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).

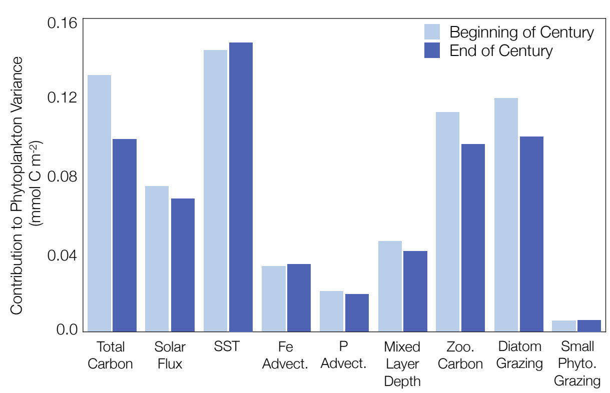
Linear decompositions should be applied for "change/anomaly", not for "climatology (10-year mean)."

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),

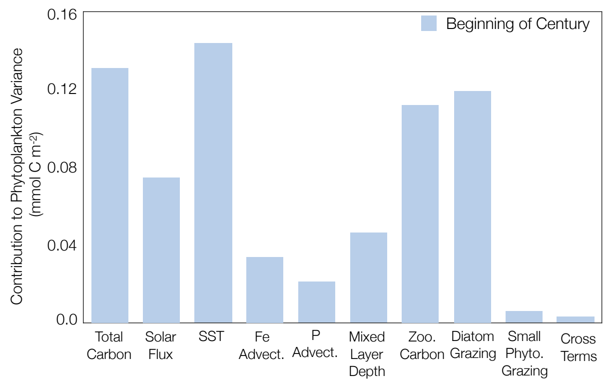
$$\Delta F = (\partial F/\partial X)\Delta X+(\partial F/\partial Y)\Delta Y+(\partial F/\partial Z)\Delta Z + \text{(Residuals from high-order and cross terms)}.$$

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:

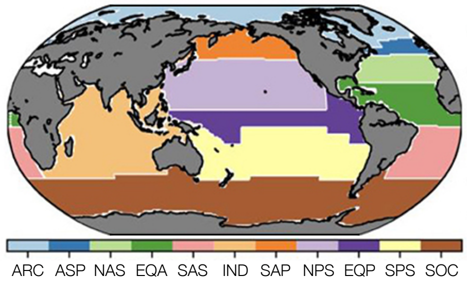
Line 95: "The CESM1-LE simulates the evolution of the climate system from 1920 to 2100 with multiple ensemble members, each expressing different phasing of internal climate variability while responding to a shared external forcing prescription (Kay et al. 2015)."

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.

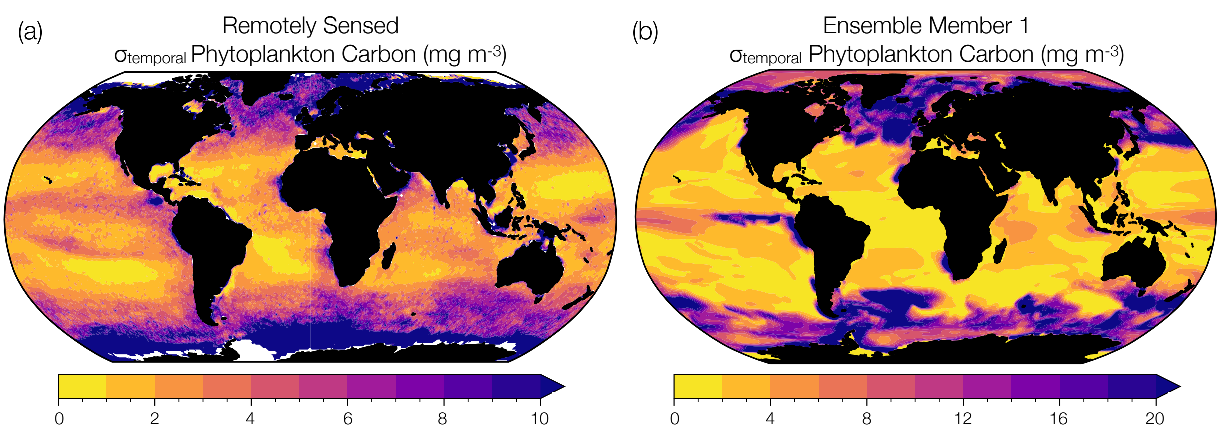


"Figure S2: The 11 ecological provinces defined in Tagliabue et al., (2021) and Vichi et al., (2011). Provinces were aggregated using multivariate statistical analysis of physical (i.e., salinity, temperature, mixed layer depth) and biological (i.e., chlorophyll concentration) ocean parameters to group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

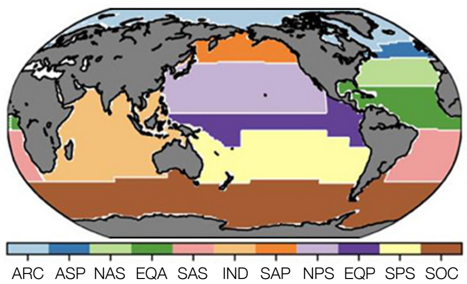
Line 287: "While the CESM1-LE represents the overall spatial pattern of observed interannual variability in phytoplankton carbon, the model overestimates the magnitude of observed interannual and internal variance in phytoplankton on regional scales…"

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.