North Atlantic patterns of primary production and phenology in two Earth System Models.

bg-2023-54

Jenny Hieronymus et al.

Reviewer summary:

In this work, the authors use daily integrated net primary production data from two CMIP6 models and compare them against an observational dataset derived from satellite. They use change points analysis to highlight changes in the mean NPP and MLD time series. The model and observational data are compared but the comparisons are often subjective, leaving it up to the reader to decide by eye whether the models successfully reproduce the observed behaviour.

However, the authors do not describe the limitations of their work, leaving several key questions unasked or unanswered. In addition, the absence of a discussions section, means that results are often described but not fully interpreted or put into a wider context. This means that the overall conclusions

On the other hand, the language in this paper is excellent. It is very clearly written and there are very few spelling and typographical or grammar errors – at least that I have spotted.

This is a very interesting topic and there is a lot that could be done with this high temporal resolution model data. With a bit more effort, this could be a very exciting and impactful paper and I'd encourage the authors to keep going!

Major points:

Here are some questions that may encourage the authors to explore new ground. I don't expect all of these questions to be answered, but they may open up some new areas of interest.

Earth System Models s tend to have relatively simple representations of the marine Ecosystem, relative to dedicated ocean models. PISCES has two phytoplankton functional types and NorESM2 has only one. More complicated BGC models exist, such as BFM, Planktom, ERSEM and others, who have four or more phytoplankton classes. Furthermore, it is widely known that blooms are formed from cascades of multiples species. See for instance Kleparski 2021, which uses 24 species assemblages in CPR data to investigate the spatial distribution of north atlantic blooms in observations. How does the low PFT number of these models impact their ability to model blooms?

> Kléparski, L, Beaugrand, G, Edwards, M. Plankton biogeography in the North Atlantic Ocean and its adjacent seas: Species assemblages and environmental signatures. Ecol Evol. 2021; 11: 5135– 5149. https://doi.org/10.1002/ece3.7406

Similarly, marine BGC models are often run as standard-alone Ocean only runs. What do we gain by including the rest of the earth system when investigating phenology? Is there some feedback between the ocean and the atmosphere that improves bloom timing modelling?

The main technical criticism is: why only focus on these two models and this one scenario when all of CMIP6 is available? This work either needs a convincing argument as to why it uses this two model limit, or it needs to include other models from CMIP6. Similarly for the SSP5-8.5 scenario. A quick check of ESGF shows that while there is no daily primary production data, there are 380 datasets for daily Chlorophyll (chlos) from 12 different models with picontrol, historical, or Tier 1 ScenarioMIP

data. A tool like ESMValTool could help accumulate and prepare all the data, even if you don't actually use ESMValTool for the analysis. This would allow a comparison of the phenology (for chlorophyll) between several future scenarios and models.

You've located several change points for bloom timing and MLD. What are the differences (if any) between the period before this point and after this point? For instance, you could revisit fig 1, 2 or 4 comparing the "pre- change" to the post-change for these models.

On L41, you mention that NPP is affected by precipitation, wind patterns, temperature and light, but do not investigate any of these fields. You only investigate MLD as a proxy for reduced nutrient availability. I'm not fully convinced that the argument has been made that MLD shallowing is the cause of the bloom in these models. Daily data for several of these variables should be available in CMIP6 as well.

The paper is also missing a discussions section. This could include things like:

- How well this work fits into the current state of research for this field.
- What mechanisms makes one model better than the others? Is it purely their physical behaviour or is there something about the parameterisation of the marine BGC model?
- Can you estimate how many years of data would be needed to detect these change points in CAFE?
- Could you use longer data sets (ie Continuous Plankton Recorder data) to detect change points in these regions?
- A limitations of this method section

A more thorough investigation of phenology may also be interesting, not just the bloom initiation or maxima, but the intensity and duration of blooms may also be changing in the future. See for instance: https://www.sciencedirect.com/science/article/abs/pii/S1470160X11002160

In addition, other sources of observational data are available, and would strengthen the arguments around the presence of change points:

Floats: https://www.frontiersin.org/articles/10.3389/fmars.2020.00139/full

CPR (in North Sea): https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.11351

Bermuda time series: https://bats.bios.edu/

Specific comments:

Title: To get more impact, state your main result in the title. Peak Net Primary Production will shift earlier in the year over the 21st century. (or something like that).

Abstract:

L11: Similarly to the Title – if you open the abstract with the main result, it can be more eye catching. Try not to hide your exciting result at the bottom of the abstract!

L16: “The low spatial resolution of the earth system models can explain much of such difference”: Can it? How so? I’m not convinced that this argument has been made.

L20: Using SSP585 as a forecast needs to be treated carefully, as this is not a “realistic” scenario. SSP585 is the scenario with enhanced fossil fuel usage – meaning that the rate of fossil fuel emissions accelerates beyond “business as usual” - something we have fortunately not seen in the previous 8 years since the end of the CMIP6 historical period.

Introduction:

L27: See Le Quéré et al for more up to date reference on the carbon budget.

> Le Quéré, C., et al : Global Carbon Budget 2018, Earth Syst. Sci. Data, 10, 2141–2194, https://doi.org/10.5194/essd-10-2141-2018, 2018.

L35 and elsewhere: CO2: 2 should be subscript

L41: Can you add some references for this? How is NPP impacted by precipitation in the North Atlantic?

L55: “Depending on the onset…” This sentence needs to be more explicit. Ie, If thermal stratification occurs, then spring bloom may start earlier…” or similar.

L57: One alternative theory about the causes of bloom timing changes is the switch from positive net heat flux to negative net heat flux:

> Smyth TJ, Allen I, Atkinson A, Bruun JT, Harmer RA, Pingree RD, et al. (2014) Ocean Net Heat Flux Influences Seasonal to Interannual Patterns of Plankton Abundance. PLoS ONE 9(6): e98709. https://doi.org/10.1371/journal.pone.009870

L72: What makes it unique? Why only these two models? Why not a full CMIP6 ensemble? What do you gain by using the years 1750-1850?

Methods:

L76: Just to confirm, are you are using the CMIP6 dataset, intpp, from ESGF? When you calculate the mean, are you taking the cell area weighted mean?

L87: Nemo should be NEMO

L89: pCO2: 2 should be subscript

L96: Primary production is indeed growth of phytoplankton, but elsewhere you talk about net primary production. Net PP usually does include some loss terms – otherwise you mean Gross Primary Production (GPP).

134: Modis should be MODIS

136: The Change point analysis method description section (sect. 2.3) does not sufficiently explain how the method works, and gives the impression that the authors have used the Ruptures package

as a “black box”. Can you give more detail on how this method works here (or perhaps in an appendix)? Same for L1 & L2 methods.

Results:

L159 and figure 1: Is it sensible to compare depth-integrated Net Primary production to satellite (surface) Net Primary Production? Are these data comparable?

L169: Some statistical tools would help give an objective estimation of model data fit, something like some pattern statistics or even a linear regression?

L176 – If the mask is important, you may need to indicate it in this figure (or elsewhere).

L180: “Reasonable”. Once again, an objective measure of goodness of fit may be useful here.

L186 – is a version of the model with higher resolution and better agreement with observations exists, then why not use data from that one?

Fig1: The MAM bloom in EC-Earth-CC seems unrealistically high, but I’m not convinced that either model is able to reproduce the observed behaviour over this time period. A statistical comparison would allow you to state how well these models reproduce observed behaviour. We can’t expect ESMs to reproduce specific observations perfectly, but broad scale decadal means should be feasible.

Fig2. The coloured lined are the multi-year mean, but what does the lightly shaded area represent?

Section 3.2:

L193 – Try to avoid single sentence paragraphs like this.

L195: CAFE (uppercase)

Fig 3: Are you actually showing 8 day means on a figure that spans over three centuries? It looks like the shaded areas are the 8 day mean’s annual minimum and maximum. Naively, from figure 2, I would expect the minimum value of NorESM2-LM to be around 50 mgC/m2/day, but it appears to be lower than that? Similarly, the range in EC-Earth-CC has a minimum around 200 in figure 2, but it looks closer to 150 in figure 3. I’m not convince that showing the range is useful here, and it’s not clear to be me what it represents. Perhaps it may be easier to show the 5-95 percentile ranges (once again – weighted by area) instead, to avoid erroneously high or low values?

Figure 4 hides a lot of the important information, only showing a little of the earlier years underneath the later years. Perhaps you could instead show some decadal averages (or various change point regimes) as semitransparent bands?

208: Is the peak NPP the best metric for this? In the past, I’ve seen bloom timing calculated using the maximum in the first derivative, ie,when phytoplankton is growing the fastest, or when the chlorophyll concentration rises above the long term median: See for instance:

> Philippart, C.J.M., van Iperen, J.M., Cadée, G.C. et al. Long-term Field Observations on Seasonality in Chlorophyll-a Concentrations in a Shallow Coastal Marine Ecosystem, the Wadden Sea. Estuaries and Coasts 33, 286–294 (2010). https://doi.org/10.1007/s12237-009-9236-y ,

Marie-Fanny Racault, Corinne Le Quéré, Erik Buitenhuis, Shubha Sathyendranath, Trevor Platt, Phytoplankton phenology in the global ocean, Ecological Indicators, Volume 14, Issue 1, 2012, Pages 152-163, ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2011.07.010.

Figure 5: The pane labelled 1850 is the mean of 1850-1879. Perhaps these labels should be 1850-1879 mean, 1970-1999 anomaly, and 2070-2099 anomaly. I also think that a discrete colour scale would be useful here. I'd also like to see the CAFÉ data peak NPP day.

L241: I'm not convinced how representative the mean over the whole region is here, especially after seeing how heterogenous the phenology behaviour is in fig 5! Perhaps it would be better to define sub regions within the domain and see how they behave (ie Labrador Sea, Gulf Stream, Southern NA, Central NA, Northern NA... etc.)?

L254: The fact that several methods agree that 2010 is a change point for EC-Earth-CC should give you confidence that this is a real change. However, there isn't the same agreement in NorESM2-LM. How would you interpret this? Are the changes perhaps more real in EC-Earth-CC than in NorESM2-LM? Can this shift at 2010 be seen in any observations? How do the phytoplankton bloom and the physical drivers of the bloom differ in EC-Earth-CC before and after 2010? The second EC-Earth-CC change point is around the year 2090 – how many years after the change are required to register the change?

Figure 6: What's the value of using 8 change points? I can't see them mentioned anywhere in the text, I recommend removing them from this figure if they aren't discussed.

Figure 6: What does it look like if you apply this same method to CAFE? There isn't as much data but there is still a couple decades. Is that enough to detect changes?

L271 and figure 7: How do you interpret change points in the pre-industrial period? If they can occur using this method, then it's hard to justify that later change points are linked to climate change without additional analysis.

Figure 7: I really like this figure, but I think it could be more effective. Perhaps you could focus in on the recent past and the future regions by limiting the colour scale to (ie) the years 2000-2100? Do we really expect change points to occur earlier than say 1950? If white means no change single point could be found, the land colour needs to be a different colour – light grey perhaps. (You also have some ocean points occurring over the land mask, so make sure you set land zorder to be higher than the pcolormesh here and elsewhere. Similarly, the contour lines are the same style, thickness and colour as the same surface, and this is confusing.

Fig 7 caption: rephrase "White spaces are areas where a single change point could not be found.**"**

L278: I find the Smythe theory about Net heat Flux (linked above) particularly compelling – even if it may not be applicable to the open ocean. Could heat (or temperature as there is a lot of CMIP6 SST data!) play a role in bloom timing?

L279: Please be cautious with using "more and less" to describe depths. As closer to the sea floor means larger values of depth, "40m or more" can mean: "deeper than 40m" or "shallower than 40m"!

L287: You have tested for the first day below 40m but comparing figures 6 and 8 makes it look like this threshold generally occurs after the peak NPP in EC-Earth-CC. The MLD lines average around day

140, while the peak NPP seems to be around 130-135. How can MLD shallowing occur after the bloom peak if it is the main cause of the bloom initialization?

Fig 8: This figure could be merged with figure 6, and it would drive the MLD discussion earlier in the results section.

Conclusions:

L300: Unresolved Eddies? How do you know this? I'm not convinced that you've demonstrated this conclusion.

L313: "1/11 day/s in NorESM2-LM/EC-Earth3-CC" should be: "1 day in NorESM2-LM and 11 days in EC-Earth3-CC"

L314: 31/33 should be "31 and 33"

L320: First mention of fish! Maybe put something in the introduction or the discussion sections.

L322: First mention of ecosystem structure! Maybe put something in the introduction or the discussion sections.

Code Availabililty:

L331: What about the Ruptures python package?

L340: This link did not work for me, nor could I find it using the zenodo search bar.