We thank the reviewer for the helpful feedback on improving our manuscript. Most of the minor changes were included in our revised paper as suggested. Two of the three suggested major points were straightforward to address. To address point R2-1), which asked for clarification regarding our selection of marine heatwaves, we clarified the description of methods following the reviewer suggestions. To address point R2-3), which requested clarification regarding N and Fe cycles, we clarified the interactions between iron and nitrate limitation, and updated the methods sections 2.1 and 2.3.

Major point R2-2), which questioned the magnitude and significance of the change to the size class distribution of primary production, was also brought up in the comments from Reviewer 1. We are confident that we can strengthen the manuscript with a discussion of the significance of our results following both reviewers’ suggestions. In particular, we propose to add a new figure (inserted below) which shows the magnitude of the biological response to marine heat waves in comparison to interannual variability in both the Alaska gyre and transition zone, and allows to discuss the changes that are significant vs. the one that are relatively mild (i.e. amplitude lower than one standard deviation, see details in comment #XXX below). This figure demonstrates that while marine heatwaves are classified as extreme events, the response of biological production to these events is relatively mild compared to each region’s interannual variability.

Please see below for a detailed response to each comment. Note: Reviewer comments are in plain text, while manuscript quotes are in italics. We hope these proposed changes are satisfactory for us to move forward with the publication in Biogeosciences.

-----

I think this is a good paper that is publishable with more or less minor revisions. Some aspects of the methodology are insufficiently explained. The terminology is confusing in some places, and some unnecessary jargon is used (see details below).

Major points:
R1-1) There are some important details missing from the description of the methods and the data. Most importantly, a marine heatwave is defined as "anomalies that exceed 1 standard deviation for 5 months or more". But standard deviation of what and anomalies relative to what? The obvious answer is relative to a climatology calculated over the period of the ERSSTv4 data product, but that needs to be stated explicitly, and which years of this data product were used does not appear to be stated anywhere.

And is there an area threshold? Is the criterion applied point-by-point, or only to the regional mean? Would it be a heat wave if only 1 grid point exceeded the threshold?

Thank you. The reviewer is correct that we reference the climatology of the ERSST product when calculating the standard deviation and anomalies. We have clarified that we use an area mean SST anomaly, so that a single grid point exceeding the 1-standard deviation for 5 months threshold would not qualify in our study as a MHW. We also clarify that this anomaly is relative to the climatology of the specified region from 1958-2020. Section 2.1 now reads.

- Following the method of Xu et al. 2021, we calculate the area mean sea surface temperature anomalies (SSTA) relative to the climatology of the region 35° to 46° N, 150° to 135° W using the monthly data from 1958 – 2020 of the Extended Reconstruction SST dataset (ERSSTv4, Huang et al., 2015). Northeast Pacific marine heatwaves are defined as periods when the monthly deviation relative to the climatology exceeds 1 standard deviation for 5 months or more. The same method is used to detect marine heatwaves in the ocean model (see Sect. 2.3 for model details).

(And why does the "MOM6-COBALT climatology" in Figure 10 appear to have interannual variability?)

Figure 10 depicts net community production (NCP) calculations derived from the float trajectories (lat, lon, and time) as sampled in the world ocean atlas (WOA) and the model climatological nitrate field vs the interannually varying model field. In the case of the WOA and model climatology, only the month and day were used to sample the field, while in the interannually variable model field, the year was also used. The fact that this figure appears to show interannual variability while using climatological nitrate fields highlights the observation bias tied to the float sampling trajectories. We conclude that 70% of the NCP collapse calculated during MHWs is actually attributable to the sampling bias and not interannual variability. We've updated the caption to read:

- Fig 10. Net community production (NCP) calculated as nitrate drawdown from winter supply (February) using BGC-Argo float data (dark red). These float trajectories were then used to sample the WOA climatological nitrate field (black dashed), the interannually variable MOM6-COBALT nitrate field (light blue) and the MOM6-COBALT climatological nitrate field (blue dashed) with an apparent NCP calculated similarly.

And we have clarified this at the end of section 4.2

- We quantified the effect of the float shifting from the NPTZ to the AG on the NCP estimate by recomputing the NCP along the same float trajectories (latitude, longitude and time) but sampling the climatological World Ocean Atlas nitrate concentrations to create simulated "profiles". (Fig. 10). In the case of the WOA and model climatology, only the month and day were used to sample the fields, while in the interannually variable model field, the year was also used. We find that most (>70 %) of the NCP reduction derived from this float can be explained by sampling the climatological nitrate field, and that the apparent ecosystem collapse in 2015 is in fact a feature of the float trajectory (red dashed line vs. black line).
"The model was spun-up using three repetitions of ... 1958 to 1985" (112). But spun up from what? From rest? 81 years doesn't seem very long to spin up a global ocean model. And why go to the trouble of initializing short-lived (i.e., insensitive to initial conditions) biological tracers from an ESM piControl (116-117), but not the physical ocean? I think it would make more sense to use the ESM piControl data to initialize the physical ocean, or 1958 of the historical run.

We use WOA to initialize temperature and salinity (T&S) for the physical dynamics as this run uses realistic historical atmospheric forcing (Jra) and we want to use a realistic initial ocean. Using ESM piControl for the physical initialization would not necessarily allow us to have realistic T&S fields matching observations.

We have clarified this in section 2.3:

- The model was spun-up from rest using three repetitions of the 1958 to 1985 Japanese atmospheric reanalysis v1.4 (JRA55do v1.4, Tsujino et al. 2018) for a total of 81 years. Nutrient initializations, including temperature and salinity fields that establish a realistic physical ocean state, come from the 2013 World Ocean Atlas (WOA, Boyer et al. 2013) concentrations. Dissolved inorganic carbon (DIC) and alkalinity are from the global ocean data analysis project v2 climatologies (GLODAPv2, Olsen et al. 2016) with DIC corrected to 1958 using anthropogenic carbon concentrations from Khatiwala et al. (2013).

I find it difficult to believe that there are no Line P chlorophyll data before 2011 (175-176). Line P is one of the longest-running ocean time series programs, and the basic methodology for chlorophyll concentration has not changed in half a century.

As pointed out by reviewer #1, the files published by the Line P community were incomplete we have reached out to Marie Robert and she will send us the updated files. Figure 4g will include these data in the revised manuscript. Preliminary analysis of that data is consistent with the model results and confirm our conclusions.

- ...the Line P and simulated Chl in the model strongly suggests a decline (<0.3 mg m⁻³) during the 2014-2015 period, consistent with the satellite observations (Fig. 1).

Satellite chlorophyll data should be available back to 1996 or 1997. "GlobColour" is referred to several times in the figure captions but never in the main text.

This oversight was corrected and a description of the data product was added to section 2.5.

- Satellite chlorophyll observations (1997-2020) are from the GlobColour dataset (http://globcolour.info) which has been developed, validated, and distributed by ACRI-ST, France (Maritorena et al. 2010).

Are these really all of the Argo floats available in this region? Or is there some other selection criterion being applied that is not spelled out here (e.g., availability of nitrate data or data within a certain area)? I find it hard to believe that these are the only Argo floats deployed in this region over an 11 year period.

While there are other CORE-Argo floats that have sampled the region, the selected floats are the only available BGC-Argo floats that also sampled nitrate. We have updated section 2.5 to specify that we use BGC floats with the nitrate sensors.
This study makes use of the 2008–2018 series of BGC-Argo floats with nitrate sensors deployed near OSP (e.g., Fig. 9).

R2-2) The Abstract ends by saying that "primary production anomalies modify the allometric phytoplankton distribution, resulting in a 2 % decrease in the ratio of large to small phytoplankton in both regions". Firstly, this seems like a very small change to emphasize as a key point in the Abstract: I am wondering if it is a mistake and it should be 2X or 20%. Secondly, it isn't easy to tell whether this passage is talking about production or biomass, and seems to shift arbitrarily between the two. Finally, where exactly in the main text is this assertion substantiated? Figures 7 and 8 illustrate the seasonal decoupling of large and small phytoplankton production, but can not be used to directly infer the Large/Small ratio of either biomass or production. Figure 9 shows only summer data.

To address the significance of our results, as was also requested by Reviewer #1, we have proposed a new figure (attached) which compares the size of the marine heat wave composite biological anomalies to the interannual variability of each region. This figure shows that chlorophyll and phytoplankton production anomalies tied to marine heatwaves can for some of these events reach relatively high values compared to the interannual variability exhibited in each region (i.e. of the order of 1 to 2 standard deviations). For instance, in year 1965 there are strong negative production anomalies (>2.2 σ for large phytoplankton production; >2.1 σ for small phytoplankton production) in the NPTZ. Yet, on average across the 9 events, and in contrast to prior work using satellite-based chlorophyll (e.g. Whitney 2015), we find that chlorophyll, phytoplankton and zooplankton production respond relatively modestly to marine heatwaves in both regions (variability of the order of 1 standard deviation or lower, new Figure). Notably, however, we find a relatively robust decrease in the ratio of large phytoplankton to small phytoplankton production across all events and in both regions (meets or exceed 1 standard deviation), suggesting that marine heatwaves in the northeast Pacific result in a shift of the phytoplankton assemblage favoring small phytoplankton production.

Thus we propose to complement the result section and discussion using this figure and have rewritten the abstract as follows:

- Marine heatwaves (MHWs) are a recurrent phenomenon in the Northeast Pacific that impact regional ecosystems and are expected to intensify in the future. These events, including the 2014–2015 "warm blob," are associated with widespread surface nutrient declines across the subpolar Alaska Gyre (AG) extending south into the North Pacific Transition Zone (NPTZ) with reduced chlorophyll concentrations confined to the NPTZ only. Here we explain the contrast between these two regions using a coupled global ocean-biogeochemical model (MOM6-COBALT) with Argo float and ship-based observations to investigate how MHWs influence marine productivity. Our study finds that chlorophyll, phytoplankton and zooplankton production respond relatively modestly to MHWs in both regions (variability of the order of 1 standard deviation or lower), with the stronger response in the NPTZ. Differences in the reaction of the two primary phytoplankton size classes (large >10 μm, small <10 μm) to changes in seasonal iron and nitrate limitation explain the differences in ecosystem response to MHWs across the two biomes. The reduced nutrient supply during MHWs strongly influences large phytoplankton in the NPTZ (-13 % annually), whereas it has a limited impact on the usually iron-limited large phytoplankton population in the AG (-2 %). In contrast, we find that MHWs yield a small springtime increase in small phytoplankton population in both regions due to shallow mixed layers and higher mean irradiance. These modest primary production anomalies, however, modify the phytoplankton size distribution, resulting in a significant decrease in the ratio of large to small phytoplankton
production that meets or exceeds the time series interannual variability. This shift in the assemblage towards small phytoplankton production is associated with reduced secondary and export production, especially in the NPTZ.

And the new figure will be captioned:

- Caption: Composite anomalies during MHWs compared to the standard deviation of the spatially averaged interannual variability (in units of sigma) for a) the AG (red) and b) the NPTZ (black). Chl is shown as a monthly mean while large phytoplankton production, small phytoplankton production, the ratio of large to small phytoplankton production, total zooplankton production and export production are annually integrated. Individual heat wave years shown chronologically from left to right as faded blue rectangles.

(BTW "allometric phytoplankton distribution" here is a good example of unnecessary jargon: "phytoplankton size distribution" would suffice. And if one wishes to get dogmatic, the anomalies do not "modify" the size distribution. This sort of quasi-teleological confusion of subject and object is characteristic of inexperienced authors receiving inadequate guidance

This has been corrected to:

- These primary production anomalies modify the phytoplankton size distribution...

(see also 208, "Salinity maintains ...").

This has been corrected from “salinity maintains a lateral gradient” to:

- There is a lateral gradient in salinity across the region,

R2-3) The interaction of the N and Fe cycles is sometime characterized in superficial terms, although I think the overall conclusions are mostly sound. It might help to spend a few sentences in the Introduction sketching out a conceptual model of how the authors think the overall system works.

We address these dynamics in the introduction when we introduce the two study regions. In particular we discuss the transition that occurs seasonally in the NPTZ from spring, when nitrate is abundant, to summer, when nitrate is depleted. It reads as follows:

- The AG is a high nutrient, low Chl (HNLC) region, characterized by high nitrate concentrations, but moderate primary production throughout the year due to iron limitation that prevents the development of a strong spring bloom (Martin and Fitzwater 1988; Harrison 2002; Boyd et al. 2004, Peña and Varela 2007). In contrast, the NPTZ is a region characterized by strong seasonality in nitrate and Chl due to the seasonal biological consumption and the Ekman transport of nutrients (Polovina et al. 2008, Chai et al. 2003; Ayers and Lozier 2010). As a result, the NPTZ evolves from a subpolar-like, iron-limited biome when nitrate is abundant in spring to a nitrate-depleted, subtropical-like biome in summer,

On 164-165, would not a prolonged period of stratification also result in depletion of surface iron concentrations? In the absence of significant aeolian sources I think it would. However, it would also tend to drive the system towards N limitation even in the absence of new aeolian Fe.

While we note that we do have climatological aeolian iron deposition that affects
the iron limitation, stratification would, however, generally reduce both nutrients. Therefore we have changed the wording to the more general “nutrient” instead of “nitrate”.

- ...which posited reduced surface nutrient concentrations as a driver of reduced primary production and Chl concentrations during the “warm blob”.

It also seems to be implied that only large phytoplankton are subject to iron limitation (130-135), which I think is questionable. Iron is potentially limiting for nanophytoplankton even if iron limitation is the main driver of the dominance of diatoms or nanophytoplankton. On 268 it is stated that "small phytoplankton are not simulated with iron limitation" so possibly the lack of Fe limitation is by construction in this model. If this is the case it should be stated up front in the Methods.

Yes, this was unclear. In this model, the iron deficiency for each size class of phytoplankton is calculated explicitly, however in the study regions, iron is never limiting for small phytoplankton.

We have updated section 2.3 to state this more clearly.

- Phytoplankton growth is explicitly modeled as size-dependent functions of light, temperature and nutrient limitations (nitrate, ammonia, phosphate, etc.). Small phytoplankton are simulated to be efficient nutrient and light harvesters (Munk and Riley 1952; Geider et al. 1997) in contrast to large phytoplankton, which are parameterized to grow quickly in response to abundant nutrients. Notably, in the study regions, this results in large phytoplankton being sometimes iron limited while small phytoplankton are not. The limitation factors are output from the model as a number between zero and one, with zero indicating complete limitation, i.e. no phytoplankton growth. There are also three zooplankton size classes of which large (>2000 𝜇m) and medium (200 to 2000 𝜇m) make up the mesozooplankton pool with a third, separate small zooplankton class (<200 𝜇m) all of which consume phytoplankton using size-related predator-prey relationships.

The model does include iron limitation for nanophytoplankton, which are simulated as diazotrophs, however, we do not discuss impacts of marine heat waves on that size class in this study as they are a small proportion of total primary production in the study regions.

The limitation factors are never really explained. I assume this means a number between 0 and 1 where 1 means N or Fe replete and 0 means no growth, but this should be clearly stated in the Methods. (On a terminological note, I think "nitrate limited" and "nitrate limitation" should be changed to "nitrogen limitation" instead of "nitrate limitation". Changing to the more general nitrogen limitation slightly changes Fig 7h and Fig 8h, however, the result remains the same: In the NPTZ (Fig 7) iron is limiting in the spring only while in AG (Fig 8) iron is always limiting.

In the last paragraph of section 3.2, the terminology is sometimes vague or confusing, wrt what is meant by a "boundary". On 223, the “2 uM nitrate boundary” could be "2 uM nitrate contour”. In the next sentence, "nitrate boundary" occurs without any context. I assume this means the boundary between regions of N and Fe limitation, but it could be spelled out more clearly. This is an example of a place where adding a few more words could increase clarity substantially. The last few sentences (226-229) read like a
description of the model solution, and this seems like a missed opportunity to state what
the authors think is happening in terms of physical processes (see also 339-343).

We apologize for the confusion here, which Reviewer 1 also pointed out. This
paragraph references a figure showing limitation boundaries that was removed
prior to the original submission. This paragraph will be cut to reference the Line
P data analysis only:

- The Line P data support the model result and show an expansion of the nitrate-
depleted region during the 2014–2015 “warm blob” (Fig. 4), leading to a westward shift
of the 2 μM boundary to 140o W in 2014 (vs a location of ~130o W in the other
years). In the model, this westward shift of the nitrate boundary is overestimated,
extending past 140o W. Thus, in both the observations and model this implies that
nitrate becomes depleted inside the climatological boundary of the HNLC AG. The HNLC
region can therefore be considered to contract while the nitrate-depleted region
expands.

Some details:

10 and elsewhere I would change "Alaskan gyre" to "Alaska gyre" across the board

15 change "limitations" to "limitation"

17 delete "climatologically" or change it to e.g., "usually" or "chronically"

18 "Contrastingly, we find that ..." conversely? in contrast? by contrast?

19 maybe change "lower light limitation" to "higher mean irradiance"

20 change "allometric phytoplankton distribution" to "phytoplankton size distribution"

26 not sure "recorded" is the appropriate term here; how many of these were recognized
as such when they occurred?

26 Changed to: on record

31 " a redistribution of marine biogeography " ???

32 delete "geographical"

Deleted
Corrected to “Chl” in agreement with rest of paper

36 change "demarks" to "demarcates"

Changed

37 delete "Pacific"

Deleted

38 change "nitrate surface concentrations" to "surface nitrate concentrations"

Changed

48-50 this sentence is very awkwardly worded

Changed to:

- This bottom-up explanation does not explain why the decrease in Chl was highly localized (confined to the NPTZ) while anomalously low nitrate concentrations extended 600 km north of any significant Chl anomalies (Peña et al. 2019) (into the AG).

57 change "Ekman-driven transport" to "Ekman transport" or "Ekman flow driven transport"

Changed

60-61 I would consider also citing Glover et al 1994 (10.1029/93JC02144) here (Bograd et al appears to be missing from the ref list)

Glover citation has been added and Bograd citation added to reference list

67, 387 change "contrasted" to "contrasting"

Changed

127 delete "re-

Corrected

132-133 delete "and are efficient ... Geider et al., 1997)"

This is significant to the resulting response between the two size-classes. We’ve moved this detail, however, to the biogeochemical model description in section 2.3 where it’s better suited.

158, 160 mmol kg^-1 should be umol

Changed. Figure 3 units also corrected to umol

159 add a ' on "floats"

Corrected
171-172 "nitrate concentrations are near-zero for most stations (P4–P20)" Is this unusual? Don't some of these stations always see drawdown in summer? (e.g., Pena and Varela 2007).

**Figure 4a** shows the stations (P4–P8) that usually exhibit depleted nitrate during the summer cruises, while during the MHW, the nitrate depletion extends to P20. We have updated the text to the following:

- *During this period, observed Chl data reached concentrations below 0.3 mg m$^{-3}$ (Fig. 4g) while nitrate concentrations are near-zero west of P8 (P4–P20, Fig. 4a).*

186 add "North" before "American"

**Added**

191 change "biophysical" to "biogeochemical"

**Changed**

205 change "values" to "concentrations"

**Changed**

207 "> 5 mg m$^{-3}$" Is this a mistake? This is an extremely high concentration for an open-ocean environment.

**Yes. Typo was corrected to 0.5 mg m$^{-3}$**

211-214 this assertion seems disconnected from the preceding text; not clear what its relevance is

Here we wish to make it clear that while there are similar spatial patterns of low/high nitrate regions, the model exhibits a bias towards lower nitrate concentrations across this region in comparison to the observations. This text has been corrected to the following:

- *However, we note that the modeled surface nitrate concentration is generally lower in comparison to the Line P data, with maximum values rarely exceeding 8$\mu$M versus 15$\mu$M in the observations (Fig. 3a–b) consistent with the annual mean nitrate bias mentioned above.*

217 add a "~" before "130 W"?

**Added**

219 "nitrate becomes more depleted" more than what? (unclear antecedent)

**This was corrected to the following:**

- *Thus, in both the observations and model this implies that nitrate becomes depleted inside the climatological boundary of the HNLC Alaskan Gyre.*

250 not clear what is meant by "in this region of the model"

**This paragraph references the NPTZ region, which is the region to which this sentence refers.** Removed “of the model”.
267 "the limitation factor is significantly lower (-0.06)" significant by what criterion? P<what?

Changed to:
- and the limitation factor is .06 significantly lower (~ 1 standard deviation).

285-287 "Lg Chl" and "Sm Chl" appear only in this one place,

Corrected to be “large phytoplankton Chl” and “small phytoplankton Chl” respectively.

as does "chl" (elsewhere Chl)

Corrected to "Chl"

286-296 "southern-like" and "northern-like" appear only in this paragraph and are not defined or explained

southern-like was changed to “subtropical-like” and northern-like to “subpolar-like”

306 specify mmol of C or N

Corrected to mmol C

311-313 another very awkwardly worded sentence

Changed to:
- This is consistent with the slightly negative annually integrated Chl anomaly observed in satellite data (-0.02 mg m-3, integrated green line) though those data exhibit a greater compensation between a large negative spring anomaly and a positive summer anomaly.

333-336 Does this sentence make sense? It reads like it is sort of arbitrarily combining different levels of causation. If there is a clear hypothesis as to "A leads to B leads to C", it would be better to express it that way.

This sentence combined too many ideas and has been broken into two parts.

- During the “warm blob” atmospheric blocking by an atmospheric ridge (Le et al. 2019) decreased the wind-driven Ekman transport that carries nitrate from the northern AG southward, a process that normally supports up to 40 % of new production (Ayers and Lozier 2010). Further, nitrate concentrations were reduced by warmer upper ocean conditions which drove a reduction in winter mixing (Amaya et al. 2021).

350 add "concentration" after "nitrate"

Added

360 "changes sampled along the floats" along the floats' trajectories?

Changed to "along the floats' trajectories"

364 change "this data" to "these data"
Corrected

398 comma in wrong place

Corrected

416-417 I’m not sure this sort of editorializing is necessary, and I doubt that it is discussed by Frölicher and Laufkötter.

In the 2018 Nature Communications paper that is cited, the section “Impacts on physical, natural, and humans systems” discusses these issues. Here we’ve changed the verbiage to:

- ...can create challenging social and political environments stemming from the associated economic impacts.

As for the following sentence (418-420), the intended meaning is fairly clear but the wording could be improved.

Changed to:

- In the future, we should anticipate these ecosystem shifts as MHWs are expected to persist (Xu et al. 2021) and the atmospheric pressure systems associated with extreme events will increase in frequency (Giamalaki et al. 2021).

Figure 7d, 8d unit should be nM?

Yes. Corrected

Figure 9 unit needs a space between mg and m-3

Corrected

Figure 9 caption: There are a bunch of details about this Figure that are not really explained in the caption: the meaning of the vertical bars (probably mean, but needs to be stated, and panel (b) is different from the other 3), the vertical position of the symbols (arbitrary, but again should be stated), and the meaning of the symbol colours (obvious from the positions, but in this case is having two colours even necessary?) And there appear to be more years than there are symbols.

The new caption will read:

- Observed and modeled summer (May–Aug) Chl (mg m-3) contained in the large (left) and small (right) phytoplankton size fraction in two regions: (a,b) Alaska Gyre and (c,d) North Pacific Transition Zone. Model data are shown as normalized probability density functions for the MHW composite (red) and the climatology (gray) with the mean of each shown as a vertical bar on the x-axis. Observations from the six OSP cruises in the Alaska Gyre are shown as symbols on panels a and b for large (purple) and small (blue) phytoplankton respectively, (note: y-axis placement is arbitrary) The data for years 2000, 2001, 2008, and 2018 are filled circles while data from the 2015 warm blob (star) and 2013 volcanic eruption (hollow circle) have varying symbols due to their association anomalous events. See also method Section 2.4.

Please also note the supplement to this comment: