Responses in bold

Anonymous Referee #2
Received and published: 2 March 2020

The manuscript by Kwiatkowski and Coauthors discusses results from the latest Coupled Model Intercomparison Project Phase 6 (CMIP6), focusing on the biogeochemical component of ocean change in response to a variety of future radiative forcing (RF) scenarios. The CMIP6 models are further compared with the previous generation of simulations and scenarios, CMIP5. The analysis indicates similar ocean biogeochemical responses to anthropogenic RF as for CMIP5: warming, acidification, deoxygenation and reduction of surface nutrients (i.e., ecosystem "stressors"), with varying magnitudes depending on the scenario, and various spatial patterns underlying interactions between changes in ocean circulation, chemistry, and ecosystem.

The analysis shows that responses in the CMIP6 simulations are very consistent with the responses in CMIP5, with two notable differences. First, the RF scenarios used in CMIP6 imply higher CO2 concentrations in the atmosphere than the equivalent RF scenarios in CMIP5. This results in stronger C uptake, and thus acidification, for the same level of RF. Second, the climate sensitivity of CMIP6 models appear to be stronger than CMIP5. Thus, for a given RF scenario, the warming response in CMIP6 is overall more intense. This is reflected for example in stronger stratification, the ensuing decrease in O2 ventilation and surface nutrient supply, etc.

Together with these main differences, the Authors detail a series of changes that were not documented at the same level of detail in the CMIP5 equivalent of this paper (Bopp et al., 2013, Biogeosciences; B13 hereafter). First, they detail responses in the benthic layer, showing similar but muted effects, and larger inter-model variability. Second, they detail changes in the seasonality of specific stressors, focusing on surface warming and acidification.

The results are worth publication. These CMIP6 experiments encapsulates the latest "consensus" of the climate and ocean community on future climate change, and are the result of an impressive scientific undertaking. Results from CMIP5 scenarios have guided much of the research on anthropogenic ocean biogeochemical/ecosystem change over the past decade, and CMIP6 experiments are likely to guide the next round of studies. Thus, there is a need to document the main features of these simulations, and provide a reference for future work. In this sense, the manuscript by Kwiatkowski et al. is needed and welcome. I should add that, by themselves, the results documented by Kwiatkowski et al., are not particularly novel: they mirror results from CMIP5 and previous work, with the appropriate (somewhat minor) differences related to changed scenarios and sensitivities. That said, the Authors strive to add a sense of novelty to the paper wherever possible, e.g. by presenting a few new analyses, and interpreting differences with CMIP5.

As the outcome of a community effort, the results in the paper appear robust, although it is impossible for a single reviewer to assess the vast amount of information that went into this synthesis. However, all models utilized are presumably documented independently and thorougly, and I am sure that the literature describing them will continue growing over the coming years, allowing more in-depth evaluation of individual models, or of specific responses.

Overall, the manuscript is informative and well written, and will be useful. The figures are clear, and the main results supported by the evidence provided – with the caveats discussed above. Overall, I am supportive of publication in Biogesciences, after the following comments are addressed.

-A comparison with observations is missing. Such a comparison was part of B13, and provided needed background to discuss anthropogenic changes. I suspect this comparison is missing for conciseness, but Biogesciences should allow for a few additional figures "for the record". I also suspect that some of these comparisons will be shown in other papers – although they would be helpful here too.

Specifically, it would be useful to see maps of present-day mean properties relating to the various stressors (SST, pH, O2, etc.) at the relevant depths, for the best observational products and the CMIP6 model ensembles. This would allow the reader to contextualize the maps of changes. I would suggest adding this comparison to observations for the properties shown in Fig. 1 and 8, and perhaps 10, 11.

A sister manuscript focussed on ocean biogeochemistry model performance in CMIP5/CMIP6 is currently in revision in Current Climate Change Reports (Seferian et al.,). This manuscript focuses on comparing mean-state ocean biogeochemistry variables to observations and includes mean state maps, observation-model anomaly maps, Taylor Diagrams, etc. We obviously do not want to repeat figures that are central to this manuscript and are happy to provide a current version to the reviewer/editor.

We propose to add the CMIP6 climatological mean values (historical period used to calculate anomaly maps) to Figures 1, 8 and 10. This is already shown for figure 11.

- Related to the above, a figure showing individual model performances vs. observations would also be useful, to provide context on overall model biases and spread. Again, I suspect that such a comparison will be presented in more detail in other publications, but it would be beneficial in this paper too. My suggestion would be adding a figure along the lines of Fig. 2 in B13, i.e. a "Taylor Diagram" of individual model performances vs. observations. (Taylor Diagrams provide a good "summary" compromise.)

While we do not wish to replicate figures/analyses in Seferian et al., (see response above), we agree that some observational comparison would be very useful. We propose to add an observation-model comparison of upper-ocean mean state trends (SST, pH, O2, NO3. NPP). Such an analysis is not included in Seferian et al,.

- A discussion of the robustness of spatial patterns of impacts, e.g. of the agreement between model spatial changes, is glaringly missing. Different stressor changes are obviously associated with different degrees of uncertainty in different regions, as the Authors discuss in the text – with pH changes being very robustly constrained, and other changes (e.g. in the benthic layer) being much more uncertain. B13 addressed this point by plotting a measure of model agreement on maps of changes – i.e. the "stippling" on Figs. 5, 6, etc. in B13. This is extremely valuable information that contextualizes the magnitude of the stressors and our knowledge of how they will likely play out. I strongly encourage the Authors to address this point by revising the relevant figures.

We will add stippling to multi-model maps in the revised manuscript.

- The paper focuses on physicochemical changes, possibly to limit the amount of information that needs to be discussed, but it stops short of addressing major ecological changes predicted by the models. In particular, a discussion of NPP changes (which again was included in B13) is missing. Again, I suspect this will form part of a more ecologically-focused publication, but at the same time I feel that NPP changes are an integral part of the story told in this paper – e.g. they are the real implication of including stratification and declining surface nutrients, and in turn may drive more or less important changes in the other stressors discussed. I see how discussing NPP could open up discussion of an entire new set of (complex) processes (export, recycling, remineralization), but if a demarcation should be arbitrarily imposed, I would suggest it includes NPP in the current manuscript, at least as the major ecological change (and potentially stressor), and a "tease" for future, in-depth analysis of other ecosystem implications.

The reviewer is correct that it was a conscious decision to not include NPP in the initial manuscript, as much of the NPP response requires additional processes to be assessed. We will add a brief assessment of the NPP projections as they suggest.

- Related, the introduction could do a somewhat better job rationalizing the scope and rationale of this paper. E.g., it is clearly not a comparison of different models, and can not go into too much detail on the effects of model structure, or resolution, or representation of different processes. Yet, all of these aspects underlie the imputes discussed in the paper, or at least their uncertainty. Similarly, it stops at mostly physicochemical changes, but ecological changes (NPP) are part of the picture, even when considering the stressors discoed here.

We will revise the introduction, to make the link between physicochemical and ecological changes clearer.

- At times, I wished for more details on the models than are summarized in the two tables (perhaps including an additional table), mostly to avoid having to go to the primary references, or to other syntheses (e.g. often the reader is referred to Seferian et al., in prep.; I did not have access to that publication, and I would have preferred to see the relevant information in this paper). The information that I think would be useful, at least when contextualizing individual model results, inter-model spread etc., includes in particular model resolution (horizontal and vertical; atmospheric), and biogeochemical complexity (e.g. functional groups, ecosystem model structure, stoichiometry, etc.).

These details and more are provided in Seferian et al., in revision. We are happy to provide the reviewer with a copy of this manuscript as stated above.

- Parts of the paper (e.g. abstract; Section 3.4, and others), read at times like "laundry lists" of changes and uncertainties for different stressors, scenarios, and Intercomparison Phases. This doesn't make for a particularly engaging read of those sections, and the reader could be easily referred to Table 3, where changes are summarized, while discussion could rather focus on new findings (e.g. consistency with CMIP5, etc.) or processes. Along these lines, I think the abstract could be made much more incisive, and could highlight novelties compared to previous studies (e.g. B13), rather than reporting lists of numbers.

We will address this.

- Section 1.2: I suggest summarizing in a few sentences the relevant information from Seferian et al., in review, mostly to provide the required context without referring the reader to another publication.

We will do this.

- Section 2: I am a bit confused by the use of the word "integral". I tend to think about the mathematical definition of integral, although here the term is used to refer to an average (related, but not identical).

This will be clarified.

- Line 264: I suppose what declines is the relative contribution of structural uncertainty, rather than the absolute value. This could be clarified.

We will do this.

- Section 3.5. The discussion of benthic changes is a useful addition tot he paper. However, a discussion of deep-ocean model resolution and other sources of uncertainty could be included. Ocean models are usually not designed to resolve deep ocean properties (and processes) as well as in the surface ocean, so the caveats may be different and more important here.

The reviewer is correct. We will include this, discussing for example the rationale behind the drift correction of benthic projections.

- Lines 304-307: More detail could be given on these processes, in particular the effect of freshwater dilution. Also, changes are quite hard to see on the maps, especially for the low RF scenario, e.g. Fig. 2c. I wonder if on the maps, contour lines and labels could improve readability.

We will discuss the potential importance of freshwater fluxes here and will work on improving map clarity.

- Lines 310-315: Looking at O2 changes in the N Pacific, I cannot help wondering what the importance of marginal seas is on some of the stressor changes – for example, in this specific case the role of the Sea of Okhotsk in ventilating NPIW (other examples can be thought of, e.g. Persian Gulf, Red Sea, etc.). I suspect global models have significant biases in marginal sea circulation, but sometimes these poorly-resolved regions can disproportionally affect the open ocean. Perhaps a discussion of these issues can be included somewhere, with some indication of obvious biases and possible directions for improvement.

The manuscript is global in scope and focussed on projections. Discussion of model biases and the underlying model features that may drive these is provided in Seferian et al., in revision. With respect to marginal seas, Seferian et al., assesses the potential role of external boundary conditions (e.g. sedimentary and riverine inputs) as well as vertical and horizontal resolution as drivers of model biases.

- Lines 339-340, and 350-351: I have a hard time seeing the effect discussed in practice, e.g. by comparing Figs 2h and 4b.

While the pattern of absolute NO3 anomalies in CMIP6 is similar to that published for CMIP5 projections (Fu et al., 2016; Cabré et al., 2014), we do not currently show the relationship between relative stratification index anomalies and relative NO3 anomalies. We will perform additional analysis of this during revision and modify the manuscript appropriately.

- Section 3.3. This is a nice section, and it's clearly important to look at the compound effect of multiple stressors. That said, the Authors could do a better job in discussing the (arbitrary) thresholds selected. Especially for O2, picking a change threshold may not be that informative – a change by 30 mmol/m3 may be negligible in waters close to saturation, and would be massive in waters close to suboxia. I realize a best summary threshold that encompasses a heterogeneous range of stressor responses may not exist, but some rationalization (and caveats) would be useful for context.

We take the reviewers point. There is no perfect solution to this as using an absolute threshold for O2 typically just highlights the present distribution of OMZs, as opposed to changes in O2. We will revise this figure to improve clarity and further discuss our rationale.

- Line 381: I am not sure I get the referent to the MAGICC7 model – I couldn't find it mentioned elsewhere in the manuscript.

The MAGICC7 model is a reduced complexity climate model that emulates the response of complex Earth system models and can provide probabilistic projections. As opposed to the ESMs, which have changed substantially between CMIP5 and CMIP6, the same MAGICC7 model projects marginally greater twenty-first century warming in RCP8.5 than SPP5-8.5 (Meinshausen et al., 2019). We will make this clearer in the revised manuscript.

- Lines 415-420: I was surprised by the inconsistency in the bottom water O2 changes, which are in fact larger in SSP1-2.6 than SSP5-8.5 (although indistinguishable given the uncertainty). Maybe this can be commented on. This also bring up an additional thought: bottom ocean ventilation, especially in the Southern Ocean, may be strongly affected by another sets of processes poorly captured by current climate model, namely, open-ocean polynyas. I wonder if different RF scenario result in somewhat non-trivial changes in SO deep ventilation, which in turn affect bottom water O2 and other properties.

The reviewer is correct that the differences between mean O2 changes are not significant. Moreover the model ensemble differs between SSP1-2.6 and SSP5-8.5. We will comment on this. While analysis of physical ocean biases such as Southern Ocean ventilation is beyond the scope of this paper, we will see if there has been any analysis of this in the recent literature and include it in our discussion accordingly.

- Line 542: the relationship between RF scenario and impacts is shown in the paper in a somewhat indirect way: i.e. there is not a single figure (e.g. along the lines of Fig. 6) that relates RF to impact. I think the closest would be a figure relating stressors to SST, somewhat along the lines of Fig. 6a.

We will revised figure with NPP included and pH/O2/NO3/NPP anomalies shown against SST anomalies in a four panel figure.

- Fig. 6b: I'm puzzled by the fact that SSP3-7.0 shows more dramatic changes here than RCP8.5. I suppose this may have to do with the stronger climate sensitivity of CMIP6 compared to CMIP5. But the SST response is similar in the two scenarios (Fig.6a).

This is likely partially driven by the higher climate sensitivity in CMIP6. Moreover, the ensemble of models that provide SST and O2 in CMIP5 and CMIP6 are different (see tables 1 and 2). As such, so a direct comparison between the points in panels 6a and 6b is not straightforward.

- Figure 7: This is a useful figure, but I wonder how straightforward the interpretation actually is, since it conflates changes at very different depths. I.e. most of the ocean sits at around 4km depth, where impacts are muted, but much stronger impacts would occur in shallower benthic waters (which it should be noted host more important benthic resources). I wonder if an additional figure showing depth-dependent changes (i.e. profiles for benthic grid boxes only), e.g. for the end of century, could be a useful addition to Figs. 7-8.

We will add depth profiles of benthic grid cells to figure 8.

- Line 156: "bacteria" -> "heterotrophic bacteria"? (I'm thinking that classic picoplanktonic functional groups already include bacteria).

We will clarify this.

- Line 272: "follows" -> "follow".

This will be corrected.

- Line 428: "in" -> "from"?

This will be corrected.