Interactive comment on "Increase in ocean acidity variability and extremes under increasing atmospheric CO$_2$" *by* Friedrich A. Burger et al.

James Orr (Referee)

james.orr@lsce.ipsl.fr

Received and published: 11 February 2020

General comments:

This manuscript has the potential to become the first published peer-reviewed study on extreme events in ocean acidification, something that would nicely complement recent studies focused on projected changes in marine heatwaves. The subject is highly relevant for publication in Biogeosciences, the analysis is original, and the authors have clearly devoted considerable effort. Before it can be published though, more work seems needed to make the analysis more accurate and to better communicate these results to the larger community.

For the analysis, my main concern is that in the deconvolution of drivers, the authors'

equation for the Taylor expansion of the variance (Equation 2) is flawed. The bad news is that the units for the first 4 terms on the right-hand side (RHS) do not check. Those units should each be identical to the units for the sole term on the LHS, i.e., the total variance in (nmol/kg)^2, but they are not. To have the right units, the sensitivities (partial derivatives) in the first 4 terms would each need to be squared. That modification will change the balance between terms. Given that error, it is not surprising that the authors say that "these contributions can not be separated into summable terms" (line 186). The good news is that the 6 final terms in that equation do have the right units; they are correct as is. Moreover, by making these modifications, the authors should be able to get the terms to add up. With the squared sensitivities, this equation has already been given correctly in previous work such as for uncertainty propagation of CO$_2$ system variables (Dickson and Riley, 1978; Orr et al., 2018) and for analysis of the variance of seasonal to interannual variability (Ericson et al., 2018).

When using the corrected equation, the authors will need to demonstrate quantitatively that all the terms on the RHS add up to the value on the LHS. That could be done very clearly by showing zonal means on the same plot for each RHS term, the sum of all RHS variance terms, and the actual simulated value for the total variance of [H+].

A second flaw with the analysis is that when comparing different contributions, the authors usually compare standard deviations, not variances (e.g., in Figs 7, 8, 9, and A5, and the 4 equations in Appendix C). Although perhaps more intuitive because of the units, comparing the standard deviations of the different components leads to a false impression of relative importance. It is only the variances of the components that linearly add up to the total variance. When minor components are compared in terms of their standard deviations, they appear overly important in ters of their contribution to the total variance. The authors should make all comparisons in terms of variances, not standard deviations.

A third flaw with the analysis is that changes in the mean state (trend) have been removed and seldom enter into the discussion. Because most of the future change

in both [H+] and Ω_A will be due to changes in the mean state, this neglect leads the authors to make statements that make little sense, such as the following:

L152: "changes in different extreme event characteristics are only caused by variability"

L301-302: "extreme [H+] days are projected to disappear in the RCP8.5 scenario by the end of the century"

L309: "the occurrence of extremes is projected to decrease"

L314: "Ω_A extreme events are projected to disappear by 2081-2100"

L316-317: "No extreme events are projected for most of the ocean during 2081-2100 under RCP8.5"

Hence there is a communication problem that is directly tied to the authors' peculiar meaning for "extreme event". Unless the authors can bring the mean state back into the picture, they cannot legitimately use the term "extreme event". They are currently focusing only on a diagnostic for changes in variability. To help remedy the problem, I would like to see the authors provide a quantitative analysis of the contributions of each of the temporal components, including the change in the mean state (trend), to the overall change in the maxima. They might be able to do this without repeating their entire analysis procedure, simply by computing the variance due to the change in the mean state and adding that to the total of the other variance contributions that have already been computed. This analysis would clearly demonstrate the dominance of the change in the mean state and properly put the authors' other results into context.

The study of marine heatwaves by Froelicher et al. (2018) included the change in the mean state as part of the analysis, so it is even more unclear to me as to why the same was not done here for the analysis of extreme events in [H+] and Ω_A. Moreover, the relative importance of the change in the mean state relative to other temporal fluctuations (subannual, monthly, interannual) seems to be much larger for [H+] and Ω_A relative to SST. Its prominent role needs greater emphasis in this study by Burger et al.

C3

Specific comments:

L2: Please define what is meant by by "acidity". Acidity is a term used by aquatic chemists to refer to base neutralizing capacity, just as alkalinity is used to refer to acid neutralizing capacity. Acidity does not a priori refer to [H+] just as alkalinity does not refer to [OH-]. The meaning of the authors ([H+]) differs and has to be defined. If acidity is to be used, I would be more comfortable with the term "free acidity" when referring only to [H+] because that is only part of the total acidity that is not bound up in other ions that react with [OH-].

L3: It is unclear what is meant by "mean ocean acidification". Ocean acidification involves changes in many CO_2 system variables simultaneously.

L28: Please define "short term". Also add a hyphen (see Global changes section below).

L61-62: The sentence refers to changes in [H+] and Ω_A and cites references most of which only addressed the seasonal cycle of pCO_2. Please separate the references so that the readers know which one(s) actually addressed [H+] and Ω_A.

L76: "daily output" is ambiguous. Please say "daily mean output" if that is what you mean. If more frequent, say something like "6 hourly output". This is an important point because the analysis does not seem to include the potentially large diurnal cycle if it is based on daily mean output.

L77: It seems ambiguous to write about "daily variability". That could be misunderstood by readers to mean "diurnal variability", which the authors did not consider because they are presumably using daily mean output.

L79-81: Please delete the last 3 lines. It is much nicer to end the Introduction with the aim of the study. Moreover, subsequently diluting the aim by following that with an outline of the paper makes no sense, particularly when that outline follows the standard IMRAD format (Introduction, Methods, Results, and Discussion) which is what is

C4

expected anyway.

L98: I suspect that the GFDL-ESM2M model does not simply use the K1 and K2 directly from Mehrbach et al. (1973), unlike what is stated, but rather the K1 and K2 from the Mehrbach data after being converted to the total hydrogen scale by either Dickson and Millero (1987) or Lueker et al. (2000). Please clarify. Another important point that is not mentioned which concerns the model's total alkalinity equation. Does the total alkalinity equation in the GFDL-ESM2M model include contributions from phosphoric and silicic acid systems. Many models do not, and this can bias pCO_2 in the high latitudes by 10 ppm or so. [H+] would also be affected.

L100: This mention of diurnal variability seems irrelevant if the authors are using only daily mean output, which seems to be the case. Please delete.

L140: The mention of the software seems too vague. What function was used from the library?

L148-149: It is not clear to me why each ensemble member is detrended with the ensemble mean trend rather than the trend in the individual ensemble member. I understand that it is easy conceptually just to calculate the ensemble mean trend and use that. But one could also use a spline to detrend each ensemble member. Anomalies relative to the ensemble mean trend will be larger than those relative to the individual member trend. They would contain differences due to the different trends in the ensemble members as well as differences due to the other variability.

L151: Please reword. It is not that the mean state is constant. Rather, the trend was removed.

L159: It says that both the standard deviation and variance are used. As mentioned in the general comments, I think that the comparison should be made only with the variance.

Figure 1: caption - change "subtracted" to "removed". If the ensemble mean were

simply subtracted, the mean of the preindustrial period would be zero rather than 7.3. Using "removed" gives more leeway in the meaning.

L164: "daily variability" is a confusing term. It could be taken to mean "diurnal variability" by some readers. I think the authors should stick with the term "subannual" and define that carefully as meaning "with the seasonal cycle removed but ignoring diurnal variability."

L166: It is dangerous to simply compare contributions from the different standard deviations, which do not add up linearly and give a biased impression of their contributions to the total variability. Rather, use the variance of the different components, which add linearly to make up the total variance.

L168: Please clarify what is meant by "changes".

L172: Equation 1 does not seem to be used. Please just delete it and remove any mention of it in the text.

L182: Equation 2 is wrong. In the first 4 terms, the partial derivatives should be squared. This fix will change the balance of their contributions.

L186: Since your Eq. 2 is wrong, it is not surprising that the terms do not sum up to match the total simulated variance. This match is key if we are to believe the deconvolution analysis.

L209: Table A1 should be brought into the main body of the text and given as Table 1.

L215-221:

- It is unclear how trends and the error bars were computed for the data based estimates

- What is the statistical significance of the model vs. data-based comparisons? In some cases, they do not appear to differ statistically, e.g., for Ω_A in the northern

high latitudes.

L229: Is there not some question about the data-based estimate of seasonal variability in the Southern Ocean where winter data is sparse?

L230: I would recommend being more cautious, by saying something like "it appears that the GFDL-ESM2M model is adequate to assess..."

L238: What is meant by intensity?

L239: Table A2 should be brought into the main body of the paper and given as Table 2. This would allow the authors to simplify the text to mention only the mean changes and avoid giving the uncertainty ranges in parentheses). That simplification would make reading easier. Alternatively they could just give the ranges and not the mean changes. More generally, the first 4 paragraphs of this section (3.1) suffer from trying to cram too much information into every place where any numerical fact is mentioned. Writing complex things such as "0.20 nmol kg-1 (0.19-0.21 nmol kg-1; 18

L240: Please tell us what fraction of the total volume above 200 m is represented by the 2.7 x$10^3 km^3$.

L254-255: "is connected to" is vague. The increased contribution of interannual variability is not a mechanistic explanation. Previous work has discussed why changes in [H+] in the subsurface are larger than those at the surface (Orr, 2011; Resplandy, 2015). Perhaps their chemical explanation explains the longer duration of extreme variability mentioned here as well.

L270-271: Given the large uncertainty between ensemble members, can the authors really say that there is a statistically significant increase in extreme days per year, intensity, and volume of events? I see no subplot for the Volume for the 200-m analysis.

Section 3.2 and 3.3: There are many statements particularly in these sections that make no sense because when addressing "extreme events", one must include the

dominant contribution from changes in the mean state.

Section 3.4: This section will need to be completely rewritten after the fixing the first 4 terms in Eq. 2 and reevaluating the associated contributions. In addition that analysis should include a more quantitative assessment than just showing maps of the different contributions. A comparison of zonal means would be a good start. One could also compare contributions to the variance with stacked bar charts with different parts of each bar indicating different temporal components, similar to those shown in Kwiatkowski and Orr (2018).

L390: It is not clear what "nonlinear dependence" refers to. The authors use a 1st order Taylor expansion, which is linear in terms of the variances.

L396-405: The comparison of the results for the 99th and 99.9 percentiles of "extremes" is also confused by the general neglect of the effect of the mean state, which is dominant. This paragraph should be rewritten with that in mind or otherwise deleted.

L410-415: There seems a missed opportunity to compare the authors' results in terms of the effect of changes in variability to those from McNeil and Matear (2008) who assumed that seasonal variability was unchanged but advanced the time at which the Ω_A=1 threshold was reached relative to previous work that did not address seasonal variability.

L416-418: The discussion about arbitrary thresholds for [H+] is too vague. More details would be needed if it is to be kept.

L424-425: The text states, "It is therefore critical to uses daily temporal output to assess extreme events in ocean biogeochemistry." I believe that this statement tends to oversell the importance of the variability of the daily-mean output variability. Currently the authors results indicate that daily-mean variability of [H+] and Ω_A is generally a minor, second-order concern in surface waters. That is actually good news for future analysis, because saving daily mean output is not feasible on a routine basis, especially

considering the many CMIP models and scenarios.

On the other hand, the contribution to the total variance from diurnal variations, a component that the authors do not quantify, may be much larger especially in coastal regions. An assessment of that variability in the observations and in high-resolution models should remain a priority.

L447-449: The discussion about the "partial" diurnal cycle in TOPAZv2 (GFDL-ESM2M) does not seem relevant because by using daily means, the authors do not assess diurnal variability. This sentence should simply be deleted. It will only confuse the reader.

L455-456: It seems misleading to lead off this sentence with "While coastal species may be adapted to large variability in ocean acidity...". Although already large variability of [H+] is seen in coastal regions, that variability will also grow as atmospheric CO_2 continues to invade the ocean and coastal-water buffer capacities also decline.

L464-465: As written, this first statement in the final paragraph was known already before this study was undertaken based on previously published papers concerning the future increase in the mean state and the future growth in seasonal variations of [H+]. This new study confirms those findings. It should not say "our analysis reveals", although it could say something like it "confirms previous findings ..."

L467: What is "mean" ocean acidification?

Technical concerns:

Global changes:

- change "200 m depth" to "200 m"

- change "point in time" to "time"

- change "points in time" to "times"

- MISSING HYPHENS: the terms "short term", "deep water", and "sea ice" should all include a hyphen when they modify a noun

L38: change "The vast majority of" to "Most of the". Delete "so far"

L44:

- delete "one might expect that".

- change "exhibit" to "have"

- It is not "carbonate chemistry" that will cross critical thresholds. That is like saying "physics will cross thresholds". You could instead say "key CO_2 system variables"

L46: change "negatively impacted" to "adversely affected".

L49: change "undergo a decline in" to "exhibit reduced"

L63:

- change "suggest" to "project".

- change "is projected to" to "will"

L69:

- change "extreme" to "extremes"

- Please separate citations so that we know which ones address heatwaves and which ones address sea-level rise.

L78: change "imprint on the occurrence" to "affect"

L84: change "are performed" with "were made"

L104:

- change "1000 year" to "1000-year".

- "a new computing infrastructure" is vague. Please clarify your intended meaning.

L108: insert "the average" before "atmospheric"

L111: insert "average" before "atmospheric"

L120:

- change "500 year" to "500-year"

- delete "long"

- what is "potential vegetation"

L121: make the same first 2 changes as in the previous line but for "220 year long"

L124:

- write "daily mean" instead of "daily"

- add "the aragonite saturation state" before the symbol Ω_A

- delete the 2nd sentence

L202: change "is" to "appears to be"

L238: change "typical" to "average"

L266: delete "blue lines in". That color information should only be given in the Figure or its caption.

Figure 2: The choice of colors for the lines could be improved. It is hard to distinguish light green vs. dark green and blue vs. purple. Yellow is difficult to see on a white background. Are these colors good for colorblind people?

Figure 3: There is too much white space and repeated information. To remedy these problems, I'd suggest to start by labeling the subpanels as a matrix, i.e., with row labels and column labels. Thus you could remove the title for each subpanel and use that as the row label of the matrix. The row labels should only be given on the left (the connection with the right column is obvious). And column headings "Surface" and "200 m" should only be given at the top. The numbers on the tick marks can be kept for all subpanels, but the x-axis label should only be kept on the bottom row. The current y-axis labels should all be removed. The same approach should be taken for all subsequent figures with more than 2 subplots. The figures will then be more compact, less verbose and redundant, and readers will intuitively understand the setup of your multi-panel figures (without going to the figure caption, repeatedly).

L283-284: delete "but show distinct spatial patterns" (redundant).

L295: the meaning of "reoccur" is unclear.

L421: Please provide separate citations for the papers that addressed pCO_2 and those that addressed [H+].

L476: I like the choice of the authors in Appendix A to use "subannual" instead of "daily" variability. Please do the same in the main body of the paper.

Appendices: Please provide an equation number for each equation listed in the appendices.

Appendix C: The equations should be written in terms of variances rather than standard deviations.

Tables A1 to A3 do not follow the formatting standards used by Biogeosciences. See the author guidelines and previously published BG papers.

Colorbar: On all the maps with diverging scales, there seems to be an excessive amount of yellow and it is difficult to assess where the "zero" line is. A better choice would be a color bar with very pale blue and very pale red next to each other in the center of the color distribution.

References

Dickson, A.G., Millero, F., 1987. A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. Deep-Sea Res. Vol. 34, 1733–1743.

Dickson, A.G., Riley, J.P., 1978. The effect of analytical error on the evaluation of the components of the aquatic carbon-dioxide system. Mar. Chem. 6, 77–85.

Ericson, Y., Falck, E., Chierici, M., Fransson, A., Kristiansen, S., Platt, S. M., et al., 2018. Temporal variability in surface water pCO_2 in Adventfjorden (West Spitsbergen) with emphasis on physical and biogeochemical drivers. J. Geophys. Res.: Oceans, 123, 4888–4905. doi:10.1029/2018JC014073

Frölicher, T. L., Fischer, E. M., Gruber, N. (2018). Marine heatwaves under global warming. Nature, 560(7718), 360-364.

Kwiatkowski, L., and Orr, J. C. Diverging seasonal extremes for ocean acidification during the twenty-first century, Nature Climate Change 8, 2, 141–145, doi:10.1038/s41558-017-0054-0, 2018.

Lueker, T.J., Dickson, A.G., Keeling, C.D., 2000. Ocean pCO_2 calculated from dissolved inorganic carbon, alkalinity, and equations for K1 and K2: validation based on laboratory measurements of CO_2 in gas and seawater at equilibrium. Mar. Chem. 70, 105–119.

McNeil, B. I., Matear, R. J., 2008. Southern Ocean acidification: A tipping point at

450-ppm atmospheric CO_2. Proc. Nat. Acad. Sci., 105(48), 18860-18864.

Orr, J. C. Recent and future changes in ocean carbonate chemistry, In: Gattuso, J.-P. and Hansson, L. (eds.), Ocean Acidification, Oxford Univ. Press, 41–66, 2011.

Orr, J. C., Epitalon, J. M., Dickson, A. G., Gattuso, J. P., 2018. Routine uncertainty propagation for the marine carbon dioxide system. Marine Chemistry, 207, 84-107.

Resplandy, L., Bopp, L., Orr, J. C., Dunne, J. P., 2013. Role of mode and intermediate waters in future ocean acidification: Analysis of CMIP5 models. Geophys. Res. Lettr., 40(12), 3091-3095.