

This is an interesting, well-written article that explores the incidence and drivers of compound marine extreme events (high temperature and low primary production conditions) in observations and two ESMs. The authors focus on the drivers of low NPP extremes during marine heat waves and use an interesting approach to decompose the relative contribution of these drivers in each model. Their work further highlights the progress that needs to be made to robustly “observe” and model marine primary production.

My main comments are that I think the authors could:

1. Better explain the decomposition of phytoplankton biomass anomalies. Specifically, the advection/mixing terms which appear to be very important to the decomposition of NPP anomalies in Figure 6.
2. Explore and further discuss the mechanisms behind some of the non-intuitive NPP responses. I find the increase in light limitation in the high latitudes particularly non-intuitive in ESM2M-LM.
3. Assess how the drivers of low NPP extremes differ in a MHW as opposed to non-MHW conditions. I.e. Is there anything different/unique about the drivers of low NPP extremes when they occur within a MHW?

In addition to this I have a number of minor comments below that will hopefully improve manuscript clarity and accessibility. Subject to these comments being addressed, I am happy to recommend publication.

L40 Is it fair to characterize ‘the blob’ as solely a MHW/NPPX event what about cooccurring deoxygenation and acidification?

L74 A word seems missing after “phytoplankton”. Growth, biomass, both?

L157 interval(s)

L161 Maybe clarify you mean “ocean ecology/biogeochemistry. The models presumably also have differing land ecosystem models/DGVMs.

L166 Are you sure this is correct? In the appendices it appears that MARBL uses higher half saturation constants. Maybe there’s an error in the appendix legends or units. To aid comparison I would recommend using the same units for parameters that are present in each model where possible.

L167 Are nutrient levels similar between models? I ask because this is an interpretation that the authors come back to.

L182 Perhaps add something here on the rationale behind deseasonalizing. Do observations support this approach? It seems contrary to the thermal stress DHW estimates that are used for coral reefs.

L186. Over what period are percentiles computed? Does this correspond to a 1-in-10-day event? This seems like a low threshold for defining an extreme that probably need to be discussed/justified.

L193. I'm not sure I would interpret these as hotspots of compound extremes, rather simply a lack of independence between the drivers of MHW and NPPX events.
Warming/stratification/reduced nutrient supply?

L196 I find it a little strange to have model evaluation in the methods but I suppose there's a certain logic to not having this in the results.

Figure 1. Maybe mention that Std is normalized on the Taylor plot axes.

Figure 1d. One wonders if an overestimation of durations is a resolution issue. E.g. a result of non-eddy models failing to capture short-lived extremes that maybe associated with mesoscale processes.

L232 This may be an overstatement for NPP where the models seem a long way from the ref (even if there is very little confidence in the ref).

L266 Is this advection/mixing term explicitly calculated or is this a residual? If the former can the authors say how well the decomposition works? Related to this, are the driver decompositions calculated online or using models outputs at a certain resolution?

L269 The method could be clarified here. Do you mean the maximum absolute anomaly? And if so, relative to what, the climatological mean?

Results- Lots of the interpretation and speculation that is currently in the results section is really "discussion". I would recommend having a combined results and discussion section to avoid substantial reformatting.

Figure 3. It's not entirely clear to me how to interpret the contours in this figure. Should there be values associated with the contours?

L332. You're showing NPP not "abundance" (although they are presumably closely related in the climatological mean).

L345-347 I suggest removing the "contribution" phrasing here which is confusing.

L356 Any such NPP increases are very difficult to see with the current color scale of Figure 3. Maybe you want to white-center your color scale to avoid overinterpretation of very low anomalies.

L364 Is the "per day" unit correct? I.e is this a rate of decline in nutrient limitation? I would expect a mean nutrient limitation anomaly to be unitless.

L373-374 I find it odd that the impact of nutrient limitation isn't exacerbated in the larger phytoplankton given their higher half saturation constants. Could the authors comment on this?

L366 As mentioned previously, why are MHW-NPPX events associated with enhanced light limitation in the high latitudes? Presumably these waters are more stratified or are sea-ice dynamics interfering with the expectation that mixing of phytoplankton below the euphotic zone declines during a MHW?

Figure 4. Perhaps the authors could add a comment in the legend on how well the growth rate decomposition works, assuming it is not perfect.

L381 See earlier comment on this.

Figure 5. The units in the figure and legend don't match.

L421. Maybe clarify you mean the grazing of phytoplankton by zooplankton.

Figure 6. I really like the idea behind this figure but it would be nice to see a sum of the contributions to see how well the decomposition works in each region. What does the "+res." term stand for (residual?). As previously stated, I am uncertain how to interpret these large circulation anomalies that offset NPP declines in all regions.

Figure 6. I'm quite surprised by the very small Nlim term in CESM2, particularly in the low latitudes and the eastern equatorial Pacific. Can the authors comment on this.

L443-444 Don't the models basically show no light limitation effect in the low latitudes? Perhaps best to refer to light limitation and not light levels. Presuming it's the vertical distribution of phytoplankton biomass and not incoming shortwave which is the main driver of light limitation during MHWs (unless authors have assessed this?).

L450 See previous comment.

L457 Is there causality here? Do higher temperatures enhance loss or are they simply associated with enhanced loss?

L478 I don't think this citing of the study "goals" is needed and is a bit distracting.

L496-497 Presumably nutrient and light limitation only influence production not loss. Can the authors discuss how loss is actually affected? Is there greater thermal sensitivity of grazers than phytoplankton? how do changes in physical dynamics influence loss during MHWs?