Sensitivity of 21st-century projected ocean new production changes to idealized biogeochemical model structure

Author response to reviewers. Note that original text of the reviews is in italics.

We thank the reviewers for their thoughtful notes, their attention to detail, and their efforts to help us improve this manuscript.

Review 1

1. *Section 2.3*

This decomposition is a clever analysis technique, however there are some unaddressed issues that make the interpretation of the results questionable.

There is an error in equation 6, which forms the foundation for much of the analysis. Specifically, there is a sign error and the cross term between the changes in light and nutrients should be negative rather than positive. The equation should be Δ(QL) = QΔL + LΔQ - ΔQΔL. Because the components in the figures (e.g. Figure 6) look like they sum to the total, it is possible that this is only a typo and does not affect the results that are presented in this manuscript, but I would ask the authors to verify the nature of this error. This difference in sign may help the reader to understand why this term has a negative covariance with the total change.

It may help the reader if equation 6 is written out more explicitly (i.e. (Q_2100-Q_2000) rather than ΔQ) to show which differences are being presented. One reason this is helpful is that it may not be immediately obvious to the reader that L and Q on the right hand side are the values at the later time and that this is a difference, not a product rule derivative.

Thank you for checking the math here. Indeed, we formulated this equation differently than you assumed, so we now write it out more explicitly. The equation is:

d(QL)=QdL +LdQ +dQdL =Q0(L1-L0) + L0(Q1-Q0) + (Q1-Q0)(L1-L0)

where d is used rather than delta, Q0 is Q in 2000s, Q1 is Q in 2100s, and so on; equation 6 (now labeled 7) has been updated in the text to include these details. The alternate formulation which you expected is:

d(QL)=QdL +LdQ +dQdL =Q1(L1-L0) + L1(Q1-Q0) - (Q1-Q0)(L1-L0).

The pattern correlation coefficients are similar but not quite the same for these different formulations; we do not see qualitative changes. In the below table, dQdL in the new equation is - (Q1-Q0)(L1-L0).

Correlations with d(QL)	Slow	Fast	Slow, new eqn	Fast, new eqn
LdQ	0.73	0.64	0.78	0.52
QdL	0.44	0.19	0.37	0.17
dQdL	-0.035	-0.058	0.035	0.058

The change in light is calculated from R and Q instead of being calculated independently from equation 3. In fact, based on equation 3 the light would not change at all under climate change. Instead, the change in L seems to combine changes in mixed layer depth (because light is averaged over the mixed *layer) and changes in the vertical structure of nutrients (and particularly covariance of light and nutrients). Which of these factors is most important? Which is most important in regions where \Delta L \Delta Q cancels L\Delta Q and which is most important in regions where \Delta L \Delta Q largely cancels Q\Delta L?This should be made more explicit in the presentation of the methods, the results, and the discussion.*

As for the computation of L, our climate change scenario directly impacts the incoming radiation and the sea-ice coverage of the ocean, changing the light incident at the sea surface. See Fig 1a, which shows changes in the minimum incoming short-wave radiation of 10-25W/m^2 over most of the ocean. Then the mixed layer depth changes further impact the light availability through the averaging you noted. The reason for computing L from production rate and Q(N) is that the mixed layer depth, and thus L, can change rapidly. Production rate and N are averaged on-line in our model, and thus computing L from them allows us to have a time-averaged L that would not be possible to compute from the time-averaged mixed layer depth and incoming radiation. Note, however, that sub-monthly covariability of Q and L projects onto L using this approach. This explanation has been added to the text in the first paragraph of Section 2.3. For future development of models, it is a good idea to have as output an online-average of nonlinear functions like L and Q.

2. Section 3.1

One of the main quantitative metrics used in this study is pattern correlations. The correlation coefficients are presented but there is no assessment of the statistical significance of these correlations. The authors should compute the statistical significance or provide another context in which these correlations should be interpreted. The values of the pattern correlations are difficult to assess in isolation. Moreover, there is an over-reliance on pattern correlation in some sections. This is particularly true when discussion the \Delta L \Delta Q term (line 309), which has a small correlation with \Delta (QL) but can be a large contributor in some locations.

Determining statistical significances for ocean pattern correlations was not originally attempted because the Q and L functions and their differences are not distributed similarly to a standard statistical distribution. If we treat these terms as normally distributed, there is the challenge of deciding on an appropriate estimate of the degrees of freedom in the model global ocean. There are 82,565 ocean gridcells used in these correlations, but these are not independent. If we suppose a moderate correlation lengthscale of about 10 gridcells in both directions, which will be about 6-9 degrees of longitude and latitude depending on location, we have perhaps 825 degrees of freedom and $p<0.01$ for $r>0.09$. Empirically, the lagged autocorrelation of d(QL) drops from 1 to ½ after 18 gridcells (slow case, longer), which suggests a single-direction 36 gridcell correlation lengthscale, 2294 degrees of freedom, and $p<0.01$ for $r>0.06$. As the first estimate is a stronger requirement, we use that in the manuscript, and have added a paragraph at the end of Section 2.3 to explain this reasoning.

We have added text after equation 6 (now 7) on how we will analyze pattern correlations. We have also added that small pattern correlations, like those of dQdL with d(QL), are insignificant. Finally, we have collected our correlations into a table (Table 2 in the manuscript) as requested in the second reviewer's comment and note those which are significant.

3. Section 3.2

This section concludes that there are different mechanisms that govern changes in the seasonal cycle of production than that govern changes in the annual average of production. However, I remain unconvinced by this conclusion. The results shown seem to be driven more by regional differences than by seasonality. The largest seasonal cycles are in the high latitudes while the low latitudes have weak seasonality. The results show that changes in the seasonal cycle and the mechanisms driving those changes are similar to the arctic and sub polar North Atlantic, perhaps with more influence from the Southern Ocean in the slow timescale case.

Furthermore, the statement in lines 344-346 is only true for the fast case but not the slow case, again likely due to different responses in the high latitudes with each of the timescales.

The reviewer seems to infer that we are making more sweeping claims than we intend to convey. We agree with the reviewer in that global annual-average production is reduced in the warmer climate mainly due to reduced nutrient availability (LdQ), and this is reflected in the seasonal cycle results as well. Looking at the seasonal cycle, production is reduced mainly in latter half of the growth season, which is shortened. In both the fast and slow cases have different contributions from light availability and light-nutrient covariance during the latter half of the growth season. We rewrote the final paragraph of this section to make these points clearer, including the fact that the noted statement is only in regards to the fast case. The final sentence now reads, "In both our cases, reduced nutrient availability is a major contributor to the shortened growing season, indicating a consistent mechanism for reduced total production in response to the climate perturbation."

4. *Methods: How effectively is the nutrient mixed within the mixed layer? The light is averaged over the mixed layer, however there is a comment about productivity being enhanced below the mixed layer depth (line 469). Does this mean that light is more effectively homogenized than nutrients due to the mixed layer not being an actively mixing layer? How does this affect the results about the mechanisms that drive changes in productivity?*

Line 469 should (and has been changed to) read "enhances production in the lower portions of the mixed layer"; the averaged light enhances light availability, and thus production, in the deeper parts of the mixed layer.

The extent to which nutrient is mixed within the mixed layer is a very good question. Looking at the values of monthly-mean N, we computed the range and mean across the shallower of the mixed layer depth or the top 100m, as analyses in the text are limited to the top 100m. The median range of N and of range(N)/mean(N) are in the table below. Broadly speaking, there is a gradient of N within the mixed layer which is larger in the summer, near the equator, and in the present climate. The range of N is smaller than the mean N, but these are more similar in the fast case when the mean is quite small.

	Slow 2000	Slow 2100	Fast 2000	Fast 2100
Median range(N)	0.0151	0.0125	6.39 10^-6	1.29 10^-6
Median range(N)/mean(N)	0.119	0.145	0.447	0.447

The effect of stratification of N is that production is lower. This is because Q(N) is concave down, such that Q(mean(N))>mean(Q(N)). Nonetheless, these effects are likely small. As a simple test, we compared the mean of Q over the top 100m and Q computed from the mean of N over the top 100m. The spatial patterns of the annual mean of both fields are very similar. Quantitatively, across the slow and fast cases in both climates, about half the ocean gridcells have Q(mean(N)) within 10% of mean(Q(N)).

We have added a sentence to the text discussing the global spatial patterns of production and nutrient concentration: "We note that the variations in nutrient concentrations within the mixed layer are typically small, less than half the mean concentration and much less than k_N; these variations lead to slightly lower production rates than those that would occur if the nutrient concentrations were constant from the surface to the mixed layer depth."

5. *I am concerned about the low deep nutrient concentration and the implications that has on the high latitudes, particularly the Southern Ocean. Some of the largest changes are in the high latitudes and especially the Southern Ocean. The authors state that the model is a very poor fit in these regions due to the low nutrient concentrations in these regions, which would otherwise be larger than the 20 uM deep nutrient concentration used in this model. Could the authors justify this choice and make the implications of this choice more clear? Would the global average statistics differ if the Southern Ocean were excluded? How do these results then relate to mechanisms for changes in primary production such as Southern Ocean nutrient trapping and the predominance of the Southern Ocean is global carbon export?*

The choice of value for the deep nutrient concentration was based on global observed deep nitrate values. The observed concentrations of nitrate at depth varies considerably, from about 13mmol/m^3 in the Arctic to about 38mmol/m^3 in the South Pacific, so 20mmol/m^3 is within the range of observations. We now include this range of observed concentrations in the text for context. The deep nutrient pool is likely to change with the climate and associated shifts in deep water formation and ventilation. We chose a constant value of nutrient at depth across space and time for simplicity and to keep our focus on changes in the upper ocean.

With regards to the Southern Ocean, modeling production here with fidelity to real processes requires the inclusion of iron; we now note this in our section 2.2.2, paragraph 3. While one could get closer to observations with our idealized model, the deep concentration and half-saturation coefficient, at minimum, would need to change. The production in the Southern Ocean is close to half the global annual production (see appendix figure B1). The reductions in global annual production are 8.5-11% without the Southern Ocean, as opposed to 9.5-19.5% when it is included. This has been added to the end of the first paragraph in Section 3.1: "A large portion of this variability in global new production is related to the Southern Ocean, which is the basin with largest production and which our model does not represent well. Without the Southern Ocean the reductions in global annual production are 8.5-11%, which is smaller than the range of export decreases in CMIP5 and within the range of net primary production changes in both CMIP5 and CMIP6."

The causes for reduced annual new production remain mainly reduced nutrient availability without the Southern Ocean. We have similar results in the pattern correlations between d(QL) and its components regardless of including the Southern Ocean (table below). All correlations between d(QL) and either LdQ or QdL are significant; those with dQdL are not. We have added to the end of the paragraph on the pattern correlations a sentence: "These pattern correlations are qualitatively similar without the Southern Ocean, which we do not represent well with this model."

Correlations with d(QL)	Slow	Fast	Slow, no SO	Fast, no SO
LdQ	0.73	0.64	0.66	0.52
QdL	0.44	0.19	0.45	0.14
dQdL	-0.035	-0.058	-0.016	0.016

Minor comments

Introduction:

--*Paragraph at line 30: it is unclear what is meant by "essential properties" here although the phrase is repeated multiple times in this paragraph*

We have eliminated this phrase. The goal is to elucidate the sensitivity of the global-warming response of biological productivity to nutrient and light limited productivity dynamics, which we do using a single particularly simple form of joint single-nutrient and light limitation.

--*Line 65: add citations to the specific prior work to which you are referring*

We have added here references to Kriest et al 2012 and Levy 2015, which were noted earlier in the discussion.

--*This study is not the first to use simplified models to discuss biophysical coupling. The introduction should cite more of the relevant theoretical literature such as: Smith, K. M., Hamlington, P. E., & Fox-Kemper, B. (2016). Pasquero, C., Bracco, A., & Provenzale, A. (2005).*

We do not wish to imply we are the first to use this method, but we also wish to keep the discussed literature focused on studies similar to ours, which are generally those using global models. These papers are now included for context alongside the OCMIP-2 studies.

Methods:

Line 95: Are the tracers initialized in the model at the beginning of the spin up or only at the beginning of the 10 year timeslice? What is the tracer initial condition?

This clause now reads "we perform a 20-year physics-only spin-up, which is sufficient to reduce interannual drift in the physical state, and then use 10 further years, which include our biogeochemical model, as our early-century timeslice." Around line 150 (section 2.2.1) we have added the initial condition: "The initial conditions for N, which are used at the start of both 10-year timeslices, are a linear interpolation between 20mmol/m3 at 1000m and 1mmol/m3 in the surface gridcell."

Line 111: Figure 1 shows more changes than are outlined briefly in this section. There are some prominent changes like a speeding up of the ACC that are not mentioned.

The changes we present here are those we consider most important for the remainder of the text. We have added your point about the ACC and a few other details. These changes are not exceptional compared to the CESM-LE results and so we felt they did not merit much time.

Equation 1: what is the functional form of the restoring S? Is there a parameter that relates to the nutrient restoring rate?

We do not have a restoring process in the typical sense, but rather a reset process, which adjusts all N values below 1000m to 20mmol/m3 every timestep. The text here has been changed from "restored" to "continually reset" to be more accurate.

Line 154: please explicitly state the advection and mixing methods, which form part of the results later in the paper.

We are using default values for this model. We will adjust the relevant sentence to provide the appropriate reference: "The physical transport and mixing are done by the same mechanics as existing passive tracers in CESM-POP, with a third-order upwind scheme for advection and diffusive mixing that is spatially variable due to parameterizations of mixed-layer, submesoscale, and mesoscale isopycnal processes (see Section 2.2 of Danabasoglu et al., 2020, for details)."

Line 165: This is a key point about equilibration, but as it is currently written it is confusing what is being compared.

Thank you for noting this. We have split this sentence into two, with the second now reading "We do not see substantial differences in our climate change results when using year 5 or year 10 of the timeslices in our computations."

Section 2.2.2 is a results section rather than a methods section

We disagree. The question under consideration is the sensitivity of new production projections, not the sensitivity in the current climate. The purpose of 2.2.2 is to determine which parameter choices are worth including in the main analysis, and so we see it as a continuation of the description of the model.

Results:

Use consistent terminology for averages. Sometimes average is used and sometimes mean is used. If these are the same, please use just one term.

Thank you for noting this, we have replaced 'average' with 'mean' whenever referring to our results.

Line 277: The text makes it sound like the patterns don't change much, but the the correlation seems very low (r = 0.26).

The same is true at line 481-2.

With about 82,500 ocean gridcells, a reasonable decorrelation in both directions of 10 gridcells (6-9 degrees latitude or longitude) allows 825 degrees of freedom. Correlation coefficients above r=0.09 are significant. This has now been noted in the methods section following equation 6 (now eqn 7).

The correlation values are presented for multiple combinations of variables and averages of those variables but it is at times unclear what is being correlated. One example is line 293.

We have made a table (which will be Table 2) of all such correlations to make this clearer.

Line 305: this sentence has confusing wording

Thank you, we have split this into two sentences: From the spatial fields, QL and Delta(QL) show the same spatial patterns as production and its changes, respectively, as expected by definition. However, Delta(QL) looks somewhat different from the percent changes in production shown before (Fig.4).

Line 330: do you mean to repeat qualitative twice in this sentence?

Yes.

Line 403: Is the Arctic region defined using two criteria or are these two criteria equivalent?

These are equivalent, and we have added ‘equivalently’ following ‘or’ to make this clear.

Conclusions:

Line 466: “other processes” is confusing because the next sentence discusses productivity processes that were not included.

This now reads “certain aspects”, in order to keep the focus on production.

An overall note on the grammar is that the sentence structures can be repetitive with a few sentences in a row beginning with the same clauses (e.g. “here”).

We have removed 3 of 11 instances of the use of “here” throughout the manuscript.

Figures:

Figure 5 caption: Expand on what is meant by “Annual and 100m mean nutrient concentration.” Is this an integral over the upper 100 meters? Include all information in the caption (what is shown in panels a, b, c, d).

Updated caption: (a,b) Nutrient concentration, mmolN/m3, averaged over a year and the top 100m. (c,d) Percent change of this field from 2000s to 2100s climate. (a,c) slow case, (b,d) fast case.

Figure 6: the color scales appear to be saturated. Could the minimum and maximum values be includes as annotations on the figures?

Yes, they are saturated. The min and max values have been added below each panel here.

Figure 6: What is meant by 100m average? Is that an integral over the upper 100 meters?

It is the average over the top 100m, or the integral over 100m divided by 100m. The first sentence of the caption now reads “QL for 2000s climate (a,f), its change (b,g), and its components, all averaged over 1 year and the top 100m, then normalized by the maximum of QL in the 2000s.”

Figure 7: This figure would be easier to interpret if panels c and d were both underneath b.

This has been implemented.

Figure 9: Over what depth range are these values averaged?

The vertical fluxes are at 100m depth, not averaged over any range of depths. We have updated the caption to include this point. This is consistent with the flux in Fig 8a.

Review 2

(1) *Assumption of "constant phytoplankton": The model implicitly assumes that phytoplankton is constant in space and time, with the concentration being included in mu_0. Therefore, one might expect a strong*

dependence of surface nutrients, new production, or climate sensitivity on the model parameters and biological time scale. Given this, and the fact that the structure of the model is in contrast to models applied in CMIP6 (as noted briefly in the conclusions), I think this assumption and the consequences, that might arise from it, should be discussed a bit more.

We are assuming here that the total phytoplankton has an approximately constant effect on the new production rate, such that it may be subsumed into mu_0. This assumption does indeed lead to a strong dependence of near-surface nutrient concentration, new production, and climate sensitivity to the model parameters/biological timescale, as discussed in sections 2.2 and 3.1 (figures 2 and 3). Models which include the phytoplankton concentration generally are interested in simulating the total phytoplankton, not just the new production, and will also include remineralization. Our choice to simulate only new production and new phytoplankton will over-estimate growth when total concentration is low and under-estimate growth when total concentration is high in comparison to a production function including the new phytoplankton concentration. We thereby prevent exponential growth and limit spatial differences in behavior.

The second paragraph of section 2.2.1 has been expanded to include the expected effects of our simplifications and now reads as follows: "In designing the nutrient tracer, we make three simplifying assumptions. First, we assume that the deep nutrient pool has a fixed concentration, not dependent on explicit remineralization, which decouples the nutrient tracer from the export tracer. This assumption will create different vertical nutrient gradients than models with remineralization included. Second, we assume that new production depends on the availability of this nutrient and light alone, not on the water temperature or on the existing plankton population that may be sustained by recycling of nutrients; this omits processes thought to be important in bloom-type events (Behrenfeld and Boss, 2014) but again keeps the nutrient and export tracers decoupled. One may reframe this choice as subsuming an effectively constant total phytoplankton concentration into mu_0, which leads to an over-estimate of growth when and where total concentration would be low and an under-estimate of growth when and where total concentration would be high in comparison to a production function including the phytoplankton concentration. Finally, we assume that the light available for new production in the mixed layer is the mean of the light levels within the mixed layer (as done in McGillicuddy et al. 2003); below the mixed layer, productivity depends on the light at only the depth in question. This choice increases subsurface growth within the mixed layer and decreases near-surface growth, while allowing growth below the mixed layer depth."

This also concerns the normalization of alpha by alpha_0 (Line 195). What could be the (biological) meaning and implicit assumptions of alpha_0? The authors note that tau_bio shows a similar correlation for alpha_0 between 0.1 and 2, but not for a wider range. What range or value would be plausible? As alpha_0 is used to unify the rate constants k_N and alpha for the evaluation of tau_bio, I think this is very much at the heart of the paper, and should be discussed more.

Alpha_0 quantifies the relative effect of changes in nutrients and light on productivity in a single constant. The slopes of the production curves for N and I both contribute to the response of the production rate, which we describe as tau_bio. Alpha_0 is an expression of how we need to stretch the I coordinate so that the slope of Production(I) is equivalent to a Production(N) slope. We could imagine doing this using the ratio of expected values of I and N. Large values of both might be 200W/m2 and 20mmol/m3, leading to alpha_0 of 0.1. Any equivalence like this will likely lead to alpha_0<1. We could also expect slopes to be the same order of magnitude, so alpha_0/alpha would be similar to kn; this suggests alpha_0 similar to kn*alpha, which ranges between 0.0031 and 0.8 for our parameters. Since we are interested in comparing across parameter cases, a constant value of alpha_0 is best. Based on these considerations, our choice of alpha_0=1 is an upper bound.

We have added to the paragraph introducing the biological timescale: "Our alpha_0 is an expression of how we stretch the light coordinate so that the initial slope of production with respect to I is in the same units as the initial slope of production with respect to N, suggesting alpha_0 as a ratio of N/I. Given the relative values of I and N, alpha_0<1mmolN/Wm is likely to be the most fruitful." We have also added to the next paragraph, following "This correlation is best for this and similar values of alpha_0, e.g. 0.1 or 2 mmol N/WM, but is lower for e.g. 0.01 or 100 mmol N/Wm" the sentence "This is consistent with our understanding of alpha_0 as a ratio of N/I."

(2) *Lack of subsurface remineralization and its potential feedback on surface nutrients: By setting w_s=0, the model assumes no particle sinking and subsurface remineralization below the euphotic zone. But: wouldn't we regard any nutrients remineralized in depths > 100m as new nutrients, which could then be injected back again into the mixed layers or euphotic zone? The effect of sinking is mentioned briefly in line 219: "For instance, with ws of 5m/day, sigma =1/yr has 95% of annual production sink below 100m." I assume that this value of 95% arises for the equilibrium case, i.e. from exp(-100/(5x365)), correct? Then, using these parameters, exp(-900/(5x365))=61% of the export at 100m would sink below 1000m, and 39% would be remineralized in the water column between 100-1000m, adding to the nutrient pool. The effect would be even stronger with the parameter for sigma applied in the study: with sigma = approx. 6/y (1/60days) only exp(-100*6/(5x365))=72% of the particles produced in the surface would leave the upper 100m and the remaining 28% would be remineralized within. While the latter nitrate, by definition, would not add to new production, it might nevertheless affect the gradients and thus supply of nutrients. Further, 95% of the flux leaving the surface would be remineralized above 1000m, thereby increasing the concentration of subsurface nutrients, and their potential re-injection into the surface. This feedback (which is possibly included in all global models run in CMIP6) may have considerable consequences for the model's sensitivity. Especially, since the "slow" case is considered as small phytoplankton (which might even sink more slowly), it may reduce the importance of term Delta Q in the subtropical southern Pacific (Fig 8) quite strongly. I think the specific setup of the model, and its consequences for the different terms should be discussed in much more detail.*

The definitions of new nutrients and new production are indeed flexible and could include nutrients that were remineralized below 100m and above 1000m. Such nutrients will be more important in cases where particles/detritus sink slowly and remineralize quickly. While the equilibrium calculation above is an accurate way to approach this point, we in fact did this work empirically, using our slow case with the above w_s and sigma values.

It is indeed the case that remineralized nutrients affect the vertical gradients of the nutrient field. This feedback mechanism is one that contributes to the challenges of understanding the CMIP class of models' response to climate change: reduced near-surface nutrient in a warmer climate cannot be solely attributed to reduced physical supply from depth, but may also be affected by changes production, sinking, and remineralization rates that affect the nutrient concentrations from the surface downward. These effects will certainly change the sensitivity of new production to climate change. The advantage of not coupling the nutrient field to the particulate field allows us to more easily and definitively attribute causes to the changes in new production with climate.

We have added a relevant sentence when introducing this assumption, as noted in the previous point, "This assumption will create different vertical nutrient gradients than models with remineralization included.” We also added a sentence to the first paragraph of the conclusions, “Second, the lack of nutrient remineralization contributions to the nutrient field and the constant value of the deep nutrient pool remove a mechanism of production feedback which can affect its climate sensitivity.”

(3) *Given the differences to other models mentioned in (1) and (2) (and also in the conclusions), I wonder how one could apply or adapt this analysis to CMIP6 models (Lines 501 to 506). Would it be possible to apply this analysis to the BEC model, to which the present model is compared, and which is simulated in the same circulation? This might indeed be a good proof of concept!*

We agree that the first target for applying this analysis to a CMIP model would be the BEC model. The empirical methods discussed in 501-506 would determine tau_bio for the combined production of all phytoplankton, as opposed to the tau_bio for each phytoplankton class which can be determined analytically from the governing equations. Unfortunately, it is outside the scope of this work to include such an effort, as we anticipate it taking several months. Below, we describe in more detail plausible analysis options and the challenges involved.

The option to use a single column with plenty of all but one limiting factor and an injection of a burst of the remaining one at a few levels to determine the production slope is most straightforward but requires implementing BEC into a column ocean. The option to fit the production curve over the nutrients and lights, using the global ocean’s variety to provide the range of data, does not require new simulation development but would require a novel analysis. The goal would be to identify the slopes of production at the edge cases where a single nutrient, or light, has a low level and the rest are plentiful. A challenge is that the community composition in different regions varies, such that it is not clear whether a single tau_bio may be an accurate representation. In all cases, combining the production slopes will require choosing normalizing factors, like alpha_0, for each term.

We have added a sentence to the relevant paragraph in the conclusions, “While computing an effective tau_bio is outside the scope of this work, we believe developing a reusable procedure for these intermediate-complexity models to be a useful next step toward interpreting climate change production projections.”

Specific comments:

Line 5: "and export via sinking organic particles": as the sinking speed is set to zero, I don't think this is strictly correct, but implies that new production=export production.

This is exactly what we wish to suggest. The choice of zero sinking was for simplicity, as we do not discuss the export in detail in this work. We have experimented with sinking speeds of 1-10m/day, and future work will include a detailed examination of export.

Lines 23-26: The meaning of two sentences is not clear to me: What are the differences between "structural differences" and "a variety of different ocean biogeochemical models".

The first is meant to refer to the differences in the way physics is represented across models, including differences of numerical methods of implementation. The relevant section of the sentence has been expanded to “structural differences in the models' representations of physical processes produce”.

Lines 36-27: "The effects of both biogeochemical model structure and physical circulation–biogeochemical model interactions have been examined in isolation." - What is meant with this? That studies examined either physical or biogeochemical effects? (But see, for example, Romanou et al., 2014, https://doi.org/10.5194/bg-11-1137-2014 or Kriest et al. (2020), https://doi.org/10.5194/bg-17-3057-2020, who both investigated the effects of physical model and biogeochemical setup at the same time.)

Many papers do focus on one aspect or the other of the problem, but we have updated this sentence to no longer imply that there are no papers doing both. Thank you for suggesting these papers in particular; we have included them alongside Loptien and Dietze (2019) in a new paragraph of the introduction focused on examples of papers considering both parts of the problem.

Line 42-44: "(...) differences in mixing that cause small changes in temperature and salinity or global production and biomass, respectively, create large differences in primary and export production (...)" - What is the difference between "global production" and "primary and export production"?

We have split this sentence into two in order to clarify that these are referring to fields in the two different studies.

Line 75: Wrong section number?

Yes, all sections numbers here have been updated, thank you.

Line 95: "to minimize interannual drift" - I assume that with a longer simulation time drift could be even smaller; perhaps better: "To reduce annual drift"?

Yes, we have replaced ‘minimize’ with ‘reduce’.

Line 145: POP has not been defined before. I assume that it is particulate organic phosphorus. But, given that the basic unit of the model seems to be nitrogen, shouldn't it rather be PON (particulate organic nitrogen)?

POP is the Parallel Ocean Program, the ocean component of CESM. Both uses are now noted as CESM-POP to avoid confusion and we define POP near the beginning of the methods section.

Eqn. 4: If w_s is set to zero (line 159), why mention this loss term at all?

This term is included so that this model can be used more broadly. As mentioned above, we have varied this parameter in our efforts and future work will include analyses of export. We decided including variations in w_s here would over-extend this piece.

Line 153: "is the specific mortality rate of particles". Particles (as a general term) don't have a mortality rate. I would suggest to choose a different, more general expression, such as "decay rate".

Yes, we have changed this as you suggest.

Line 159: What was the reason for choosing w_s = 0? Please specify the units of w_s.

We have added the units, m/day. W_s=0 gives the best comparison between P and BEC’s phytoplankton near the surface; BEC does not have a similar sinking term, but instead a vertical redistribution process.

Line 162: Please specify the units for mu_0, k_N, alpha.

These are noted when the terms are introduced and are now included when giving their values for the sensitivity study.

Line 167-168: It would be interesting to also see the comparison of the model results to those of BEC (as this is also done for particles).

We have added the BEC surface nitrate concentration curve to figure 2a. This model also misses the location of the northern hemisphere peak, suggesting this may be due to the model circulation, but is much closer to observations in the Southern Ocean.

Lines 171-172: "Although our particles represent both living matter and detritus, near the surface this P is most like newly-produced plankton which we expect to have the same spatial patterns as total phytoplankton." - Why not compare directly against BEC's phytoplankton and detritus?

As stated in the quoted sentence, near the surface our P is typically newly-produced and thus more similar to phytoplankton than to detritus. We felt the additional process of the transition in BEC from plankton to detritus would make it a poor comparison.

Line 171: Should be 16/117 mol N/mol C (units), correct?

Yes, we have added the units.

Lines 172-175: But then why compare the model against WOA nitrate? (See my above comment.)

Now that we have added the BEC nitrate to figure 2, we hope it is clear that BEC misses some of the observed patterns. We felt it best to use observations for N and model data for P in part due to the relative sparseness of in-situ phytoplankton concentration observations; the text has been updated here to better express this idea.

Line 195: Eqn. number is missing.

The equation number has been added, thank you.

Fig. 2: Why use units of gN/m3 Is this averaged over the upper 100m? Shouldn't it rather be mmol N/m3?

Yes, these are properly mmolN/m3, this has been fixed. These are the surface fields, not the mean over the top 100m.

Line 240-241: "As μ0, the maximum growth rate, is held constant," - Does this mean "As μ0, the maximum growth rate, does not vary in space or time"?

Yes. We have updated the text as you have written here.

Eqn 6: Shouldn't the sign of the third term on the RHS be negative?

No. This equation has been updated to include the detailed definition of each delta term; a more detailed response was written to reviewer 1's first point.

Fig 3: Again, why gN/m3? Also, I would suggest to label the panels with (a) - (b) -(c) ... in reading direction, as for Fig 4, to avoid confusion.

The panels are labeled in the order they are referred to in the text, to assist in following along. The nutrient units were incorrect, and have been fixed to mmol N/m3.

Line 262-263: "but the reduction percentages highlight that the changes are somewhat insensitive to the varied light- and nutrient-limitation choices." - This is unclear to me: does this refer to panel (d) of Figure 3? They seem to decline considerably (by more than 10%) between tau_bio=4.5 and tau_bio =

160.5. But I might be miistaken. Perhaps using a % scale (as in panel (b)) would be better to compare the two diagnostics?

We show absolute values of the change in mean N concentration rather than percent change because the absolute values do have a notable range that varies with tau_bio while the percentages do not. We have rewritten this sentence: Both initial concentrations and absolute reductions are smaller for shorter tau_bio, but the small range of reduction percentages (15-22%) highlight that the changes are somewhat insensitive to the varied light- and nutrient-limitation choices.

Lines 288-298: Here, and elsewhere, the correlations are mentioned. I would suggest to combine those in one or two tables, and highlight those that are significant.

Yes, we have created table 2 for this purpose.

Lines 314-315: "while the global differences are consistent with the chosen parameters in that the faster cases, which have a higher nutrient utilization, have new production more correlated with the reductions of near-surface nutrient and its vertical supply." - What does "higher nutrient utilization" mean? More production? Could the stronger correlation between new production and nutrient supply and concentrations be caused by the higher k_N of the fast case (with k_N=1 mmol N/m3 and nutrient concentrations in that range for large parts of the tropical and subtropical ocean)?

Your intuition has done well interpreting this point. We have adjusted the middle clause to read "which have higher production and lower near-surface nutrient". The correlation cannot be explained by k_N alone; we think tau_bio is a better explanation.

Line 357: "The"

Yes, thank you.

Line 375: "austral winter and spring"?

Yes, we have added 'austral'.

Lines 383-393: This paragraph introduces (the effects of) KPP, Redi and GM. These should be explained in more detail, as not every reader is familiar with them.

We adjusted this discussion to use more-accessible language rather than the jargon of the names of these parameterizations. The sentence introducing the parameterized mixing now will read: "These fluxes are comprised of downward advective fluxes, upward fluxes from parameterized vertical mixing (Large et al., 1997), and upward fluxes from vertical effects of parameterized along-isopycnal mixing (Gent et al 1995)." These adjustments will also be made to the labels of figure 9ab.

Fig. 8-11: To my opinion, nutrient supply should not be given in gC/m2, which somehow weird. If the unit should be comparable to new production, I'd rather suggest to give both in units of mmol N (which seems to be the basic unit of the model).

We appreciate your insight. We are using gC as the production unit throughout in order to have the rates be easily referenced to other works in the mind of the readers. While this is odd for nutrient supply, we do indeed want the units to match, as you surmised. We have decided against remaking these figures.

Line 403-404: "The Arctic region is defined by being within the Arctic circle (north of 66.5N), or having at least one day per year with no incoming solar radiation." - The logics of this sentence is not clear to *me: Does this mean "either-or" (i.e., even including regions south of 66.5, if they have at least one day without insolation), or is the insolation criterion a consequence of phi>66.5?*

These criteria are equivalent, and ‘equivalently’ has been added after ‘or’.

Line 408: "Early century" - Do you mean the year 2000?

Yes, this sentence now begins “Seasonal cycles in the 2000s”.

Fig 9: "Profiles of change" - Year 2100-year2000?

Yes, and we have added this to the caption.

Fig 10: Is the magenta line really Delta Q Delta L, and not rather L Delta Q?

Yes, it is deltaQdeltaL.

Fig 7,8,10,11: What are the units of the denominators for normalization (0.055, 0.14, etc.)?

These normalizations are values of QL, which is unitless (as are both Q and L). This is now noted in the text, Section 2.3: “As mu_0, the maximum growth rate, does not vary in space or time, we can examine simply the nutrient availability, Q, and the light availability, L, both of which are nondimensional and have values between zero and one, as does their product... These difference terms are also nondimensional and have values between negative one and one.” We have also noted that these are nondimensional terms in the captions for figures 6 and 7.