

Reviewer comments in blue

Author comments in black

We thank the editor and both reviewers for their detailed and helpful feedback. The comments were very useful at improving the clarity of our paper and strengthened our results. In particular, we have added to and clarified the discussion of the significance of our results, adding a new section 2.2. We hope that the current version is acceptable for publication in Biogeosciences. Line by line responses to the reviewers are detailed below.

Referee #1

These authors have mostly done a good job of addressing the previous reviews. Some aspects of the statistical analysis are inadequately explained, and there are a number of ambiguities in the figure captions /legends.

Main points:

(1) I applaud the authors' efforts to add tests of statistical significance, but the description of what was done remains incomplete and unconvincing. The description on 87-94 is too terse to be useful. How exactly we get from what is described here to a significance level of 0.03 (253), 0.04 (330), or 0.001(331) is quite mysterious. Let's consider Figure 7b for example: there are 9 MHW years, and the Large Phytoplankton anomaly is negative in all of them. So it would certainly seem plausible that the probability of that occurring by chance is <5% or <1%. But one can not tell from the present text how that was demonstrated. We need more detail e.g., about the effective sample sizes. If we have 63 years of monthly data, but we are considering only a seasonal mean anomaly for the MHW years, then $N_2=9$ and $N_1=$ at most 63. If we are considering monthly data then $63*12$ is almost certainly an inflated estimate of N_1 because of autocorrelation within the time series (e.g., 10.1175/1520-0442(1997)010<2147). If we are using annual or seasonal mean data then autocorrelation should be weak and $N_1\sim 63$ will not be too far off. In any case, whatever assumptions the authors are making, they need to spell them out in more detail.

We thank the reviewer for emphasizing the need for clear statistics. We have decided to include a separate section to discuss our statistical method. See section 2.2.

In response to this specific comment, we have used the annual anomaly (or seasonal mean anomaly as specified) for MHW ($N=9$) and non-MHW years ($N=53$), as now described in section 2.2.

(2) Two issues that are partially unresolved from the previous review (there are few minor ones as well, see below) are the description of iron limitation of small phytoplankton and question of cell-size-dependence of carbon-to-chlorophyll ratios. I still find the text ambiguous regarding whether the lack of iron-limitation of small phytoplankton is an emergent property or is programmed into the model a priori (e.g., 118, 283).

We have rewritten this part of the methods section. An iron deficit term is calculated for all size-classes; however, model results show it is only the “limiting nutrient” for large phytoplankton in this study. L131

“For macronutrients (e.g. nitrogen, phosphate), limitation factors are calculated using saturating kinetics while for iron, an internal iron deficiency term is calculated based on an internal cell quota (see supplementary materials in Stock 2014 for details). These limitation factors are output from the model as a number between zero and one, with zero indicating complete limitation, i.e. no phytoplankton growth, with the lowest value considered the limiting nutrient (Liebig 1840). Notably, in the study regions iron is only a limiting nutrient for large phytoplankton.”

And the repeated claims that large phytoplankton have a systematically higher Chl:C ratio (e.g., 264, 290, 323) are unconvincing. Chl:C is function of temperature, irradiance, and cell nutrient status. There may be taxonomic differences, but they are generally much smaller than variation with environmental conditions. The repeated invocations of this idea sound glib and superficial and explain very little.

On L140-142 we specify how the simulated Chl:C ratio is calculated within the model as a function of temperature, light and nutrients, in agreement with the reviewer. When we reference the higher Chl:C ratio of large phytoplankton, however, we are referencing the model output Chl:C ratio for the 2 study regions. On L283 we specify what that ratio is for both size classes in the NPTZ.

“Seasonal Chl largely follows the large phytoplankton production due to a higher Chl:C ratio for large phytoplankton (0.022 vs. 0.014) simulated in this region (Geider et al. 1997; Stock et al. 2020).”

We have also added the simulated Chl:C to the AG section (L343-345) for completeness.

“The negative Chl anomaly that starts in the spring (April) is due to the decreased large phytoplankton, which have a higher simulated Chl:C (0.027), offset by the increased small phytoplankton (Chl:C = 0.016) production anomaly”

For clarity, we have added “simulated” to our references to the Chl:C ratio to make it clear that we are here talking about the modeled ratio.

(3) I think section 3.3 is too long and contains some superfluous text (e.g., 259-271). If this material is important enough to keep in the Results then the data are important enough not to be relegated to Supplementary. I would move this to Discussion or delete.

We have trimmed down that section of text to focus on the need to understand the seasonal cycle before discussing the MHW anomalies. However, these figures are best suited to the supplementary material to keep the main text streamlined.

Some details:

14 change "Yet" to "However" (see also 254, 350, 363)

Changed to “However” in all instances

33 delete "in the Northeast Pacific"

Removed

37 change "and/or" to "or"

L36 Change to “or”

51-52 "surface Chl alone provides little information on food web changes beyond primary production" Actually it doesn't necessarily tell us anything about production, just biomass.
L52 Yes, we have deleted this comment.

60 Polovina not in ref list

Thank you. Reference has been added.

111 "Tropics" should be "Trophics"

Yes, corrected.

113 not sure what the stray "12" means

Typo removed

140 "Mt. Pavlof eruption" is there a literature reference for this?

L158 Reference to Waythomas et al., 2014 added

142 "(2018)" something missing here

"2018" removed

145 italicize taxonomic names

L165 *Synechococcus* is now italicized

187 "west of P8 (P4–P20" P4 is east of P8

L208 Changed to "west of P4 (P4-P20)

188 "climatological" maybe not best word here

L209 Rephrased to "despite a model bias toward lower climatological surface nitrate"

190 "due to a lack of data prior to 2011" figure shows data before 2011

Yes, this statement should have been removed in the last iteration. It has been removed now.

203 add "North" before "American"

L224 Added

210 "intrusion of the nitrate-depleted region from the south into the NPTZ" expansion of the nitrate-depleted region northward into the NPTZ

L231 Corrected as suggested

217-220 I would change "the temporally variable, high-nitrate near-shore region" to "the highly variable but generally nutrient-rich near-shore region" and "the depleted nitrate region" to "the seasonally nitrate-depleted region"

L239-240 Corrected as suggested

221 change "Observations show" to "Ship-based Line P observations show"

L242 Corrected as suggested

223 delete "lateral"

Deleted

244 micromolar should be nanomolar?

L266 Yes, units corrected to nM

246 change "primary production" to "phytoplankton production"

L268 Changed to "phytoplankton"

275 change "dropping" to "declining"

L295 Changed to "declining"

281 "starts nearly a month earlier (early April vs late April)" how do they know this if working with monthly data?

The onset of nitrogen limitation is indicated by the intersection of the iron and nitrogen limitation shown in figure 8h, referenced immediately prior to this sentence. This sentence has been changed to the following:

L300 "the onset of nitrogen limitation, which occurs when the nitrogen limitation factor (dotted red line) intersects and drops below the iron limitation factor (dotted blue line), happens nearly a month earlier (early April) than the climatology (solid lines, late April) and the limitation factor is significantly lower"

284 something missing after "-2 mmol/m²/d"

L303 "for large phytoplankton" added

329 "andwithin"

L351 Space added

339 "further supports this shifting of the phytoplankton assemblage" further supports the hypothesis of a shift in the phytoplankton assemblage

L362 Corrected as suggested.

360 "Our results agree with this literature" Our results support these previous studies

L385 Corrected as suggested.

383 change "depleted nitrate" to "nitrate-depleted"

L407 Changed to nitrate-depleted

397 "three different nitrate ..." something missing here

Corrected to "three different nitrate fields"

400 " Note that for ... the year is also used." On the one hand this seems like unnecessary belaboring of the obvious. But I understand the impulse behind it, as reviewers are often dense and demand explication of things like this. Maybe it can be worked into the figure caption.

L421 We have opted to keep this sentence for clarity.

414-415 This sentence borders on tautology. Maybe just delete everything after "both biomes".

We have removed the clause after “both biomes” as suggested.

419 change "relieved" to "reduced"

L442 Changed to reduced

428-429 " iron returns to near-climatological levels by summer" this needs a data reference (e.g., Fig. 9d)

L 453 “(Fig. 9d)” Added

438 "Evidence of this shift has been observed in the AG during the “warm blob” (Peña et al. 2019) which found higher concentrations of cyanobacteria in the nitrate-depleted region of Line P and data presented in this paper, which show higher Chl a concentrations in the smaller size classes at OSP (Sect. 3.3, 3.4)." Break this into two sentences.

L462 – 464 hanged to:

“Evidence of this shift has been observed in the AG during the “warm blob” (Peña et al. 2019) which found higher concentrations of cyanobacteria in the nitrate-depleted region of Line P. Further, the data presented in this paper show higher Chl *a* concentrations in the smaller size classes at OSP (Sect. 3.3, 3.4).”

446 change "persist" to "recur"

L470 Changed to “recur”

Figure 1 - The meaning of the rectangular box does not appear to actually be stated in the caption (I assume it is the region averaged in panel a). The contours in panels d and e are stated to be gray but appear black, and there is an additional black line whose meaning does not appear to be stated. Anyway, why use the same color for the contours and the Argo float trajectories? and why not use the same contour levels for b/c and d/e? There are only 5 panels (a-e), but there is a reference in the text to "Fig. 1 e & f" (168).

Caption has been updated to include box description (which is the box averaged in panel a, as assumed). All contours in b-e are colored dark gray, distinct from the Argo floats, and are the same values for SST panels (b,c) and chlorophyll panels (d,e). Text reference has been corrected to “Fig 1. d & e).”

Figure 3 - Caption is incoherent and some labels appear incorrect. (a-c) and (d-f) should be a, c, e and b, d, f? Title above panel b should be MOM6-COBALT?

Caption has been changed to correctly reference panels. Right side panels were incorrectly showing a similar analysis from the WOA. The WOA figures have been moved to the Supplementary Info and replaced with the Model sampling as described.

Figure 4 - The map of the station positions is in Figure 1 not Figure 2. The map does not show the station numbers, so the reader is being asked to "read between the lines" here, although in this case the inference is fairly obvious. However, there are many more ticks on the plot than there are labels, or symbols in Figure 1 (i.e., all of the minor stations are shown) and the caption should explain what these are. I still think "2 μ M nitrate contour" is more appropriate than "2 μ M nitrate boundary" (see also e.g., 211, 215, 232).

Caption changed to reference station map in Figure 1. Caption in figure 1 has been updated to explicitly mention the numbering of Line P stations.

“Line P (P1-P26) shown as yellow circles at every fourth Line P station with the black star denoting P26, also known as ocean station papa (OSP, 50° N, 145° W).”

“Boundary” has been replaced by “contour” in caption and throughout (L...).

Figure 5 - "boxes for AG and NPTZ are shown as described in Fig 1" Actually, Figure 1 has only one box and it is not explicitly defined.

Descriptions of the two study zones shown as boxes has been added to caption.

Figure 6 - MLD, POC not defined. Iron should be in nM? Change "Modeled composite of the 9 marine heatwave anomalies" to "Modeled composite anomaly of the 9 marine heatwaves".

MLD and POC now spelled out in the caption. Units for iron corrected to nM.

"Modeled composite of the 9 marine heatwave anomalies" changed to "Modeled composite anomaly of the 9 marine heatwaves" as suggested.

Figure 7 - The abbreviations "Lg" and "Sm" do not appear in the figure itself, and nowhere else in the text, so I would say they are expendable. Since the AG and NPTZ data are shown on separate panels, why does each warm composite need its own color?

Lg and Sm abbreviations have been dropped. We agree distinct coloring in this figure isn't necessary, but it follows the coloring of the boxes in Figures 5, 6 & 10 for continuity.

Figure 8/9 - Again the averaging box is misspecified: the box in Figure 1 is not black and its boundaries don't fit what is stated here. Also the y axis units do not appear to be stated anywhere. And there is something wrong with the legend to panel (h): I think "Mod MHW climatology" is not correct. And the main legend has a solid line for individual years while dashed lines are actually used.

Figures 8, 9 now correctly reference the black box representing the NPTZ and the red box representing the AG described in Figure 5. Y-axis labels have been added to all variables.

Legend in panel h has replaced “Mod MHW climatology” with “Mod climatology”

Figure 10 - Definitions of regions need to be stated (e.g., by reference to Figure 5 once its problems are fixed). Meaning of vertical bars still not explained.

Reference to Fig 5 added as well as the meaning of the vertical bars.

Figure 11 - y axis label should be anomaly in multiples of sigma?

I believe this comment references Figure 7? We have changed “units of sigma” to “multiples of sigma”

Maybe change "winter supply" to "winter inventory".

Caption for Fig 11 changed as suggested.

Terminology:

I have doubts about the usage of "Chl" as synonymous with "chlorophyll". When this abbreviation is defined, it refers to satellite chlorophyll (31). But then it is used to refer to field

measurements (e.g., 222), and there are several references to "Chl a". While oceanographers often use "chlorophyll" and "chlorophyll a" interchangeably, chlorophylls are a family of pigments of which chlorophyll a is only the most common (e.g., 101/m101p307). Satellite chlorophyll is an operationally defined quantity calculated from outgoing shortwave radiation at the top of the atmosphere, and implicitly includes absorption by other pigments. I think "Chl" should be reserved for satellite chlorophyll, and "chlorophyll" spelled out where it refers to field (bottle) data or specific pigments.

We agree with the reviewer and have changed all instances of chlorophyll to the full spelling "chlorophyll" except when referring specifically to Chl a, which is introduced in section 2.5, L155.

It is customary to choose a discrete significance threshold like 0.01 or 0.05. When I see a number like 0.03 (253) it makes me think the authors just chose the smallest value that all of their data exceeded. Not that this is wrong necessarily, or that it implies the test is artificial (but see my major comments above). But the convention is to e.g., use $P < 0.05$ if $0.025 < p < 0.05$.

In section 2.2 we have selected $p < 0.05$ to be our threshold for when the MHW years differ significantly from non-MHW years. This value is used throughout the text.

"nitrate limitation" still appears in several places (e.g., 15, 375)

Changed to "nutrient" limitation on lines 15, 375, 385, 386

"sigma" is sometime spelled out, sometimes a symbol

Changed to the symbol σ throughout.

I don't like the terms "subtropical-like" and "subpolar-like". If the outstanding characteristic of the subtropical biome is that nutrients are depleted year round (371), why does a mid-latitude seasonal convection/stratification regime become "subtropical-like" just because seasonal drawdown becomes a bit more widespread or intense? And the use of the word "mode" (e.g., 310-311) is not consistent with its usual meaning.

Here we are specifically referencing the bimodal nature of the NPTZ chlorophyll distribution shown in Figure 10c & 10d. For both size-classes there is one mode that is consistent with the distribution shown for the subpolar AG, which we describe in lines L301-303. We now refer to this as a "high-chlorophyll mode", similar to the distribution of the subpolar gyre. The other mode is consistent with the subtropical gyre chlorophyll distribution which has a near-zero mean (L303-305) and we now refer to this as the "low-chlorophyll mode".

In in this section, we have changed the references from "peaks" to "modes" for added clarity.

Formatting:

Multiple references within a parenthesis should be arranged either alphabetically or chronologically; pick one and apply it consistently. This paper uses both, and sometimes neither (e.g., 60).

We have modified this using chronological order.

The figures are not cited in order in the text (e.g., 90).

Corrected

Always leave a space between a number and its unit (e.g., 228)

Corrected

Referee #2

This paper focuses on quantifying and explaining ecosystem impacts of marine heat waves in the NE Pacific, using a modelling approach supported by observational comparisons. The focus is on anomalies in chlorophyll and phytoplankton and their drivers. The topic is of high interest to the community. Some statistical tests for significance have been included in the revised manuscript, but in many places the significance is still unclear. This makes the conclusions of the paper unconvincing. The reader is left wondering whether there is any significant difference between the collection of MHW-years vs. normal years besides the large:small phytoplankton ratio. The most novel aspect of this paper is its consideration of the collection of MHW years, as previous work has investigated the 2014-15 event in some detail, so clarity in what is significantly different between the two populations of years is important. A few of my other previous comments have also not yet been fully addressed or revisions not fully incorporated throughout the paper.

The paper presents two new tests for significance: the Welch's T-test and a test for whether the mean of the MHW years falls outside the standard deviation of the climatology.

- For both tests, it's not clear whether the population that the MHW years are being compared to includes those years or not. I would recommend comparing the MHW years to the collection of years that are not MHW years. The text discusses comparison to climatology which would presumably include all years, MHW and non-MHW. Suggest adding statistics section in the methods and clarifying what tests were used and how they were done. In some cases, the conclusions of these two tests conflict with each other. For example, the p-value from the T-test suggests that lower small phytoplankton production, large phytoplankton production, and zooplankton production in the NPTZ during MHWs is significant but Figure 7 shows that the MHW mean is within 1 standard deviation of the climatology variability. I think it would be most straightforward to choose a single metric for significance and apply this across the board. Also, present an argument for why this is the best metric.

It is correct that this paper initially compared MHW years to the climatology, which did include all years.

We have updated our methods (Section 2.2) to now use the Kolmogorov-Smirnov (K-S) test instead of the Welch's T-test to evaluate ,using a significance threshold of p-value < 0.05 , this test enables us to conclude that the ecosystem changes are a result of the MHW events.

Following this suggestion, we have updated our significance analysis so that MHW years are now compared to non-MHW years.

The comparison of MHW anomalies to the interannual standard deviation, is however not used to evaluate significance but to quantify the amplitude of the perturbation. As we show in Fig 7, some changes tied to MHW are significant (K-S T-test p -value < 0.05) but small (amplitude lower than 1 std). This suggests that there are sources of interannual variability that contribute more than MHWs (e.g., ENSO ref). Following the reviewer's suggestion, we have clarified our method by including a new, separate section (Methods 2.2) to discuss our statistical method. In there, we also added a sentence that explains the use of the standard deviation and modified the text in result/discussion to clarify.

There are multiple places in the paper where significance is not discussed at all when differences between all MHW and non-MHW are presented or where apparent differences are presented, discussed, and interpreted but then shown to be insignificant almost as a caveat. Where differences between MHW and non-MHW years are not significant, I don't think they should be presented and discussed at all, except to say that they are not significantly different. Below, I point out areas where statements of significance are required or removal of discussion of differences if they are not significant.

When it was determined that differences were not significant, we have either removed the quantification of the change, unless relevant to discussion.

e.g. L294, decrease in spring large phytoplankton production has been removed.

- Lines 209-213, 629-630: differences in the northward displacement of the 2 μ M nitrate boundary

The argument we make here is qualitative. We have updated figure 5 to show that during MHW this limit is located north of the climatology.

For all other instances below, we have added "significant" or clarified the K-S's t-test p values in parenthesis (following the statistical methods description in section 2.2).

- Lines 167-168: differences in chlorophyll in each region
- Lines 240-244: differences in NPTZ chlorophyll, mixed layer depth, winter surface nitrate, and iron
- Lines 276-284: differences in NPTZ small phytoplankton production late winter / early spring, zooplankton production through May, iron limitation factor Jan-Apr, onset of nitrate limitation, nitrate limitation factor in ?late spring / summer?, small phytoplankton in June, total zooplankton production in June.
- Line 288: differences in NPTZ annual mean surface Chl anomaly
- Lines 303-310: differences in NPTZ large and small phytoplankton chl, % area of NPTZ with low large phytoplankton chlorophyll, % area of NPTZ with low small phytoplankton chlorophyll
- Line 320: differences in AG small phytoplankton production. The text later in this paragraph suggests this difference is not significant, in which case I don't think it should be presented and discussed.
- Lines 343-345: differences in AG large phytoplankton chlorophyll
- Lines 361-363: differences in AG and NPTZ nitrate and iron

- Lines 375-376: differences in NPTZ nitrate limitation factors
- Lines 426-432: differences in AG large phytoplankton production, annual small phytoplankton production

Treatment of the large:small phytoplankton ratio. Sometimes special statistics are needed when dealing with a ratio, because the variance may not be normally distributed in the ratio even when it is in the individual numerator and denominator. Please consider whether you can perform the same significance tests on the ratio.

To address this concern, we decided to do our statistical analysis of all variables between the non-MHW years and MHW years using a two-sample Kolmogorov-Smirnov test, which does not require a normal distribution or equal variances between the two datasets. We make note of this in section 2.2.

Comment from previous review: In general, the iron concentrations in the model (Figure S2) seem very high with modelled values around 100-150 nM iron in winter, whereas typical values for dissolved Fe in the region appear to be < 1 nM (see <https://doi.org/10.1016/j.marchem.2015.04.004> for example).

This difference raises concerns that the model does not handle iron well in this region. Suggest that the authors explicitly discuss the model's simulation of iron, including comparison to observations, and the biases in that, especially since some of their conclusions rest on changing iron limitation.

This unit was incorrectly labeled, as noted by Reviewer 1. Surface iron concentrations throughout the paper are in units of nM (not μM). We have made sure this unit is correctly given in figures 6, 8, 9, S2, S3, S4 as well as correctly identified in the text L266. This corrects the model values to <1 nM (0.1 – 0.15 nM) throughout the region, in agreement with the cited reference.

How are the limitation factors defined? The authors have included a new statement that “limitation factors are output from the model as a number between zero and one, with zero indicating complete limitation, i.e. no phytoplankton growth”. However, this does not clarify how the factor is calculated. Are these limitation factors the Michaelis-Menton values? If so, it would be useful to include the equation and limitation coefficients assumed for each species (in the supplemental material would be fine for this with a reference in the main text).

We have not included the equations as that is previously published in detail in Stock 2014. However, we have rewritten this section of the methods to include more details and clarify the equation number to look for in Stock et al L131-135.

“For macronutrients (e.g. nitrogen, phosphate), limitation factors are calculated using saturating kinetics while for iron, an internal iron deficiency term is calculated based on an internal cell quota (see supplementary materials in Stock 2014 for details). These limitation factors are output from the model as a number between zero and one, with zero indicating complete limitation, i.e. no phytoplankton growth, with the lowest value considered the limiting nutrient (Liebig 1845). Notably, in the study regions iron is only a limiting nutrient for large phytoplankton.”

Minor suggestions:

Lines 100 and 190-191: Author response states that data from earlier years has now been obtained and incorporated into the paper. Revise text accordingly.

L114 Thank you, this oversight has been corrected and the references to the missing data have been removed.

Line 113: Phytoplankton12 ?

Typo removed

Lines 191-192: Figure 1 doesn't really support this statement. Modelled and satellite chlorophyll shown are for four years, not just the 2014-15 anomaly. The chlorophyll decline shown in Figure 1 is much less than the stated 0.3 mg m⁻³ and is found a lot farther south than Line P.

Reference to Fig. 1 has been removed

Lines 254-255: Confusing wording. "for all events" implies that this is true of each event, when the sentence refers to the composite.

Changed to "for the composite"

Figure 1: Suggest including the same contours in 1c as in 1b. Colour bar and axis labels overlap in 1c. Caption should specify what grey boxes indicate. Line 606: BGC-Argo floats with nitrate sensors (not all Argo floats).

Contours in 1c have been specified to match 1b. Color bars shifted. Caption updated as suggested.

Figure 3: The text around lines 173-175 say that the right hand side of this figure shows model data, but this part of the figure is labeled WOA – a climatology. High temperatures and low nitrate concentrations associated with the Blob are not visible in the WOA labeled panels.

Yes, this is an error. The WOA profile has been moved to the supplementary material and Figure 3 now appropriately has model data shown on the right-hand side, as described in lines 173-175.

Indeed, figure S2 showing the WOA profiles does not have the high temperature and nitrate concentrations associated with the warm blob because the sampled WOA nitrate field is climatological. However, the apparent shoaling of the thermocline and nitracline that occurs in 2015 is present, indicating that these features are not a result of the warm blob. Instead, this supports our argument that float #5904125 sampled a new water mass due to its northwest trajectory across the spatial nitrate gradient.

Figure 4: The lack of Line P chlorophyll observations near the coast seems incorrect. These regions are sampled on nearly every Line P cruise. The lack of surface salinity data also seems incorrect, particularly where there is temperature data.

This was an artifact of our colorbar selection. In the updated Figure 4, both salinity and chlorophyll show full coverage at the coast.

Figure 5: The boxes in this figure are not the same as Figure 1, and so should be described in the caption. Adjust caption for Figure 6 for the same reason.

Figure 5 caption has been changed to include box descriptions.

“The North Pacific Transition Zone (NPTZ, 39°–45°N and 160°–135°W) shown as a black box and the Alaskan Gyre (AG, 48°–54°N and 160°–145°W). Line P stations and OSP are shown as described in Fig 1.”

Figure 6 caption now references the description in Figure 5.

“Line P, OSP as shown in Fig. 1; boxes for AG and NPTZ are shown as described in Fig. 5.”

Figure 8: units for the values on the y-axis should be included in the figure or at least the caption. First (h) in the caption should be (g).

Units now provided along y-axes in Figure 8 (and Figure 9).

Figure 10: Caption should state what the vertical lines are.

Caption edited to include

“The mean of each is shown as a short vertical line on the x-axis (red, black respectively)”