REF surface distribution of the POC and PIC:POC ratio for March 2008

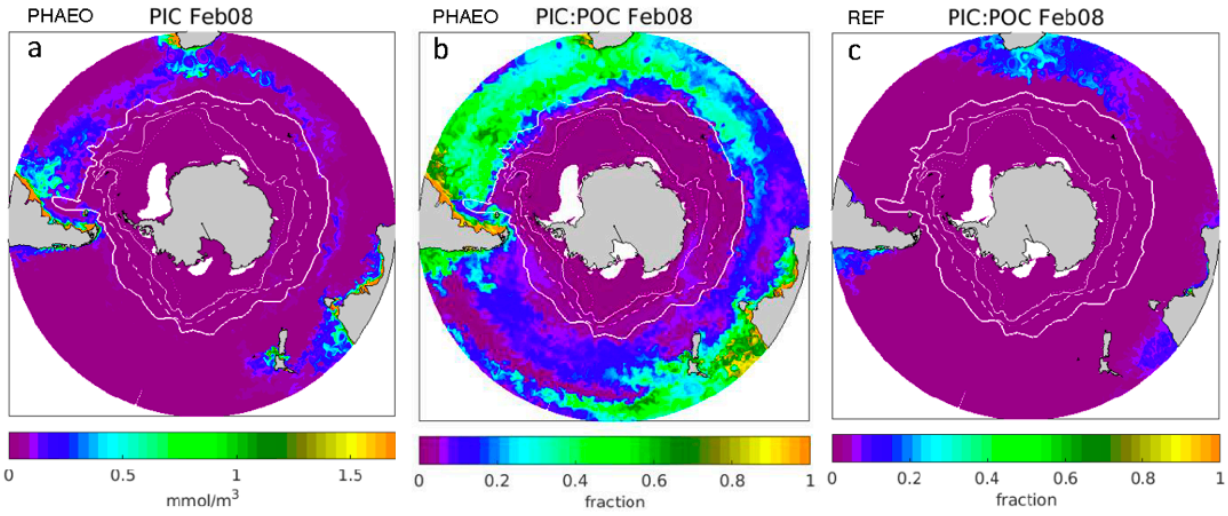


Figure 7 in the revised manuscript: Spatial distribution of the model surface particulate inorganic carbon (PIC, mmol m^{-3}) for experiment PHAEO (left panel), ratio of PIC to total particulate (dead) organic carbon (PIC:POC) for experiment PHAEO (middle panel) and PIC:POC for experiment REF (right panel) in February 2008. White contours denote the Southern Ocean fronts (Orsi et al., 1995; Orsi and Harris, 2001) as in Figure 1.

Figure 7: Same comments as Fig 2. Winter patterns unrepresentative, and very likely very tricky to model. Where is the "Great Calcite Belt"?

R: You can expect to see the "Great Calcite Belt" depicted in Figure 7 that shows dominance among haptophytes, diatoms and prokaryotes. The simulated winter PFT Chla patterns are clearly seen (but hardly being evaluated with available data) in the supplemented videos.

Fig 8: Merge with Fig. 5.

R: Figures 8 and 5 can hardly be merged since illustrate different aspects. While Figure 8 shows the PHAEO Chla distribution for diatom and coccolithophores compared to SynSenPFT retrievals in support to the discussed diatom *vs.* coccolithophores dominance as predicted by Darwin-MITgcm and observed by Smith et al. (2017), Figure 5 (6 in the revised version) is introduced to discuss, in particular, on drivers of nano-phytoplankton biogeography in the Great Calcite Belt and limitations of the satellite retrievals.

Fig. 9 - 11: These figures do not contain any quantitative information. Omit.

R: The figures are replaced by the reference to three supplemented videos (Simulated distribution of diatom (small + large) chlorophyll concentration in the Southern Ocean, https://doi.org/10.5446/42871; Simulated distribution of haptophytes chlorophyll concentration in the Southern Ocean, https://doi.org/10.5446/42873; Simulated distribution of prokaryotes chlorophyll concentration in the Southern Ocean, https://doi.org/10.5446/42872). Although this video material does not

provide quantitative assessment, we still think it a valuable evaluation. It clearly shows the simulated distribution and Chla phenology of the key Southern Ocean PFTs against the observed distribution. Moreover, the statistical analysis of model to data matchups was also, performed, shown and discussed.

In the text we write:

"Although these videos only allow visual comparison, they do show that the *in situ* observations (indicated by circles) match well the model Chla of diatoms and haptophytes in the area close to the Antarctic Peninsula and in the Southwest Atlantic Shelves biogeochemical province (FKLD, Longhurst, 1998), which illustrates a good agreement between the model and observations. In the Ross Sea, however, the model performance is less accurate: our simulated Chla for Phaeocystis as haptophytes in Ross Sea are underrepresented in comparison with HPLC-derived estimates."

The shown model and HPLC-based PFT Chla matchups were/are supported by statistical analysis. The matchup statistics were/are presented in Tables 3-5 and Tables S7-S9. We have slightly restructured and edited the text so that the reader does not miss the provided quantitative assessment:

"We have obtained matchup statistics for the comparison of our PHAEO model results against the *in situ* HPLC-based PFT Chla observations by Soppa et al. (2017). The mean absolute deviation (mean absolute error, MAE) of collocated model and *in situ* PFT-Chla over the considered time frame (August 2002 – April 2012) and the entire Southern Ocean is 0.74 mg m^{-3} and 0.22 mg m^{-3} for diatoms and haptophytes, respectively. Tables 4 and 5 present the statistics of model and in situ PFT-Chla comparison at several Longhurst's biogeochemical provinces (Longhurst 1998, see Figure 1). The highest disagreement was obtained for diatoms in the Atlantic Sector of the ANTA province, where the simulated diatom Chla is systematically overestimated by ~0.5 mg m^{-3}. The best agreement with the HPLC based diatom Chla (excluding small provinces, see Figure 1) was obtained at the SSTC and SANT. For the haptophytes, the highest systematic error towards overestimation has been found at two small provinces east of Africa and Australia (EAFR and AUSE) with the bias = 0.57, 0.48 (mg m^{-3}), respectively. The highest random error is (RMSE = 0.62, 0.44 mg m^{-3}) at EAFR and APLR. The lowest differences between predicted and observed haptophytes was at the FKLD, SSTC provinces where haptophytes are mostly presented by coccolithophores, and at the SANT biogeochemical province, where both coccolithophores and *Phaeocystis* co-exist. As additional information on the agreement between model and observations, Figures S9 and S10 in the Supplementary Material present frequency distributions of diatoms and haptophytes Chla for the simulations and measurements as well as the frequency distribution of the model and data differences. The latter shows that statistical criteria, such as MAE and root mean squared error (RMSE) give statistical meaningful metrics with respect to "model minus *in situ* Chla data" and the evaluation does not necessarily require a logarithmic transformation, as it is often done in ocean colour product validation (Brewin et al., 2010; Losa et al., 2017).

With respect to the agreement between model and observed *in situ* Chla for prokaryotic pico-phytoplankton (Soppa et. al 2017) depicted in Figure S11 (Supplementary Material) one can conclude that the frequency distributions of the simulated and

observed pico-phytoplankton are different, and the frequency distribution of the differences confirms that MAE and RMSE given absolute (Table 6) or logarithmically transformed values can hardly provide satisfactory estimates. Nevertheless, it is worth emphasizing that the largest differences between model and observed in situ prokaryotic pico-phytoplankton are located along the Antarctic Peninsula.

It is worth mentioning that the statistical estimates between model and observation PFT-CHla were carried out using matchups within ±1 week. Moreover, the model does not explicitly represent sea-ice algae and, therefore, might work less well in the region around the sea-ice. In this respect, we have to point out that all the statistics are presented for a qualitative assessment of the model rather than for a quantitative estimate of model uncertainties, since the representation error (Janji´c et al., 2018) related to the differences in spatial and temporal scales considered and sampled by the model vs. observations as well as to the mismatch in grouping phytoplankton (Bracher et al., 2017) are quite large."

4 Concluding remarks and outlook

Same as above. It is impossible for me to evaluate this section. Since the modeling work does not seem to be up to the standard in the field, the model evaluation is missing and the analysis of the results is severely flawed, it is impossible for me to judge the scientific interpretation of the findings. Any conclusions based on this work must remain mere speculations at this point in time.

R: Our conclusions were drawn based on the experiment PHAEO evaluated with available in situ and satellite observations quantitative (when possible) and qualitatively. The required by the reviewer additional model evaluation is provided in the revised version but does not change the conclusion.

In a sense it is not a standard study because we look on biogeochemical modeling of phytoplankton diversity from observational measurement perspective: towards a synergy between different types of observations (given their assumptions, principles and limitations) and numerical models which have now a possibility to account for different measurement principles, different aspects of the diversity. Such models might guide the further retrieval algorithm developments that in turn would allow to get observations for constraining better the biogeochemical models, since there are still lots of uncertainties in the model parametrizations and parameters representing differences in plankton traits (e.g. phytoplankton growth and lost).