

Phenological shifts in the North Atlantic net primary production detected in the 21st century. Results from two Earth system models.

Jenny Hieronymus, Magnus Hieronymus, Matthias Gröger, Jörg Schwinger, Raffaele Bernadello, Etienne Tourigny, Valentina Sicardi, Itzel Ruvalcaba Baroni, and Klaus Wyser

Submitted to Biogeoscience, bg-2023-54

Review by Dr. Lee de Mora, Plymouth Marine Laboratory, Plymouth, Devon.

Summary

In this work, the daily Net Primary production of two Earth System Models in the Northern Atlantic are described, compared against satellite data and analysed using change point and cross correlation analysis for several regions of the North Atlantic. The timing of the peak of the bloom will shift earlier in the year in the Northern parts of the North Atlantic. The models disagree for the Southern North Atlantic, but it is less of a shift than in the northern regions. The change point analysis highlights that several regions are likely to have pass the change already and that nearly all regions will cross the change point in the 21st century. However, it's not clear how significant the scale of the change point will be.

The text is well written, the underlying science is well introduced in a clear way, the results are presented and described accurately, I did not spot any spelling mistakes and the grammar is almost always fine. At times, the style is a little colloquial and would be improved if certain parts were written in a more formal style. There's a few run-on sentences which need to be pruned. There were many formatting issues, described below, and there are likely many more that I missed.

The figures are generally clear, but I suggest a few improvements below which I think would benefit the paper as a whole.

I found that there were a few discussion points that were hinted at in the abstract and introductions, but never made it to the final draft. I have a few questions below, a few suggestions and a few possible additions.

In my opinion, the biggest weakness of the paper as it currently stands is that the key results are not sufficiently well articulated. It's crucial that the revision of this paper focuses on its unique results and explains why they are important as clearly as possible.

I would also recommend a careful and thorough readthrough (including the references section) before re-submission. Sometimes it's better to share this task with a more distant co-author, as the lead author is often too close to the work to spot these issues.

As the list of changes below has become rather long, I recommend major corrections before reconsideration. However, I don't want to come across as being harsh. This is a good, well written paper, with good scientific content, it fits within the remit of the journal, and most of these changes should be resolvable with minimal effort.

Specific Comments

I expect a more forceful and direct tone in the title, abstract and the conclusions. At the moment, the abstract focuses on the methods, but it should effectively read: "This is our main result. This is why it

is important.” Then once you’ve said that, only then you can describe the methods and models that were used to found out.

Similarly, the title could be a more direct and effective. Something like: “Net Primary Production Annual Maxima in the North Atlantic projected to shift in the 21st century” or something like that. (As an aside, I don’t think you need to mention ESMs in the title – it’s obvious that models were used if you’re making projections of the future!)

I’m not convinced that either model captures the observations over the historical period. The model-data comparisons in figure 1 and table 1 are subjective. I’d like to see a robust statistical comparison of the model and data over the historical period. A graphical version like a pair of Taylor or Target diagrams would be a solid improvement. Alternatively but probably less effective, you could add pattern statistics (bias, deviation, and correlation) to Table 1. Please bear in mind that you have made a new and unique piece of work, and people in the future will be glad to have a robust statistical benchmark to cite that they can compare their model quality against.

It would be valuable to include the analysis of the mean of the whole North Atlantic region in some of these figures (ie Figure 3, Tables 1, 2 and 3 as well?) I understand the value of the individual regions, but a clear result for the whole region would be a good headline result.

I would be interested in seeing how different the phenologies of the various regions are on either side of the change point. Basically, a version of figure 3 which compares the climatological mean of ten (or some useful number of) years each side of the change point. This would be a clear and effective way of showing that there is indeed a real change between those two periods.

There is no daily net primary production data from CMIP6 on ESGF, but there is quite a lot of surface chlorophyll (chl_{oc2}) and surface phytoplankton carbon (phy_{oc2}). It would be interesting to place these two models against the rest of CMIP6 in the context of the phytoplankton carbon or chlorophyll. Are they typical or are they outliers? This would be a lot of additional work, so I leave it up to the authors to decide whether they can perform the additional analysis. If not, then maybe add it as a suggested extension in the discussion.

As I mentioned, several modelling centres have contributed daily surface chlorophyll and phytoplankton carbon to CMIP6, but no one has contributed daily NPP. Do you not want to make the case to include daily NPP (intpp) as a standard variable in future CMIP experiments? What do we gain from including NPP that we don’t get from chlorophyll and carbon?

At the end of the discussion section, I found myself asking several follow up questions like these. I have listed these below with the label “L394”.

Typesetting & style comments

There is a tendency for sentences to be too long and complex, which makes them harder to read and parse - particularly in the abstract. While they may be accurate, they take more effort to understand. For this reason, I personally have a strong preference for simpler shorter sentences. I have pointed out a couple in the abstract and made some suggestions on ways to shorten and split them. However, I’ll leave it up to you from that point.

Please try to be consistent with hyphenation and capitalisation. Change-point, time-series, cross-correlation, North-west, PI-control are all written in several different ways throughout the text.

There are several places where the superscript is lost for both the degree symbol, units and centuries (especially in the title!). Please be more careful with subscript and superscripts.

There are a few places where the text is stretched: L157, L543, L691.

For references in the body of the text, there are a few places where the name in the text either doesn't match the reference, or a reference does not exist. Similarly, there is some variability in typography of the "et al.", sometimes the period is missing (L336) and sometimes there's no space before the year (L140).

In the reference section, there are a lot of inconsistencies:

- Several references with strange characters that need to be corrected. Ie L471, 474.
- Some DOI's are links in blue and some are not, ieL 462 vs L464.
- Most references have the author initials, but L475 includes author's first names.
- Some are missing DOI's ie L480.
- Some place the year at the start (L461) and some at the end (L466).
- Some times the journal name is in italics (L468) and sometimes it is not (L464).
- Some references include et al. (L494 & L574) but most do not. Some instead use "... " (L501, L556, L562 & L594)

To me, this suggests that the author is manually writing the bibliography – which just makes your life more difficult! If you are, I strongly recommend instead using some kind of reference management tool to keep track of the bibliography and ensure that references are done properly and consistently. (I use bibtex for latex).

Without being an expert on the North Atlantic, I find the 8 different region names to be confusing. I'm constantly having to refer to figure 1 to check what region is being discussed. It may be clearer to write where things are happening more descriptively than just relying on the four letter acronyms. For instance, in line 253: "The largest standard deviation is found in NASE and the lowest in NWCS" could be clearer as "The largest standard deviation is found in the southeast of our domain in NASE and the lowest is in the Northwest Atlantic Shelf (NWCS)." This is closer to how you have described figure 5 in lines 264-274, which is much clearer.

In the figures, I think there's scope for some additional consistency which would make it easier to interpret. For instance, you can use the same regional colours from figure 1 again in figure 3. Instead of the blue/orange colour scheme in figures 4, 6 and 7, you could use the regional colours again, but have a lighter one for EC-Earth and a darker one for NorESM2 (for instance). Alternatively, you could keep the same two line colours, but change the regional labels to match figure 1. Similarly, I see no reason for figure 9 to be different to figures 6 and 7. Finally, and this is purely subjective, but I'm not crazy about the blue and orange colours – I think a more aesthetic colour pairing could be found (<https://colorbrewer2.org/> is a good resource for something like this).

As a added suggestion, daily data looks great in video format – much better than monthly and annual data. Have you considered including a supplementary video of the daily climatological mean for the two models and the observations? This would be a great resource to show when presenting this work in person. Alternatively, it could contribute to a great video abstract.

Specific Points

Abstract:

L14: This is a long sentence which could be clearer: "The majority of the region displays the largest change point in the day of peak NPP occurring after the year 2000 indicating a shift towards earlier peak NPP with the most change occurring in the northern parts of the domain."

Suggested change: "Most of the region has the largest change point in the day of peak NPP after the year 2000. This indicates a shift towards earlier peak NPP and the most change occurs in the northern parts of the domain."

L18: long sentence: "Furthermore, the occurrence of the first day with MLD shallower than 40 m shows positive correlation with the occurrence of the day of peak NPP for most of the domain and, similar to the day of peak NPP, displays the largest change points occurring around or after the year 2000."

Suggested change: "Furthermore, the occurrences of the first day with a MLD shallower than 40 m and the day of peak NPP are positively correlated over most of the domain. As was the case for day of peak NPP, the largest change points occur around or after the year 2000."

L20 and elsewhere: Is it **change point** or **change point** or **change-point**? Beaulieu (2012) uses change-point so maybe that is the best option?

Introduction:

L26: Turnover time is not defined and never referenced again. Why is it important?

L27: "Almost equals"? What is the land NPP value?

L27 – and elsewhere: I prefer Pg instead of Gt. Ton is not an SI unit and there are many definitions of tons/tonnes/imperial tons and it can be confusing for international readers. Also, /yr should be yr^{-1}

L28: "constitutes" isn't the right word. NPP is the act of fixing the carbon, while the phytoplankton themselves are the basis of the food chain.

L30-31: Add a reference for this.

L35: north -> North

L37: remove "here, ", but also consider simplifying this sentence.

L53: "The seasonal cycle of phytoplankton blooms has been explained with various theories" -> "Several mechanisms have been hypothesized to explain the seasonal cycle of phytoplankton blooms." (Suggestion)

L53-59. There's a bit of a confusion here about theories vs hypotheses. A theory is specifically a widely accepted and tested hypothesis (like gravity, evolution or similar). So by definition, these competing explanations can't all be theories! Also, I don't think that the critical depth hypothesis can be both a hypothesis and a theory. I recommend rephrasing this paragraph so that these are called hypotheses, explanations or mechanisms, instead of theories.

L62: Please be more explicit with your definition of phenology. It's the core of the paper and its in the title. It deserves a full definition.

L76: “ a maximum temporal resolution of not more than 20 days is required.” -> “a temporal resolution of 20 days or less is required.”

L79 “In this paper...”: long sentence

L88: “highlights at which leads and lags” is there a missing word here? Consider simplifying this sentence

Methods

L91: Is the NPP integrated to the sea floor or to some other depth?

L92: “100 years pi-control” -> “100 years of Pre-industrial Control (piControl)” Note that there several spellings of PI-control here, you should choose one.

L92: “Kriegler et al., 2017” doesn’t exist in the references.

L92: Can you justify why only one scenario and why you chose SSP5-8.5?

L94: Please check the submission guidelines for referencing sections. This should be: “The models are described in Sect. 2.1. Section 2.2 describes the observational data set and Sect. 2.3...”
<https://www.biogeosciences.net/submission.html>

L96: “which is calculated as a simple max of NPP” -> “which is calculated as the annual maximum of the daily means NPP, in units of $\text{mgC m}^{-2} \text{d}^{-1}$.”

L96: Do you calculate the regional mean first and then the annual maximum? Or do you calculate first the spatial distribution of the annual maximum and then the regional means?

L100: comma after EC-Earth3-CC and NorESM2-LM.

L100 and L104: unit superscripts: cm^2 should have a superscript cm^2 , s^{-1} and kg m^{-3} .

L104: I don’t think that these two methods are compatible with each other. While I note that you never compare the two MLD datasets, do you expect the differences here to impact your conclusions? If not, please explain why.

L110: This reference is authored by “Gurvan Madec and the NEMO team”, so perhaps should be an et al, or the NEMO collaboration or similar.

L114: please use colons: “PISCES is a mixed Monod-quota model simulating two different phytoplankton functional types: diatoms and nanophytoplankton, two size classes of zooplankton: micro and meso, and the nutrients: nitrate, ammonium, phosphate, iron and silicate.”

L116: Add a reference to Redfield.

L124: remove comma after EC-Earth3.

L125: North-west -> Northwest.

L124 and elsewhere: “ocean only” -> “ocean-only”

L124: "Skyllas et al. (2019) validated EC-Earth3, in an offline ocean only NEMO-PISCES version, for a north-south (29-63°N) transect in the North-west Atlantic using cruise data of temperature, salinity and nutrients and chlorophyll-a and found a good agreement with observations."

Suggest re-writing this as: "Skyllas et al. (2019) showed a good agreement between EC-Earth3 and temperature, salinity and nutrients and chlorophyll-a observations in an offline ocean-only version of NEMO-PISCES, for a north-south (29-63°N) transect in the Northwest Atlantic."

L138: 2o -> 2°

L140: Assman et al.(2010). This should be Assmann, and add a space after et al.

L141: replace "with one phytoplankton and one zooplankton compartment and" with "with one phytoplankton functional type, one zooplankton functional type and"

L142: Is this nitrogen and phosphorus or nitrate and phosphate?

L158 and L168: move the weblink to a reference.

L159 please define MODIS.

L166: "Division of the global ocean into biogeochemical provinces has been done in a number of references" -> "The division of the global ocean into biogeochemical provinces has been attempted several times" (you can probably find more recent attempts as well.)

L176: Replace "and" after (NADR) with a comma.

L184: C-> Carbon

L188: "Generally speaking" is colloquial.

L191: "we directly pick the number of change points to find" -> "we directly pick the desired number of change points"

L191: Remove "in fact," ... This should all be facts, lol.

L193-194: extra space between "to instead" and "that is"?

L200: remove "in the following"

L201: "all sorts" is colloquial.

L206: ruptures was previously capitalised: Ruptures.

L207: You haven't described the cross-correlation method yet.

Results

L211: "for March" -> "for the March"

L213: "the internal variability of the climate system as modelled by the two ESMs is not in sync with that in reality or with each other." -> "the internal variabilities of the two ESMs climate systems are not synchronised with nature or each other."

L222: Is the yellow blob of overestimated NPP in the Spring EC-Earth3 related to the Döscher et al (2022) result that the model has too much active convection in the Labrador Sea?

L240: "In general, EC-Earth3-CC is closer to the CAFE data in size but NorESM2-LM is closer in timing." Can you back this up with some kind of objective statement?

L246: New paragraphs is missing a blank line.

L247 and elsewhere: Please fix the units.

L257-258 and elsewhere: yr-> year. Please use "year" in prose and use "yr" when it is a unit.

L265 & 309: Fig -> Fig. (Maybe elsewhere too)

L296: What is going on in EC-Earth3-CC ARTC in figure 7? There's such a wide range of behaviour there, it's hard to see why the charge point is placed where it is. Is there any way to quantify the quality of the fit – because this looks like a poor fit!

L315: I don't fully understand the value of the cross correlation analysis. Surely it's obvious that the current MLD is going to have the most impact on the current year's NPP? Maybe there's more to it here, but I think if you must include this analysis in your final paper, it needs to be explained with a bit more precision, details and specificity for those of us that don't use it on a regular basis.

L324: "lurking variable" could use a bit more explanation.

Discussions

L329: "the size of peak NPP was well captured by the ESMs". I'm not sure that this case has been made. You could just as easily argue from Figure 1 and Table 1 that neither model captures the observations very well. I'd like to see a robust statistical comparison of the model and data over the historical period. A graphical version like a pair of Taylor or Target diagrams or add the bias, deviation, and correlation of the annual time series to Table 1. (I repeat this in the general comments above).

L336: north -> North

L345: Several modelling centres have contributed daily surface chlorophyll to CMIP6, but no one has contributed daily NPP. Do you want to make the case to include daily NPP (intpp) as a standard variable in future CMIP experiments?

L351: Remove "A noteworthy observation is that"

L352 and elsewhere: 21st should be 21st

L355: IPCC 2022 is not an appropriate reference here. It's 3000+ pages long! At the very least, please cite an individual chapter. Also, this citation is for WG2, and the information you cite is more likely to be in WG1. You should probably instead cite O'Neill 2016

<https://gmd.copernicus.org/articles/9/3461/2016/> and Riahi 2017
<https://www.sciencedirect.com/science/article/pii/S0959378016300681>.

L372: "NPP and its timing is, of course, both in the models and in reality dependent on many other factors in addition to the MLD. Some examples are light availability, nutrient concentrations and temperature." -> "In both models and in nature, NPP and its timing is dependent on many other factors beyond the MLD, include light availability, nutrient concentrations and temperature."

L374: remove "it is clear that"

L379 & L387: earth -> Earth

L379-385: Is there any indication of a difference between PISCES's phytoplankton functional types in terms of how climate change will impact their bloom onset phenology? Either here, in the monthly data, or in literature?

L391-393: While I more or less agree with this sentiment, I think it's a bit of a reach. You may need more to back up this statement. I'm not convinced that understanding bloom phenology will be better served by ESMs with only one PFT than an ocean-only model with multiple PFTs. This could be my personal bias talking – but I think there's definitely scope for both approaches.

Additional discussions to consider:

L394: The abstract concludes with "This highlights the need for long term monitoring campaigns in the North Atlantic.", but this is never discussed or mentioned again. Please add some discussion around this idea.

L394: I'd also like to see a discussion around the consequences of the changes these models have projected. As you mention in L35-L39, "the north Atlantic is a region of particular importance for carbon sequestration in the deep ocean." How will the changing phenology impact Carbon sequestration and deep mixing? Which regions are the most important and how will they change? Is the drop in NPP likely to affect higher trophic levels? How does this interact with biodiversity and marine policy?

L394: Several of the figures are not explicitly mentioned in the discussion. Please add links to the relevant figures where you discuss them, and make sure that all figures (except maybe fig 1) are mentioned in the discussion.

L394: What are the limitations of change point analysis? What does it mean when we pass a change point? How should this be interpreted? Is it like a regime change?

L394: Figures 2 and 3 shows that neither model is amazing at representing the historical behaviour. How much can we trust the projections of models that fail to capture historical observations? (I realise that we have no other tools available – but allow me to play devils advocate here!)

L394: Can you discuss the long-term projection of NPP in figure 4? In most regions (except NASE), it looks like both models project a rise in NPP? Can you compare these two models against the CMIP6 mean, for instance from <https://www.frontiersin.org/articles/10.3389/fclim.2021.738224/full>, which shows a decline in North Atlantic NPP in both CMIP5 and CMIP6 multi-model means.

L394: Similarly, considering that SSP5-8.5 is the most extreme climate change scenario (where future fossil fuel emission grows to 5x current values!), it barely impacts the NPP in figure 4. Is it possible that these two modelled ecosystems are particularly insensitive to climate change? Are they suitable for this type of analysis or are more flexible BGC models necessary for projecting the impact of climate change on the marine ecosystem?

L394: I'd like to see some discussion about the suitability of the SSP5-8.5 scenario. It has extremely high fossil fuel emissions and subsequent warming, which is likely to move these change points earlier in the simulation than you would see in other scenarios.

L394: NorESM2-LM has a particularly low (but feasible) Effective Climate Sensitivity or 2.54K (<https://gmd.copernicus.org/articles/13/6165/2020/>), while EC-Earth3 has an ECS of 4.3K

(<https://gmd.copernicus.org/articles/13/3465/2020/>). Does the difference in their sensitivity to carbon impact the overall conclusions? For instance, is it possible that the surface waters of EC-Earth3 will warm more than NorESM2-LM, and this may shift the locations of habitats in these models.

Conclusions:

L402: This is the first mention of the growing season. Please add a description of these around your figure 2 results in lines 220.

L416-419: Please explain why this is important?

Figure captions:

L655: Seasonal mean vertically integrated NPP-> "Vertically integrated seasonal mean NPP"

L686: There are several issues with white space here. Including an extra space in (L1) and (L2), here and in the caption for figure 7.

L694: remove period from ".Figure 7".

L698: No period at the end.

Figure 2:

- This colour bar should be made with a pointed end at the maximum value (use `extend = 'max'` in `matplotlib`), to indicate that the highest values shown are beyond the end of the scale.
- It's a shame you don't include DJF, as it looks like both models really struggle to capture the behaviour then in figure 3.

Figures 1, 3 and 9:

- It would be nice if the line colours matched the colours in the map (and the region labels in figure 4). See my comments earlier.

Figure 6 & 7:

- While I understand why all figures share the same y axis, perhaps this figure would be better served by each region showing it's own bespoke range so that the change point is easier to see. If you have to, you can move the region label above the axes.
- Please add ticks to the x axes of the top 6 panes.
- Move the legend outside the figure – ideally below the main figure.
- Can you also the solid and dashed lines to the legend.
- Replace the sideways pointing triangle with a vertical pointing triangle.
- Is there any observational data for this that can be added? If you can't find anything, WOA has monthly MLD: <https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=M>

- In the past, I've seen change point analysis that included a trend line either side of the change point. Is there any reason why this was not included here?

Figure 8

- You don't need the thick black contours here, just the colour scale might be more readable. As it stands, the figure really emphasizes the regions where the first change point is before the year 2000.

Figure 9:

- Can you remake this plot with the same style as figure 6 and 7?
- Can you highlight the times when you're within the 95% confidence bands, perhaps by making the line or dots thicker, or making them thinner when you're outside the confidence bands?

Supplementary data

There's no readme to describe the contents of each file.

Supplementary Model

The directory structure is not straightforward to understand. The structure described in the Readme doesn't match up with the directories in the zip. Please make the leading directory names more explicit.