Review 1 of Elsworth et al. (2022) "Anthropogenic climate change drives non-stationary phytoplankton variance", submitted to Biogeosciences.

The authors investigated future changes in interannual variability of phytoplankton carbon biomass by using the CESM1 large ensemble simulation. Their results seem to indicate highly spatially heterogeneous response of interannual variability in the biomass to the global warming by the end of the 21st century and relatively important contribution from changes in "top-down" control of phytoplankton growth.

I totally agree with the authors' initial point that, especially in the context of ocean biogeochemistry, future changes in variabilities have not been paid much attentions compared to those in the climate mean states, although these are critically important on decisions of mitigation and adaptation policy. I don't think that the study has no potential for being a step to help our understanding about the ocean ecosystem (including from lower to higher trophic levels) response to climate changes. However, I can not recommend the editor to publish the current manuscript to BioGeosciences, because of the following two concerns: (1) Model validity and (2) authors' conceptual misunderstanding about MLR analysis.

(1) Model validity: The model ability to represent observed variability is critical on judging if projected future changes are valid. The author must show 1) "additional model-observation comparisons" with choosing the variables which are relevant to this study's focus (i.e., phytoplankton biomass) and 2) "evidence" on which results projected from the model with biases can be considered conclusive.

The authors showed the model-observation comparison of variability of annual mean phytoplankton carbon biomass (main target of this study) in Figure 1 and mentioned "Similar spatial patterns (to observations) are apparent (in the model)" in L139. But, for me, obviously, the model special pattern has different spatial characteristics from the observations. In the high latitudes, the observation shows the maximum variance in the most pole-side latitudes in the both hemispheres, while the model shows the variance maximum in somehow equator latitude around 50-50N and 50-60S. In the equator, although there is a strong latitudinal maximum along the equator in the model, no such structure can be seen in the observation, rather higher variabilities are observed in the off-equatorial regions. Moreover, model overestimations of the observed variability can exceed 200% in the equator and the subpolar North Atlantic.

The author also showed the model validity by comparing global internal variance in chlorophyll between the model ensemble and the observational ensemble (Figure S2). However, this study's focus is the phytoplankton carbon biomass, not chlorophyll, and these two can have very different spectrums. I think that model-observation comparisons in the biomass are more suitable for the purpose and the author should assess the model in the regional scale (not global), given the spatial heterogeneous response of the biomass.

We agree that a more ideal model validation study would include an evaluation of the modeled internal variability in phytoplankton biomass as compared to that of observations. Unfortunately, such an evaluation is not possible with the tools we have on hand. As such, we opted to evaluate the modeled phytoplankton variability using two approaches: (a) assess the temporal variability in

modeled ocean phytoplankton carbon biomass as compared to that inferred from satellite observations of chlorophyll, backscattering coefficients, and phytoplankton absorption, and (b) assess the internal variability (ensemble spread) in modeled phytoplankton using a synthetic ensemble of global ocean chlorophyll concentrations derived from satellites. While neither is the perfect evaluation tool, when taken together, they provide a holistic view of modeled variability as compared to observed variability in phytoplankton. We agree with the reviewer with regard to their point about (a) – indeed, there are some regions in the ocean where there is a substantial mismatch in the temporal variability represented by the model and that estimated from observations (Figure 1). In response to this comment, we now include and make reference to a supplemental table quantifying the temporal standard deviation in each of the 11 ecological provinces (Table S1). While the differences can be quite large in some regions, we note that this is an evaluation of temporal variability (not internal variability, the focus of our paper results), and that the estimates from the satellites are derived from a collection of data products and may also have biases. We include a new paragraph in the methods section describing these caveats.

Biome	Name	$\sigma_{temporal}, model$	$\sigma_{temporal}, obs$
ARC	Arctic	2.7	4.5
ASP	Atlantic subpolar	9.7	4.1
NAS	North Atlantic subtropical gyre	2.8	1.7
EQA	Equatorial Atlantic	1.3	1.4
SAS	South Atlantic subtropical gyre	1.1	1.2
IND	Indian Ocean	0.81	2.0
SAP	subarctic Pacific	3.7	4.0
NPS	North Pacific subtropical gyre	0.85	1.5
EQP	Equatorial Pacific	5.8	1.8
SPS	South Pacific subtropical gyre	0.60	0.93
SOC	Southern Ocean	2.7	2.7

Table S1: " The temporal standard deviation of phytoplankton biomass ($\sigma_{temporal}$) for ensemble member 1 of the CESM1-LE and the satellite observations (Bellacicco et al., 2020) from 1998 to 2019 averaged across the 11 ecological provinces defined in Vichi et al. (2011) and Tagliabue et al. (2021). Units are mg m^{-3}."

Line 163: "Some regions of the global ocean display a substantial mismatch in temporal variability between the model and that estimated from observations (Figure 1, Table S1). While the differences can be quite large in some regions, we note that this is an evaluation of temporal variability (rather than internal variability, the focus of this study), and that estimates from the satellite product derive from a collection of data products which may also display biases (Table S1)."

Line 176: "Taken together, our model validation exercises demonstrate that the model tends to overestimate both the temporal (interannual) variability and the internal variability in phytoplankton as compared to satellite observations on both global and regional scales. Thus, we must interpret our findings with this caveat in mind. The change in variance that we model is likely an upper-end estimate."

Line 287: "While the CESM1-LE represents the overall spatial pattern of observed interannual variability in phytoplankton carbon, the model overestimates the magnitude of observed interannual and internal variance in phytoplankton on regional scales…"

(2) MLR analysis: The methodology is unclear and it seems wrong.

The author tried to reconstruct the contribution of each driver variable to phytoplankton biomass using the MLR coefficients (equation 3 and 4). However, it obviously failed. As shown in Figure S6, the reconstructed "Total Carbon" is not equal to the sum of the other terms (i.e., equation 3 and 4 are not correct), maybe because of inaccurate MLR coefficients, neglecting offset term or strong multicollinearity between variables (e.g., MLD and SST, SST and Solar).

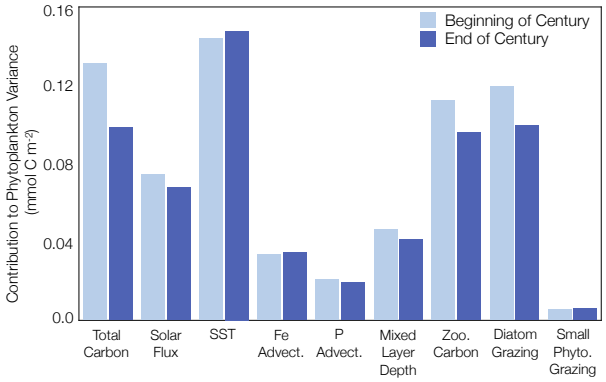
Linear decompositions should be applied for "change/anomaly", not for "climatology (10-year mean)."

Given a function F(X,Y,Z), in general, the first order Taylor expansion is robust only for a small change in the F (ΔF),

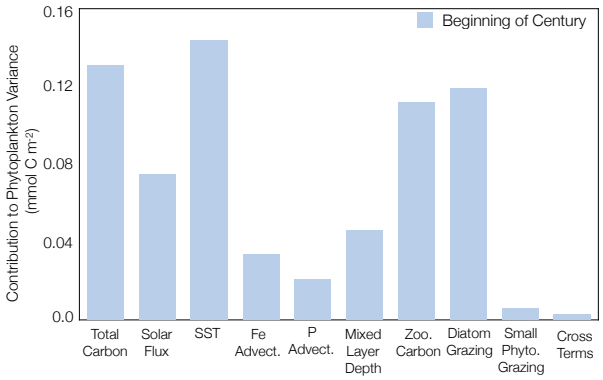
$\Delta F = (\partial F/\partial X)\Delta X+(\partial F/\partial Y)\Delta Y+(\partial F/\partial Z)\Delta Z$ + (Residuals from high-order and cross terms).

The author should apply such analysis for "change" (not "climatology") considering residual terms. As the authors also mentioned, the partial differential coefficients are time-varying. The authors should be able to calculate the coefficients analytically using the model equations of phytoplankton carbon biomass.

Thank you for bringing this to our attention. We agree and have removed reference to the mean state analysis. We have removed the mean state MLR analysis subplots in Figures S5a and S6a (now labeled S7 and S3 respectively), as well as the regional mean state analysis shown in Figure S7. This helps draw the reader's attention to the changes in internal variance, the main focus of our paper.



"Figure S7: Reconstructed global changes in the contribution of each variable to changes in phytoplankton biomass variance across the RCP8.5 forcing scenario (2006 to 2100). The beginning of the century is shown in light blue and the end of the century is shown in dark blue. The change in variance is calculated using averages across the first (2006 to 2016) and last (2090 to 2100) decades of the RCP8.5 forcing scenario."



"Figure S3: Reconstructed contributions of each variable to phytoplankton biomass variance at the beginning of the RCP8.5 forcing scenario (2006 to 2016). The contribution of cross terms to the MLR reconstruction is shown in the rightmost bar. The variance is calculated using the average across the first (2006 to 2016) decade of the RCP8.5 forcing scenario."

Specific comments:

All line numbers refer to those in the revised manuscript.

L29-: Any reference? And, does this mean the CESM1 shows the opposite response of the high-latitude biomass to the global warming? (Figure 3a shows increase in biomass only in the sea-ice biome).

Thank you for this suggestion. We have included the following citations in the text for clarity.

Steinacher, M., Joos, F., Frölicher, T., Bopp, L., Cadule, P., Cocco, V., Doney, S., Gehlen, M., Lindsay, K., Moore, J., Schneider, B., 535 and Segschneider, J.: Projected 21st century decrease in marine productivity: A multi-model analysis, Biogeosciences, 7, 979–1005, https://doi.org/10.5194/bg-7-979-2010, 2010.

Bopp, L., Resplandy, L., Orr, J., Doney, S., Dunne, J., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and Vichi, M.: Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models, Biogeosciences, 10, 6225–6245, 350 https://doi.org/10.5194/bg-10-6225-2013, 2013.

Line 28: "A majority of Earth System Models (ESMs) project an increase in phytoplankton abundance in the high latitude ocean as light limitation is alleviated by stratification, increasing temperature stimulates photosynthesis, and sea ice cover declines (Steinacher et al., 2010; Bopp et al., 2013)."

The response of the CESM1-LE is consistent with other CMIP5 simulations which show increasing phytoplankton biomass in the high latitude regions of the Arctic and Southern Oceans with warming. However, the exact distribution of increasing phytoplankton biomass varies between different models. In some models, the increase in phytoplankton biomass is confined to

the sea ice biomes (as seen in the CESM1-LE), while in others there is a broader distribution throughout parts of the Northern Atlantic and Southern Ocean.

L49-: Please elaborate "Clarifying how variance in phytoplankton biomass may be changing over long time scales with climate change is important for fisheries management, especially at regional scales." What kind of impacts on fisheries by changing in variance in Phytoplankton biomass can one expect?

Understanding how variance in phytoplankton biomass is changing in a warming climate is important because it affects our ability to make accurate near-term predictions of fisheries production. We have clarified this point in the text.

Line 48: "… Clarifying how variance in phytoplankton biomass may be changing over long time scales with climate change is important for fisheries management, especially at regional scales, as it affects our ability to make accurate near-term predictions of fisheries production."

L82-85: I could not understand clearly. Please clarify with showing equations.

This is an excellent point. The existing Equation (5) in the Supplemental Information describes the parameterization of zooplankton grazing rate in the CESM1-LE. We have referenced this Equation (5) in the text to clarify.

Line 82: "Grazing rate is computed using a Holling Type III (sigmoidal) relationship and is a function of both prey density and temperature (Figure S1, Equation 5)."

L94-97: The author's description of the experimental setting of CESM1 large ensemble is inaccurate. Please describe it correctly.

We have changed the text to read:

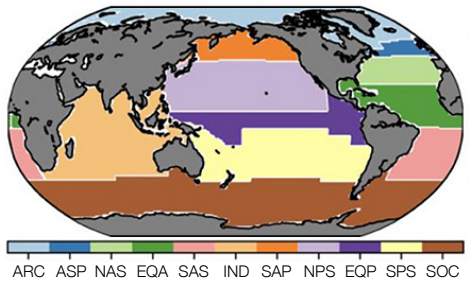
Line 95: "The CESM1-LE simulates the evolution of the climate system from 1920 to 2100 with multiple ensemble members, each expressing different phasing of internal climate variability while responding to a shared external forcing prescription (Kay et al. 2015)."

L99-101: Show figure as an example.

We have removed this sentence for clarity.

L118-120: Please provide the map of the aggregated biological provinces used in this study as supplementally figure or superpose the biome boundary on the main figures (e.g., Figure 3).

This is a great suggestion! We have included a map of the aggregated biomes as cited in Tagliabue et al., 2021 in the Supplemental Information as Figure S2. We have also referenced this map in the text.

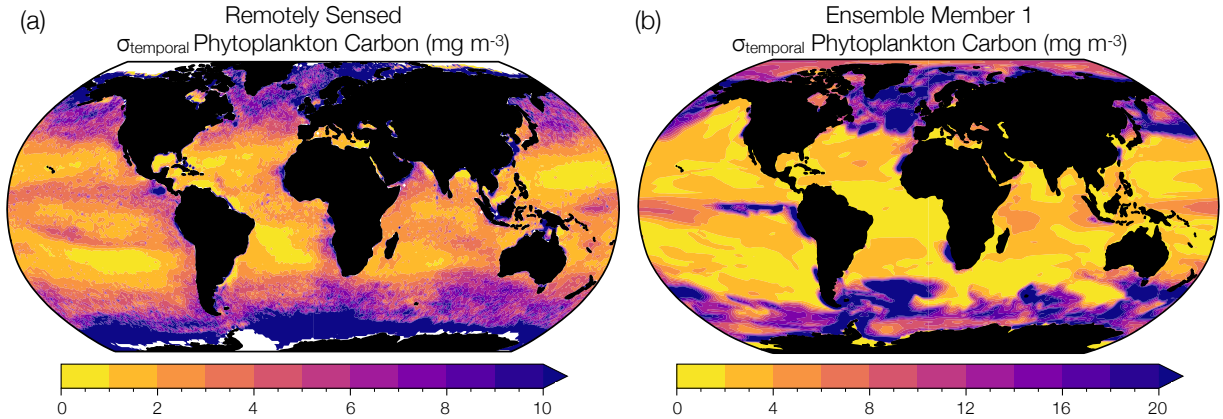


"Figure S2: The 11 ecological provinces defined in Tagliabue et al., (2021) and Vichi et al., (2011). Provinces were aggregated using multivariate statistical analysis of physical (i.e., salinity, temperature, mixed layer depth) and biological (i.e., chlorophyll concentration) ocean parameters to group ocean regions with similar physical and environmental conditions. Figure adapted from Tagliabue et al., (2021)."

Line 113: "We classified the marine environmental into 11 ecologically cohesive biomes as in Tagliabue et al., (2021) and Vichi et al., (2011) (Figure S2), which are a consolidation of the 38 ecological regions defined in Longhurst et al., (2007)."

Figure 1: Please use the same colormap and same value range for fair comparison. And, it is better to show the ensemble mean of the σtemporal with a rank analysis (to show whether the observational σ is inside the ensemble spread grid by grid).

Thank you for this suggestion. We have revised Figure 1 to display both maps on the same colormap. However, when the colormaps are the same range it is not possible to see spatial nuances in the remotely sensed plot (Figure 1a). We now note for the reader that we are using different magnitude colorbars in the figure caption, and we added text to reference Table S1 showing the temporal standard deviation differences between model and observation in each province.



"Figure 1: Comparison between observed and modeled phytoplankton biomass interannual variability. (a) Temporal standard deviation in annual mean phytoplankton carbon concentration (mg m^{-3}) reconstructed from remotely sensed chlorophyll concentrations, backscattering coefficients, and phytoplankton absorption (1998 to 2019) (Bellacicco et al., 2020). (b) Temporal

standard deviation in annual mean phytoplankton carbon concentration (mg m^{-3}) simulated by ensemble member 1 of the CESM1-LE over the same observational period (1998 to 2019). Note the different magnitudes on the colorbars."

Biome	Name	$\sigma_{temporal}, model$	$\sigma_{temporal}, obs$
ARC	Arctic	2.7	4.5
ASP	Atlantic subpolar	9.7	4.1
NAS	North Atlantic subtropical gyre	2.8	1.7
EQA	Equatorial Atlantic	1.3	1.4
SAS	South Atlantic subtropical gyre	1.1	1.2
IND	Indian Ocean	0.81	2.0
SAP	subarctic Pacific	3.7	4.0
NPS	North Pacific subtropical gyre	0.85	1.5
EQP	Equatorial Pacific	5.8	1.8
SPS	South Pacific subtropical gyre	0.60	0.93
SOC	Southern Ocean	2.7	2.7

Table S1: " The temporal standard deviation of phytoplankton biomass ($\sigma_{temporal}$) for ensemble member 1 of the CESM1-LE and the satellite observations (Bellacicco et al., 2020) from 1998 to 2019 averaged across the 11 ecological provinces defined in Vichi et al. (2011) and Tagliabue et al. (2021). Units are mg m^{-3}."

Line 163: "Some regions of the global ocean display a substantial mismatch in temporal variability between the model and that estimated from observations (Figure 1, Table S1). While the differences can be quite large in some regions, we note that this is an evaluation of temporal variability (rather than internal variability, the focus of this study), and that estimates from the satellite product derive from a collection of data products which may also display biases (Table S1)."

Line 176: "Taken together, our model validation exercises demonstrate that the model tends to overestimate both the temporal (interannual) variability and the internal variability in phytoplankton as compared to satellite observations on both global and regional scales. Thus, we must interpret our findings with this caveat in mind. The change in variance that we model is likely an upper-end estimate."

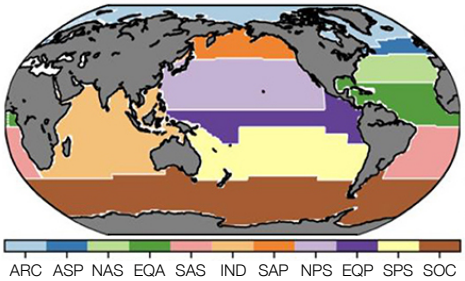
Line 287: "While the CESM1-LE represents the overall spatial pattern of observed interannual variability in phytoplankton carbon, the model overestimates the magnitude of observed interannual and internal variance in phytoplankton on regional scales…"

L179: Figure 2d?

We believe this to be a mistake on the part of the reviewer. We reference Figure 2b (not 2d) to illustrate the changing coefficient of variance in biomass carbon concentration. There is no Figure 2d.

L213-216: Which regions did the author chose? Please show these on map.

We have included a map of the aggregated biomes as cited in Tagliabue et al., 2021 in the Supplemental Information as Figure S2. We have also referenced this map in the text.



"Figure S2: The 11 ecological provinces defined in Tagliabue et al., (2021) and Vichi et al., (2011). Provinces were aggregated using multivariate statistical analysis of physical (i.e., salinity, temperature, mixed layer depth) and biological (i.e., chlorophyll concentration) ocean parameters to group ocean regions with similar physical and environmental conditions. Figure adapted from Tagliabue et al., (2021)."

Line 113: "We classified the marine environmental into 11 ecologically cohesive biomes as in Tagliabue et al., (2021) and Vichi et al., (2011) (Figure S2), which are a consolidation of the 38 ecological regions defined in Longhurst et al., (2007)."

Technical corrections:

I don't list any small technical/editorial corrections at this time. Above-mentioned conceptual/major comments should be addressed or fixed by the authors before going into details.

Review 2 of Elsworth et al. (2022) "Anthropogenic climate change drives non-stationary phytoplankton variance", submitted to Biogeosciences.

The manuscript "Anthropogenic climate change drives non-stationary phytoplankton variance", summarizes projected changes in global and regional phytoplankton variability using the NCAR CESM1 Large Ensemble under a high emissions scenario. The authors explore the key drivers of declining phytoplankton variability, highlighting the importance of top-down, zooplankton grazing in potentially driving future phytoplankton response.

Generally, the article concisely represents its findings but there are several points of clarification I would recommend. In particular, the use of specific statistical terminology could be more accurate. Multiple times throughout the text, the term "variance" is used when, I think, "variability" is intended. In many cases this "variability" is being assessed via the standard deviation of the large ensemble members which is similar to the variance by not the same. Additionally, I am not proficient in MLR, but the comments made in the prior review are troubling especially considering the results are key to the paper's conclusions regarding top-down controls but these results seem underrepresented in the primary manuscript text. I've included several additional minor comments and suggestions below pertaining to clarity and organization.

We have clarified the use of "variance" and "standard deviation" in the text.

We have addressed Reviewer 1's comments regarding the MLR approach in our response above.

Specific Comments and Suggestions:

Lines 49-52: *Clarifying how variance in phytoplankton biomass may be changing over long time scales with climate change is important for fisheries management, especially at regional scales. Near- term predictions of phytoplankton biomass may also benefit from knowledge of the projected magnitude of internal variability, as the chaotic nature of internal variability hampers near-term predictions (Meehl et al., 2009, 2014).*

I think it's worth noting that the internal variability quantified using a large ensemble is Internal variability specific to the model and indicative of our uncertainty that results from its simplified representations of the real world processes and numerics. It doesn't necessarily have any bearing on real world manifestations of variability. Its primary utility to management and fisheries is in guiding our level of confidence in disentangling model signals from the noise.

This is an excellent point. We have included additional text to clarify this point.

Line 53: "… However, modeled internal variability may differ from that observed in the real world."

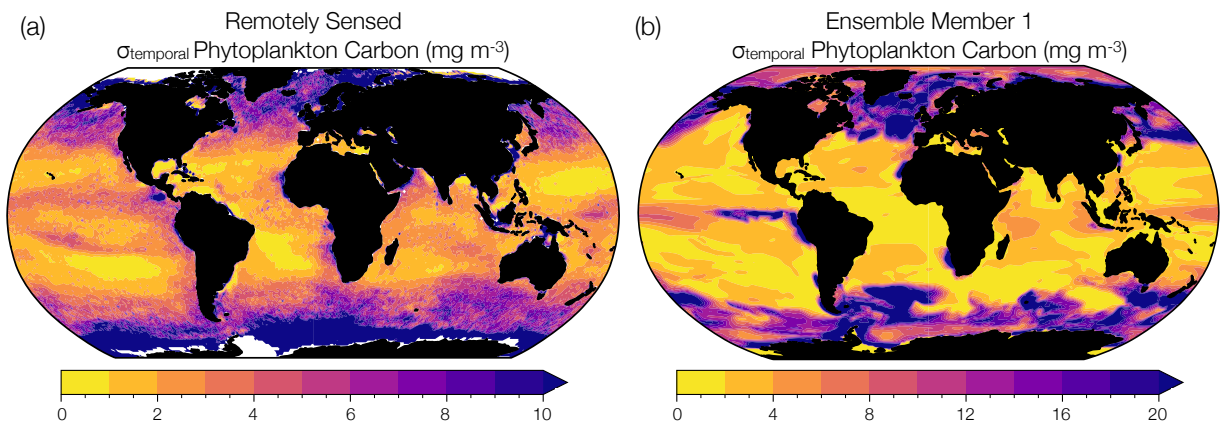
Lines 103-104: *Six CESM1-LE members had corrupted ocean biogeochemistry*

I'm curious, what does "corrupted ocean biochemistry" mean? it might help to explain what makes an ensemble member usable versus not.

The ocean biogeochemistry output fields of ensemble members 3 though 8 were corrupted during the saving process. Therefore, no information on biogeochemical variables is available for these ensemble members. However, the corruption of these ensembles affected only the biogeochemical output and, thus, other Earth system variables is preserved. Details are referenced here.

Figure 1. Add units: standard deviation should have the same units as the variable being assessed (i.e., phytoplankton carbon) but none appear in figure 1.

Thank you for bringing this to our attention. We have revised Figure 1 to include the units of mg m^{-3} and have modified the text of the figure caption.



"Figure 1: Comparison between observed and modeled phytoplankton biomass interannual variability. (a) Temporal standard deviation in annual mean phytoplankton carbon concentration (mg m^{-3}) reconstructed from remotely sensed chlorophyll concentrations, backscattering coefficients, and phytoplankton absorption (1998 to 2019) (Bellacicco et al., 2020). (b) Temporal standard deviation in annual mean phytoplankton carbon concentration (mg m^{-3}) simulated by ensemble member 1 of the CESM1-LE over the same observational period (1998 to 2019). Note the different magnitudes on the colorbars."

Lines 121-122: *Internal variability at each location (x,y) is approximated as the standard deviation across ensemble members (EMs) at a given time (t)*

The method described here indicates that the standard deviation is being used to quantify variability. However, throughout the paper, the authors reference the "variance" when I think they mean "variability". This is problematic because "variance" and "standard deviation", while related, are two different values and the way they are interchanged throughout the text is confusing. Please check all instances of "variance" in the paper for intended meaning and replace with "variability" where appropriate. I suggest including a description of the "coefficient of variance" method here, too.

We agree. We have included a description of the coefficient of variance in the methods section.

Line 130: "The coefficient of variance (CoV) is calculated as the standard deviation across the ensemble members divided by the ensemble mean,

$$\text{CoV}(x, y, t) = \frac{\sigma(\text{EM}(x, y, t))}{\text{LE}}\text{"}$$

Lines 142-143: *However, while the model ensemble captures regional patterns of observed variability, the CESM1-LE overestimates the magnitude of observed interannual variability.*

I may be mistaken but it seems this was determined using only a single ensemble member - is it appropriate for conclusions to be drawn for the full ensemble when only considering one ensemble member?

We tested the temporal standard deviation for all ensemble members and report the (small) difference for the reviewer. The figure below shows the temporal standard deviation of 34 ensemble members of the CESM1-LE across the observational period (1998 to 2019). The difference in temporal standard deviation between ensemble members is small over this period.

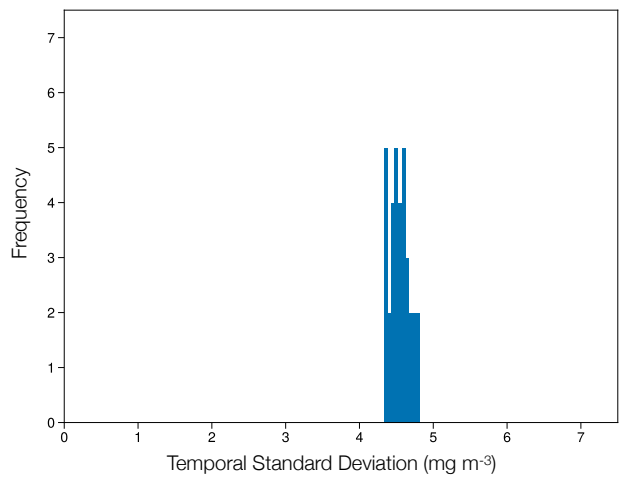


Figure: Histogram of global average temporal standard deviation (mg m^{-3}) for each of the 34 ensemble members of the CESM1-LE over the observational window (1998 to 2019).

Lines 147: *A synthetic ensemble is a novel technique*

I don't think this technique can be called "novel" if it appears in two prior references

This is a valid point. We have removed the word "novel" from the text on Line 147 and Line 590.

Lines 149-151: *Compared to the internal variability over the observational period (2002 to 2020) (purple circle, (Figure S2), the model ensemble slightly overestimates the magnitude of internal variability in chlorophyll observed in the real world.*

This seems like a result/ should appear in the result section. Also, it makes an assessment of the ensemble as a whole, but isn't it still based on the results from the single ensemble member? If not, this was a point of confusion for me, and I suggest clarifying.

This is our second model validation exercise, and thus we opted to keep it in the methods section. We note for the reviewer that this is an assessment of internal variability (ensemble spread), as compared to a synthetic ensemble generated from observations. In response to this comment, we added a paragraph describing the interpretation of the results from the interannual and internal variance validation exercises to the methods section.

Line 163: "Some regions of the global ocean display a substantial mismatch in temporal variability between the model and that estimated from observations (Figure 1, Table S1). While the differences can be quite large in some regions, we note that this is an evaluation of temporal variability (rather than internal variability, the focus of this study), and that estimates from the satellite product derive from a collection of data products which may also display biases (Table S1)."

Line 176: "Taken together, our model validation exercises demonstrate that the model tends to overestimate both the temporal (interannual) variability and the internal variability in phytoplankton as compared to satellite observations on both global and regional scales. Thus, we must interpret our findings with this caveat in mind. The change in variance that we model is likely an upper-end estimate."

Lines 153-154: *Annually averaged, global mean, upper-ocean (top 150m) integrated phytoplankton biomass across the model ensemble decreases from 76.1 mmol C m-2 to 66.2 mmol C m-2.*

It's not clear what timeframes these values represent. Is it 2006 vs. 2100? If so, it seems that such a narrow, 1-year window would risk aliasing higher frequency variability and potentially under- or overestimate the change in mean state. This is somewhat compensated for by the size of the ensemble but differs from the 10-year averaging described later in Line 223

This is an excellent point. The decline in phytoplankton biomass is calculated as the difference between the average of the first (1920 to 1930) and last (2090 to 2100) decades across the historical and the RCP8.5 forcing scenario. We have clarified the time windows used in this calculation in the manuscript text.

Line 181: "The change in the mean is calculated as the difference between the first (1920 to 1930) and last (2090 to 2100) decades across the historical and the RCP8.5 forcing scenario."

Lines 177-178: *we calculated the coefficient of variance as the standard deviation across the ensemble members for a given year (ensemble spread) divided by the ensemble mean.*

I suggest including this description in the methods section rather than the results.

Thank you for this suggestion. We have included a description of the coefficient of variance in the methods section.

Line 130: "The coefficient of variance (CoV) is calculated as the standard deviation across the ensemble members divided by the ensemble mean,

$$\text{CoV}(x, y, t) = \frac{\sigma(\text{EM}(x, y, t))}{\text{LE}}\text{"}$$

Lines 178-180: *Figure 2b illustrates the change in the coefficient of variance from the historical period through the RCP8.5 forcing scenario (1920 to 2100).*

The results seem to jump from Figure 2a, to Figure 3, then back to 2b which is a bit confusing.

This is an excellent point. We have modified the text to enhance the flow of the manuscript.

Line 180: *The coefficient of variance is relatively constant across the historical period (1920 to 2005), and then significantly declines by ~20% from 2006-2100.*

I'm not sure I agree with the assessment that the coefficient of variance is relatively constant across the historical period. 1920-1980 appears to have a positive trend with a range of about 6.1 to 7.3, which appears similar to the range of the time period covered by the dashed line in Figure 2b. I suggest testing the significance of the 1920-1980 trend. Also, could the drop in coefficient of variance instead be explained by temporal distance from the perturbation that differentiates the ensemble members? If the 34 ensemble members differ in initial air temperature conditions, would the spread perhaps be expected to decrease as the simulation integrates further away from that initial discrepancy (i.e., solutions start to converge)?

This is a good point. We have tested the significance of the 1920 to 1980 trend and find that it is not significant.

However, the decrease in the coefficient of variance over the course of the simulation is not due to an increase in the time since the ensemble members were perturbed. This has been demonstrated in the study "An Ensemble Covariance Framework for Quantifying Forced Climate Variability and Its Time of Emergence" published by Yettella et al., 2018. This is illustrated by different responses of ocean and land variance over the 21st century, with ocean variance declining and land variance increasing over time.

Lines 190-193: *From 2006 to 2100, the coefficient of variance decreases by 3.3 x 10-5 yr-1 in the CESM1-LE, 2.0x10-4 yr in the MPI-ESM-LR1, 5.2x10-5 yr-1 in the CanESM2, and 3.9 x10-4 yr-1 in the GFDL-ESM2M. These declines are statistically significant in all model ensembles with the exception of the MPI-ESM-LR1 (Figure S2).*

It's not clear how these values across models are calculated, whether the end points of the time series or a range of years - the latter would be more appropriate (as done in Line 223) to avoid higher frequency variability and thus under- or overestimating the nature of the change. I also

suggest reporting the specific statistical testing methods in the text if stating that the changes are significant.

Thank you for mentioning this. Changes in the coefficient of variance are calculated using averages of the first (2006 to 2016) and last (2090 to 2100) decades of the RCP8.5 forcing scenario. We have clarified this in the caption of Figure S2 (now Figure S4) and in the manuscript text. Significance of the trends are determined by a t-test with a p-value less than 0.05. We have also included this in the caption of Figure S2 (now Figure S4).

Line 220: "… The change in the coefficient of variance is calculated using averages across the first (2006 to 2016) and last (2090 to 2100) decades of the RCP8.5 forcing scenario."

Figure S2 (now S4): "… Trend significance is determined by a *t*-test with a p-value less than 0.05."

Line 201: *We observe the largest magnitude decline in total phytoplankton carbon variance*

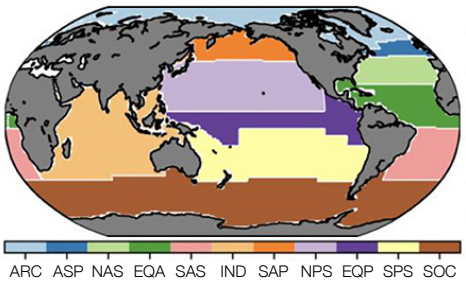
The table is reporting change in standard deviation, not variance. Standard deviation is expressed in the same units as the analyzed variable while variance is reported in the square of those units.

Thank you for clarifying. We have changed "variance" to "standard deviation" in the text to be more precise.

Line 229: "Global changes in total phytoplankton biomass standard deviation are a manifestation of changes in diatom and small phytoplankton variability (Table 1). We observe the largest magnitude decline in total phytoplankton carbon standard deviation in the subpolar Atlantic (ASP) region, where diatom standard deviation declines by ~10 mmol C m^{-2} and small phytoplankton standard deviation declines by ~2 mmol C m^{-2} (Table 1). The CESM1-LE simulates a moderate magnitude decline in total phytoplankton standard deviation in the subarctic Pacific (SAP) region, driven by a decrease in small phytoplankton standard deviation (~2 mmol C m^{-2}) with minor contributions from declines in diatom standard deviation (~1 mmol C m^{-2}) (Table 1). Moderate declines in standard deviation are also simulated in the Arctic (ARC), North Atlantic subtropical gyre (NAS), Southern Ocean (SOC), and Equatorial Pacific (EQP) regions, driven by declines in diatom carbon standard deviation in the SOC region and declines in small phytoplankton variance in the EQP region (Table 1)."

Figure 4: It's not clear what this figure adds to the discussion - it seems to be redundant with information in Figure 5. Perhaps if the outlines of the ecological regions were included?

This is an excellent suggestion. We have included a map of the aggregated biomes as cited in Tagliabue et al., 2021 in the Supplemental Information as Figure S2. We have also referenced this map in the text.



"Figure S2: The 11 ecological provinces defined in Tagliabue et al., (2021) and Vichi et al., (2011). Provinces were aggregated using multivariate statistical analysis of physical (i.e., salinity, temperature, mixed layer depth) and biological (i.e., chlorophyll concentration) ocean parameters to group ocean regions with similar physical and environmental conditions. Figure adapted from Tagliabue et al., (2021)."

Line 113: "We classified the marine environmental into 11 ecologically cohesive biomes as in Tagliabue et al., (2021) and Vichi et al., (2011) (Figure S2), which are a consolidation of the 38 ecological regions defined in Longhurst et al., (2007)."

Lines 219-221: *We quantified the relationship between phytoplankton carbon and the variables which contribute to changing phytoplankton biomass and its internal variability by performing a multiple linear regression (MLR) analysis. The MLR analysis was performed on linearly detrended annual anomalies using the ordinary least squares function of the Python package statsmodels.api*

This and the associated equations seem to belong in the methods section.

Thank you for this suggestion. We have moved this and the associated equations to the methods section.

Line 134: "… We quantified the relationship between phytoplankton carbon and the variables which contribute to changing phytoplankton biomass and its internal variability by performing a multiple linear regression (MLR) analysis. The MLR analysis was performed on linearly detrended annual anomalies using the ordinary least squares function of the Python package statsmodel.api."

Line 274: *…and important global biogeochemical regions…*

What is considered an important biogeochemical region? This seems somewhat vague - I suggest elaborating to be a bit more specific.

We appreciate this comment. An important biogeochemical region is an ocean region characterized by coherent physical and environmental conditions, which support unique marine ecosystems with an outsized role on ocean biogeochemical cycling. We have added a sentence in the text to clarify this point.

Line 123: "… Important biogeochemical regions are those characterized by coherent physical and environmental conditions, which support unique marine ecosystems with an outsized role on global ocean biogeochemistry."

Lines 278-280: *As such, the magnitude of changes in phytoplankton internal variance derived from the model ensemble should be interpreted as an overestimate when considering changes in phytoplankton internal variance driven by anthropogenic warming.*

Again, my impression was that this conclusion was derived from analyzing a single ensemble member which seems insufficient for assessing the entire ensemble.

This conclusion was derived from our second model validation exercise, where we compare the spread across all modeled ensemble members with that of a synthetic ensemble derived from satellite observations.

Review 3 of Elsworth et al. (2022) "Anthropogenic climate change drives non-stationary phytoplankton variance", submitted to Biogeosciences.

In this manuscript, the authors use the Community Earth System Model I Large Ensemble to evaluate the impacts of anthropogenic climate change on long-term variability in phytoplankton distributions within the global ocean. The authors additionally use a multiple linear regression to evaluate the ecological drivers of this change, reporting zooplankton grazing as being a major factor in reducing variability in phytoplankton biomass.

The analysis of earth systems models is well outside my area of expertise. So while the authors' main finding that variance in phytoplankton biomass is anticipated to decrease in the future ocean seems informative from my perspective, I defer to the first reviewer's comments regarding best practices in model interpretation. I was interested to see the multiple linear regression results, which seem to highlight a particularly strong coupling between phytoplankton biomass and grazing in model results. However, by the authors' admission on L265, it does not seem possible to establish cause and effect regarding the nature of this interaction. With this, it seems like an overstatement to suggest (as in the abstract and elsewhere) that these results provide evidence for grazing-driven declines in phytoplankton biomass.

More importantly, insufficient documentation is provided for the reader to interpret the MLR results. Critically, it is not immediately clear from the text how contributions to phytoplankton/diatom variance were calculated. Equations should be provided, and associated details on the MLR analysis should be moved to the methods section to make this information easier to locate in the manuscript. Moreover, the MLR results themselves seem insufficiently documented. No details are provided on the overall model fit nor on uncertainties associated with the MLR coefficients. The relationship between the parameters 17quationns 3 and 4 and the larger set of parameters included in figure 5 is unclear as well.

The discussion should also be expanded to provide more context on the authors' interpretation of these results. Altogether, even after reading the manuscript several times, I'm not sure why the results shouldn't be interpreted as a weakening of top-down control in the future ocean (with the decrease in contributions to phytoplankton biomass variance due to grazing in Figure 5 reflecting a reduced coupling of phytoplankton biomass and grazing and, by extension, a strengthening of bottom-up controls). If this interpretation is beyond what can be determined based on the analysis (for instance because of large uncertainties in coefficient errors), this is not evident from the information provided.

Without this information on the MLR results, it is impossible to critically evaluate some of the the manuscript's main conclusions. With this, and in light of the comments made by the first reviewer regarding issues with the authors' analysis of the CESM results, I cannot recommend this manuscript for publication without major revisions. A few specific comments are provided below.

We agree with the reviewer. We have made multiple changes to the wording in the manuscript to address this point. These modifications are listed below.

We have changed the text on Line 8: "In these high-latitude regions, bottom-up controls (e.g., light, nutrients) have only a small effect on biomass variance. Rather, the declining internal variance in phytoplankton biomass co-occurs with a reduction in zooplankton grazing variability. Similar patterns emerge in the biogeochemically critical regions of the Southern Ocean and the Equatorial Pacific."

We have changed all wording about "driver / drivers" to "contribution / contributions" throughout the manuscript.

We have modified the topic sentence of the paragraph Line 9: "The declining internal variance in phytoplankton biomass co-occurs with a reduction in zooplankton grazing variability".

We have changed the language on Line 270: "The declining internal variance in phytoplankton biomass co-occurs with a reduction in zooplankton grazing variability."

We have modified the text on Line 285: "Statistical analysis of these specific regions reveal the decline in phytoplankton biomass variance to co-occur with a reduction in zooplankton grazing variability, consistent with previous studies (Bopp et al., 2001; Laufkötter et al., 2015)."

We have changed Line 289: "our study demonstrates a strong connectivity between phytoplankton and zooplankton grazing variance…."

We have changed Line 300: "Our study demonstrates a strong connectivity between phytoplankton and zooplankton grazing variance in the subpolar North Atlantic and the subpolar North Pacific."

We have clarified Line 318: "However, our regional analyses suggest that both phytoplankton and zooplankton grazing variance are likely to change with anthropogenic warming."

Specific comments:

L114 – 115 – A quick review of the method used in Tagliabue et al. would be useful here. What were the multivariate statistical methods used? How were they applied? A map of the biomes would be informative as well.

This is an excellent suggestion. We've included a description of the methods used by Vichi et al. 2011.

Line 113: "We classified the marine environment into 11 ecological cohesive biomes as in Tagliabue et al., 2021 and Vichi et al., 2011, which are a consolidation of the 38 ecological regions defined in Longhurst et al., 2007. The provinces were aggregated using multivariate statistical analysis of physical (i.e., salinity, temperature, mixed layer depth) and biological (i.e., chlorophyll concentration) ocean parameters to group ocean regions with similar physical and environmental conditions (Vichi et al., 2011). Analyses were performed by randomly selecting from a combination of model and observational datasets and testing for statistical significance using analysis of similarities (ANOSIM) (Vichi et al., 2011)."

L159 – 161 - This text feels more appropriate for conclusion/discussion.

We agree. We have removed this text from the manuscript.

"In the North Atlantic subpolar gyre, the phytoplankton biomass declines by 40-50% of its mean (Figures 3a, S3a)."

L215 - FAO citation and the associated reference seem to be improperly formatted

Thank you for this comment. We have reformatted the FAO citation in the references. The references is cited parenthetically as (*FAO*, *2020*). We will ask the editorial staff for clarification if/when the manuscript is accepted.

"FAO. 2020. *The State of World Fisheries and Aquaculture 2020*. *Sustainability in action*. Rome."

L219 – 234 - This text feels more appropriate for the methods section

Thank you for this suggestion. We have moved this text to the methods section.

L289 – 291 - Is this conclusion inconsistent with the disclaimer provided at L264 – 266?

Equations 3 & 4 — Why are the terms in the equations (e.g., Solar, SST, Nutrient, etc.), different from those included in figure 5? Were the equations in the text just providing a summary of the actual equations used? If so, this should be made explicitly clear, with some description of all the variables included.

Yes, the terms in the equations (e.g., Solar, SST, Nutrient, etc.) are the same as those included in Figure 5. We have added the terms parenthetically below the variable names in Figure 5 to clarify this point.

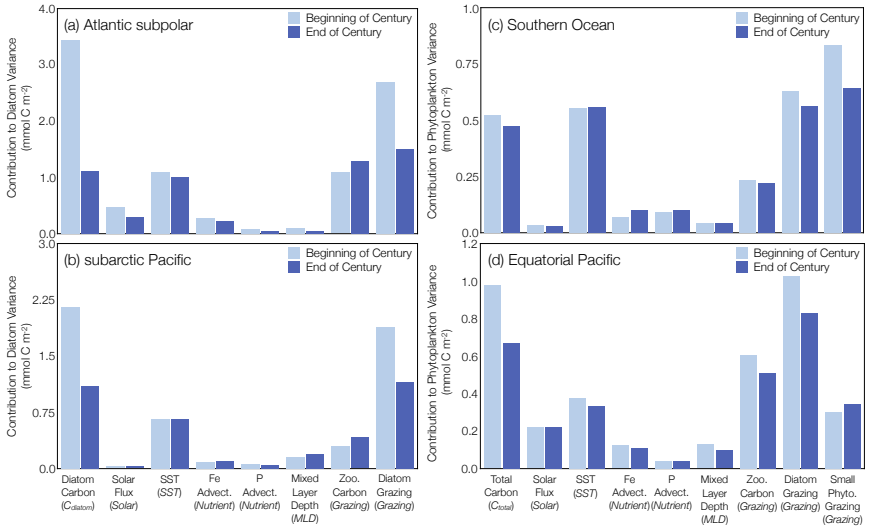
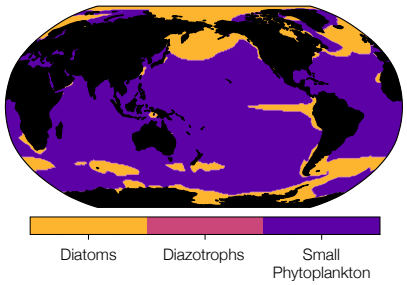


Figure 5: “Reconstructed changes in the contribution of each variable to phytoplankton biomass standard deviation across the RCP8.5 forcing scenario (2006 to 2100) with the beginning of the century shown in light blue and the end of the century shown in dark blue. Marine ecological regions are defined in Tagliabue et al. (2021). Regions were selected which aligned with the highest fisheries catch in the (a) Atlantic and (b) Pacific basins and the biogeochemically important (c) Southern Ocean and (d) Equatorial Pacific regions. The dominant phytoplankton functional type is considered in each region. In regions with a mixed ecological assemblage, total phytoplankton carbon is considered. The change in the coefficient of variance is calculated using averages across the first (2006 to 2016) and last (2090 to 2100) decades of the RCP8.5 forcing scenario.”

Figure 4 — Minor tick marks not necessary on color scale; difficult to see regions dominated by diazotrophs. Maybe use color palette with more contrast?

Thank you for this suggestion. We have removed unnecessary tick marks from the color scale. Diazotrophs do not dominate in any regions of the global ocean and are not visible on Figure 4 for this reason.



“Figure 4: Distribution of the dominant phytoplankton functional type in biomass carbon averaged across the RCP8.5 forcing scenario (2006 to 2100). The CESM1-LE simulates three phytoplankton functional types: diatoms, diazotrophs, and small phytoplankton. Regions where diatoms dominate are shown in yellow, regions where diazotrophs dominate are shown in pink, and regions where small phytoplankton dominate are shown in purple.”

Figure 5 —Note inconsistent capitalization of biomes in subplots; Are units correctly labeled? Are the units for "contribution to phytoplankton/diatom variance" really mmol C m^{-2}? On a related note, where did the values on the Y axis come from? Based on the axis label they don't correspond to the MLR coefficients, but I didn't see any details in the text.

Thank you for clarifying. We have changed the text in the figure caption to clarify that we show the phytoplankton biomass standard deviation.

We follow the province labels set forth in Tagliabue et al., 2021 in both Table 1 and Figure 5. Proper nouns are capitalized (e.g., Equatorial Pacific, Southern Ocean) while adjectives are lowercase (e.g., Atlantic subpolar, South Pacific subtropical gyre).

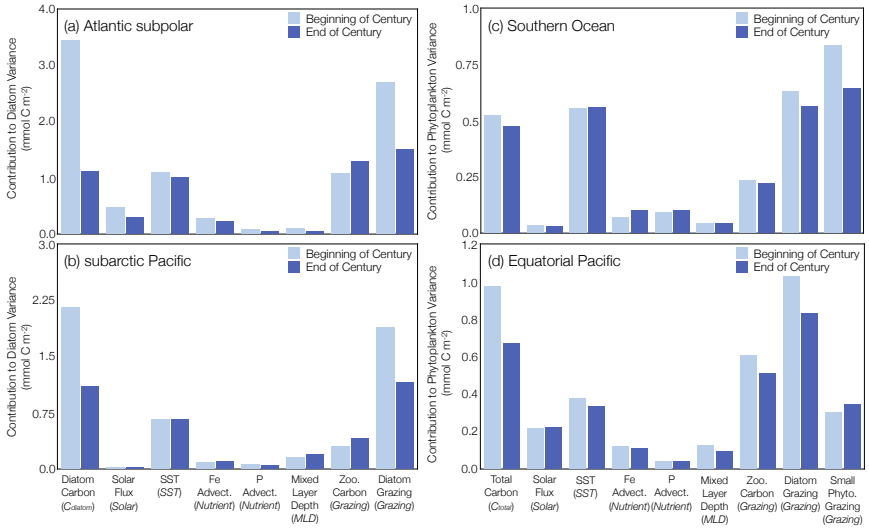


Figure 5: "Reconstructed changes in the contribution of each variable to phytoplankton biomass standard deviation across the RCP8.5 forcing scenario (2006 to 2100) with the beginning of the century shown in light blue and the end of the century shown in dark blue. Marine ecological regions are defined in Tagliabue et al. (2021). Regions were selected which aligned with the highest fisheries catch in the (a) Atlantic and (b) Pacific basins and the biogeochemically important (c) Southern Ocean and (d) Equatorial Pacific regions. The dominant phytoplankton functional type is considered in each region. In regions with a mixed ecological assemblage, total phytoplankton carbon is considered. The change in the coefficient of variance is calculated using averages across the first (2006 to 2016) and last (2090 to 2100) decades of the RCP8.5 forcing scenario."