Increase in ocean acidity variability and extremes under increasing atmospheric CO_2
Response to reviewers' comments

Friedrich A. Burger, Thomas L. Frölicher, and Jasmin G. John

May 2020

Response to Editor

Dear Author,

Thank you for providing thorough replies to the referees' comments. Please revise the manuscript along the lines of your replies. Please do consider the suggestion to add a figure highlighting the various steps of the approach you followed. It would be extremely useful to the readership. You mentioned your concern of having too many figures already. Some of them could be moved to the supplementary information.

The revised manuscript will be sent to reviewer #2 who indicated his willingness to review the revised version.

Best regards,

Jean-Pierre Gattuso

→ Dear Editor,

we have revised the manuscript according to our replies to the reviewers' comments. We thank for the editorial work and for the helpful comment to add the figure that highlights the steps taken to decompose $[H^+]$ variability. We have added this new figure to the manuscript. In order not to increase the amount of figures, we have now merged Figures 4 and 6 (in the previous manuscript version) into one figure (Figure 5 in the revised manuscript).

Best regards,

Friedrich Burger

Reviewer 1

General comments:

This is a nice manuscript that assesses extreme chemistry variability in ensemble projections of an Earth system model. The manuscript is both interesting and timely.

→ We thank the reviewer for the positive and encouraging feedback.

My comments mainly relate to improvements I think the authors could make in understanding the different drivers of carbonate chemistry variability. Particularly, I'd like to see more on the physical processes driving the differences between the RCPs and the projected frequency/intensity/duration of extreme variability events.

→ In this study, we show for the first time how extreme variability events change under increasing CO_2 and we also identify the individual drivers of changes, i.e. changes in temperature, salinity, carbon and/or alkalinity. An in-depth quantification of the physical processes that cause the changes in the individual drivers would clearly be an interesting additional analysis, but this is beyond the scope of this study. Nevertheless, we have now addressed some particular questions on physical processes driving the extremes events, such as that raised for lines 255-256, and we have added a sentence to the discussion section highlighting that an analysis of the physical processes would be an important next step.

Another general issue I think is how the authors choose to define extreme events. I would be more comfortable calling these "extreme variability events" given that mean trends have been removed. This is particularly an issue when they discuss saturation state and often give the impression that extreme events are projected to decline.

→ We thank the reviewer for this suggestion. We have changed throughout the manuscript "extreme events" to "extreme variability events". We also added a new Figure 11 that compares the changes in extreme events due to changes in variability with changes in extreme events that are caused by both changes in variability as well as changes in long-term ocean acidification.

Finally, given that the carbonate chemistry decompositions apparently do not sum, I'm not convinced of their value. I suggest removing this analysis if it can't be properly validated.

→ We think that our statement on lines 185-186 "Unfortunately, these contributions can not be separated into summable terms because $[H^+]$ standard deviation is a nonlinear function of those." may have led to confusion. Therefore, we clarify our decomposition approach in the following:

Variance of $[H^+]$ as a function of the sensitivities (partial derivatives), standard deviations, and correlations of the drivers involves terms of the form $s_i s_j \sigma_i \sigma_j \rho_{ij}$, so products of five variables (s_i denotes the sensitivity with respect to variable i here. i could be for example C_T and j could be for example A_T). Because of this large degree of nonlinearity, one can not confidently estimate changes in such a product by applying a first-order Taylor expansion on it. To state this was the intention in lines 185-186. Instead we added changes in sensitivities, standard deviations, and correlations in three steps and analyzed how much of the total variability change can be explained additionally by e.g. also taking into account variability changes on top of the sensitivity changes. This procedure included the nonlinearities in Equation 2 by construction and was thus exact. We have now streamlined the text to make the method more transparent.

However, in response to reviewer 3, we have also extended our approach and decompose variance change by applying a full fifth-order Taylor decomposition on Equation 2. While the previous approach did not allow to track how much variability change in $[H^+]$ arises from variability changes in the drivers alone and how much from the interaction between sensitivity and variability changes, the new approach allows to quantify this aspect. The new decomposition is exact since we take into account all non-vanishing orders of the Taylor series. We then group these terms of the Taylor series into four groups: (1) contributions of sensitivity changes to the overall change in variance, (2) contribution from standard deviation changes in the drivers, (3) simultaneous changes in the sensitivities and standard deviations (these can neither be attributed to sensitivity changes nor to variability changes alone), and (4) all terms that include changes in the correlations. The method is explained in the revised section 2.2.3 and in more detail with the full decomposition in appendix C.

Nevertheless, the results of this decomposition are consistent with the previous one in the sense that (1) is the same as 'All Means' in previous Figures 8+9, (2)+(3) (the sum) is identical to 'All Variabilities', and (4) is identical to 'Phasing' in previous Figures 8+9. The partitioning of 'All Variabilities' into (2) and (3) allows to understand more accurately what role variability changes in the drivers play for $[H^+]$ variance.

In response to a comment by reviewer 3, we now also decomposed variance changes for Ω_A and added a plot showing the zonal mean decomposition (new Figure 10).

In summary, we have revised method section 2.2.3, results section 3.4 and appendix C to clarify our carbonate chemistry decomposition approach.

Specific comments:

L21-23. This could be better explained. I suggest a sentence or two more, including a full definition of omega.

→ We now write: 'The rise in $[H^+]$ is partially buffered by the formation of $[HCO_3^-]$ from $[CO_3^{2-}]$. The associated decline in $[CO_3^{2-}]$ reduces the calcium carbonate saturation state $\Omega = [Ca^{2+}]\,[CO_3^{2-}] / \left([Ca^{2+}]\,[CO_3^{2-}]\right)_{sat}$, i.e. the product of calcium and carbonate ion concentrations relative to the product at saturation. Undersaturated waters with $\Omega < 1$ are corrosive for calcium carbonate minerals.'

L43. I don't think Hofmann et al., 2011 is really relevant here as they don't assess organism adaptation/acclimation under variable chemistry regimes. Many other papers do, with mixed findings, for example see: (Rivest et al., 2017; Cornwall et al., 2020)

→ Thank you for these relevant references. We now cite Rivest et al. (2017) and Cornwall et al. (2020).

L45-48. The authors are being too concise here. Explain what you mean by undersaturation. I understand that you might have to refer to aragonite versus calcite but that should probably be already mentioned anyway.

→ We changed 'undersaturation' to 'aragonite undersaturation' and added additional explanation above on different calcium carbonate minerals and their saturation states, also introducing aragonite and calcite.

L77. Its not very clear what "residual" means in this context, as it hasn't been defined yet.

→ Following the recommendation by reviewer 2, we renamed 'residual daily variability' to 'subannual' variability throughout the manuscript.

L90. What is the depth of the first ocean level? This will useful to know, as it will be a major determinant of surface variability.

→ The MOM4p1 model has a free surface and the depth of the first ocean grid level is centered at around 5 m. We clarified: 'The MOM4p1 model has a free surface and the surface level is centered around about 5 m depth and the spacing between consecutive levels is about 10 m down to a depth of about 230 m (Griffies,2009).'

L120. What is potential vegetation? Are you referring to a coupled terrestrial carbon cycle? It is not clear what relevance this has.

→ Potential vegetation refers to a terrestrial carbon cycle setup without land-use change. However, we agree with the reviewer that this is not relevant in the context of the paper and we deleted this additional information.

L142. Do you mean that extremes that last over a change in year are split in two?

→ Yes, this is correct. No changes are made to the manuscript.

L143-145. It would be worth saying something about why this upper ocean region is so important e.g. location of most reef forming corals, calcifying phytoplankton etc.

→ We added a sentence to motivate our choice: "We focus our analysis not only on the surface, but also on 200 m depth to study changes in extreme events within the thermocline, where most organisms susceptible to ocean acidification are found, such as reef-forming corals and calcifying phytoplankton."

Fig 1. Legend. This second line of this needs clarifying. I guess you're subtracting the ensemble mean change not the ensemble mean. You should say what reference years are used to calculate this ensemble

mean change.

→ Thank you for the pointer. We have modified the sentence to: '(b,d) Same as (a,c), but the ensemble-mean change with respect to the average of the 500-year long preindustrial control simulation has been subtracted.'

L160-165. The methods here are quite convoluted. An illustrative figure highlighting the different steps in the approach would really benefit readers. This could go in the main text or appendices.

→ We we agree and we have added an illustrative figure. Furthermore, we clarified that we used three steps to assess whether changes in low or high frequency variability cause changes in extreme variability events and their characteristic. To keep the number of figures at a reasonable level, we have combined the previous figures 4 and 6 into one figure (now Figure 5).

L185-186. This suggests something is wrong with the decomposition. At any rate I'm not sure you can call this a decomposition if the separate terms don't sum. How far off summing is the decomposition? If it's not working its value is highly questionable and probably shouldn't be included.

→ We have revised the carbonate chemistry decomposition. Please see more details above.

L206-209. Has something like Table A1 been published elsewhere? I would move Table A1 into the main text. It's a nice finding that your observation-based product shows $[H^+]$ seasonal amplitude increases and variable trends in omega seasonal amplitude. This forms a nice link between this work and the Landschützer et al., 2018 and Kwiatkowski & Orr, 2018 papers.

→ We thank the reviewer for this comment and we agree that this has, to our knowledge, not been published elsewhere yet. Reviewer 2 has made a similar comment. We have moved the Table from the Appendix into the main text.

L239. What is meant by coherently here?

→ We clarified in the method section the calculation of the volume of individual extreme events: 'mean volume of clusters of connected grid cells that are above the 99$^{\text{th}}$ percentile'. We therefore deleted the word coherently here.

L255-256. Is this true? In Fig 3c/d this appears true under RCP8.5 but the opposite seems to occur under RCP2.6. Indeed the projections at 200m under RCP2.6 are very interesting and quite different across metrics. Any idea why the duration is so much more responsive? Some more detail here would be great. Are there reductions in stratification post 2040 in RCP2.6? I wonder if greater vertical mixing under mitigation might be driving the halt/slight decline in the duration of events and the difference in lags across metrics.

→ Yes, the changes (in a relative sense) are larger at the surface than at 200m depth both for the RCP8.5 and for RCP2.6, as can be seen from Table A2 (now in the main text as Table 2). We clarified this by adding the word 'relative': "However, projected relative changes over the historical period and the 21$^{\text{st}}$ century are smaller at 200 m than at surface...". Global mean event duration at 200m depth is declining under RCP2.6 in the second half of the 21$^{\text{st}}$ century mainly due to the decrease in event duration in the subtropics. There, the reduction in duration is connected to reductions in the contribution from interannual variability to total variability and partially also to reductions in total variability. We have added this to section 3.1 and write there: "This decrease in duration mainly occurs in the subtropics, where events generally last long (Figure A2b). It is connected to an increase in the contribution from high-frequency variability to total variability in those regions over that period.". We think that the mechanism proposed by the reviewer (increased vertical mixing under mitigation leading to increased high frequency variability at subsurface and thereby to reduced event duration) is interesting. However, a complete analysis of the physical processes behind the changes in duration under RCP2.6 is beyond the scope of the study. We added a sentence to the discussion section that further studies are needed which focus on the physical processes.

L275-279. More care needs to be made when making these sort of statements as the mean decline in OmegaA has been removed. I would call this variability/extreme variability not extreme events/extreme days.

→ We have changed the notation of extreme events to extreme variability events throughout the manuscript

and we also added a clarifying sentence in this section: "It should be noted that, despite this decline in extreme variability events, the long-term decline in the mean state of Ω_A still leads to more frequent occurrence of low values in Ω_A (see Discussion section)"

L299-300. Could this be because the areas of upwelling are moving polewards in the model (see Rykaczewski et al., 2015)? Poleward of these grey regions there appears to be a general increase in extremes, which fits this narrative. Can you check upwelling or some proxy of this in the model?

→ Except maybe in the California Current Region, Figure 6b does not show a general increase in extremes polewards of the EBUS. No changes are made to the manuscript.

L313. Extreme "variability" events would perhaps be more accurate (here and elsewhere in the manuscript).

→ We agree and changed it throughout the MS. Please also see reply above.

Section 3.4 As stated above, it's hard to have confidence in the decomposition if it doesn't sum to the model realisation. Maybe the authors would be better to focus on physical processes (upwelling/mixing/ice loss) and how they might explain changes in variability.

→ Section 3.4 has been revised. Please see above for detail. Although we agree that the investigation of the underlying physical processes is interesting, we believe that it is out of the scope of the present study. We added a sentence in the discussion section stating that there is more work to do on the underlying physical processes.

L394-395. Some further detail is needed here I think.

→ We rewrote the sentence to "Extreme variability events in Ω_A are projected to become less frequent in the future. It is because Ω_A, unlike $[H^+]$, becomes less sensitive to variations in the drivers with the mean increase in C_T. Furthermore, the projected reductions in the drivers' variabilities, mainly in C_T, significantly add to the reduced occurrence of Ω_A variability extremes."

Section 4. I recommend dividing this into a few small subsections.

→ Many thanks for the suggestion. However, we decided to keep the structure of the discussion and conclusion section as is.

L411-414. But presumably once omega<1 is reached, the ocean spends a greater amount of time undersaturated when this reduced variability is taken into account. Can you comment on this?

→ This is certainly true. We now discuss this aspect in the new discussion paragraph about extreme events defined by a fixed baseline. We write: "Interestingly, the GFDL ESM2M projects that surface mean $[H^+]$ overshoots the preindustrial 99^{th} percentile in year 1975 on global average. Thereafter, higher variability actually reduces the number of extreme event days that are above the preindustrial percentile. Surface mean Ω_A falls below the preindustrial 1^{st} percentile in year 1990. After that, lower variability further increases the number of extreme event days below the preindustrial percentile."

Fig 11. This is an interesting figure.

→ Many thanks!

L424-425. This is computationally a big task. Maybe the community could get by with some daily statistics output at monthly resolution?

→ Our analysis shows that the average duration of an extreme $[H^+]$ event at surface is about 10 days at preindustrial and 15 days at present-day. It can therefore not be represented by monthly data. We modified the text to: "In addition to earlier studies, we also show that changes in subannual variability contribute to changes in extreme $[H^+]$ variability events under increasing atmospheric CO_2 and that the average duration of extreme variability events at the surface and at present-day is about 15 days. It is therefore critical to use daily temporal output to assess extreme events in ocean acidity."

L448-449. This seems off topic.

→ We agree and removed the sentence.

L471. I think you need to clarify some of these definitions. To most people seasonal cycles are a form of sub-annual variability.

→ We added a clarification on the definition of the seasonal cycle: 'the seasonal cycle, here defined as the 365-day long mean evolution over the course of a year'

Technical comments:

L36. A reference is needed at the end of this paragraph. Some of those already cited in this paragraph would suffice.

→ We added Hofmann et al., 2011.

L39. The formatting of multiple references here is different to elsewhere in the manuscript.

→ We fixed the formatting issue.

Fig 1. Legend. Mention that this is for the surface ocean.

→ Mentioned at beginning of caption:' Simulated daily surface $[H^+]$ (a) and Ω_A (c)'

L221-224. It would be clearer to discuss model performance at capturing mean seasonal cycles before discussing trends in seasonal cycles.

→ Following the recommendation, we changed the order and now discuss the mean seasonal cycles before discussing the trends.

L272-273. I would rephrase this. 200m is not really the deep ocean.

→ Changed to 'surface-to-deep ocean transport" to "surface-to-subsurface transport'

Fig 5. Labels (extreme days/intensity/duration) on left of this figure would make it easier to read.

→ We have followed the reviewers recommendation, also for the analogous supplementary figure A1.

336. "during the" preindustrial

→ Changed.

Reviewer 2: James Orr

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.

→ We thank the reviewer for this positive statement and the careful and detailed review. We appreciate the suggestions on how to improve the manuscript.

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 CO2 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+].

→ We thank the reviewer for pointing us to this issue. We apologize that we have made a typo in equation 2. The sensitivities in the first four terms on the RHS should be squared (we have had applied the formula correctly, it was an unfortunate typo that made it's way into the formula in the manuscript.). In response to the reviewers' comments, we have revised all sections in the manuscript on the carbonate chemistry decomposition. More details can be found above in response to the third general comment by reviewer 1. We appreciate the proposal of adding zonal mean plots and included those in the figures. They show that the representation of $[H^+]/\Omega_A$ variance by Equation 2 works reasonably well. The Taylor decomposition of Equation 2 itself is exact since it contains all terms. The references mentioned by the reviewer are now cited in the manuscript.

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 terms of their contribution to the total variance. The authors should make all comparisons in terms of variances, not standard deviations.

→ We agree that Equation 2 suggests to use variance as variability measure. We therefore changed all variability analyses such that only variances are used throughout the manuscript.However, we would like to note here that using variances instead of standard deviations does not change one of our main conclusions that changes in mean C_T are mainly responsible for changes in extreme variability events in $[H^+]$. Furthermore, it does not changes our conclusions on the frequency decomposition of variability changes.

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"

→ We have included a new Figure 11 in the Discussion section that also shows changes in extreme events due to changes in the mean state in addition to changes in variability. We have also added a new paragraph that discusses this new figure. In addition, we have clarified throughout the manuscript that in our study, changes in different extreme event characteristics are only caused by changes in variability. We now also call these extreme events 'extreme variability events' to clearly distinguish from extreme events that are caused by long-term ocean acidification.

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

→ We have clarified: "In most of these regions, extreme variability events are projected to disappear in the RCP8.5 scenario by the end of this century"

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

→ We have clarified: "The regions in the Southern Ocean where the occurrence of extreme variability events is projected to decrease largely overlap with those for RCP8.5, at surface and at depth."

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

→ We have clarified: "extreme variability events in Ω_A 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"

→ Changed to: 'no extreme variability 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.

→ As pointed out above, we have now included a new Figure 11 in the Discussion section that also shows changes in extreme events due to changes in the mean state in addition to changes in variability. In addition, we now use the term 'extreme variability events' to clearly distinguish extreme events that are caused by only variability changes from extreme events that are caused by long-term ocean acidification and variability changes.

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.

→ We are well aware of earlier studies that investigate changes in marine heatwaves under global warming. A co-author of our study is first author of the mentioned heatwave study. In general, there are different approaches how to define extreme events: (i) defining these events relative to ocean conditions during a fixed period of time (i.e. as was done in the Froelicher et al. study), or (ii) relative to a shifting baseline as was done here. When following approach (i) and taking the secular change into account, the increase in the number of extreme days is much larger than following approach (ii) owing to the high signal to noise ratio in the ocean's carbonate chemistry under anthropogenic carbon uptake (see for example Froelicher et al. 2016). We have now clarified this with an additional paragraph and Figure 11 in the discussion section. The figure puts the changes in extreme variability events in context with changes obtained when analyzing extreme events defined with respect to preindustrial thresholds.

We included following paragraph in the discussion section: "In this study, we analyze changes in extreme variability events that are defined relative to a shifting baseline. If the long-term increase in ocean acidity and decrease in Ω_A is taken into account, i.e. defining the extremes with respect to a fixed preindustrial baseline (here the preindustrial 99^{th} percentile for $[H^+]$ and the preindustrial 1^{st} percentile for Ω_A), the changes in $[H^+]$ and Ω_A extremes are much larger (cyan lines in Figure 11). Under the RCP8.5 scenario, every day becomes an extreme event day in year 2051 at surface and in year 2067 at 200 m depth (Figure 11a). The model also projects year-round extreme conditions for Ω_A at the surface and at 200 m by the end of the 21^{st} century under RCP8.5 (Figure 11b). Comparing the two frameworks for surface $[H^+]$ extremes under present-day conditions, the annual number of extreme event days as defined in this study (i.e. with shifting baseline; black line in Figure 11) is on global average only 3.8 % of that also including the mean changes (i.e. with fixed preindustrial baseline; cyan line in Figure 11). This fraction differs regionally and reaches more than 10 % in the North Pacific, the North Atlantic, and the Arctic Ocean. Interestingly, the GFDL ESM2M projects that surface mean $[H^+]$ overshoots the preindustrial 99^{th} percentile in year 1975 on global average. Thereafter, higher variability actually reduces the number of extreme event days that are above the preindustrial percentile. Surface mean Ω_A falls below the preindustrial 1^{st} percentile in year 1990. After

that, lower variability further increases the number of extreme event days below the preindustrial percentile. "

One reason why we have focused our main analysis on extreme variability events is the fact that even changes in variability can have deleterious consequences for marine organisms as we state in the introduction section.

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-].

→ We thank the reviewer for bringing this up. When writing acidity, we actually refer to the hydrogen ion concentration on the total scale that includes sulfate ions

$$[H^+]_T = [H^+]_F + [HSO_4^-].$$

We added a clarifying sentence to the manuscript: "$[H^+]$ is on the total scale and hence the sum of the concentrations of free protons and sulfate ions."

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

→ We replaced 'mean ocean acidification' by the hopefully less ambiguous expression 'long-term ocean acidification'.

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

→ We clarified: 'Superimposed onto the long-term decadal- to centennial-scale ocean acidification trend are short-term extreme variability events on daily to monthly timescales, during which ocean pH and/or Ω are extremely low (Hofmann et al., 2011; Joint et al., 2011; Hauri et al., 2013)'

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 .

→ Also in response to reviewer 3, we now write: 'In addition to the changes in the mean, recent studies suggest that the seasonal cycles in $[H^+]$ and Ω are also strongly modulated under elevated atmospheric CO_2. Higher background concentrations of dissolved inorganic carbon and warmer temperatures produce stronger departures from mean state values for a given change in pertinent physical or chemical drivers for $[H^+]$ and weaker departures for Ω (Kwiatkowski and Orr, 2018; Fassbender et al., 2018). Other studies have also addressed the changes in the seasonal cycle of pCO_2 (Landschützer et al.,2018; Gallego et al., 2018; McNeil and Sasse, 2016; Rodgers et al., 2008; Hauck and Völker, 2015).'

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.

→ Yes, we meant daily mean output. We now refer to daily mean output and daily mean data throughout the manuscript.

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.

→ We replaced 'residual daily variability' by 'subannual variability' throughout the manuscript.

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 expected anyway.

→ Following the reviewers suggestion, we deleted the last three lines of the paragraph.

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.

→ We thank the reviewer for this pointer. The model indeed uses K1 and K2 from Dickson and Millero (1987) (Table 4 therein) based on Mehrbach et al. (1973). Total alkalinity in ESM2M includes contributions from phosphoric and silicic acid and their conjugate bases. We clarified in the manuscript: 'The ocean carbonate chemistry is based on the OCMIP2 parametrizations (Najjar and Orr, 1998). The dissociation constants for carbonic acid and bicarbonate ions are from Dickson and Millero (1987), which are based on Mehrbach et al. (1973), and the carbon dioxide solubility is calculated according to Weiss (1974). Total alkalinity in ESM2M includes contributions from phosphoric and silicic acids and their conjugate bases.'

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.

→ We decided to keep this, as we do not want to hide this information from the reader, who might be interested in this caveat. No changes are made to the manuscript.

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

→ We specified the software in the text: 'calculated using the *measure.label* function from the *scikit-image* library for *Python*'

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.

→ Our approach is based on the assumption that all ensemble members share the same anthropogenic trend in $[H^+]$ and Ω_A (they share the same prescribed atmospheric CO_2 evolution) and that deviations from this trend are only a consequence of internal variability. Therefore, we chose to subtract the ensemble mean (with the seasonal cycle removed by an additional running mean) instead of splines because we did not want to make an assumption about the shape (linear, cubic, etc.) of the secular trend in $[H^+]$ over the 240yr period from 1861 to 2100. For a large ensemble, averaging the individual ensemble members yields an ensemble mean that only contains the forced evolution. In our case, due to the relatively small ensemble size, we likely also remove some variability from the ensemble members as discussed in appendix A. However we also think that, when subtracting an individual polynomial spline from each ensemble member, one also introduces some error by either a) adding interannual to decadal variability by subtracting a too-low-order polynomial that deviates from the actual forced evolution or b) subtracting a too-high-order polynomial that misinterprets interannual to decadal variability as forced evolution. Therefore, we kept our method as is.

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

→ We clarified: 'The removal of the secular trend ensures that the mean state in the processed data stays approximately constant while day-to-day to interannual variability can change over the simulation period (depicted for one grid cell in Figure 1).'

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.

→ Also in response to the general comment above, we removed any reference to standard deviation from the manuscript.

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.

→ Also in response to reviewer 1, we clarified: '(b,d) Same as (a,c), but the ensemble-mean change with respect to the average of the 500-year long preindustrial control simulation has been subtracted'.

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."

→ Following the reviewer's proposal, we replace the term 'residual daily variability' by 'subannual variability' throughout the manuscript. Many thanks for this suggestion.

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.

→ Following the general comment above, we removed any reference to standard deviation from the manuscript.

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

→ In response to a comment by reviewer 1, we rewrote this sentence. It now states: 'In order to assess whether changes in low or high frequency variability cause changes in extreme variability events and their characteristics, we use three steps to decompose the total variability in $[H^+]$ into interannual, seasonal, and subannual variability.'

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

→ We prefer keeping Equation 1 in the text for clarity as it helps to understand Equation 2.

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.

→ We thank for bringing up this typo and squared the partial derivatives.

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.

→ This comment is discussed above in the general comments.

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

→ We thank the reviewer for this comment and we agree that this has, to our knowledge, not been published elsewhere yet. Also reviewer 1 has suggested the same thing. Therefore, we move the Table from the Appendix to the main text.

L215-221: It is unclear how trends and the error bars were computed for the data based estimates

→ We added an additional appendix section on the methodology for estimating trends and their confidence intervals as well as on testing for difference between trend estimates.

L215-221: 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.

→ We now tested all trend estimates for difference. Furthermore, we modified the uncertainty of the estimates for the simulated trends, as we had a mistake in our calculations for the uncertainty of the simulated

ensemble-mean trend. With the corrected simulation trend uncertainties, it turns out that the difference for Ω_A in the northern high latitudes is in fact significant. Further details on the methodology can be found in the added Appendix D.

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

→ This is a good point. We added an additional comment on line 203: 'An exception is the Southern Ocean where data-based pCO_2 products are uncertain due to sparse data in winter (Gray et al.,2018).' We also included at the end of the section: 'Nevertheless, the observation-based trends in the northern and especially southern high latitudes are rather uncertain because winter time data is sparse there.'.

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

→ Many thanks. We followed the recommendation and changed the sentence accordingly: "Even though we lack the daily observational-based data to undertake a full assessment, it appears that the GFDL ESM2M model is adequate to assess changes in open ocean ocean acidification extreme events."

L238: What is meant by intensity?

→ We clarified that we mean 'maximal intensity' as introduced in the method section.

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

→ We moved Table A2 to the main text. Furthermore, we followed the recommendation and removed the ensemble ranges from the text to make it more accessible. We only kept the ranges for the comparison of RCP2.6 and RCP8.5 since these are not given in the table.

L240: Please tell us what fraction of the total volume above 200 m is represented by the 2.7 x10^3 km^3.

→ We added the requested information and changed the sentence to: "Ocean acidity extremes in the upper 200 m occur with a typical volume of $2.7{\cdot}10^3 km^3$, which is about 0.004 % of the total ocean volume in the upper 200 m (Figure 4g)."

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.

→ L254-255 was intended to state that the longer duration of events at depth compared to the surface is connected to the higher relative contribution of interannual variability to total variability at depth (63% at 200m depth vs 11% at the surface). We rewrote the sentence to clarify that we compare 200m depth duration to surface during preindustrial times: "The longer duration is connected to the more pronounced contribution from interannual variability (see Section 3.3).".

Differences in changes within our simulations between the surface and 200m depth are similar to the differences reported by Resplandy et al., 2015 for the comparison between stratified tropical waters compared to mode and intermediate waters: Although the mean DIC changes between preindustrial and RCP8.5 2081-2100 are larger at the surface than at 200m depth ($184 \mu mol\,kg^{-1}$ vs $154 \mu mol\,kg^{-1}$), the change in mean $[H^+]$ is larger at 200m depth ($11.4 nmol\,kg^{-1}$ vs $9.8 nmol\,kg^{-1}$). Yet, total $[H^+]$ variance at surface quadruples between the two periods while variance at 200m depth only increases by a third. In our understanding, this is due to the fact that $[H^+]$ variability changes (in contrast to $[H^+]$ mean changes), not only depend on mean changes in the drivers (most of all DIC), but also on variability changes in the drivers. This is discussed in

section 3.4 in the manuscript and we hence do not go further into detail in section 3.1.

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.

→ We have not stated that there is a significant increase, but just say 'considerably'. However, we have now tested the significance with a simple t-test and found that the differences between 2041-2060 and 2081-2100 in maximal intensity, event days, and volume are all significant with p values smaller 1e-7 (testing the 20 ensemble-mean annual averages of the first period against those of the second). The text has been changed accordingly.

The volume of event metric describes the volume covered by events within the first 200m - it is by construction a column-integrated metric and not connected to a specific depth level. We changed the wording to clarify this.

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.

→ To clarify that we investigate extreme events that arise only from variability, we replaced 'extreme event' by 'extreme variability event' throughout the manuscript.

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).

→ We completely rewrote section 3.4. More detail on the revised decomposition can be found in the response to the third general comment by reviewer 1. As proposed by the reviewer, we added zonal mean plots to the figures that allow a more quantitative assessment.

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.

→ We use the first order Taylor expansions to decompose variance in a period. When we analyse changes in variance between two periods, we analyse the change in this first order Taylor expansion. The changes in the partial derivatives in fact represent the contribution from nonlinearity in our belief. However, we agree that the sentence, as it was in the text, might confuse and now write instead "The increase in $[H^+]$ extreme variability events is a consequence of increased sensitivity of $[H^+]$ to variations in its drivers.".

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.

→ We do not understand the reviewer comment here. Yes, as we stated above, we do neglect the changes in the mean state in our main analysis and focus on changes in extreme variability events. But here we investigate how different definition of extreme events may impact our results. No changes are made to the manuscript.

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 $\Omega_A = 1$ threshold was reached relative to previous work that did not address seasonal variability.

→ We thank for the great hint and added a comparison of our results to the study by McNeil and Matear: "Assuming unchanged seasonality, McNeil and Matear (2008) found that seasonal aragonite undersaturation

of surface waters in the Southern Ocean may occur 30 years earlier than annual mean aragonite undersaturation. However, our simulation shows that the reduction in Ω_A variability delays the onset of undersaturation by about 10 to 15 years in the Southern Ocean relative to a hypothetical simulation where variability does not change. Therefore, changes in variability need to be taken into account when projecting the onset of seasonal undersaturation, especially in the high latitudes and in the thermocline of the tropics."

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

→ We agree with the reviewer and have deleted these two sentences.

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.

→ We somewhat disagree with the reviewer given that the average duration of an acidity extreme event is less than a month. We clarified this in the text: 'In addition to earlier studies, we also show that changes in subannual variability contribute to changes in extreme [H^+] variability events under increasing atmospheric CO_2 and that the average duration of extreme variability events at the surface and at present-day is about 15 days.'

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.

→ We mention that in the caveat section: "Future studies with Earth system models that resolve diurnal processes are needed to quantify changes in diurnal variability and the impacts of these changes on extreme acidity events."

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.

→ We somewhat disagree with the reviewer. This statement about the diurnal cycle is included in caveat section as we think this is an important caveat of our study that needs to be discussed. No changes are made to the manuscript.

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.

→ We agree with the reviewer and have deleted this sub-sentence.

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 ..."

→ We have extensively discussed previous findings in the discussion section as well as in the introduction. Therefore, we would like to highlight in this last paragraph some of the main findings of our paper. We therefore changed the sentence to: 'In conclusion, our analysis shows that marine organisms and ecosystems are projected to be exposed to less stable [H^+] conditions in the future with more frequent occurrences of variability-driven short-term extreme [H^+] conditions.'

L467: What is "mean" ocean acidification?

→ We replaced 'mean ocean acidification' by the hopefully less ambiguous expression 'long-term increase in ocean acidity'.

Technical concerns:

Global Changes:

- change "200 m depth" to "200 m" → Changed throughout the manuscript.
- change "point in time" to "time" → Changed throughout the manuscript.
- change "points in time" to "times" → Changed throughout the manuscript.
- MISSING HYPHENS: the terms "short term", "deep water", and "sea ice" should all include a hyphen when they modify a noun → Changed throughout the manuscript.

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

→ Changed.

L44:

- delete "one might expect that". → Changed.
- change "exhibit" to "have" → Changed.
- 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" → Changed and acknowledged.

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

→ Changed.

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

→ Changed.

L63:

- change "suggest" to "project". → Changed.
- change "is projected to" to "will" → Changed.

L69:

- change "extreme" to "extremes" → We kept it as is.

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

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

→ Changed.

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

→ Changed.

L104:

- change "1000 year" to "1000-year". → Changed.

- "a new computing infrastructure" is vague. Please clarify your intended meaning. → Also in response to reviewer 3, we have deleted the detailed information about the spin-up procedure as it is a distraction and not needed for understanding the model simulation setup.

L108: insert "the average" before "atmospheric"

→ Changed.

L111: insert "average" before "atmospheric"

→ Changed.

L120:

- change "500 year" to "500-year" → Changed.

- delete "long" → Changed.

- what is "potential vegetation" → Also in response to other reviewer comments, we deleted this additional information as it is not key to understand the results of the study.

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

→ Changed.

L124:

- write "daily mean" instead of "daily" → Changed.

- add "the aragonite saturation state" before the symbol Ω_A → Changed.

- delete the 2nd sentence → Changed.

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

→ Changed.

L238: change "typical" to "average"

→ Changed.

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

→ We kept this information in the main text as, in our opinion, this facilities the understanding of the text and its link to the figure.

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?

→ The colors were taken from the viridis color scale and should be color-blind safe (see `https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html`). We changed the yellow color to orange color to make it more visible.

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).

→ We decided not to present the subpanels as a matrix in Figure 3 (which is Figure 4 in the revised manuscript version) since we think this might lead to confusion with subpanel g) that does not fit the structure of 'Surface' as left column and '200m' as right column, since volume is a 3D metric, describing the volume of extremes in the entire first 200 m of water column - so it is neither belonging to 'Surface' nor to '200 m'. Row labels have been introduced in Figure 6 (previous Figure 5) and Figure A1.

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

→ Changed.

L295: the meaning of "reoccur" is unclear.

→ We replaced 'reoccur annually' by 'occur each year' which is hopefully clearer.

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

→ We adressed this comment and now write "Previous studies have shown that the seasonal cycle of surface ocean pCO_2 will be strongly amplified under increasing atmospheric CO_2 (Gallego et al., 2018; Landschützer et al., 2018; McNeil and Sasse, 2016) and that a similar amplification is expected for surface $[H^+]$ (Kwiatkowski and Orr, 2018)."

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.

→ Many thanks. We have changed it throughout the text.

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

→ We followed the comment and provided numbers for all equations in the appendix.

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

→ As we have removed any notion of standard deviation throughout the manuscript, these equations are now written in terms of variances.

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

→ We changed the format of the tables and believe they follow the BG guidelines now.

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.

→ As the focus of our analysis is not on the sign of changes, but rather on the magnitude of changes, we kept the color bar as is.

Reviewer 3: Sarah Schlunegger

Burger et. al. presents a clear and logical assessment of projected changes in the variability of ocean acidity and their impact on acidic extremes. The study uses an Earth System Model (ESM) which they demonstrate has historical fidelity with observed acidification trends. The study contrasts future acidity extremes under low- and high-emissions scenarios, all relative to 99th percentile extremes as defined in the preindustrial control simulation. Changes in the duration, frequency, intensity and volume extent of extreme events are presented.

→ We thank the reviewer for the positive and encouraging review.

Drivers of change in [H+] variability are partitioned into contributions from changes in the mean state and variability of carbon concentrations, temperature, alkalinity and salinity. As noted by other reviewers, the exact decomposition used is flawed and needs revision before publication.

→ We have revised the methodology for the carbonate chemistry decomposition and the sections on the decomposition. We would like to refer to the reply to the third general comment by reviewer 1 for more details.

Other than this obvious issue, I have only three broad suggestions which would ready the manuscript for publication in Biogeosciences.

1. As a first assessment of changes in OA extremes, this paper has the opportunity to present simple, conceptual explanations for why those changes are occurring. This is done adequately for the case of $[H^+]$ but not for the case of Omega. Why is Omega variability decreasing? Is this largely driven by changes in carbonate chemistry, the way it is for $[H^+]$, or are changes in ocean dynamics involved? This has implications for the robustness of the results across other ESMs.

→ We thank the reviewer for pointing out this missing piece. We now also decompose and discuss the variance changes in Ω_A. We also added a new Figure 10 to Section 3.4 that shows the zonal mean decomposition of Ω_A variance change into the contributions from sensitivity changes, variability changes, and correlation changes. Furthermore, we show that changes in dissolved inorganic carbon are the dominant driver.

2. Further discussion is needed of why variability changes at the surface differ from those at 200m depth and of why there is strong compensation between the contributions of increasing carbon concentrations and decreasing carbon variability to $[H^+]$ variability at depth (9e vs 9f, and to a lesser extent at the surface, 8e and 8f). The striking spatial patterns in Figure 9e and 9f should also be explained.

→ With the revised carbonate chemistry decomposition, we can now understand the striking compensating patterns we saw in the previous Figures 9e and 9f as well as 9b and 9c. The large compensating contribution arises from mixed changes in the sensitivities due to changes in mean DIC and DIC variability: The large

variance increase from the mean increase in DIC is weakened by the simultaneous decrease in DIC variability there. Variability changes in DIC alone would have a much smaller effect (compare Figure 9c and 9d in the new manuscript version.). The striking pattern in (previous) Figure 9e itself is mainly a consequence of background $[H^+]$ variability: DIC mean changes have a much larger imprint on $[H^+]$ variability in regions that are already more variable (see preindustrial variability in $[H^+]$ in Figure A3a). We revised section 3.4 accordingly.

3. Include discussion of model uncertainty. For example, results that are a direct chemical consequence of invading anthropogenic carbon (e.g. increasing sensitivity of $[H^+]$ to drivers) are more likely to be more robust across models than results that are a consequence of changes in ocean physics. ESM2M has less warming and a stronger ocean carbon sink than the majority of CMIP5 models – how might the results presented compare to a model with higher climate sensitivity and less ocean carbon uptake models?

→ We included a discussion of model uncertainty in the discussion section: "Secondly, our results, in particular at the local scale, might depend on the model formulation. As the mean increases in C_T mainly drive the increases in extreme $[H^+]$ variability events (see Figure 8f), we expect that models with larger oceanic uptake of anthropogenic carbon show larger changes in extreme variability events than models with lower anthropogenic carbon uptake. The GFDL ESM2M matches observation-based estimates of historical global anthropogenic CO_2 uptake relatively well, but still has difficulties in representing the regional patterns in storage (Frölicher et al., 2015). Therefore, the exact regional patterns of C_T changes may differ from model to model and further studies focusing on the physical processes that lead to the regional C_T changes may help to better constrain the regional patterns in changes of acidity extremes. In addition, it is currently rather uncertain how $[H^+]$ and Ω_A variability changes as a result of changes in the drivers' variabilities. We have demonstrated that this factor is particularly important at depth for $[H^+]$ and for Ω_A. It is well known that current Earth system models have imperfect or uncertain representations of ocean variability over a range of timescales (Frölicher et al., 2016; Resplandy et al., 2015; Keller et al., 2014). A possible way forward would be to assess changes in ocean acidity extreme events within a multi-model ensemble, which would likely provide upper and lower bounds of future changes in these events."

Additional comments and stylistic suggestions

L5: number of days where? At each point? On average? Overall occurrence anywhere?

→ Changed to 'Globally, the number of ...'

L28: replace 'or' with 'and/or'

→ Changed and acknowledged.

L33: define saturation horizon – e.g. Saturation generally decreases with increasing depth, with transition from saturated to undersaturated referred to as the saturation horizon.

→ We clarified: 'saturation horizon (i.e. the depth between the supersaturated upper ocean and the undersaturated deep ocean (Feely et al.,2008; Leinweber and Gruber, 2013))'

L33: add 'can' in front of 'also' . . . 'can also'

→ Changed.

L38: remove 'vast'

→ Removed.

L40: add 'also' in front of necessary. . . 'also necessary'

→ Included.

L45: . . .'critical threshold, like undersaturation.'

→ Changed to: 'e.g. from calcium carbonate saturation to undersaturation'

L50: sentence rearrangement for clarity: 'In laboratory experiments in which deep-water corals are exposed to low-pH waters for a week. . .'

→ Changed and acknowledged.

L53: Start sentence with 'Therefore,'

→ Changed.

L61: '...at which higher background concentrations of dissolved inorganic carbon (DIC or C_T) and warmer temperatures produce stronger departures from mean state values for a given change in pertinent physical or chemical drivers.'

→ Many thanks for this suggestion which we added to the text. Also taking into account a comment by reviewer 2, we now write: "In addition to the changes in the mean, recent studies suggest that the seasonal cycles in $[H^+]$ and Ω are also strongly modulated under elevated atmospheric CO_2. Higher background concentrations of dissolved inorganic carbon and warmer temperatures produce stronger departures from mean state values for a given change in pertinent physical or chemical drivers for $[H^+]$ and weaker departures for Ω (Kwiatkowski and Orr, 2018; Fassbender et al., 2018). Other studies have also addressed the changes in the seasonal cycle of pCO_2 (Landschützer et al.,2018; Gallego et al., 2018; McNeil and Sasse, 2016; Rodgers et al., 2008; Hauck and Völker, 2015)."

L103: Detail about spin-up and control is distracting.

→ Also in response to reviewer 2, we have deleted the detailed information about the spin-up simulation.

L110: omit 'and', replace with comma

→ Changed.

L120: 'year-long' – however these detail is distracting so potentially remove all together.

→ We agree. We have modified and shortened it accordingly.

Figure 1. The figure caption (and methods) say that the ensemble mean is removed – however what is removed is actually the departure of the ensemble mean from the preindustrial state. If the ensemble mean itself were removed, then all values at year 1861 would be $\sim$ zero. Update description or replot.

→ We thank for the pointer. We have clarified: "(b,d) Same as (a,c), but the ensemble-mean change with respect to the average of the 500-year long preindustrial control simulation has been subtracted."

Figure 1. Panel b is presumably derived from panel a, however it is confusing that the variability in panel a contracts during the 21st century, however it is shown to increase in panel b. Please add an explanation in the caption or text.

→ As the reviewer correctly says, panel b is derived from panel a. However, we would like to note here that the variability is actually increasing and not decreasing in panel a. No changes are made to the manuscript in response to this comment.

L230: remove 'we consider'

→ Removed.

L240-245: Further and continued clarification that values given are a global average of local changes. With each presentation of a value, indicate it is a global average.

→ We clarified at the beginning of section 3 that global values are grid cell based characteristics that are aggregated globally.

L244: change 'single events' to 'individual events'

→ Changed.

L255: Why is high-frequency variability larger at the surface whereas low-frequency variability larger at depth? ... Presumably because the surface has direct contact with the atmosphere, and its chaotic, high-frequency (weather) variability, whereas the deeper ocean experiences stronger low-frequency variability as the ocean acts as a low pass filter on the atmospheres' stochastic forcing (e.g. Hasselmann 1976). Add/include explanation.

→ At preindustrial, subannual variance is five times larger at depth than at surface (0.062 $\mathrm{nmol^2 kg^{-2}}$ vs 0.012 $\mathrm{nmol^2 kg^{-2}}$), also in percent of total variability (14.6% (200m depth) compared to 8.3% (surface)). So the model does not simulated larger high-frequency variability at surface than at depth. In fact, the seasonal variance is most important at surface (81% of total) while interannual variability is most important at depth (63% of total) (at preindustrial). Therefore, the stronger relative contribution from interannual variability at depth is leading to longer-lasting events. We added a short paragraph on preindustrial variability in section 3.3: "For the preindustrial, the model simulates overall larger $[H^+]$ variance at depth than at the surface ($0.42\,\mathrm{nmol^2 kg^{-2}}$ vs. $0.15\,\mathrm{nmol^2 kg^{-2}}$, not shown). Seasonality has the largest contribution at the surface (81 % of total variance). At 200 m, interannual variability has the largest contribution (63 %), and also subannual variability is more important compared to the surface (15% vs. 8%)."

L298: insert 200 meters... ' In contrast to the surface, 200 meter [H+] extremes'...

→ Changed and acknowledged.

L357-361: C_T increases due to invasion of anthropogenic carbon. But why does ALK change? ALK is not influence by gas-exchange, but by biology or circulation. Briefly explain ALK changes and give reference.
→ We added an explanation: "The changes in A_T are largely due to changes in freshwater cycling that also manifest in salinity changes (Supplementary Figure A4, Carter et al., 2016)."

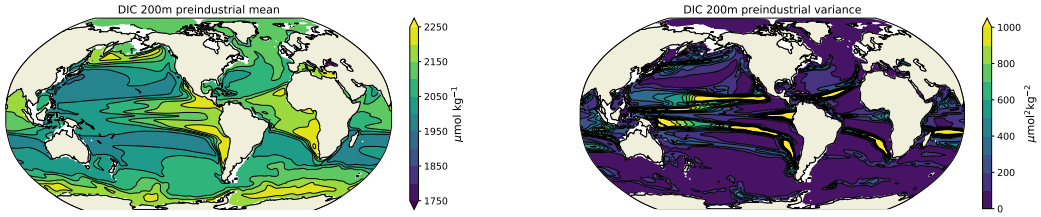
378-379: [H+] variability changes at depth are larger than at surface – is this mostly due to the fact that background C_T concentrations increase with depth, however increased C_T itself is associated with decreased susceptibility to external variability? Figure 8 e) vs f) indicates this is the case. Provide commentary.

→ Yes, the $[H^+]$ variability changes (from PI to 2081-2100 under RCP8.5) from mean changes in C_T are larger at 200m depth compared to the surfaced due to the higher preindustrial background carbon concentration (global mean: $2078\,\mu\mathrm{mol\,kg^{-1}}$ (200m depth) vs. $1959\,\mu\mathrm{mol\,kg^{-1}}$ (surf.)), although changes in C_T are smaller at depth (global mean: $154\,\mu\mathrm{mol\,kg^{-1}}$ (200m depth) vs. $184\,\mu\mathrm{mol\,kg^{-1}}$ (surf.)). The impression that increased C_T is associated with decreased variability in C_T was an artefact of the method. We thank the reviewer for bringing this up. We rewrote most of section 3.4.

Figure 9. Panel 9b and 9c bear strikingly similar patterns and magnitudes – where the changes in the mean-state of the drivers is largest and positive, changes in the variability are also largest and negative. And this is mostly coming from C_T mean (increasing) and C_T variability (decreasing). Is this directly understandable through carbonate chemistry? In a higher DIC world, DIC variations are actually smaller? Or is it that the physical, surface to deep gradient in DIC is weakening with surface invasion of anthropogenic carbon and therefore mixing-related variability is reduced? Explain this near-perfect anti-correlation. Also, the structures of strong mean and variability changes look dynamical. Potentially the expansion of the subtropical gyres. Explain the spatial structure of the patterns of strong drivers.
→ As we have modified the Taylor decomposition method, these patterns have changed. Please see above.

In the appendix, include figures of the pre-industrial and year 2100 concentrations of C_T at surface and 200m depth. The patterns in 9e and 9f are likely related to both the starting concentrations of C_T and the storage of C_T. Figure A4a and A4e show the change in C_T, which resembles the strong patterns of 9e and 9f. Synergies of anthropogenic carbon and background concentrations of C_T may help explain strong structures of 9e and 9f. The strongest temperature changes (A4g) also resemble the strong patterns of 9e and 9f, indicating either dynamical changes or thermal changes are involved.
→ We agree that the patterns of variability changes in 9e and 9f are related to the preindustrial starting points. The preindustrial variability in $[H^+]$ (see Figure A3a - event intensity is a good indicator for variability) already shows the patterns that are also visible in in 9e and 9f. But what's the reason for these preindustrial patterns of high $[H^+]$ variability? Focusing on DIC, this could have in principle two reasons -

(a) because of large sensitivity w.r.t drivers variations arising from high mean DIC and (b) because of large variations in DIC. It is actually more connected to (b), the patterns resemble the preindustrial patterns of DIC variability, not so much the patterns of mean DIC (see figure below). We decided not to add an analysis of preindustrial $[H^+]$ variability, which would likely add even more complexity to the manuscript. But we added a sentence in section 3.4 to highlight that the patterns seen in Figure 9 are connected to the preindustrial background variability in $[H^+]$: "There, the preindustrial background $[H^+]$ variability is also the largest (Figure A3a). As a result, an increase in the sensitivities due to an increase in mean C_T has the largest effect there.".



Line 393: Is the decrease in Omega variability a direct consequence of carbonate chemistry? Or of changes in the variability of the drivers or physical state (e.g. vertical gradient hypothesis stated before)? This is an important distinction, as carbonate chemistry is well constrained (little model uncertainty) whereas dynamical changes have much larger model uncertainty. If reductions in Omega variability are chemically deterministic, then this feature will be relatively robust across models, as most use the same (e.g. OCMIP2, Najjar et al. 2007) protocols for carbonate chemistry calculations. But if it is dynamical, it may not be. Furthermore, if it is chemically driven, and if it is related to the storage of anthropogenic carbon and/or concentration of natural carbon, then contemporary observations and reconstructions of the distribution of natural and anthropogenic carbon in the ocean could reveal where changes in Omega variance are likely to occur.

→ We added a new paragraph in section 3.4 that shows the carbonate chemistry decomposition of variability changes in Ω_A and a new Figure 10. Changes in the variability of the drivers have a larger impact on Ω_A compared to $[H^+]$. Furthermore, Ω_A variability changes are even more determined by changes in DIC. Changes in the other potential drivers are less important (dashed vs. solid lines in Figure 10). As a result, projected changes in Ω_A variability largely depend on changes in DIC variability. We added a sentence to the Discussion section: "In addition, it is currently rather uncertain how $[H^+]$ and Ω_A variability changes as a result of changes in the drivers' variabilities. We have demonstrated that this factor is particularly important at depth for $[H^+]$ and for Ω_A.".

Increase in ocean acidity variability and extremes under increasing atmospheric CO_2

Friedrich A. Burger[1,2], Thomas L. Frölicher[1,2], and Jasmin G. John[3]

[1]Climate and Environmental Physics, Physics Institute, University of Bern, Bern, Switzerland.
[2]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.
[3]NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA.

Correspondence: Friedrich A. Burger (friedrich.burger@climate.unibe.ch)

Abstract. Ocean acidity extreme events are short-term periods of extremely high $[H^+]$ concentrations. The uptake of anthropogenic CO_2 emissions by the ocean is expected to lead to more frequent and intense ocean acidity extreme events, not only due to ~~mean~~ long-term ocean acidification, but also due to increases in ocean acidity variability. Here, we use daily mean output from ensemble simulations of a comprehensive Earth system model under a low and high CO_2 emission scenario to isolate and quantify the impact of changes in variability on changes in ~~ocean acidity extremes. We show that~~ variability-driven ocean acidity extreme events. Globally, the number of days with ~~extreme~~ variability-driven extremely high $[H^+]$ conditions for surface waters is projected to increase by a factor of 14 by the end of the 21st century under a high CO_2 emission scenario relative to preindustrial levels. The duration of individual variability-driven extreme events is projected to triple, and the maximal intensity and the volume extent in the upper 200 m to quintuple. Similar changes are projected in the thermocline. Under a low emission scenario, the large increases in ocean acidity extreme event characteristics are substantially reduced. At surface, the changes are mainly driven by increases in $[H^+]$ seasonality, whereas changes in interannual variability are also important in the thermocline. Increases in $[H^+]$ variability and variability extremes arise predominantly from increases in the sensitivity of $[H^+]$ to variations in its drivers due to the increase in oceanic anthropogenic carbon. In contrast to $[H^+]$~~extremes~~, the occurrence of ~~short-term~~ variability-driven extremes in low aragonite saturation state ~~due to changes in variability~~ is projected to decrease. ~~An~~ The increase in $[H^+]$ variability and ~~an~~ associated increase in extreme variability events superimposed onto the long-term ocean acidification trend will enhance the risk of severe and detrimental impacts on marine organisms, especially for those that are adapted to a more stable environment.

1 Introduction

Since the beginning of the industrial revolution, the ocean has absorbed about a quarter of the carbon dioxide (CO_2) released by human activities (Friedlingstein et al., 2019). Oceanic uptake of anthropogenic CO_2 ~~mitigates~~ slows global warming by reducing atmospheric CO_2, but also leads to major changes in the chemical composition of seawater through acidification (Gattuso and Buddemeier, 2000; Caldeira and Wickett, 2003; Orr et al., 2005; Doney et al., 2009). When CO_2 dissolves in seawater, it forms carbonic acid that dissociates into bicarbonate ($[HCO_3^-]$) and carbonate ions ($[CO_3^{2-}]$), releasing hydrogen ions ($[H^+]$) and thereby reducing pH (pH = -log($[H^+]$)). The rise in $[H^+]$ is partially buffered by the formation of $[HCO_3^-]$

, thereby decreasing from $[CO_3^{2-}]$ and . The associated decline in $[CO_3^{2-}]$ reduces the calcium carbonate saturation state (Ω) that describes whether water is supersaturated or undersaturated with respect to calcium carbonate $\Omega = [Ca^{2+}][CO_3^{2-}] / ([Ca^{2+}][CO_3^{2-}])_{sat}$, i.e. the product of calcium ($[Ca^{2+}]$) and carbonate ion concentrations relative to the product at saturation. Undersaturated waters with $\Omega < 1$ are corrosive for calcium carbonate minerals. The calcium carbonate saturation state Ω differs between different mineral forms of calcium carbonate, such as aragonite and calcite, which differ in their solubilities. Over the last four decades the surface ocean pH has declined by about 0.02 pH units per decade (Bindoff et al., 2019). Continued carbon uptake by the ocean will further exacerbate ocean acidification in the near future (Caldeira and Wickett, 2003; Bindoff et al., 2019) with potential major consequences for marine life (Doney et al., 2009) and ocean biogeochemical cycling (Gehlen et al., 2012).

Superimposed onto the long-term decadal- to centennial-scale ocean acidification trend are short-term extreme eventsvariability events on daily to monthly timescales, during which ocean pH or the calcium carbonate saturation state and/or Ω are extremely low (Hofmann et al., 2011; Joint et al., 2011; Hauri et al., 2013). These events may can be driven by ocean mixing processesa range of different processes, such as ocean mixing, biological production and remineralization, mineral dissolution, temperature changes, and air-sea gas exchange variations, or a combination thereof (Lauvset et al., 2020). In eastern boundary upwelling systems, for example, short-term upwelling events and mesoscale processes can lead to low surface pH events and to short-term shoaling of the saturation horizon (Feely et al., 2008; Leinweber and Gruber, 2013)(i.e. the depth between the supersaturated upper ocean and the undersaturated deep ocean (Feely et al., 2008; Leinweber and Gruber, 2013)). Ocean pH also rapidly changes can also rapidly change as a consequence of microbial activity (Joint et al., 2011). Phytoplankton blooms and accompanying respiration drastically increase the partial pressure of CO_2 (pCO_2) and reduce pH in the thermocline (Sarmiento and Gruber, 2006). Such extreme variability events may have pH levels that are much lower than the mean pH conditions projected for the near future (Hofmann et al., 2011).

The vast majority of Most of the scientific literature on ocean acidification has focused so far on gradual changes in the mean state in ocean chemistry (e.g. Orr et al. (2005); Bopp et al. (2013); Frölicher et al. (2016))(Orr et al., 2005; Bopp et al., 2013; Frölicher et a . However, to understand the full consequences of ocean acidification for marine organisms and ecosystem services, it is also necessary to understand how ocean acidity extremes variability extremes in ocean acidity change under increasing atmospheric CO_2 (Kroeker et al., 2020). The ability of marine organisms and ecosystems to adapt to ocean acidification may depend on whether these the species have evolved in a chemically stable or a highly variable environment (Hofmann et al., 2011). If (Rivest et al., 2017; Cornwall et al., 2020). Furthermore, if the frequency and intensity of short-term ocean acidity extreme events strongly increases, one might expect that extreme variability events in ocean acidity strongly increase, some organisms may exhibit have difficulties to adapt, especially if carbonate chemistry crosses key CO_2 system variables cross some critical thresholds, e.g. from calcium carbonate saturation to undersaturation. Key plankton species such as coccolithophores (Riebesell et al., 2000), foraminifera and pteropods (Bednaršek et al., 2012) were found to be negatively impacted adversely affected by low carbonate concentrations. After only several days of aragonite undersaturation, some species such as pteropods already show reduced calcification, growth and survival rates (Bednaršek et al., 2014; Kroeker et al., 2013). Carbonate system vari-

ability also plays a role in shaping the diversity and biomass of benthic communities (Kroeker et al., 2011; Hall-Spencer et al., 2008). In laboratory experiments ~~, some deep water corals undergo a decline in calcification for~~ in which deep-water corals are exposed to low-pH ~~conditions over~~ waters for a week, some corals exhibit reduced calcification, while recovery may be possible when the low-pH condition persists for six months, stressing the importance of high-frequency variability and short-term acidification events (Form and Riebesell, 2012). There is also growing evidence that the organism response to variability in ocean acidity could change with ocean acidification (Britton et al., 2016). ~~Understanding~~ Therefore, understanding the temporal variability of ocean carbonate chemistry and the changes therein is ~~therefore~~ of critical importance for understanding the impacts of ocean acidification on marine organisms and ecosystems (Hofmann et al., 2011).

Changes in extremes arise from changes in the mean, variability, or shape of the probability distribution (Coles, 2001). Under ~~ongoing~~ continued long-term ocean acidification (i.e. changes in the mean), one can expect that extreme events in $[H^+]$ and Ω ~~may~~ will become more frequent and intense (Hauri et al., 2013). In addition to the changes in the mean, recent studies suggest that the seasonal cycles in $[H^+]$ and Ω are also strongly modulated under elevated atmospheric CO_2 ~~due to non-linear carbonate chemistry processes (Kwiatkowski and Orr, 2018; Fassbender et al., 2018; Gallego et al., 2018; Landschützer et al., 2018; McNeil and Sas~~ . Higher background concentrations of dissolved inorganic carbon and warmer temperatures produce stronger departures from mean state values for a given change in pertinent physical or chemical drivers for $[H^+]$ and weaker departures for Ω (Kwiatkowski and Orr, 2018; Fassbender et al., 2018). Other studies have also addressed the changes in the seasonal cycle of pCO_2 (Landschützer et al., 2018; Gallego et al., 2018; McNeil and Sasse, 2016; Rodgers et al., 2008; Hauck and Völker, 2015) . Over the 21st century and under a high greenhouse gas emission scenario, Earth system model simulations ~~suggest~~ project that the winter-summer difference in surface $[H^+]$ ~~is projected to~~ will increase by 81%, whereas the seasonal amplitude for aragonite saturation state (Ω_A) is projected to decrease by 9 % on global average (Kwiatkowski and Orr, 2018). Recent observational-based estimates as well as theoretical arguments support these projected increases in seasonality for $[H^+]$ and pCO_2 (Landschützer et al., 2018; Fassbender et al., 2018). We can therefore expect that changes in variability may also impose changes in the frequency and intensity of extreme ~~events in ocean acidity~~ acidity events.

Unlike for marine heatwaves (Frölicher et al., 2018; Collins et al., 2019) and extreme sea level events~~(Frölicher et al., 2018; Collins et al~~ (Oppenheimer et al., 2019), little is known about the characteristics and changes of extreme ocean acidity events and if so, only on seasonal timescales (Kwiatkowski and Orr, 2018). A global view of how extreme events in ocean chemistry due to changes in variability will unfold in time and space and a mechanistic understanding of the relevant processes is currently missing. This knowledge gap is of particular concern as it is expected that ~~ocean acidity extreme events~~ extreme variability events in ocean acidity are likely to become more frequent and intense under increasing atmospheric CO_2. Given the potential for profound impacts on marine ecosystems, quantifying trends and patterns of ~~ocean acidity extreme events~~ extreme variability events in ocean acidity is a pressing issue.

In this study, we use daily mean output of a five-member ensemble simulation under a low and ~~a~~ high CO_2 ~~emission~~ emissions scenario with a comprehensive Earth system model to investigate how changes in interannual, seasonal, and ~~residual daily~~ subannual variability under rising atmospheric CO_2 levels ~~imprint on~~ affect the occurrence, intensity, duration and volume of $[H^+]$ and Ω extreme ~~events. The remainder of this paper is organized as follows: Section 2 presents the coupled carbon-climate Earth system model, the ensemble simulations, analysis methods, and a brief model evaluation. The global and regional changes in ocean acidity extremes and their drivers are examined in Sections 3.1-3.4. The discussion and conclusions are given in Section 4.~~ variability events.

2 Methods

2.1 Model & experimental design

The simulations used in this study ~~are performed with the CMIP5-generation~~ were made with the fully coupled carbon-climate Earth system model developed at the NOAA Geophysical Fluid Dynamics Laboratory (GFDL ESM2M) (Dunne et al., 2012, 2013). GFDL-ESM2M consists of ocean, atmosphere, ~~sea ice~~sea-ice, and land ~~models~~modules, and includes land and ocean biogeochemistry. The ocean component is the Modular Ocean Model version 4p1 (MOM4p1), with a nominal 1° horizontal resolution increasing to 1/3° meridionally at the equator, with a tripolar grid north of 65°N, and with 50 vertical depth levels. The MOM4p1 model has a free surface and the surface level is centered around about 5 m depth and the spacing between consecutive levels is about 10 m down to a depth of about 230 m (Griffies, 2009). The dynamical ~~sea ice~~ sea-ice model uses the same tripolar grid as MOM4p1 (Winton, 2000). The Atmospheric Model version 2 (AM2) has a horizontal resolution of $2° \times 2.5°$ with 24 vertical levels (Anderson et al., 2004). The Land Model version 3 (LM3) simulates water, energy, and carbon cycles dynamically and uses the same horizontal grid as AM2 (Shevliakova et al., 2009).

The ocean biogeochemical and ecological component is version two of the Tracers of Ocean Phytoplankton with Allometric Zooplankton (TOPAZv2) module that parametrizes the cycling of carbon, nitrogen, phosphorus, silicon, iron, oxygen, alkalinity, lithogenic material, and surface sediment calcite (see supplementary material in Dunne et al. (2013)). TOPAZv2 includes three explicit phytoplankton groups: small, large, and diazotrophs, and one implicit zooplankton group. The ocean carbonate chemistry is based on the OCMIP2 parametrizations (Najjar and Orr, 1998)~~, with the~~ . The dissociation constants for carbonic acid and bicarbonate ions ~~from Mehrbach et al. (1973)~~ are from Dickson and Millero (1987), which are based on Mehrbach et al. (1973), and the carbon dioxide solubility ~~from Weiss (1974).~~ is calculated according to Weiss (1974). Total alkalinity in ESM2M includes contributions from phosphoric and silicic acids and their conjugate bases. TOPAZv2 also simulates diurnal variability in ocean physics as well as in phytoplankton growth. While diurnal variations in open ocean pH are therefore simulated to some extent, we do not expect the model to fully capture the high diurnal variability in seawater chemistry in coastal regions with large biological activity (Kwiatkowski et al., 2016; Hofmann et al., 2011).

We ~~performed a 220 year spin-up simulation at prescribed preindustrial CO_2 concentration started from a quasi-equilibrated 1000 year preindustrial control simulation to ensure stability of the model under a new computing infrastructure. We then~~ ran a five-member ensemble simulation covering the historical 1861-2005 period, followed by a high (RCP8.5; RCP: Representative ~~concentration pathway~~Concentration Pathway) and a low greenhouse gas emission scenario (RCP2.6) over the 2006-2100 period with prescribed atmospheric CO_2 concentrations. RCP8.5 is a high emission scenario without effective climate policies, leading to continued and sustained growth in greenhouse gas emissions (Riahi et al., 2011). In GFDL ESM2M, global atmospheric surface temperature in the RCP8.5 ensemble is projected to increase by 3.24 (ensemble minimum: 3.17 - ensemble maximum: 3.28) °C between preindustrial and 2081-2100. The RCP2.6 scenario represents a low emission~~and ,~~ high mitigation future (van Vuuren et al., 2011) ~~. The RCP2.6 ensemble simulated here undergoes an atmospheric warming~~ with a simulated warming in GFDL ESM2M of 1.21 (1.18-1.26) °C by the end of the 21st century relative to preindustrial levels. The five ensemble members over the historical period were initialized from ~~an extension of the spin-up~~ a multi-century long preindustrial control simulation, that ~~includes~~ was extended with historical land-use over the 1700-1860 period (Sentman et al., 2011). The five ensemble members were generated by adding different very small SST disturbances of the order 10^{-5} K to a surface grid cell in the Weddell ~~sea~~ Sea at 70.5° S, 51.5° W on January 1st 1861 (Wittenberg et al., 2014; Palter et al., 2018). Although the ocean biogeochemistry is not perturbed directly, $[H^+]$ and Ω differences between the ensemble members spread rapidly over the globe. On average, the ensemble members can be regarded as independent climate realizations after about three years of simulation for surface waters and about eight years at 200 m~~depth~~ (Frölicher et al., 2020). Neither the choice of the perturbation location nor the choice of the perturbed variable has a discernible effect on the results presented here (Wittenberg et al., 2014). In addition, an accompanying ~~500 year long~~ 500-year preindustrial control simulation ~~with potential vegetation was performedthat extends the 220 year long spin-up simulation~~was performed.

2.2 Analysis methods

2.2.1 Extreme event definition and characterization

We analyze daily mean data of $[H^+]$ and ~~Ω~~ the aragonite saturation state Ω_A in the upper 200 m of the water column. ~~In this study, we focus on the aragonite saturation state Ω_A. The calcium carbonate saturation state Ω differs between different mineral forms of calcium carbonate, such as aragonite and calcite, that differ in their solubilities. Aragonite is the most soluble form and important for many calcifying organisms such as pteropods (Bednaršek et al., 2012)~~$[H^+]$ is on the total scale and hence the sum of the concentrations of free protons and sulfate ions. We define an event as a $[H^+]$ extreme event when the daily $[H^+]$ exceeds the 99th percentile, i.e. a one-in-a-hundred days event. Similarly, we define a Ω_A extreme event when the daily Ω_A falls below the 1st percentile. The percentiles are calculated for each grid cell from daily mean data of the 500-year ~~long~~ preindustrial control simulation. In contrast to absolute thresholds, relative thresholds, such as those used here, allow the characterization of extreme events over regions with different statistical properties. In addition, biases in the simulated variables already alter the definition of relative thresholds and should thus have a smaller effect on projections of changes in extreme

events based on these thresholds compared to projections based on absolute thresholds (see also Frölicher et al. (2018)).

We calculate four extreme event metrics: (a) the number of extreme event days per year (in days; number of days above the 99th percentile for [H^+] and below the 1st percentile for Ω_A), (b) the annual mean duration (in days; the average number of days above the 99th percentile for [H^+] and below the 1st percentile for Ω_A of single events within a year), (c) the annual mean maximal intensity (in nmol kg^{-1} or Ω_A unit; maximum [H^+] or Ω_A anomalies with respect to the percentile threshold over the duration of a single ocean acidification extreme event and then averaged over all events within a year), and (d) the mean volume covered by individual extreme events in the upper 200 ~~meters~~ m (in km^3; mean ~~of~~ volume of 3D clusters of connected grid cells that are above the 99~~th~~ th percentile for [H^+] or below the 1~~st~~ st percentile for Ω_A, calculated using the *measure.label* function from the *scikit-image* library for *Python* for each day, these daily means are then averaged annually). The number of days, duration, and maximal intensity are calculated for individual grid cells at the surface and at 200 m~~depth~~. While the truncation of extremes between years alters the results for duration and maximal intensity, it allows for the calculation of annual extreme event characteristics. We focus our analysis not only on the surface, but also on 200 m ~~depth~~ to study changes in extreme events within the thermocline~~. In Section 3.1, these grid cell based characteristics are then aggregated globally~~, where most organisms susceptible to ocean acidification are found, such as reef-forming corals and calcifying phytoplankton.

The aim of this study is to assess how changes in [H^+] and Ω_A variability lead to changes in different extreme event characteristics. Therefore, we isolate the effect of changes in variability by subtracting the secular trends at each grid cell and in each individual ensemble member prior to the calculation of the different extreme event characteristics (Figure 1). The secular trend is calculated as the five-member ensemble mean, which has been additionally smoothed with a 365-day running mean to keep the seasonal signal in the data (further information in Appendix A). ~~This calculation process ensures that~~ The removal of the secular trend ensures that the mean state in the processed data stays approximately constant while day-to-day to interannual variability can change over the simulation period ~~while the mean state stays approximately constant, as~~ (depicted for one grid cell in Figure 1). Thus, in our study, changes in the different extreme event characteristics are only caused by changes in variability and we call these events extreme variability events. In the Discussion section, we compare the simulated changes in extreme variability events to the total changes in extremes, which include changes due to secular trend in ocean acidity.

2.2.2 Decomposition of [H^+] variability into different variability components

In order to assess whether changes in low or high frequency variability cause changes in extreme variability events and their characteristics, we use three steps to decompose the total variability in [H^+] into interannual, seasonal, and ~~residual daily variability. First~~subannual variability (Figure 2). In a first step, we calculate the climatological seasonal cycle from the daily mean data by averaging each calendar day over all years in the time period of interest. Seasonal variability is then identified with the time-series variance ~~and standard deviation~~ of this 365-day long seasonal cycle. As described above, the secular trend in the daily mean data has been removed with the five-member ensemble mean before doing the analysis. In a ~~next~~ second

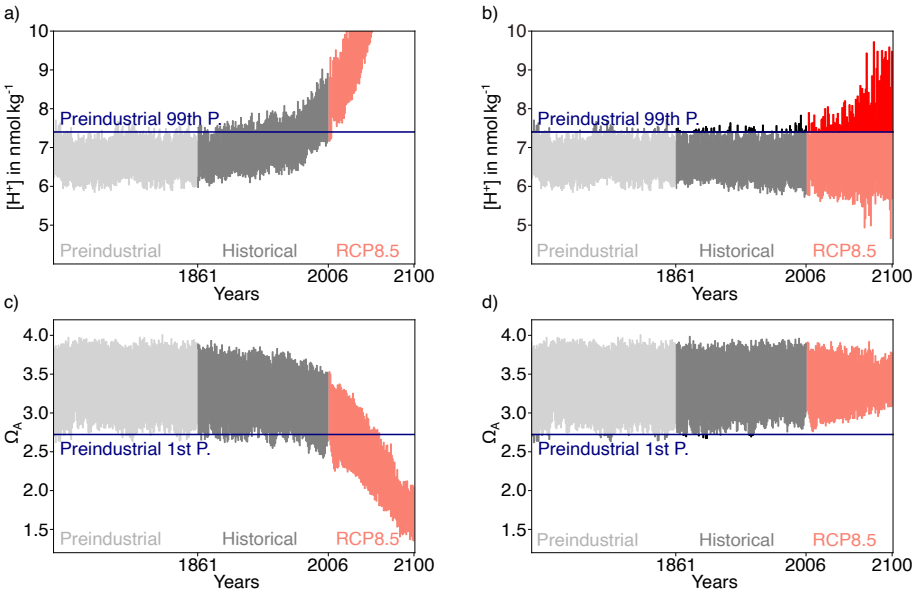


Figure 1. Simulated daily surface $[H^+]$ (a) and Ω_A (c) at 40°N and 30°W in the North Atlantic for one ensemble member over the preindustrial, the 1861-2005 historical period, and the 2006-2100 period under RCP8.5. (b,d) Same as (a,c), but the ~~ensemble mean~~ ensemble-mean change with respect to the average of the 500-year long preindustrial control simulation has been subtracted. For $[H^+]$, the preindustrial 99th percentile threshold (horizontal blue line in panels a) and b) is increasingly exceeded even when subtracting the ensemble mean change, because $[H^+]$ variability increases. In contrast, a reduction in Ω_A variability leads to a reduced undershooting of the preindustrial 1st percentile (panel d).

step, we subtract the seasonal cycle from the data and estimate the spectral density (Chatfield, 1996) of this residual time series using the *periodogram* function from the *scipy.signal* python library. ~~From the spectral density we then~~ In a third step, we calculate the variance ~~and standard deviation~~ arising from variations on interannual and ~~sub-annual timescales~~ subannual timescales from the spectral density to obtain interannual and ~~residual daily~~ subannual variability (further information is given in Appendix B). ~~In section 3.3, we use the variance to report the contributions from interannual, seasonal, and residual daily variability to global-mean variability change, but for the spatial changes in Figure 7, we use the standard deviation~~Following this methodology, subannual variability comprises all variations in daily mean data with periodicities of less than a year that are not part of the seasonal cycle.

2.2.3 Taylor deconvolution method to identify mechanistic controls of $[H^+]$ and Ω_A variability changes

To understand the processes behind the simulated changes in $[H^+]$ variability and variability extremes, we decompose ~~the~~ these changes into contributions from changes in temperature (T), salinity (S), total alkalinity (A_T), and total dissolved inorganic carbon (C_T). Assuming linearity, the difference of $[H^+]$ from its mean at time step i can be decomposed into contributions from the drivers by employing a first order Taylor expansion ~~:~~

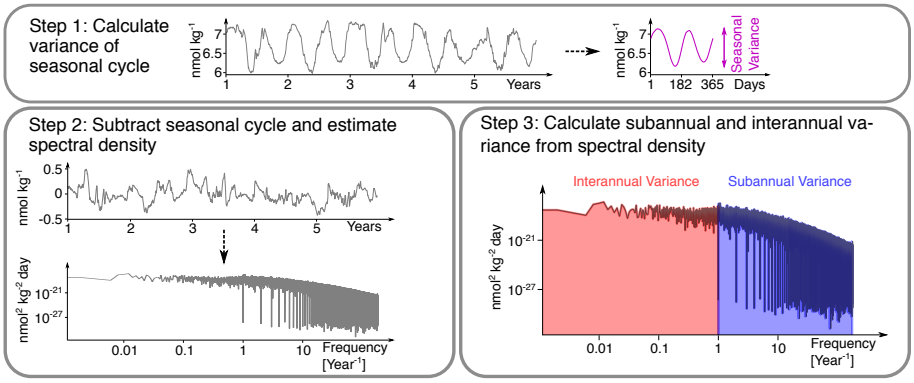


Figure 2. The three-step decomposition of $[H^+]$ variance into interannual, seasonal, and subannual variance, exemplified for a surface grid cell at 40° N and 30° W in the North Atlantic at preindustrial. In a first step, the climatological seasonal cycle is determined (over the whole period, only five years are depicted here) and its variance is calculated. Note that the seasonal cycle in this grid cell has two minima and maxima. In a second step, the spectral density of the anomalies with respect to the seasonal cycle is calculated. In a third step, interannual and subannual variance is estimated from the spectral density.

$$\mathrm{H}^+(i) - \overline{\mathrm{H}^+} \simeq \left.\frac{\partial \mathrm{H}^+}{\partial \mathrm{C_T}}\right|_{\overline{\mathrm{C_T}},\overline{\mathrm{A_T}},\overline{\mathrm{T}},\overline{\mathrm{S}}} \left(\mathrm{C_T}(i) - \overline{\mathrm{C_T}}\right) + \left.\frac{\partial \mathrm{H}^+}{\partial \mathrm{A_T}}\right|_{\overline{\mathrm{C_T}},\overline{\mathrm{A_T}},\overline{\mathrm{T}},\overline{\mathrm{S}}} \left(\mathrm{A_T}(i) - \overline{\mathrm{A_T}}\right) + \left.\frac{\partial \mathrm{H}^+}{\partial \mathrm{T}}\right|_{\overline{\mathrm{C_T}},\overline{\mathrm{A_T}},\overline{\mathrm{T}},\overline{\mathrm{S}}} \left(\mathrm{T}(i) - \overline{\mathrm{T}}\right) + \left.\frac{\partial \mathrm{H}^+}{\partial \mathrm{S}}\right|_{\overline{\mathrm{C_T}},\overline{\mathrm{A_T}},\overline{\mathrm{T}},\overline{\mathrm{S}}} \left(\mathrm{S}(i) - \overline{\mathrm{S}}\right), \quad (1)$$

and analogously for Ω_A. The partial derivatives are evaluated at $\overline{\mathrm{T}}$, $\overline{\mathrm{S}}$, $\overline{\mathrm{C_T}}$, and $\overline{\mathrm{A_T}}$, the temporal mean values of the drivers in the period of interest. While it is important to take into account the climatological total phosphate and total silicate concentrations for calculating the partial derivatives (Orr and Epitalon, 2015), one introduces only small errors by neglecting variations in phosphate and silicate. The partial derivatives in Equation 1 are evaluated using ~~the~~ *Mocsy 2.0* ~~package~~ (Orr and Epitalon, 2015).

Using the Taylor decomposition (Equation 1), one can for example express the seasonal variation in $[H^+]$ as a function of the drivers' seasonal variations (Kwiatkowski and Orr, 2018). In this study however, we analyze the time-series ~~standard deviation~~ variance of $[H^+]$ and Ω_A that also includes variability on other time scales (see Section 2.2.2) and the drivers of its changes. By making the Taylor approximation (Equation 1) and from the definition of variance (~~the squared standard deviation,~~ e.g. Coles (2001)), it follows that the ~~standard deviation~~ variance of $[H^+]$ can be written as a function of the partial derivatives with

respect to the drivers (sensitivities), the standard deviations of the drivers, and their pairwise correlation coefficients:

$$\begin{aligned}\sigma^2_{\mathrm{H^+}} = &\left(\frac{\partial \mathrm{H^+}}{\partial \mathrm{C_T}}\right)^2 \sigma^2_{\mathrm{C_T}} + \left(\frac{\partial \mathrm{H^+}}{\partial \mathrm{A_T}}\right)^2 \sigma^2_{\mathrm{A_T}} + \left(\frac{\partial \mathrm{H^+}}{\partial \mathrm{T}}\right)^2 \sigma^2_{\mathrm{T}} + \left(\frac{\partial \mathrm{H^+}}{\partial \mathrm{S}}\right)^2 \sigma^2_{\mathrm{S}} \\ &+2\frac{\partial \mathrm{H^+}}{\partial \mathrm{C_T}}\frac{\partial \mathrm{H^+}}{\partial \mathrm{A_T}}\mathrm{cov}\left(\mathrm{C_T},\mathrm{A_T}\right) + 2\frac{\partial \mathrm{H^+}}{\partial \mathrm{C_T}}\frac{\partial \mathrm{H^+}}{\partial \mathrm{T}}\mathrm{cov}\left(\mathrm{C_T},\mathrm{T}\right) \\ &+2\frac{\partial \mathrm{H^+}}{\partial \mathrm{C_T}}\frac{\partial \mathrm{H^+}}{\partial \mathrm{S}}\mathrm{cov}\left(\mathrm{C_T},\mathrm{S}\right) + 2\frac{\partial \mathrm{H^+}}{\partial \mathrm{A_T}}\frac{\partial \mathrm{H^+}}{\partial \mathrm{T}}\mathrm{cov}\left(\mathrm{A_T},\mathrm{T}\right) \\ &+2\frac{\partial \mathrm{H^+}}{\partial \mathrm{A_T}}\frac{\partial \mathrm{H^+}}{\partial \mathrm{S}}\mathrm{cov}\left(\mathrm{A_T},\mathrm{S}\right) + 2\frac{\partial \mathrm{H^+}}{\partial \mathrm{T}}\frac{\partial \mathrm{H^+}}{\partial \mathrm{S}}\mathrm{cov}\left(\mathrm{T},\mathrm{S}\right), \end{aligned} \tag{2}$$

where the pairwise covariances are functions of the ~~standard deviations~~ variances and correlation coefficients according to $\mathrm{cov}(x,y) = \sigma_x \sigma_y \rho_{x,y}$ and the partial derivatives are again evaluated at the temporal mean values $\overline{\mathrm{T}}$, $\overline{\mathrm{S}}$, $\overline{\mathrm{C_T}}$, and $\overline{\mathrm{A_T}}$. ~~We aim at quantifying the contribution from mean changes in the drivers , which change the sensitivities, compared to the contributions from changes in the drivers' standard deviationsand correlations to $\mathrm{H^+}$standard deviation changes . Unfortunately, these contributions can not be separated into summable terms because~~ This methodology has also been used to propagate uncertainties in carbonate system calculations (Dickson and Riley, 1978; Orr et al., 2018) and to identify drivers of potential predictability in carbonate system variables (Frölicher et al., 2020). Based on Equation 2 and the analogous result for Ω_A, a change in variance in $[\mathrm{H^+}]$ ~~standard deviation is a nonlinear function of those. However, we can analyse how much standard deviation change can be explained when considering only parts of the~~ and Ω_A can be attributed to changes in the sensitivities that arise from changes in the drivers mean states, to changes in the drivers ~~. We hence first analyze the change in $\mathrm{H^+}$standard deviation arising only from mean~~ standard deviations, and to changes in the pairwise correlations between the drivers. We do so by calculating the full Taylor series of Equation 2 that has contributions up to the fifth order. We then identify the variance change from mean changes in the drivers as the sum of all terms in the expansion that describe the contributions of sensitivity changes to the overall change in variance ($\Delta_s \sigma^2_{\mathrm{H^+}}$). Likewise, we identify the contribution from standard deviation changes in the ~~drivers. As a next step we also change the drivers' standard deviations and identify the additional change in~~ drivers ($\Delta_\sigma \sigma^2_{\mathrm{H^+}}$). We further group terms in the expansion that stem from simultaneous changes in the sensitivities and standard deviations ($\Delta_{s\sigma} \sigma^2_{\mathrm{H^+}}$) and the remaining terms that arise either from correlation changes alone or mixed contributions from correlation changes and changes in sensitivities and standard deviations ($\Delta_{\rho+} \sigma^2_{\mathrm{H^+}}$). Since these four components contain all terms in the Taylor series, they exactly reproduce a change in variance represented by Equation 2,

$$\Delta\sigma^2_{\mathrm{H^+}} = \Delta_s \sigma^2_{\mathrm{H^+}} + \Delta_\sigma \sigma^2_{\mathrm{H^+}} + \Delta_{s\sigma} \sigma^2_{\mathrm{H^+}} + \Delta_{\rho+} \sigma^2_{\mathrm{H^+}}. \tag{3}$$

However, Equation 2 itself is an approximation to the simulated $[\mathrm{H^+}]$ ~~standard deviation compared to the case of only changes in the means. Finally, we identify~~ and Ω_A variance, leading to a small mismatch between the sum of the components introduced above and simulated variance change (black lines vs. grey dashed lines in the zonal mean plots in Figures 8, 9, and 10).

We also assess the contributions from sensitivity changes arising only from mean changes in $\mathrm{C_T}$ ($\Delta_s \sigma^2_{\mathrm{H^+}}\big|_{\mathrm{C_T}}$), the contribution from standard deviation changes in $\mathrm{C_T}$ alone ($\Delta_\sigma \sigma^2_{\mathrm{H^+}}\big|_{\mathrm{C_T}}$), and the contribution from simultaneous mean and standard

deviation changes in ~~the phasing of the drivers as the additional change in H^+ standard deviation by also taking into account changes in the drivers' correlation coefficients (further information~~ $C_T\left(\Delta_{s\sigma}\sigma^2_{H^+}\big|_{C_T}\right)$. Further information on the decomposition is given in Appendix C~~).~~.

2.3 Model evaluation

The focus of our analysis is on changes in variability in $[H^+]$ and Ω_A. As ~~observational-based~~ observation-based daily data of the inorganic carbon chemistry at the global scale is not available, we limit the evaluation of the Earth system model simulation to the representation of the seasonal cycles of $[H^+]$ and Ω_A, and especially on its changes over the 1982-2015 period. ~~To do so, we created~~ We developed an observation-based dataset for surface monthly $[H^+]$ and Ω_A ~~covering the 1982-2015 period~~ using monthly surface salinity, temperature, pCO_2, and A_T fields. ~~Monthly salinity~~ Salinity and temperature data are taken from the Hadley Centre EN.4.2.1 analysis product (Good et al., 2013). A_T is then calculated using the *LIARv2* total alkalinity regression from salinity and temperature (Carter et al., 2018). For pCO_2, we use the neural-network-interpolated monthly data from Landschützer et al. (2016), which is based on SOCATv4 (Bakker et al., 2016). Although not fully capturing pCO_2 variability in regions with only few observations (Landschützer et al., 2016), the pCO_2 dataset ~~is~~ appears to be generally well suited for analyzing pCO_2 seasonality and changes therein (Landschützer et al., 2018). An exception is the Southern Ocean where data-based pCO_2 products are uncertain due to sparse data in winter (Gray et al., 2018). $[H^+]$ and Ω_A are then calculated from salinity, temperature, A_T, and pCO_2 using the *co2sys* carbonate chemistry package (van Heuven et al., 2011). Uncertainties in the derived seasonal cycles for $[H^+]$ and Ω_A that arise from uncertainties in the observation-based input variables are not quantified in this study.

In most regions, the GFDL ESM2M captures the observation-based mean seasonal cycle in $[H^+]$ and Ω_A well, in particular for Ω_A (the mean values of the seasonal amplitudes in Figure 3). However, potential biases in the mean seasonal amplitudes do not directly have an effect on projected changes in extreme events, as we base the extreme events definition on relative thresholds.

~~The~~ We then compare the simulated ensemble-mean trends in seasonal amplitude with the observation-based estimates (further information on the methodology is given in Appendix D). Similar as for the mean seasonal cycle, the GFDL ESM2M captures the observed trends in the seasonal $[H^+]$ and Ω_A amplitudes for different latitudinal bands over the 1982-2015 period relatively well (Figure 3). The ~~simulated ensemble mean~~ ensemble-mean trends in the simulated seasonal $[H^+]$ ~~amplitude (calculated as the average of the individual trends in the ensemble)~~ amplitudes are positive for ~~the northern low and high latitudes as well as the southern low latitudes~~ all latitude bands (Figure 3, ~~Supplementary~~ Table 1), consistent with the observation-based estimates. ~~In addition, both the model and the observational-based estimates show no significant change in H^+ seasonality~~ While the estimates for the simulated trends are significantly larger than zero for all latitude bands, this is not the case for the observation-based trends in the equatorial region ~~between~~ (10 °S ~~and~~ - 10°N) and the northern low latitudes (10 °N

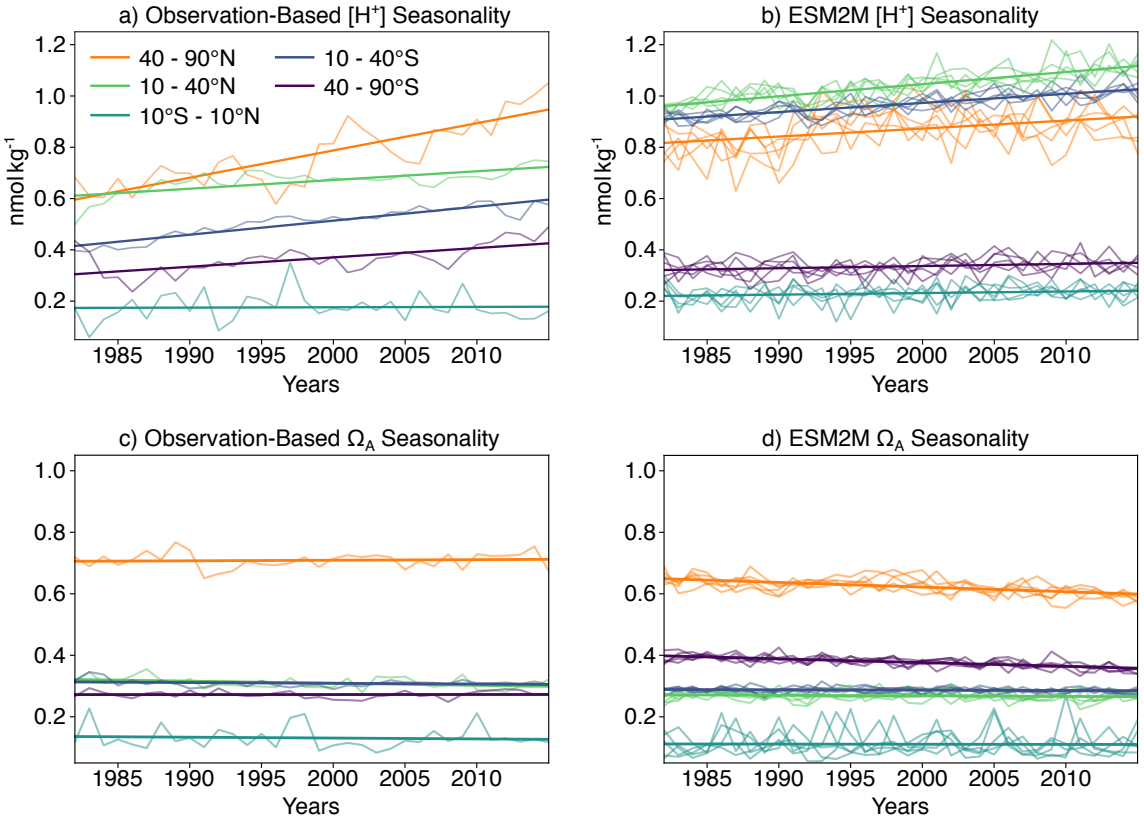


Figure 3. (a,b) Seasonal amplitude of $[H^+]$ over the period 1982-2015 averaged over five different latitude bands for a) the observation-based estimate and b) the GFDL ESM2M historical (1982-2005) and RCP8.5 (2006-2015) ensemble simulations. (c,d) The same as (a,b), but for Ω_A. Linear trends in all panels are overlaid as thick lines. The linear trend of the simulated changes is calculated as the mean of the five individual ensemble trends.

~~(purple thick lines in Figure 3a,b). However, the~~ - 40°N) (Table 1). The simulated $[H^+]$ seasonality trends ~~tend to be smaller~~ are significantly smaller (with 90% confidence level) than estimated from observations in the northern ~~(yellow~~ high (40°N - 90°N; orange thick lines in Figure 3a,b) and southern ~~high latitudes (red~~ low latitudes (10°S - 40°S; blue thick lines in Figure 3a,b), where the trends from the model ensemble are ~~0.031 ± 0.030~~0.031 ± 0.012 nmol kg^{-1} decade^{-1} ~~($\pm$ 90 % confidence interval for each ensemble member individually and then averaged) and 0.009 ± 0.010~~and 0.035 ± 0.003 nmol kg^{-1} per decade, compared to the observational-based trends of 0.106 ± 0.040 nmol kg^{-1} decade^{-1} and ~~0.037 ± 0.028~~0.055 ± 0.014 nmol kg^{-1} decade^{-1}, respectively. ~~In contrast, the simulated trend in H^+ seasonality~~ The simulated ensemble mean trends for the remaining latitude bands are not significantly different from the observation-based trend estimates.

For Ω_A, we find a significant negative trend in the observation-based data in the northern low latitudes ~~(0.047 ± 0.012 nmol kg^{-1} decade^{-1}; blue thick line in Figure 3b) is slightly larger than the observation-based trend (0.034 ± 0.032 nmol kg^{-1} decade^{-1}; blue thick~~

~~line in Figure 3a). For Ω_A, the model simulates~~ and significant negative trends in the simulations in the northern and southern high latitudes (Table 1). The negative trends in seasonal amplitude ~~that are not seen in the~~ in the simulations are significantly different from the observation-based ~~estimate~~ trends in the northern high latitudes (~~-0.015 ± 0.010 vs~~ -0.015 ± 0.004 vs. 0.002 ± 0.009 Ω_A units) and in the southern high latitudes (~~-0.012 ± 0.004 vs~~ -0.012 ± 0.002 vs. 0.000 ± 0.005 Ω_A units per decade). ~~Conversely, a slight negative trend is seen in northern low latitudes in the observation-based estimate (-0.007 ± 0.006 Ω_A units per decade) that is not present in the simulations. In addition, GFDL also captures the observed mean seasonal cycle $[H^+]$ and Ω_A well in most regions and in particular for Ω_A (the mean values of the seasonal amplitudes in Figure 3). However, potential biases in the mean seasonal amplitudes do not directly have an effect on projected changes in extreme events, as we base the extreme events definition on relative thresholds.~~

Latitude	Obs. $[H^+]$	ESM2M$[H^+]$	Obs. Ω_A	ESM2M Ω_A
40°N - 90°N	0.106 ± 0.040	0.031 ± 0.012	1.9 ± 8.7	-15.1 ± 3.8
10°N - 40°N	0.034 ± 0.034	0.047 ± 0.005	-6.7 ± 5.6	-1.8 ± 2.0
10°S-10°N	0.001 ± 0.016	0.006 ± 0.005	-2.8 ± 10.7	-0.5 ± 5.3
40°S - 10°S	0.055 ± 0.014	0.035 ± 0.003	-2.4 ± 5.1	-1.2 ± 1.2
90°S - 40°S	0.037 ± 0.028	0.009 ± 0.004	0.1 ± 4.8	-12.2 ± 1.7

Table 1. Linear trends in seasonal amplitude of $[H^+]$ (in nmol kg^{-1} $decade^{-1}$) and Ω_A (in 10^{-3} $decade^{-1}$) for five latitude bands over the period 1982-2015. Results are shown for the observational-based data (Obs.) and the five-member ensemble mean of the ESM2M simulations (ESM2M) following the RCP8.5 scenario over 2006-2015. The range ($\pm$) denotes the 90 % confidence interval.

In summary, taking into account additional evaluations not shown here of the mean states of $[H^+]$ and Ω_A and the underlying drivers (Bopp et al., 2013; Kwiatkowski and Orr, 2018), the model performs well against a number of key seasonal performance metrics. However, the model slightly underestimates past ~~changes~~ increases in seasonal amplitude of $[H^+]$, especially in the northern and southern high latitudes. ~~Furthermore, it overestimates~~ In contrast to the observation-based data, the model also projects negative trends in the Ω_A seasonal amplitude ~~in those regions~~there. Nevertheless, the observation-based trends in the northern and especially southern high latitudes are rather uncertain because winter time data is sparse there. Even though we lack the daily observational-based data to undertake a full assessment, ~~we consider our model to be well suited~~it appears that the GFDL ESM2M model is adequate to assess changes in open ocean ~~ocean~~ acidification extreme events.

3 Results

We start by discussing the simulated changes in different ocean acidity extreme variability event characteristics at the global scale (i.e. grid cell based characteristics are aggregated globally), before we analyze changes at the local to regional scale and identify the drivers of changes. We recall that the large secular increase in $[H^+]$ and the large secular decrease in Ω_A was removed for the analysis as we focus on changes in variability and their impact on extreme variability event characteristics.

3.1 Global changes in ocean acidity variability extremes

In preindustrial times, the GFDL ESM2M suggests that ~~a typical~~ an average surface [H^+] extreme ~~event had an~~ variability event had a maximal intensity of 0.08 nmol kg^{-1} (Figure 4c, ~~Supplementary~~ Table 2) and lasted 11 days (Figure 4e). Ocean acidity extremes ~~occur coherently with a typical volume~~ in the upper 200 m occur with a typical volume of $2.7 \cdot 10^3 km^3$, which is about 0.004 % of the total ocean volume in the upper 200 m (Figure 4g). Over the historical period (from preindustrial to 1986-2005), the model projects that the number of surface [H^+] extreme days ~~per year~~ increases from 3.65 days per year to 10.0 days per year (~~9.5-10.4 days per year;~~ Figure 4a, ensemble ranges are given in Table 2). The maximal intensity and duration are projected to increase to 0.12 nmol kg^{-1} and 15 days. Compared to preindustrial conditions, this corresponds to a 173% ~~(160-184%)~~ increase in number of days per year, a 44% ~~(39-47%)~~ increase in the maximal intensity and a 45% ~~(41-48%)~~ increase in the duration of [H^+] extreme variability events. The volume of ~~single~~ individual events is projected to increase by 20% ~~(14-27%)~~ over the historical period. Over the 21st century, ~~ocean acidity extreme events~~ extreme variability events in ocean acidity are projected to further increase in frequency, intensity, duration, and volume (Figure 4). By 2081-2100 under the RCP8.5 scenario, the number of [H^+] extreme days per year at surface is projected to increase to ~~50.1 days (50.0-50.3 days;~~ 50 days (corresponding to a 1273% increase relative to the preindustrial). The maximal intensity is projected to increase to 0.38 nmol kg^{-1} (~~0.37-0.39 nmol kg^{-1};~~ 371% increase), the duration to 32 days (~~31-32;~~ 199% increase) and the volume to $13.9 \cdot 10^3 km^3$ (~~13.8-14.1;~~ 414% increase).

At 200 m~~depth~~, [H^+] extreme variability events are in general more intense (0.17 nmol kg^{-1}; Figure 4d) and longer-lasting (38 days; Figure 4f) than at surface during preindustrial conditions. The stronger extreme events are caused by the overall larger variability at 200 m than at surface in the preindustrial. The longer duration is connected to the more pronounced contribution from interannual variability ~~compared to the surface~~ (see Section 3.3). However, projected relative changes over the historical period and the 21st ~~st~~ century are smaller at 200 m ~~depth~~ than at surface and with larger year-to-year variations across the ensembles. Under present-day conditions (1986-2005), the number of extreme days per year at 200 m ~~depth~~ is 4.3 days per year (~~3.7-5.1 day per year;~~ corresponding to a 18% increase since preindustrial), the maximal intensity 0.20 nmol kg^{-1} (~~0.19-0.21 nmol kg^{-1};~~ 18% increase), and the duration 46 days (~~43-50 days;~~ 21% increase). By the end of the 21st century under the RCP8.5 scenario, the number of [H^+] extreme days per year is projected to increase to 32.1 days per year~~(30.9-34.8 days per year), the~~ , the maximal intensity to 0.34 nmol kg^{-1} ~~(0.33-0.34 nmol kg^{-1})~~ and the duration to 99 days~~(95-102) days~~. Notably, extreme variability events in [H^+] ~~extreme events~~ are projected to become less intense at 200 m ~~depth~~ than at surface (0.34 nmol kg^{-1} vs. 0.38 nmol kg^{-1}) by the end of the century under RCP8.5, even though they were more intense in preindustrial times at depth. In contrast, surface [H^+] extreme variability events remain shorter in duration at the end of the century than at 200 m~~depth~~.

Under the RCP2.6 scenario and by the end of the century, the magnitude of changes in the different [H^+] extreme ~~event characteristics would be substantially reduced~~ variability event characteristics are substantially reduced compared to the

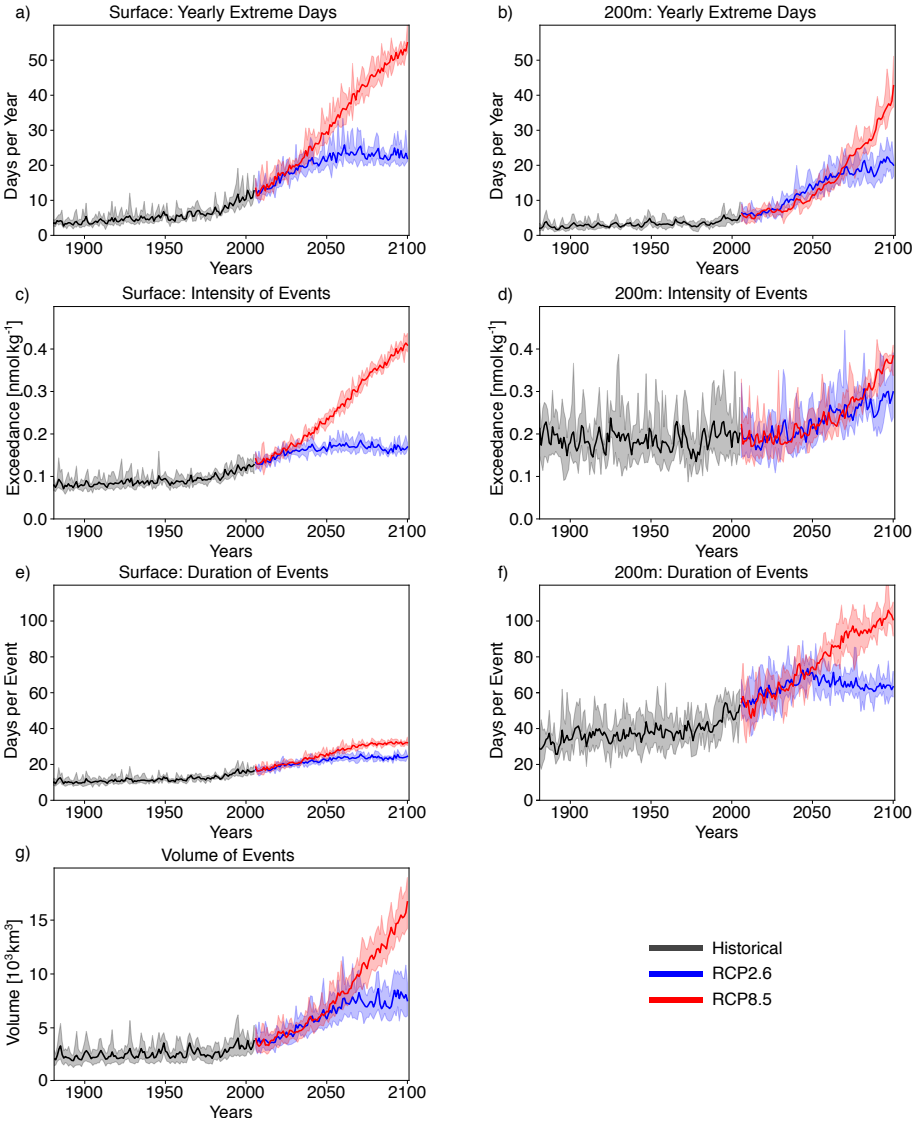


Figure 4. Simulated changes in globally averaged [H^+] extreme variability event characteristics over the 1861-2100 period following historical (black lines) and future RCP8.5 (red) and RCP2.6 scenario (blue). Frequency, maximal intensity, and duration are shown for the surface (a,c,e) and for 200 m ~~depth~~ (b,d,f). Volume is shown in (g). The thick lines display the five-member ensemble means and the shaded areas represent the maximum and minimum ranges of the individual ensemble members.

	PI	1986-2005	2081-2100 RCP2.6	2081-2100 RCP8.5
Number Surf.	3.65	9.97 (9.49-10.38)	22.87 (21.93-23.45)	50.12 (49.98-50.30)
200 m	3.65	4.32 (3.72-5.09)	19.88 (16.96-22.53)	32.10 (30.91-34.75)
Duration Surf.	10.64	15.38 (15.04-15.72)	23.79 (23.40-24.11)	31.78 (31.23-32.13)
200 m	38.00	45.95 (42.84-49.96)	62.94 (60.49-66.11)	98.66 (95.06-102.01)
Maximal Intensity Surf.	0.08	0.12 (0.11-0.12)	0.17 (0.16-0.17)	0.38 (0.37-0.39)
200 m	0.17	0.20 (0.19-0.21)	0.28 (0.25-0.30)	0.34 (0.33-0.34)
Volume	2709	3247 (3082-3451)	7654 (6873-8464)	13927 (13836-14109)

Table 2. Simulated global ensemble-mean $[H^+]$ extreme variability event characteristics for the preindustrial (PI), present day (1986-2005), and the end of this century (2081-2100) for both RCP2.6 and RCP8.5. Numbers of yearly extreme days are given in days per year, durations in days, intensities in nmol kg^{-1}, and volumes in km^3. Values in brackets denote ensemble minima and maxima.

RCP8.5 scenario. This reduction is especially pronounced at the surface (blue lines in Figure 4). There, the number of extreme days per year, maximal intensity, and duration under the RCP2.6 are projected to be only 46% (44-47), 43% (43-44) and 75% (73-77) of that under the RCP8.5 scenario. At depth, the differences between the RCP2.6 and RCP8.5 scenario are less pronounced and only emerge in the second half of the 21st century. In contrast to the surface, the number of $[H^+]$ extreme days per year ~~, the intensity, and~~ and the maximal intensity at depth as well as the volume of events ~~at depth~~ are projected to ~~considerably increase~~ increase significantly even after the atmospheric CO_2 concentration ~~stabilize~~ stabilizes in RCP2.6 around year 2050. This delayed response at subsurface is due to the relatively slow ~~surface-to-deep ocean~~ surface-to-subsurface transport of carbon. However, this is not the case for the duration, which slightly decreases in the second half of the 21st century at depth (Figure 4f). This decrease in duration mainly occurs in the subtropics, where events generally last long (Figure A2b). It is connected to an increase in the contribution from high-frequency variability to total variability in those regions over that period.

In contrast to $[H^+]$ extreme variability events, the yearly number of Ω_A extreme variability days is projected to decrease over the historical and the 21st century under both the RCP8.5 and RCP2.6 scenario (Figure 5a-b, Supplementary Table A1). The ~~simulated decrease is slightly stronger under RCP8.5 than under RCP2.6. The~~ number of surface Ω_A extreme variability days per year by the end of ~~the~~ this century is projected to be 63% ~~(-70,-54)~~ smaller under RCP8.5 and 39% ~~(-49,-20)~~ smaller under RCP2.6 than at preindustrial (ensemble ranges are given in Supplementary Table A1). Projected changes at depth are less pronounced than at surface, again with larger decreases under RCP8.5 than under RCP2.6. It should be noted that, despite this decline in extreme variability events, the long-term decline in the mean state of Ω_A still leads to more frequent occurrence of low values in Ω_A (see Discussion section).

3.2 Regional changes in ocean acidity variability extremes

Surface $[H^+]$ variability extremes are projected to become more frequent in 87% of the surface ocean area by the end of the 21st century under the RCP8.5 scenario. However, the projected changes in these ocean acidity extremes are not uniform

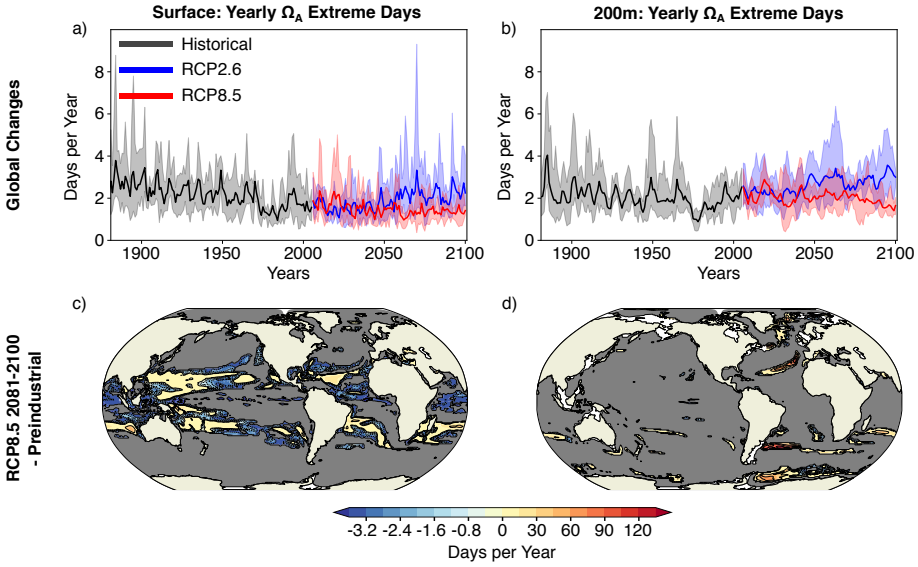


Figure 5. Simulated changes in the yearly number of Ω_A extreme variability days. Panels a-b) show the g~~G~~lobally averaged simulated number of extreme ~~Ω_A~~ variability days per year in Ω_A from 1861 to 2100 following historical (black lines) and future RCP2.6 (blue) and RCP8.5 (red) scenarios at (a) the surface and (b) 200 m~~depth~~. The thick lines display the five-member ensemble means and the shaded areas represent the maximum and minimum range of the individual ensemble members. Panels c-d) show the simulated regional changes in the number of extreme variability days per year in Ω_A from preindustrial to 2081-2100 under the RCP8.5 scenario (c) at the surface and (d) at 200 m. Shown are changes averaged over all five ensemble members. The black lines highlight the pattern structure and grey colors represent regions where no ensemble member simulates variability extremes during 2081-2100.

over the globe ~~but show distinct spatial patterns~~ (Figure 6; Supplementary Figure A2). The largest increases in the number of $[H^+]$ extreme days per year are projected in the Arctic Ocean (up to +120 days per year), in the subtropical gyres (up to +60 days per year), in parts of the Southern Ocean and near Antarctica. There are also some regions including the eastern equatorial Pacific and parts of the Southern Ocean, where the number of yearly extreme days in surface $[H^+]$ is projected to decrease. These are in general also the regions where the seasonality in $[H^+]$ is projected to decrease (see section 3.3 below). The largest changes in intensity of surface $[H^+]$ variability extremes (Figure 6c) are projected for the subtropics, especially in the Northern Hemisphere. For example, events become up to 0.8 nmol kg^{-1} more intense in the subtropical North Pacific and Atlantic. Large changes are also projected for the Arctic Ocean and around Antarctica. Regions with large increases in the number of yearly extreme days tend to show also large increases in the duration of extreme variability events (Figure 6e). The Arctic Ocean is an exception. Although the number of yearly extreme days increases sharply, the increase in duration is not as pronounced. This is because extremes are already long-lasting, but rare at preindustrial times (Supplementary Figure A2). So even though extreme variability events are projected to ~~reoccur annually~~ occur each year by the end of the century under RCP8.5, the ~~absolute~~ increase in duration is relatively small.

At 200 m~~depth~~, the projected pattern of changes in yearly extreme event days generally resembles ~~the changes~~ that at the surface (Figure 6b). The largest increases in yearly extreme event days are projected for parts of the subtropics, the Southern Ocean, and the Arctic Ocean. In contrast to the surface, $[H^+]$ ~~extremes~~ variability extremes at 200 m are projected to become less frequent in the equatorial Atlantic, the northern Indian Ocean, the North Pacific and in large parts of the Southern Ocean. The regions ~~showing~~ indicating a decline in $[H^+]$ variability extremes at depth include also some of the eastern boundary current systems, such as the the Humboldt, California, and Benguela Current systems. In most of these regions, extreme ~~H^+ days~~ variability events are projected to disappear in the RCP8.5 ~~ensemble~~ scenario by the end of this century (grey regions in Figure 6b). The largest increases in subsurface event intensity are projected in the subtropics (Figure 6d), whereas the duration of $[H^+]$ variability extremes is projected to ~~strongly increase~~ increase strongly in many regions of the mid-to-high latitudes of both hemispheres (Figure 6f). The projected increases in duration at ~~200m depth~~ 200 m are much larger than at surface.

The increase in the number of extreme days per year, the maximal intensity, and the duration is smaller under RCP2.6 compared to RCP8.5 for most of the ocean (Supplementary Figure A1). The largest increases in occurrence of variability extremes under RCP2.6 are simulated for the Arctic Ocean, similar as under RCP8.5, and for parts of the Southern Ocean. The regions in the Southern Ocean where the occurrence of ~~extremes~~ extreme variability events is projected to decrease largely overlap with those for RCP8.5, at surface and at depth. On the other hand, unlike under RCP8.5, a decrease in extreme variability event occurrence is only projected for a small fraction of the tropical oceans under RCP2.6.

~~Extreme~~ While the decline in mean Ω_A generally leads to lower values in Ω_A, extreme variability events in Ω_A are projected to become less frequent throughout most of the ocean (89% of surface area under RCP8.5 at the end of the 21st century; Figure ~~7a~~5c). In many regions, ~~Ω_A extreme events~~ extreme variability events in Ω_A are projected to disappear by 2081-2100 under the RCP8.5 scenario (grey regions in Figure ~~7a~~5c). However, the frequency of surface Ω_A variability extremes is projected to increase by 10 or more days per year in the subtropical gyres, especially in the western parts of the subtropical gyres. At depth, no extreme variability events are projected for most of the ocean during 2081-2100 under RCP8.5 (Figure ~~7b~~5d).

3.3 Decomposing $[H^+]$ variability changes into interannual, seasonal, and ~~residual daily~~ subannual variability changes

The underlying changes in $[H^+]$ variability ~~and extreme events in H^+~~ may arise from changes in interannual variability, seasonal variability, and ~~residual daily~~ subannual variability. We therefore decompose the total variability into these three components (see Section 2.2.2). ~~Figure 7 shows their contribution to the overall change in variability from preindustrial to 2081-2100 following~~ For the preindustrial, the model simulates overall larger $[H^+]$ variance at depth than at the surface (0.42 nmol^{2}kg^{-2} vs. 0.15 nmol^{2}kg^{-2}, not shown). Seasonality has the largest contribution at the surface (81 % of total variance). At 200 m, interannual variability has the largest contribution (63 %), and also subannual variability is more important compared to the surface (15% vs. 8%).

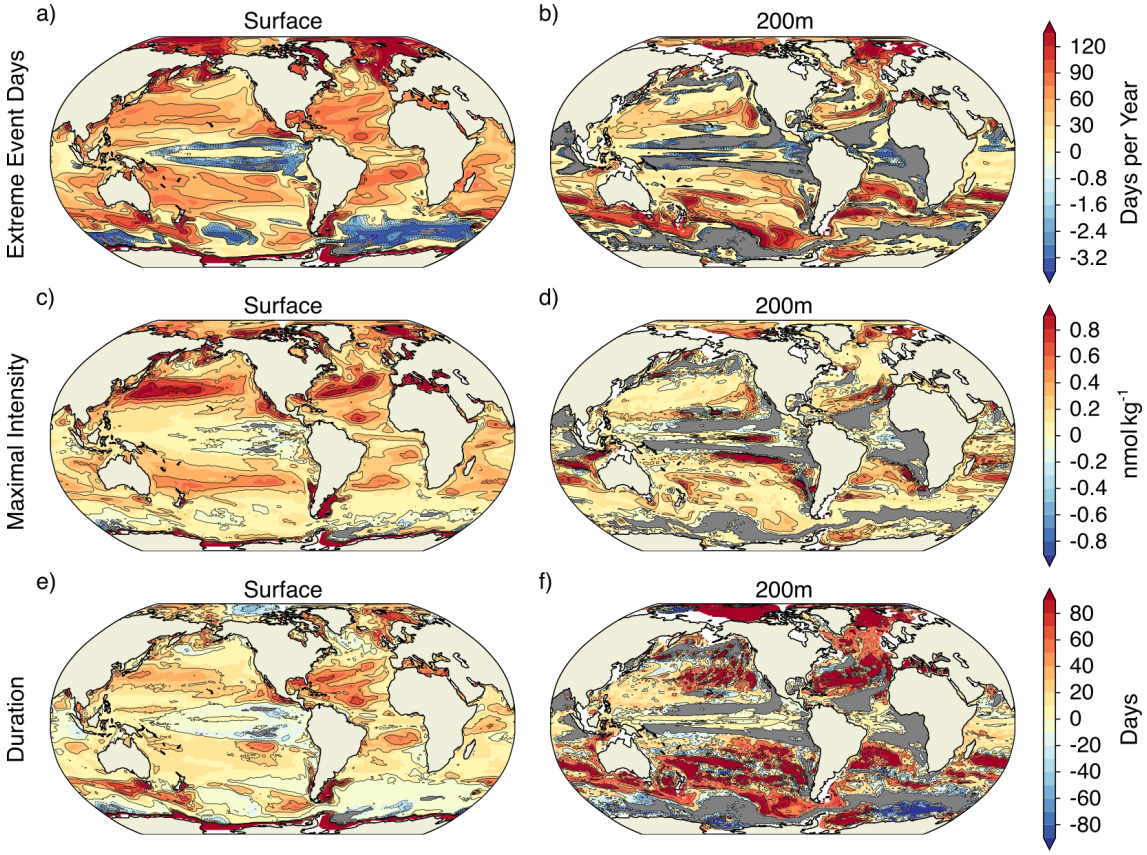


Figure 6. Simulated regional changes in $[H^+]$ extreme variability event characteristics from preindustrial to 2081-2100 under the RCP8.5 scenario at surface and at depth for (a,b) the number of extreme event days in days per year, (c,d) the maximal intensity of events in nmol kg^{-1}, and (e,f) the duration of events in days. Shown are changes averaged over all five ensemble members. Grey colors represent areas, where no variability extremes occur during 2081-2100 and the black lines highlight pattern structures.

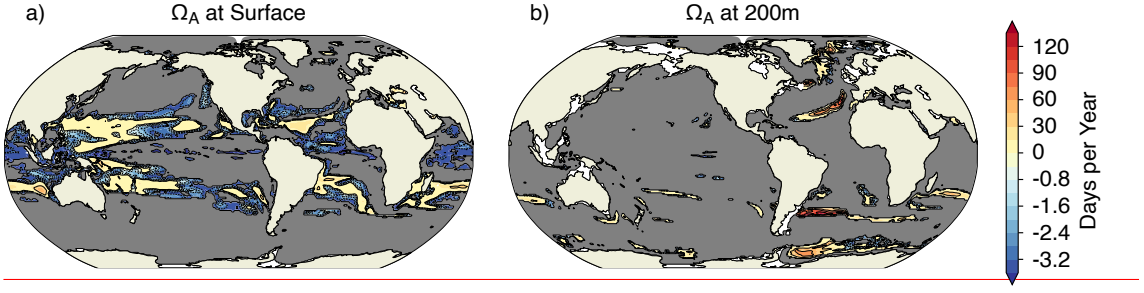


Figure 7. ~~Simulated regional changes in the number of Ω_A extreme~~ variability ~~days per year in Ω_A from preindustrial to 2081-2100 under the RCP8.5 scenario (a) at the surface and (b) at 200 mdepth. Shown are changes averaged over all five ensemble members. The black lines highlight the pattern structure and grey colors represent regions where no ensemble member simulates~~ variability ~~extremes during 2081-2100.~~

From preindustrial to the end of this century under the RCP8.5 scenario~~at surface and at depth. Changes~~, changes in seasonality clearly dominate the overall change in variability at surface with 87% contribution to the overall variance change in the global mean (Figure 7b,d). Changes in interannual variability (3% contribution to overall variance change; Figure 7a~~) and residual daily~~ ,d) and subannual variability (10%; Figure 7c,d) play a minor role. The largest increases in variability for all three variability types are projected for the northern high latitudes, where also the number of extreme variability event days increases most strongly. The increases in extreme events around Antarctica and the southern end of South America (Figure 6a) are mainly caused by increases in seasonal variability (Figure 7b). The regions that are projected to experience a decline in variability extremes (Figure 6a) coincide with those of decreasing interannual and seasonal variability (Figure 7a,b).

In contrast to the surface, changes in interannual and to a lesser extent ~~residual daily~~ subannual variability at 200 m ~~depth~~ are also important for explaining the changes in $[H^+]$ variability ~~and~~ extremes (Figure 7~~d,f). Even though the changes in seasonal variability still dominate the overall changes in variance~~ e,g,h). Changes in interannual variability contribute most to overall variance change at the global scale (with 42% contribution)~~, the interannual variability is~~. Seasonal variability changes are almost equally important (37%)~~. Changes in residual daily~~, and changes in subannual variability also contribute substantially to changes in total variability (20%). The patterns of variability changes are very similar across the three types of variability. The largest increases in $[H^+]$ variability are simulated ~~in the subtropics~~north and south of the equator. In those regions the model also projects an increase in $[H^+]$ extreme variability events (Figure 6b). Furthermore, these regions tend to be already more variable ~~at~~ during the preindustrial (see Supplementary Figure A2a). However, the model also projects an increase in variability for less variable regions at preindustrial, such as ~~the Arctic Ocean~~northern high latitudes, leading to increases in variability extremes there. All three variability types are projected to decrease in the tropics and parts of the Southern Ocean, where the occurrence of extreme variability events is projected to largely decrease (c.f. Figure 6b). The variability decrease in those regions is most pronounced for interannual variability (Figure 7~~d~~e).

3.4 Drivers of ~~changes in~~ $[H^+]$ and Ω_A variability changes

In this section, we investigate the drivers of variability changes in $[H^+]$ ~~. Changes in H^+ variability can be attributed to three factors~~ and Ω_A. We attribute changes in variability to four factors (see Section 2.2.3 for further details): (i) changes in the mean states of the drivers that control the sensitivities (~~changing the partial derivatives in Equation 2 that affect the amplitude of the variation in H^+ for a given deviation of a driving variable x from its mean $(x(i) - \overline{x})$ in Equation 1~~$\Delta_s \sigma^2_{H^+}$), (ii) changes in the variabilities of the drivers ($\Delta_\sigma \sigma^2_{H^+}$), (~~ii)~~ iii) simultaneous changes in the mean states and variabilities of the drivers (~~the drivers' standard deviations in Equation 2~~$\Delta_{s\sigma} \sigma^2_{H^+}$; this contribution arises because both mean states and variabilities change, and can neither be attributed to (i) nor (ii) alone), and (~~iii~~iv) changes in the ~~phasing~~ correlations between the drivers~~(the correlation coefficients in Equation 2 that describe the degree to which the drivers co-vary with each other) (see Section 2.2.3).~~ , also including mixed contributions from correlation changes together with mean state and variability changes ($\Delta_\rho \sigma^2_{H^+}$). In other words, (iv) describes the change in variability that arises because the correlations between the drivers also change, and not only

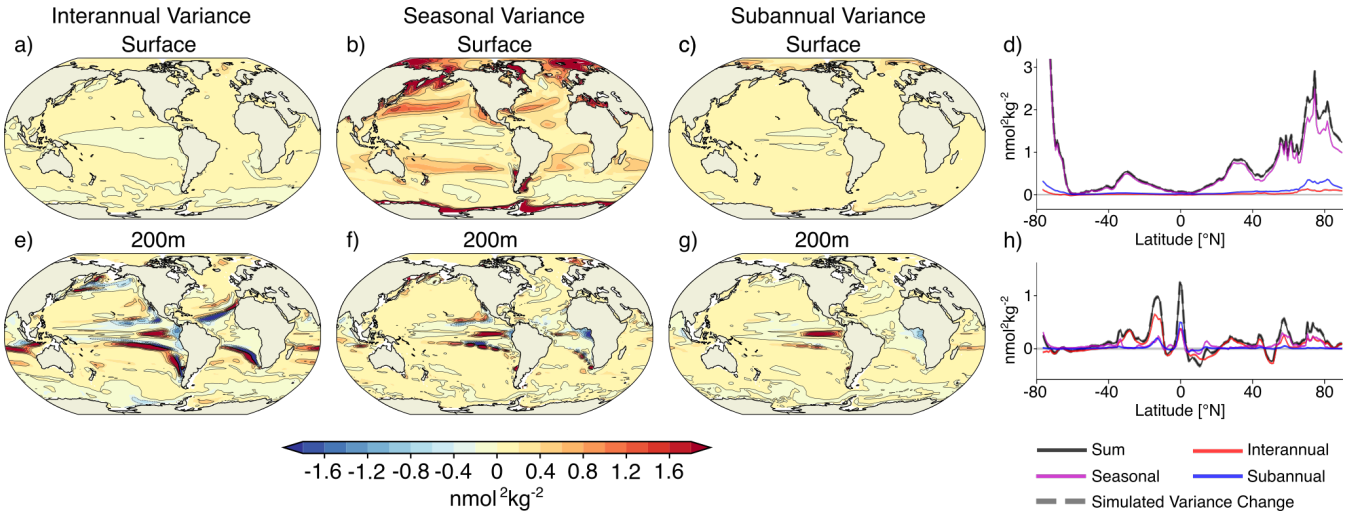


Figure 7. Contribution to projected changes in [H^+] ~~standard deviation~~ variance from (a,~~d~~e) interannual variability, (b,~~e~~f) seasonal variability, and (c,~~f~~g) ~~residual daily~~ subannual variability between the preindustrial and the 2081-2100 period following the RCP8.5 scenario at surface and at 200 m. Shown are the ensemble mean changes. The black lines highlight the pattern structure. Zonal mean contributions are shown for the surface (d) and for 200 m (h). The sum of the three components (black lines) accurately reproduces the simulated variance change (grey dashed lines).

their mean states and variabilities.

~~We investigate the first contribution (impact of mean state changes on the sensitivities) by changing only the mean states of the drivers (in Equation 2) from the preindustrial values to those of the~~ The drivers' mean changes between the preindustrial and 2081-2100 ~~period following the~~ under RCP8.5 ~~scenario. The drivers' standard deviations and correlation coefficients stay at the preindustrial levels. Figure 8b shows that the mean changes in the drivers cause an overall~~ cause a strong increase in surface [H^+] variability ~~. Global H^+ variability changes due to mean changes (0.67 nmol kg^{-1}~~ between the periods that is most pronounced in the northern and southern high latitudes ($\Delta_s\sigma^2_{H^+}$; Figure 8b~~) are more than twice as large as the total realized changes (0.30 nmol kg^{-1};~~ , pink line in Figure 8~~a). Especially in the higher latitudes, the~~ h). On global average, these variance changes due to ~~mean changes~~ the mean changes in the drivers ($\Delta_s\sigma^2_{H^+}$=1.3 nmol^{2}kg^{-2}) are much larger than the total ~~realized changes~~ simulated variance change in [H^+] ~~variability.~~

~~What causes this large increase H^+ variability?~~ ($\Delta\sigma^2_{H^+}$=0.5 nmol^{2}kg^{-2}, Figure 8a, dashed grey line in Figure 8h). In general, an increase in mean C_T, temperature, and salinity would lead to an increase in ~~H^+ variability~~$\Delta_s\sigma^2_{H^+}$, whereas an increase in mean A_T would lead to a decrease. GFDL ESM2M projects an increase in mean C_T over the entire surface ocean (Supplementary Figure A4) and therefore an increase ~~H^+ variability~~ $\Delta_s\sigma^2_{H^+}|_{C_T}$ (Figure 8~~e). Surface A_T is projected to increase globally, especially in the low-to-mid latitudes of the Atlantic (Supplementary Figure A4), and therefore dampens slightly the overall increase in~~ f, green line in Figure 8h). In the high latitudes, a relatively small increase in mean C_T leads a large increase in

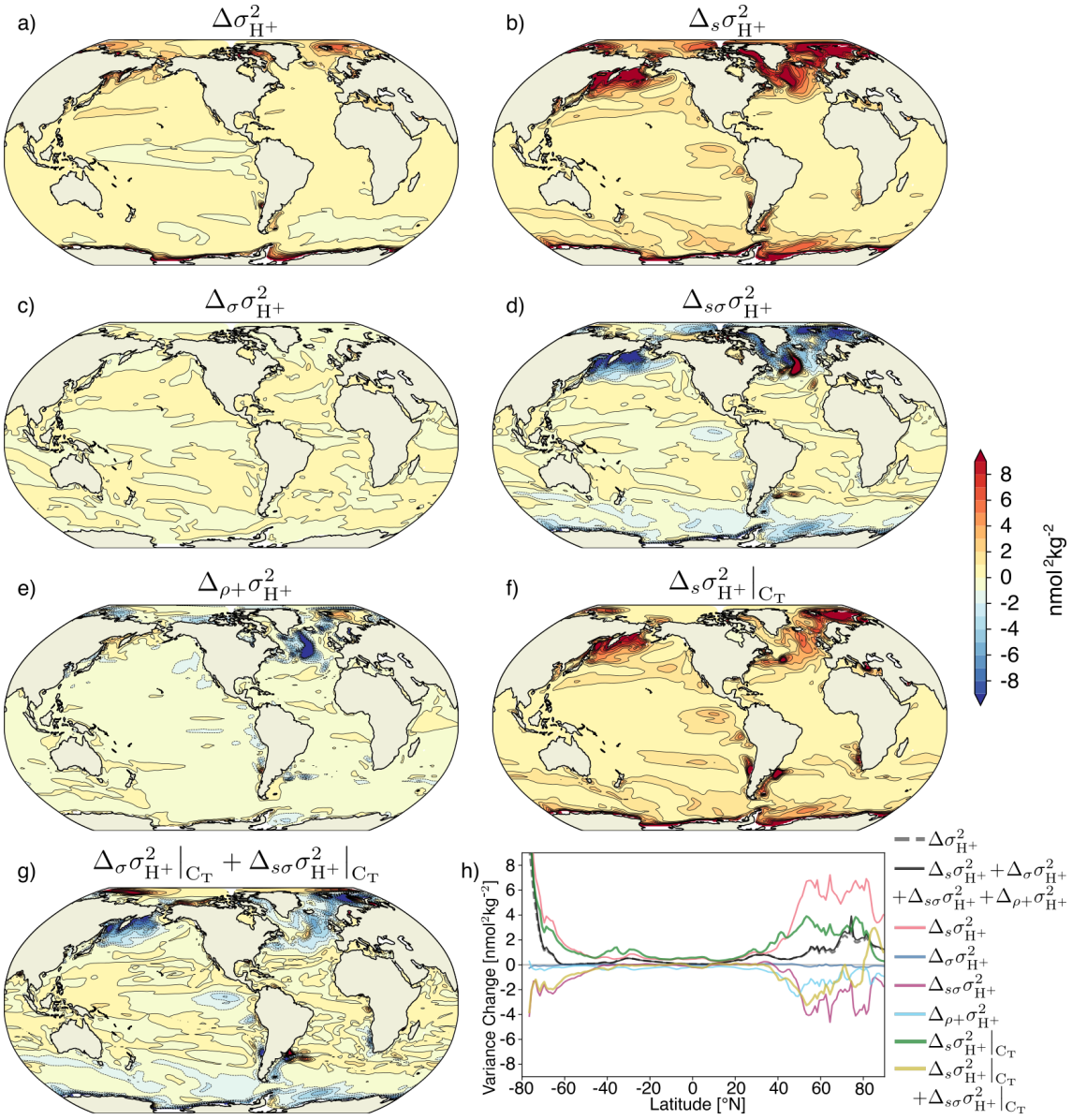


Figure 8. ~~(a) The realized change in~~ Decomposition of surface $[H^+]$ ~~standard deviation at surface~~ variability changes into different drivers. Shown are changes from preindustrial to 2081-2100 following the RCP8.5 scenario. ~~It is decomposed into (b) the~~ The simulated change in $[H^+]$ ~~standard deviation~~ variance ($\Delta\sigma^2_{H+}$) (a) is decomposed into the contribution from changes in the sensitivities that arise from changes in the drivers' mean values ~~,~~ (~~e~~$\Delta_s\sigma^2_{H+}$) (b), the ~~H^+ standard deviation change~~ contribution from ~~additionally changing~~ changes in the drivers~~variabilities, and~~ ' standard deviations (~~d~~$\Delta_\sigma\sigma^2_{H+}$) (c), the contribution from ~~additional~~ simultaneous changes in the ~~phasing of~~ sensitivities and the drivers' standard deviations (~~pairwise correlations between~~ $\Delta_{s\sigma}\sigma^2_{H+}$) (d), and the ~~drivers~~contribution from correlation changes alone and simultaneous changes in correlations and sensitivities and standard deviations ($\Delta_{\rho+}\sigma^2_{H+}$) ~~.~~ (e)~~The change in~~ . Furthermore, the contribution to $[H^+]$ ~~standard deviation obtained when only changing the~~ variance change from mean ~~state of~~ changes in C_T ~~,~~ alone ($\Delta_s\sigma^2_{H+}|_{C_T}$) (f) ~~The change~~ and that from standard deviation changes in ~~H^+~~C_T together with simultaneous changes in mean state and standard deviation ~~from additionally changing the variability~~ of C_T ~~.~~ (~~g,h~~$\Delta_\sigma\sigma^2_{H+}|_{C_T} + \Delta_{s\sigma}\sigma^2_{H+}|_{C_T}$) ~~Same as~~ (~~e,f~~g) ~~, but for A_T. Temperature and salinity contributions are small and not~~ is shown~~here~~. The black contours in a-g) highlight the pattern structures. The zonal mean contribution of panels a-g) is shown in panel h).

$\Delta_s\sigma^2_{H^+}\big|_{C_T}$, because [$H^+$] ~~variability (Figure 8g~~is very sensitive to changes in C_T due to the low buffer capacity there. Decreases in mean A_T further contributes to the increase in $\Delta_s\sigma^2_{H^+}$ (not shown). This is not the case in the ~~Arctic Ocean, where~~ low-to-mid latitudes, where mean surface A_T is projected to ~~decrease and amplifies the increase in H^+ variability caused by changes in C_T (Figure 8g). Changes~~ increase, in particular in the Atlantic Ocean (Figure A4), and therefore dampens slightly the overall increase in $\Delta_s\sigma^2_{H^+}$. The changes in A_T are largely due to changes in freshwater cycling that also manifest in salinity changes (Supplementary Figure A4, Carter et al. (2016)). Mean changes in temperature and salinity play a minor role for explaining the large increase in $\Delta_s\sigma^2_{H^+}$ (not shown~~in Figure 8~~).

~~Next, we investigate the second contribution, i.e. the impact of changes in variability of the drivers on H^+ variability (Figure 8c). At the global scale~~Why is the increase in $\Delta\sigma^2_{H^+}$ smaller than that following from the mean changes in the drivers (i.e. $\Delta_s\sigma^2_{H^+}$)? In the high latitudes, the projected change in the variability of the drivers (Supplementary Figure A5) ~~causes a decrease in~~ contributes negatively to the [H^+] variability change and counteracts to some degree the increase in ~~H^+ variability due to increase in the mean drivers. However, this decrease due to the drivers' variability is spatially not uniform. In the low latitudes (except in the eastern equatorial Pacific), changes in the variability of the drivers lead to a slight increase in H^+ variability, whereas in the high latitudes they~~ $\Delta_s\sigma^2_{H^+}$. These variability changes alone would have a small impact on $\Delta_\sigma\sigma^2_{H^+}$ (Figure 8c), but the variability changes dampen the increases from the mean changes ($\Delta_{s\sigma}\sigma^2_{H^+}$, Figure 8d). The latter contribution is large in regions where mean changes would else lead to a ~~decrease in H^+ variability. The increase in the low latitudes and the decrease in~~ strong increase (see anticorrelated patterns in Figures 8b and d). In the high latitudes~~and the equatorial Pacific is mainly caused by increases and~~, decreases in C_T variability (~~Figure 8f;~~ Supplementary Figure A5a) together with increases in mean C_T (Supplementary Figure ~~A5). Again, the contribution from changes in A_T variability~~ A4a) can explain a large part of the negative contribution from $\Delta_{s\sigma}\sigma^2_{H^+}$ (Figure 8~~h) is important in the Arctic Ocean, while temperature and salinity play a minor role~~g and golden line in Figure 8h). In the northern high latitudes, also mean and variability changes in A_T are important for $\Delta_{s\sigma}\sigma^2_{H^+}$ (not shown). The additional ~~variability change from the third contribution (~~contribution from changes in the ~~phasing of the drivers~~ correlations between the drivers ($\Delta_{\rho\pm}\sigma^2_{H^\pm}$; Figure 8~~d) is overall smaller, except~~ e) also tends to contribute negatively to [H^+] variability changes, especially in the North Atlantic.

At 200 m~~depth, the picture looks similar as at the surface. Again~~, the projected increase in ~~H^+ variability~~ $\Delta\sigma^2_{H^+}$ (Figure 9a) ~~caused by~~ is also a result of the large increase due to the mean changes in the drivers ($\Delta_s\sigma^2_{H^+}$;Figure 9b; Supplementary Figure A4) ~~is damped by the decrease~~ and the decrease due to the interplay between mean changes and decreases in the variability ~~of the drivers (Figure 9c; Supplementary Figure A5). The~~ ($\Delta_{s\sigma}\sigma^2_{H^+}$; Figure 9d). Similar to the surface, the changes in mean and variability of C_T ~~(Figure 9e,f)~~ are the most important drivers of changes ~~. Increases in mean A_T somewhat dampen the overall increase in the low latitudes~~ (Figure 9~~gf,g~~; green and golden lines in Figure 9h). Again, changes in temperature and salinity are ~~negligible~~of minor importance in most areas (not shown). In contrast to the surface, however, the individual ~~changes in~~ compensating contributions to [H^+] variability ~~due to the~~ change from mean and variability changes ~~of~~ in the drivers, in particular ~~of~~ those in C_T, are much larger at 200 m~~depth~~. The global average variance change due to the mean changes in the

drivers ($\Delta_s\sigma^2_{H^+} = 3.7$ nmol^{2}kg^{-2}) is much larger than the overall simulated variance change ($\Delta\sigma^2_{H^+} = 0.1$ nmol^{2}kg^{-2}). The largest individual changes are projected for the southern edges of the subtropical gyres in the north and for the northern edges of the subtropical gyres in the south. ~~The changes in the Arctic Ocean at 200 m depth are overall smaller than at the surface. Changes in the phasing of the drivers are overall less important than changes in the variabilities and mean states.~~There, the preindustrial background [H^+] variability is also the largest (Figure A3a). As a result, an increase in the sensitivities due to an increase in mean C_T has the largest effect there. The contribution from changes in the correlations between the drivers is overall small (Figure 9e and cyan line in Figure 9h).

Unlike for [H^+], both mean changes ($\Delta_s\sigma^2_\Omega$; red lines in Figure 10) and variability changes in the drivers ($\Delta_\sigma\sigma^2_\Omega$: blue lines in Figure 10) lead to a decrease in Ω_A variability ($\Delta\sigma^2_\Omega$; black dashed lines in Figure 10). At 200 m, variability changes are even the dominant driver for reductions in Ω_A variability. Simultaneous changes in means and variabilities ($\Delta_{s\sigma}\sigma^2_\Omega$; purple lines in Figure 10) contribute positively and dampen the reduction in Ω_A variability from mean and variability changes alone. Mean and variability changes in C_T are the main drivers for changes in Ω_A variability as indicated by the tight relation between the dashed and solid red, blue, and purple lines in Figure 10, in particular at 200 m. An exception is the northern high latitudes, where A_T changes also play a substantial role at the surface (not shown). Correlation changes in the drivers ($\Delta_{\rho+}\sigma^2_\Omega$; cyan lines in Figure 10) are of similar relative importance as for [H^+] and again have the largest imprint in the northern mid-to-high latitudes at the surface.

4 Discussion and conclusions

We provide a first quantification of the historical and future changes in ~~short-term ocean acidity extreme events at global and regional scale~~ extreme variability events in ocean acidity by analyzing daily mean 3D output from a ensemble simulation of a comprehensive Earth system model. In our analysis, we focus on changes in extreme events that arise only from changes in daily to interannual variability. Secular CO_2 emission-induced trends in the mean state were removed from the model output before ~~the extreme events analysis~~analyzing extremes. We show that extreme variability events in [H^+] ~~events~~ are projected to become more frequent, longer lasting, more intense, and spatially more extensive under increasing atmospheric CO_2 concentration, both at surface and also within the thermocline. These changes in [H^+] extreme event characteristics are substantially reduced under the RCP2.6 scenario compared to RCP8.5. The increase in [H^+] ~~variability and extreme~~ extreme variability events is a consequence of ~~its nonlinear dependence on the~~ increased sensitivity of [H^+] to variations in its drivers. It is mainly driven by the projected increase in mean C_T and additionally altered by changes in C_T variability and A_T mean and variability ~~. Extreme~~ as well as changes in the correlations between the drivers. Extreme variability events in Ω_A are projected to become less frequent in the future. ~~This~~ It is because Ω_A, unlike [H^+], becomes less sensitive to variations in the ~~physical and biogeochemical state of seawater under elevated atmospheric CO_2~~drivers with the mean increase in C_T. Furthermore, the projected reductions in the drivers' variabilities, mainly in C_T, significantly add to the reduced occurrence of Ω_A variability

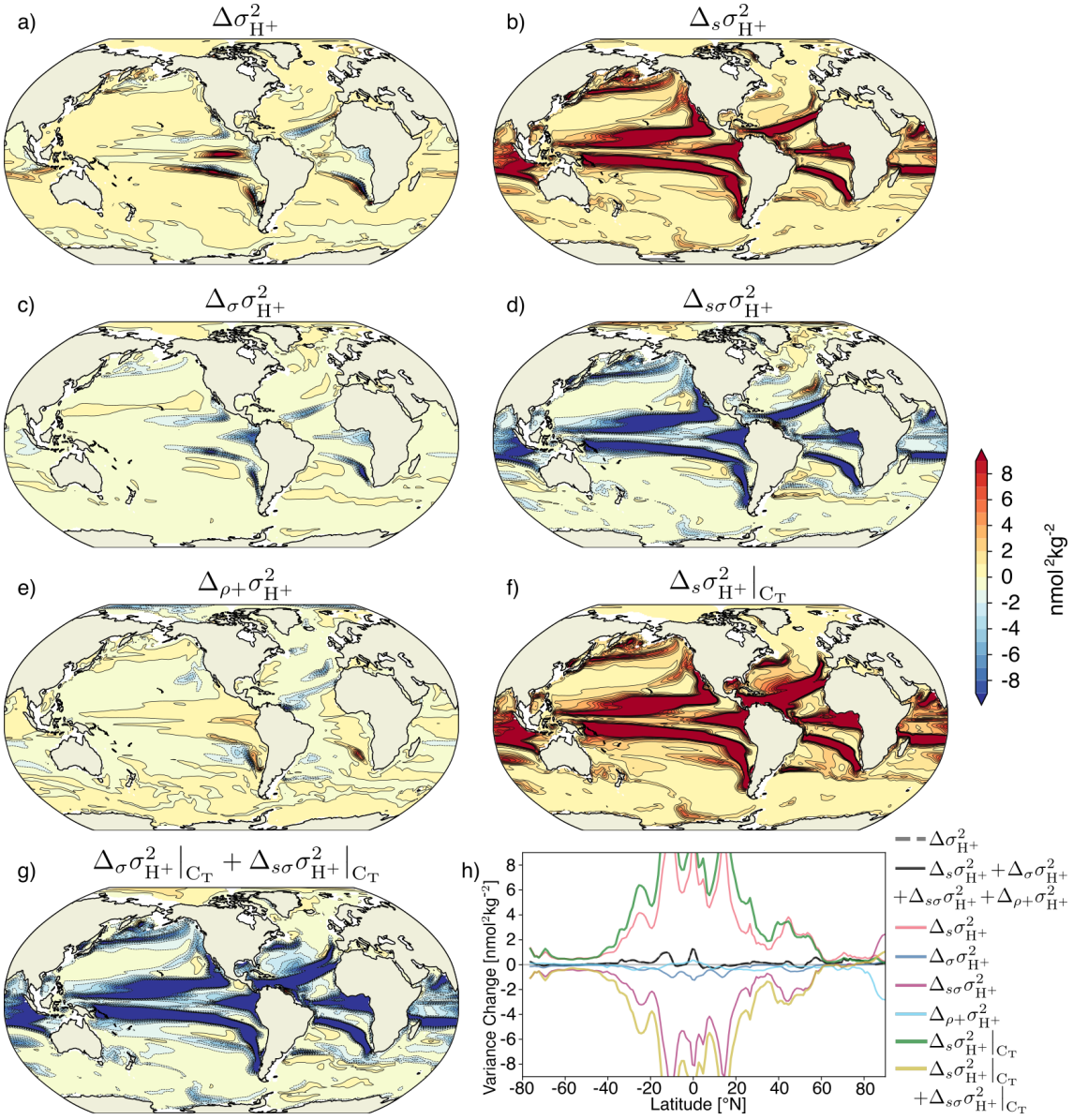


Figure 9. Same as Figure 8 but at 200 m.

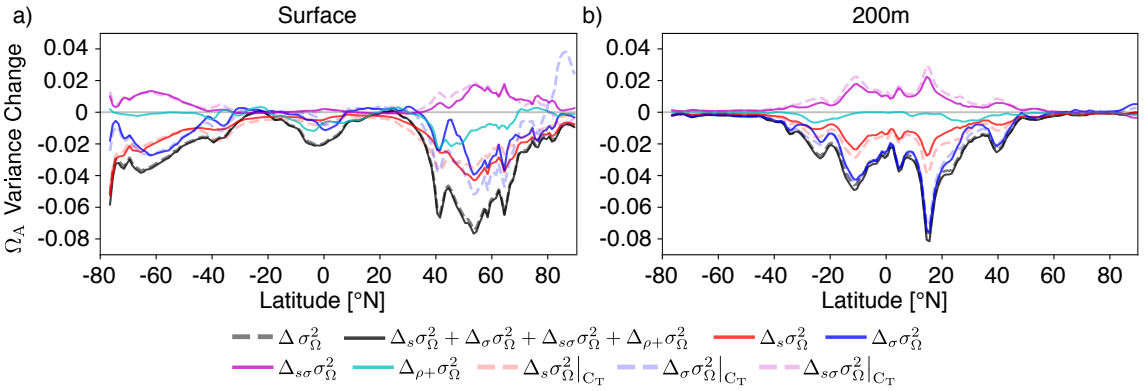


Figure 10. Decomposition of Ω_A variability changes into different drivers. The simulated zonal mean contribution to variance changes in Ω_A (black dashed lines, $\Delta\sigma^2_\Omega$) from preindustrial to 2081-2100 (RCP8.5) at the surface (a) and at 200 m (b). Shown is the contribution from sensitivity changes (due to mean changes in the drivers) (red lines, $\Delta_s\sigma^2_\Omega$), standard deviation changes in the drivers (blue lines, $\Delta_\sigma\sigma^2_\Omega$), simultaneous changes in sensitivities and standard deviations (purple lines, $\Delta_{s\sigma}\sigma^2_\Omega$), and all contributions that involve changes in the drivers' correlations (cyan lines, $\Delta_{\rho+}\sigma^2_\Omega$). Furthermore, contributions from mean changes, standard deviation changes and simultaneous mean and standard deviation changes in C_T alone are shown (dashed red, blue, and purple lines, respectively). In contrast to Figures 8 and 9 and due to their large contribution, we also show the zonal mean contribution from variability changes in C_T alone here.

extremes.

In this study, we analyze changes in extreme variability events that are defined relative to a shifting baseline. If the long-term increase in ocean acidity and decrease in Ω_A is taken into account, i.e. defining the extremes with respect to a fixed preindustrial baseline (here the preindustrial 99th percentile for $[H^+]$ and the preindustrial 1st percentile for Ω_A), the changes in $[H^+]$ and Ω_A extremes are much larger (cyan lines in Figure 11). Under the RCP8.5 scenario, every day becomes an extreme event day in year 2051 at surface and in year 2067 at 200 m depth (Figure 11a). The model also projects year-round extreme conditions for Ω_A at the surface and at 200 m by the end of the 21st century under RCP8.5 (Figure 11b). Comparing the two frameworks for surface $[H^+]$ extremes under present-day conditions, the annual number of extreme event days as defined in this study (i.e. with shifting baseline; black line in Figure 11) is on global average only 3.8 % of that also including the mean changes (i.e. with fixed preindustrial baseline; cyan line in Figure 11). This fraction differs regionally and reaches more than 10 % in the North Pacific, the North Atlantic, and the Arctic Ocean. Interestingly, the GFDL ESM2M projects that surface mean $[H^+]$ overshoots the preindustrial 99th percentile in year 1975 on global average. Thereafter, higher variability actually reduces the number of extreme event days that are above the preindustrial percentile. Surface mean Ω_A falls below the preindustrial 1st percentile in year 1990. After that, lower variability further increases the number of extreme event days below the preindustrial percentile.

~~In this study, we~~ We use the 99th percentile of the distribution from a preindustrial simulation for the definition of an extreme $[H^+]$ variability event, but the results may depend on the choice of this threshold. We tested the sensitivity of our results by

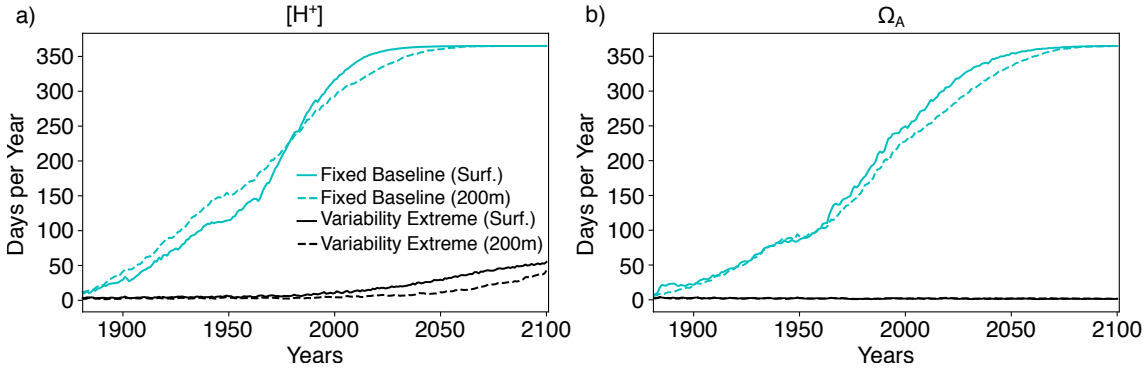


Figure 11. Simulated globally averaged number of extreme event days per year defined with a shifting baseline (black lines) and with a fixed preindustrial baseline (cyan lines) for $[H^+]$ using the 99th percentile (a) and for Ω_A using the 1st percentile (b) over the 1861-2100 period following the RCP8.5 scenario. Solid lines show results at the surface and dashed lines at 200 m.

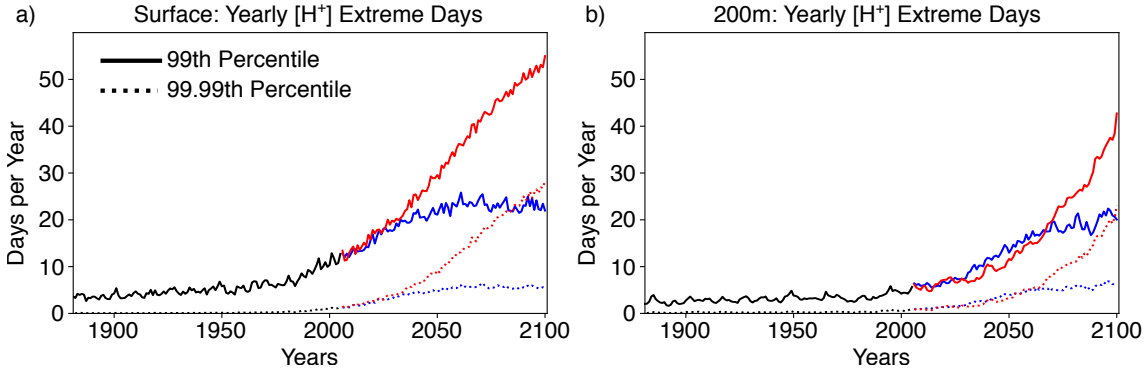


Figure 12. Globally averaged number of extreme variability event days for $[H^+]$ over the historical (black lines), RCP2.6 (blue), and RCP8.5 (red) simulations for the preindustrial 99th (solid lines) and 99.99th percentile (dotted lines) at (a) the surface and (b) ~~200m depth~~200 m.

using also the 99.99th percentile threshold. The relative increase in the numbers of extreme $[H^+]$ days per year is larger for these very rare variability extremes (Figure 12). For example, nearly every second day with $[H^+]$ exceeding the 99th percentile (red solid lines in Figure 12) is also a day with $[H^+]$ exceeding the 99.99th percentile (red dotted lines in Figure 12) by the end of the 21st century under RCP8.5, both at surface and at depth. In other words, an event that occurs every 27 years at preindustrial becomes almost as frequent in the future as an event that occurs every hundred days at preindustrial. As a result of this large relative increase in rare variability extremes, the model projects as many days with $[H^+]$ exceeding the 99.99th percentile by the end of the century under RCP8.5 (red dotted lines in Figure 12) as it projects days exceeding the 99th percentile under RCP2.6 (blue solid lines in Figure 12).

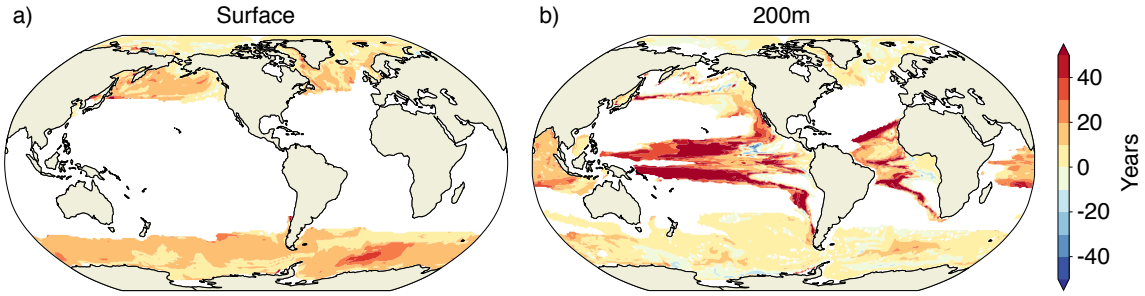


Figure 13. The temporal difference in years between the first occurrence of aragonite undersaturation in the historical and RCP8.5 ensemble and a hypothetical simulation ~~with~~ where variability does not change over the 1861-2100 period, but only the mean changes~~observed in the historical and RCP8.5 ensemble but with preindustrial variability~~. Positive values (yellow and red) indicate a delayed onset of undersaturation resulting from declines in Ω_A variability.

The projected increase in $[H^+]$ variability and decrease in Ω_A variability also alters the occurrence of extreme events based on absolute thresholds. An often used threshold is $\Omega_A = 1$ below which seawater is corrosive with respect to the calcium carbonate mineral aragonite (Bednaršek et al., 2012). We assess the influence of the general decline in Ω_A variability at the ~~point in~~ time where a grid cell falls below $\Omega_A = 1$ for the first time. To do so, we compare these ~~points in time~~ times within the historical and RCP8.5 ensemble to the ones for the hypothetical case where Ω_A variability stays at the preindustrial level but mean Ω_A undergoes the ensemble mean evolution. We find that the decline in Ω_A variability, which is observed in the historical and RCP8.5 ensemble, leads to an average delay of the first occurrence of undersaturation by about 11 years at the surface and about 16 years at 200 m~~depth~~. At surface, these delays of undersaturation occur throughout the high latitudes (Figure 13a). At depth, the delays are most pronounced in the tropics (Figure 13b), but delays also occur in the high latitudes. ~~Absolute thresholds for H^+ are generally not well established. Tests with arbitrarily set absolute thresholds suggest that these thresholds are generally exceeded a few years earlier because of increasing variability , but the temporal difference is less pronounced than that for Ω_A.~~Assuming unchanged seasonality, McNeil and Matear (2008) found that seasonal aragonite undersaturation of surface waters in the Southern Ocean may occur 30 years earlier than annual mean aragonite undersaturation. However, our simulation shows that the reduction in Ω_A variability delays the onset of undersaturation by about 10 to 15 years in the Southern Ocean relative to a hypothetical simulation where variability does not change. Therefore, changes in variability need to be taken into account when projecting the onset of seasonal undersaturation, especially in the high latitudes and in the thermocline of the tropics.

Previous studies have shown that the seasonal cycle of surface ocean pCO_2 ~~and surface H^+~~will be strongly amplified under increasing atmospheric CO_2 ~~(Gallego et al., 2018; Landschützer et al., 2018; McNeil and Sasse, 2016; Kwiatkowski and Orr, 2018)~~ (Gallego et al., 2018; Landschützer et al., 2018; McNeil and Sasse, 2016) and that a similar amplification is expected for surface $[H^+]$ (Kwiatkowski and Orr, 2018). Here we show that the changes in the seasonal cycle of $[H^+]$ translate into large increases in short-term extreme acidity events, at surface as well as at 200 m~~depth~~. In addition to earlier studies, we also show that

changes in ~~residual daily~~ subannual variability contribute to changes in extreme $[H^+]$ variability events under increasing atmospheric CO_2 and that the average duration of extreme variability events at the surface and at present-day is about 15 days. It is therefore critical to use daily temporal output to assess extreme events in ocean ~~biogeochemistry~~acidity. Currently, ocean ~~biogeochemical~~ carbonate system variables from models that participate in the sixth phase of the Coupled Model Intercomparison Project ~~Phase 6~~ are routinely stored with a monthly ~~resolution~~ frequency on the Earth system grid (Jones et al., 2016). We therefore recommend to ~~save out and use higher than monthly resolution to analyze variability in the surface and sub-surface~~ store and use high-frequency output to study extreme events in the ocean carbonate systems~~, in particular for studying extreme events~~.

Even though we consider our results as robust, a number of potential caveats remain. First, the horizontal resolution of the ocean model in GFDL ESM2M is rather coarse and cannot represent critical scales of small-scale circulation structures (e.g. Turi et al. (2018)). In addition, the biogeochemical processes included in GFDL ESM2M are designed for the open ocean, but do not capture the highly variable coastal processes (Hofmann et al., 2011). High resolution ocean models with improved process representations are therefore needed to explore extreme events in ocean carbonate chemistry, especially in coastal regions. ~~However, observation-based~~ Observation-based carbonate system data on daily time scale would also be necessary to thoroughly evaluate the models' capability to represent daily variations in carbonate chemistry. Secondly, our results, in particular at the local scale, might depend on the model formulation. ~~The GFDL ESM2M is largely able to project $[H^+]$ variability changes (Figure 3) arising from the nonlinear dependence on the drivers' mean states as the ocean carbonate chemistry is known and the mean changes in the drivers match observational records relatively well over the historical period (Bopp et al., 2013) . However~~As the mean increases in C_T mainly drive the increases in extreme $[H^+]$ variability events (see Figure 8f), we expect that models with larger oceanic uptake of anthropogenic carbon show larger changes in extreme variability events than models with lower anthropogenic carbon uptake. The GFDL ESM2M matches observation-based estimates of historical global anthropogenic CO_2 uptake relatively well, but still has difficulties in representing the regional patterns in storage (Frölicher et al., 2015). Therefore, the exact regional patterns of C_T changes may differ from model to model and further studies focusing on the physical processes that lead to the regional C_T changes may help to better constrain the regional patterns in changes of acidity extremes. In addition, it is currently rather uncertain how $[H^+]$ and Ω_A variability changes as a result of changes in the drivers' variabilities. We have demonstrated that this factor is particularly important at depth for $[H^+]$ and for Ω_A. It is well known that current Earth system models have imperfect or uncertain representations of ocean variability over a range of timescales (Frölicher et al., 2016; Resplandy et al., 2015; Keller et al., 2014). A possible way forward would be to assess changes in ocean acidity extreme events within a multi-model ensemble, which would likely provide upper and lower bounds of future changes in these events. Finally, it is assumed that physical and biogeochemical changes in the ocean will also increase diurnal variability. In particular in coastal areas, such diurnal variations can have amplitudes that are much larger than the projected changes over the 21st century (Hofmann et al., 2011). However, GFDL ESM2M does not fully resolve the diurnal variability. ~~In the ocean biogeochemistry model (TOPAZv2) that is embedded in GFDL ESM2M, for example, phytoplankton growth is not able to consume internal nutrient stores in the absence of light during night and diel vertical migration is also not~~

~~simulated.~~ Future studies with Earth system models that resolve diurnal ~~chemistry extremes~~ processes are needed to quantify changes in diurnal variability and the impacts of ~~changes in the diurnal cycles~~ these changes on extreme acidity events.

Our ~~analysis also has~~ results also have important consequences for our understanding of the impact of ocean acidification on marine ecosystems. ~~It~~ The projected increase in the frequency and the duration of ocean acidity variability extremes implies that marine organisms will have less time to recover from ~~very~~ high [H^+] events in the future~~, as the frequency and the duration of the ocean acidity extremes are projected to increase substantially over the 21st century. While coastal species may be adapted to large variability in ocean acidity, the~~ . The large projected increase in [H^+] extreme variability events in the open ocean may push organisms and ecosystems that are commonly accustomed to a more steady environment to the limits of their resilience. The risks for substantial ecosystem impacts are aggravated by the fact that the frequency and intensity of marine heatwaves ~~is~~ are also projected to substantially increase (Frölicher et al., 2018), which also negatively impact marine ecosystems (Wernberg et al., 2016; Smale et al., 2019). The interactions of intensified multiple stressors has the potential to influence marine ecosystems and the ocean's biogeochemical cycles in an unprecedented manner (Gruber, 2011). However, further research is needed to understand the combined impacts of ~~short term~~ short-term ocean acidity extremes and marine heatwaves on marine ecosystems.

In conclusion, our analysis ~~reveals~~ shows that marine organisms and ecosystems are projected to be exposed to less stable [H^+] conditions in the future with more frequent occurrences of variability-driven short-term extreme [H^+] conditions. Such extremes events are projected to last longer, to be more intense and to cover larger volumes of seawater and therefore potentially add to the stress on organisms and ecosystems from ~~mean ocean acidification~~the long-term increase in ocean acidity.

Appendix A: Identifying and removing the secular trend in the model data

In this study, we analyze the changes in ~~short-term extreme~~ extreme variability events in [H^+] and Ω_A that arise from day-to-day to interannual variability changes in these variables. We therefore need to remove the secular trends from the data prior to analysis. We estimate the secular ~~trends~~ trend in a simulation from the five-member ensemble ~~means~~mean, assuming that ~~sub-annual~~ subannual and interannual to decadal variations in the individual ensemble members are phased randomly and do not imprint on the ensemble ~~means~~ mean because they average out. A larger ensemble size would be necessary for this assumption to perfectly hold. However, this potential source of error does not qualitatively alter our results.

~~Unlike sub-annual and interannual variability, the seasonal cycle is a deterministic component within a climate time series. Its phase is not independent between ensemble members and it doesn't average out when calculating ensemble means.~~ We remove the seasonal ~~cycles~~ cycle, here defined as the 365-day long mean evolution over the course of a year, from the ensemble means by smoothing the ensemble means with a 365-day running mean filter, i.e. by calculating the convolution of the time series with a rectangular window of length 365 and height 1/365. This filter also removes variability on ~~sub-annual~~ subannual and interannual timescales and thereby also reduces the error we make due to the small ensemble size that is discussed above. We

then subtract the running-mean-filtered ensemble means from the five ensemble members to remove the secular trends in the individual ensemble members.

Appendix B: Identifying interannual and ~~residual daily~~ subannual variability

The spectral density describes how the variance in a time series is distributed over different frequencies ν_j. It is proportional to the absolute value squared of the discrete Fourier transformation (DFT) of the time series. Defining the spectral density only for positive frequencies, it is given by

$$f(\nu_j) = 2\frac{\Delta t^2}{T}\left|\sum_{k=1}^{N} x_k \cdot \exp\left(-\mathrm{i}2\pi\nu_j \cdot \Delta t\, k\right)\right|^2,$$

with N the number of time steps, x_k the values of the time series at each time step, Δt the time interval between two time steps, $T = N \cdot \Delta t$, and the frequencies $\nu_j = j/T$. The autocovariance is the inverse Fourier transform of the spectral density (Wiener-Khintchine theorem, Chatfield (1996))[1]. As a consequence, the variance within the time series, given by the autocovariance at lag zero, is obtained by integrating the spectral density over all positive frequencies, $\sigma^2 = \int_0^\infty f(\nu)\mathrm{d}\nu$. For a discrete time series, where the maximal resolved frequency is given by $\nu_{\max} = 1/2\Delta t$, the identity reads

$$\sigma^2 = \sum_{j=0}^{N/2} f(\nu_j)\frac{1}{N\Delta t}.$$

Based on this equation, one can separate the contributions to variance from low-frequency and high-frequency variations. In this study, we determine interannual variability and ~~the residual (sub-annual) daily~~ subannual variability. Interannual variability is calculated by summing over the contributions to variance from all frequencies up to a cycle of once per year, i.e. by evaluating the sum up to i_{cut} for which $\nu_{\text{cut}} = 1/365\,\text{day}^{-1}$. Accordingly, ~~residual daily~~ subannual variability is obtained by evaluating the sum from $i_{\text{cut}} + 1$ to $N/2$. Prior to this separation, the seasonal variability is removed from the data by subtracting the 365-day climatology. ~~We use the *periodogram* function from *scipy.signal* to estimate the spectral density.~~

Appendix C: Decomposition of $[H^+]$ ~~standard deviation~~ variance change

[1] In the continuous case, the theorem states

$$\gamma(\tau) = \int_{-\infty}^{\infty} \tilde{f}(\nu)\exp(\mathrm{i}2\pi\nu\tau)\mathrm{d}\nu,$$

with the autocovariance function $\gamma(\tau)$ and the spectral density $\tilde{f}$ defined for positive and negative frequencies. Since the two-sided spectral density, $\tilde{f}$, is a real and even function, one can also use

$$\gamma(\tau) = \int_{0}^{\infty} f(\nu)\cos(2\pi\nu\tau)\mathrm{d}\nu$$

with the one-sided spectral density $f = 2 \cdot \tilde{f}$ that is used in this text.

~~According to~~ Following Equation 2 in the main text, the variance in $[H^+]$ (or Ω_A) can be approximated as a function of the four ~~means~~ sensitivities

$$\boldsymbol{s} = \left(\frac{\partial H^+}{\partial A_T}, \frac{\partial H^+}{\partial C_T}, \frac{\partial H^+}{\partial S}, \frac{\partial H^+}{\partial T} \right)^\intercal \tag{C1}$$

that in turn depend on the mean values of the drivers $\mu_{1,..4}$ ~~(we assume that changes in total phosphate and total silicate are negligible)~~, the four standard deviations of the drivers $\sigma_{1,..4}$,

$$\boldsymbol{\sigma} = (\sigma_{A_T}, \sigma_{C_T}, \sigma_S, \sigma_T)^\intercal, \tag{C2}$$

and the six pairwise correlation coefficients $\rho_{i \neq j = 1,..4}$. ~~The total change in $H^\pm$ standard deviation (or variance) between two periods can thus be obtained by calculating~~

$$\sigma_H(\mu_i^1, \sigma_i^1, \rho_{ij}^1) - \sigma_H(\mu_i^0, \sigma_i^0, \rho_{ij}^0),$$

~~where the superscript 0 indicates the first period, here preindustrial , and the superscript 1 indicates the second period, here 2081-2100 under RCP8.5. Based on this relation, we then assess how changes in the drivers' means, standard deviations and correlation coefficients contribute to the change in $H^\pm$ standard deviation . Firstly, we investigate how much $H^\pm$ standard deviation change can be explained by solely adjusting the drivers' mean values to the values simulated for 2081-2100 under RCP8.5,~~

$$\sigma_H(\mu_i^1, \sigma_i^0, \rho_{ij}^0) - \sigma_H(\mu_i^0, \sigma_i^0, \rho_{ij}^0).$$

~~Changes in the mean values contribute largest to $H^\pm$ variability changes. Secondly, we investigate how much standard deviation change can be additionally explained when also taking into account changes in the drivers' standard deviations,~~

$$\sigma_H(\mu_i^1, \sigma_i^1, \rho_{ij}^0) - \sigma_H(\mu_i^1, \sigma_i^0, \rho_{ij}^0).$$

~~Overall, standard deviation changes have the second largest imprint on $H^\pm$ variability changes. Thirdly, the remaining contribution from changes in the pairwise correlation coefficient is calculated as~~

$$\sigma_H(\mu_i^1, \sigma_i^1, \rho_{ij}^1) - \sigma_H(\mu_i^1, \sigma_i^1, \rho_{ij}^0).$$

~~In analogy to the first two steps, the contributions from mean and standard deviation changes in C_T and A_T alone are also assessed.~~ , in matrix notation given by

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{AC} & \rho_{AS} & \rho_{AT} \\ \rho_{AC} & 1 & \rho_{CS} & \rho_{CT} \\ \rho_{AS} & \rho_{CS} & 1 & \rho_{ST} \\ \rho_{AT} & \rho_{CT} & \rho_{ST} & 1 \end{pmatrix}. \tag{C3}$$

Based on this notation, we can rewrite Equation 2 of the main text as

$$\sigma^2_{\mathrm{H}^+} = \sum_{i=1}^{4}\sum_{j=1}^{4} s_i s_j \sigma_i \sigma_j \rho_{ij}. \tag{C4}$$

We use Equation C4 to decompose the variability change between the preindustrial and 2081-2100 into the contributions from changes in $\boldsymbol{s}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\rho}$. Since it is a polynomial of fifth order, its Taylor series has five orders, too[2]. In the following, all terms of the Taylor series are given. We denote the sum of first order terms that contain changes in the four sensitivities $\Delta s_{1,..,4}$ by $\Delta^{(1)}_{s}\sigma^2_{\mathrm{H}^+}$, the sum of second order terms that contain changes in the sensitivities and standard deviations by $\Delta^{(2)}_{s\sigma}\sigma^2_{\mathrm{H}^+}$, and so on.

The first order is given by $\Delta^{(1)}\sigma^2_{\mathrm{H}^+} = \Delta^{(1)}_{s}\sigma^2_{\mathrm{H}^+} + \Delta^{(1)}_{\sigma}\sigma^2_{\mathrm{H}^+} + \Delta^{(1)}_{\rho}\sigma^2_{\mathrm{H}^+}$ with

$$\Delta^{(1)}_{s}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{j=1}^{4} s_j \sigma_k \sigma_j \rho_{kj} \Delta s_k$$

$$\Delta^{(1)}_{\sigma}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{j=1}^{4} s_k s_j \sigma_j \rho_{kj} \Delta \sigma_k$$

$$\Delta^{(1)}_{\rho}\sigma^2_{\mathrm{H}^+} = \sum_{k=1}^{4}\sum_{l=1}^{4} s_k s_l \sigma_k \sigma_l \Delta \rho_{kl}. \tag{C5}$$

The second order contains

$$\Delta^{(2)}_{ss}\sigma^2_{\mathrm{H}^+} = \sum_{k=1}^{4}\sum_{l=1}^{4} \sigma_k \sigma_l \rho_{kl} \Delta s_k \Delta s_l$$

$$\Delta^{(2)}_{\sigma\sigma}\sigma^2_{\mathrm{H}^+} = \sum_{k=1}^{4}\sum_{l=1}^{4} s_k s_l \rho_{kl} \Delta \sigma_k \Delta \sigma_l$$

$$\Delta^{(2)}_{s\sigma}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{l=1}^{4} (s_l \sigma_l \rho_{kl} \Delta s_k \Delta \sigma_k + s_l \sigma_k \rho_{kl} \Delta s_k \Delta \sigma_l)$$

$$\Delta^{(2)}_{s\rho}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{l=1}^{4} s_l \sigma_k \sigma_l \Delta s_k \Delta \rho_{kl}$$

$$\Delta^{(2)}_{\sigma\rho}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{l=1}^{4} s_k s_l \sigma_l \Delta \sigma_k \Delta \rho_{kl}. \tag{C6}$$

[2] We use the drivers' standard deviations instead of their variances for the decomposition. With the latter, the Taylor expansion would have infinite terms and could not be decomposed exactly as it is done in the following. However, it would asymptotically lead to the same decomposition of $[\mathrm{H}^+]$ variance change into $\Delta_s\sigma^2_{\mathrm{H}^+}$, $\Delta_\sigma\sigma^2_{\mathrm{H}^+}$, $\Delta_{s\sigma}\sigma^2_{\mathrm{H}^+}$, and $\Delta_{\rho+}\sigma^2_{\mathrm{H}^+}$ that is presented below.

The third order terms read

$$\Delta^{(3)}_{ss\sigma}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{l=1}^{4}\sigma_l\rho_{kl}\Delta s_k\Delta s_l\Delta\sigma_k$$

$$\Delta^{(3)}_{s\sigma\sigma}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{l=1}^{4}s_l\rho_{kl}\Delta s_k\Delta\sigma_k\Delta\sigma_l$$

$$\Delta^{(3)}_{ss\rho}\sigma^2_{\mathrm{H}^+} = \sum_{k=1}^{4}\sum_{l=1}^{4}\sigma_k\sigma_l\Delta s_k\Delta s_l\Delta\rho_{kl}$$

$$\Delta^{(3)}_{\sigma\sigma\rho}\sigma^2_{\mathrm{H}^+} = \sum_{k=1}^{4}\sum_{l=1}^{4}s_k s_l\Delta\sigma_k\Delta\sigma_l\Delta\rho_{kl}$$

$$\Delta^{(3)}_{s\sigma\rho}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{l=1}^{4}\left(s_l\sigma_k\Delta s_k\Delta\sigma_l\Delta\rho_{kl} + s_l\sigma_l\Delta s_k\Delta\sigma_k\Delta\rho_{kl}\right). \tag{C7}$$

The fourth order reads

$$\Delta^{(4)}_{ss\sigma\sigma}\sigma^2_{\mathrm{H}^+} = \sum_{k=1}^{4}\sum_{l=1}^{4}\rho_{kl}\Delta s_k\Delta s_l\Delta\sigma_k\Delta\sigma_l$$

$$\Delta^{(4)}_{ss\sigma\rho}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{l=1}^{4}\sigma_l\Delta s_k\Delta s_l\Delta\sigma_k\Delta\rho_{kl}$$

$$\Delta^{(4)}_{s\sigma\sigma\rho}\sigma^2_{\mathrm{H}^+} = 2\sum_{k=1}^{4}\sum_{l=1}^{4}s_l\Delta s_k\Delta\sigma_k\Delta\sigma_l\Delta\rho_{kl} \tag{C8}$$

and the fifth order is given by

$$\Delta^{(5)}_{ss\sigma\sigma\rho}\sigma^2_{\mathrm{H}^+} = \sum_{k=1}^{4}\sum_{l=1}^{4}\Delta s_k\Delta s_l\Delta\sigma_k\Delta\sigma_l\Delta\rho_{kl}. \tag{C9}$$

We identify the variance change from changes in the sensitivities as

$$\Delta_s\sigma^2_{\mathrm{H}^+} = \Delta^{(1)}_s\sigma^2_{\mathrm{H}^+} + \Delta^{(2)}_{ss}\sigma^2_{\mathrm{H}^+}, \tag{C10}$$

the change from standard deviation changes as

$$\Delta_\sigma\sigma^2_{\mathrm{H}^+} = \Delta^{(1)}_\sigma\sigma^2_{\mathrm{H}^+} + \Delta^{(2)}_{\sigma\sigma}\sigma^2_{\mathrm{H}^+}, \tag{C11}$$

the change from simultaneous changes in sensitivities and standard deviations as

$$\Delta_{s\sigma}\sigma^2_{\mathrm{H}^+} = \Delta^{(2)}_{s\sigma}\sigma^2_{\mathrm{H}^+} + \Delta^{(3)}_{ss\sigma}\sigma^2_{\mathrm{H}^+} + \Delta^{(3)}_{s\sigma\sigma}\sigma^2_{\mathrm{H}^+} + \Delta^{(4)}_{ss\sigma\sigma}\sigma^2_{\mathrm{H}^+}, \tag{C12}$$

and that from correlation changes and mixed contributions that include correlation changes as

$$\Delta_{\rho+}\sigma^2_{\mathrm{H}^+} = \Delta^{(1)}_{\rho}\sigma^2_{\mathrm{H}^+} + \Delta^{(2)}_{s\rho}\sigma^2_{\mathrm{H}^+} + \Delta^{(2)}_{\sigma\rho}\sigma^2_{\mathrm{H}^+} + \Delta^{(3)}_{ss\rho}\sigma^2_{\mathrm{H}^+} + \Delta^{(3)}_{\sigma\sigma\rho}\sigma^2_{\mathrm{H}^+} + \Delta^{(3)}_{s\sigma\rho}\sigma^2_{\mathrm{H}^+} + \Delta^{(4)}_{ss\sigma\rho}\sigma^2_{\mathrm{H}^+} + \Delta^{(4)}_{s\sigma\sigma\rho}\sigma^2_{\mathrm{H}^+} + \Delta^{(5)}_{ss\sigma\sigma\rho}\sigma^2_{\mathrm{H}^+}. \tag{C13}$$

Finally, we calculate $\Delta_s\sigma^2_{\mathrm{H}^+}\big|_{\mathrm{C_T}}$, $\Delta_\sigma\sigma^2_{\mathrm{H}^+}\big|_{\mathrm{C_T}}$, and $\Delta_{s\sigma}\sigma^2_{\mathrm{H}^+}\big|_{\mathrm{C_T}}$, the analogues for Equations C10-C12 that only take into account changes in $\mathrm{C_T}$. This is done by calculating $\Delta s_{1,..4}$ only based on mean changes in $\mathrm{C_T}$ and by setting the standard deviation changes for $\mathrm{A_T}$, S, and T to zero.

Appendix D: Comparison of simulated ensemble-mean trends in seasonal amplitude to observation-based trends

We construct confidence intervals for the observation-based slope estimates following Hartmann et al. (2013). For the simulations, we use the arithmetic average of the five ensemble-member slope estimates as the estimator,

$$\hat{\bar{b}} = \frac{1}{5}\sum_{k=1}^{5}\hat{b}_k \tag{D1}$$

with estimated variance

$$\hat{\sigma}^2_{\bar{b}} = \frac{1}{5^2}\sum_{k=1}^{5}\hat{\sigma}^2_{b_k}. \tag{D2}$$

We then construct the confidence interval for $\hat{\bar{b}}$ as

$$(\hat{\bar{b}} - q\cdot\hat{\sigma}_{\bar{b}}, \hat{\bar{b}} + q\cdot\hat{\sigma}_{\bar{b}}), \tag{D3}$$

with q the $(1+p)/2$-quantile (we use $p = 0.9$) of the t-distribution with $5\cdot(N-2)$ degrees of freedom. We correct the sample size N (34, the number of years we use for the fits) to a reduced sample size N_r when we find positive lag-one autocorrelation in the residuals of the fits (data - linear regression model). Lag-one autocorrelation is estimated as the average of the five ensemble-member lag-one autocorrelation estimates

$$\hat{\bar{\rho}} = \frac{1}{5}\sum_{k=1}^{5}\hat{\rho}_k. \tag{D4}$$

and we obtain $N_r = N\cdot(\hat{\bar{\rho}}-1)/(\hat{\bar{\rho}}+1)$. Positive $\hat{\bar{\rho}}$ is only found in the northern high latitudes. This is in contrast to the observation-based case, where we find large positive $\hat{\rho}_o$ (up to 0.7) in the residuals of all latitude bands besides the tropical region.

For testing the significance of a difference between the simulation slope estimate $\hat{\bar{b}}$ and the observation-based estimate $\hat{b}_o$, we use Welch's test that assumes different variances for the two estimates (Andrade and Estévez-Pérez, 2014). The variance of the

simulation slope estimate is calculated by dividing the ensemble-averaged slope variance by the ensemble size (Equation D2) and is hence smaller than the observation-based slope variance. If the absolute value of the test statistic

$$\frac{\hat{\bar{b}} - \hat{b}_o}{\sqrt{\hat{\sigma}^2_{\bar{b}} + \hat{\sigma}_o}} \tag{D5}$$

is larger than the $(1+p)/2$-quantile of the t distribution with (Andrade and Estévez-Pérez, 2014)

$$\frac{\left(\hat{\sigma}^2_{\bar{b}} + \hat{\sigma}^2_{b_o}\right)^2}{\hat{\sigma}^4_{\bar{b}}/(5\cdot(N_r-2)) + \hat{\sigma}^4_{b_o}/(N_{r,o}-2)} \tag{D6}$$

degrees of freedom, we consider the observation-based and simulation slope to be different from each other with confidence level $p = 0.9$.

	PI	1986-2005	~~EOC~~ 2081-2100 RCP2.6	~~EOC~~ 2081-2100 RCP8.5
Number Surf.	3.65	1.75 (1.50-2.20)	2.24 (1.86-2.93)	1.36 (1.09-1.69)
~~200m~~ 200 m	3.65	1.98 (1.51-2.77)	3.01 (2.28-3.71)	1.72 (1.38-2.02)
Duration Surf.	19.70	17.84 (16.84-18.92)	19.37 (18.07-21.13)	29.28 (27.37-32.57)
~~200m~~ 200 m	38.61	66.06 (59.74-18.92)	98.71 (89.01-109.01)	111.56 (106.62-122.70)
Maximal Intensity Surf.	2.92	3.42 (3.26-3.64)	3.21 (3.07-3.48)	1.51 (1.42-1.63)
~~200m~~ 200 m	3.26	4.96 (3.87-6.67)	7.90 (6.05-11.06)	6.02 (2.85-9.13)
Volume	3640	3158 (2888-3460)	3662 (3021-4215)	3378 (3086-3714)

Table A1. ~~The same as Table 2, but for~~ Simulated global ensemble-mean Ω_A extreme variability event characteristics for the preindustrial (PI), present day (1986-2005), and the end of this century (2081-2100) for both RCP2.6 and RCP8.5. ~~Intensity is~~ Numbers of yearly extreme days are given in days per year, durations in days, intensities in $10^{-3}\,\Omega_A$ units and volumes in km^3. ~~The remaining units are identical to those~~ Values in ~~Table 2~~brackets denote ensemble minima and maxima.

Data availability. The GFDL ESM2M simulations are available upon request.

Author contributions. FAB and TLF designed the study. FAB performed the simulations, assisted by TLF and JGJ. FAB performed the analysis and wrote the initial manuscript. All authors contributed significantly to the writing of the paper.

Competing interests. All authors declare no competing interests.

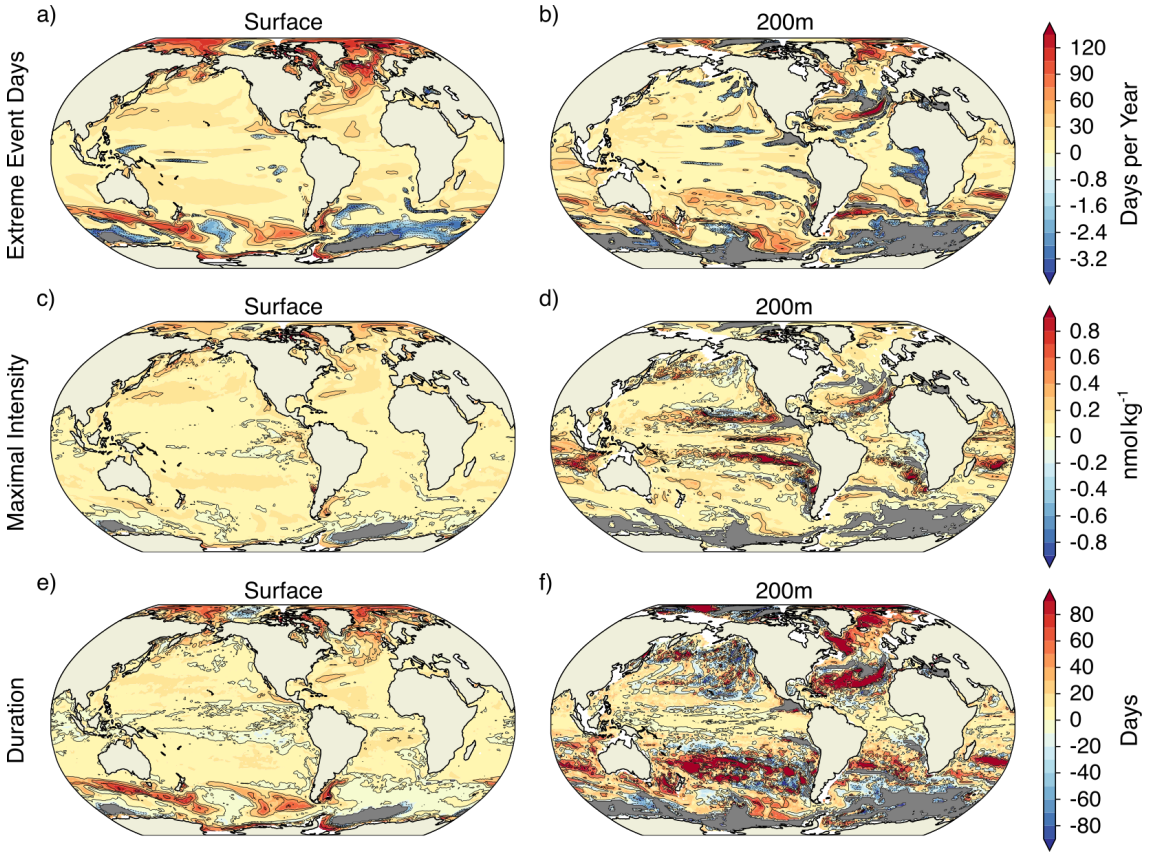


Figure A1. Simulated regional changes in (a,b) the number of extreme $[H^+]$ days per year, (c,d) the maximal intensity of extreme $[H^+]$ variability events, and (e,f) the duration of extreme $[H^+]$ variability events between preindustrial and 2081-2100 following the RCP2.6 scenario. Left panels show changes for the surface, whereas right panels show changes for 200 m ~~depth~~. Shown are changes averaged over all five ensemble members. The black contours highlight the pattern structures. Grey areas represent areas with no variability extremes during 2081-2100.

Disclaimer. The work reflects only the authors' view; the European Commission and their executive agency are not responsible for any use that may be made of the information the work contains.

Acknowledgements. FAB and TLF have received funding the Swiss National Science Foundation (PP00P2_170687) and from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820989 (project COMFORT, Our common future ocean in the Earth system — quantifying coupled cycles of carbon, oxygen, and nutrients for determining and achieving safe operating spaces with respect to tipping points). FAB and TLF also thank the CSCS Swiss National Supercomputing Centre for computing resources. The authors thank Elizabeth Drenkard, Fortunat Joos, and Jens Terhaar for discussion and comments, and Rick Slater for the help in porting the ESM2M model code to CSCS.

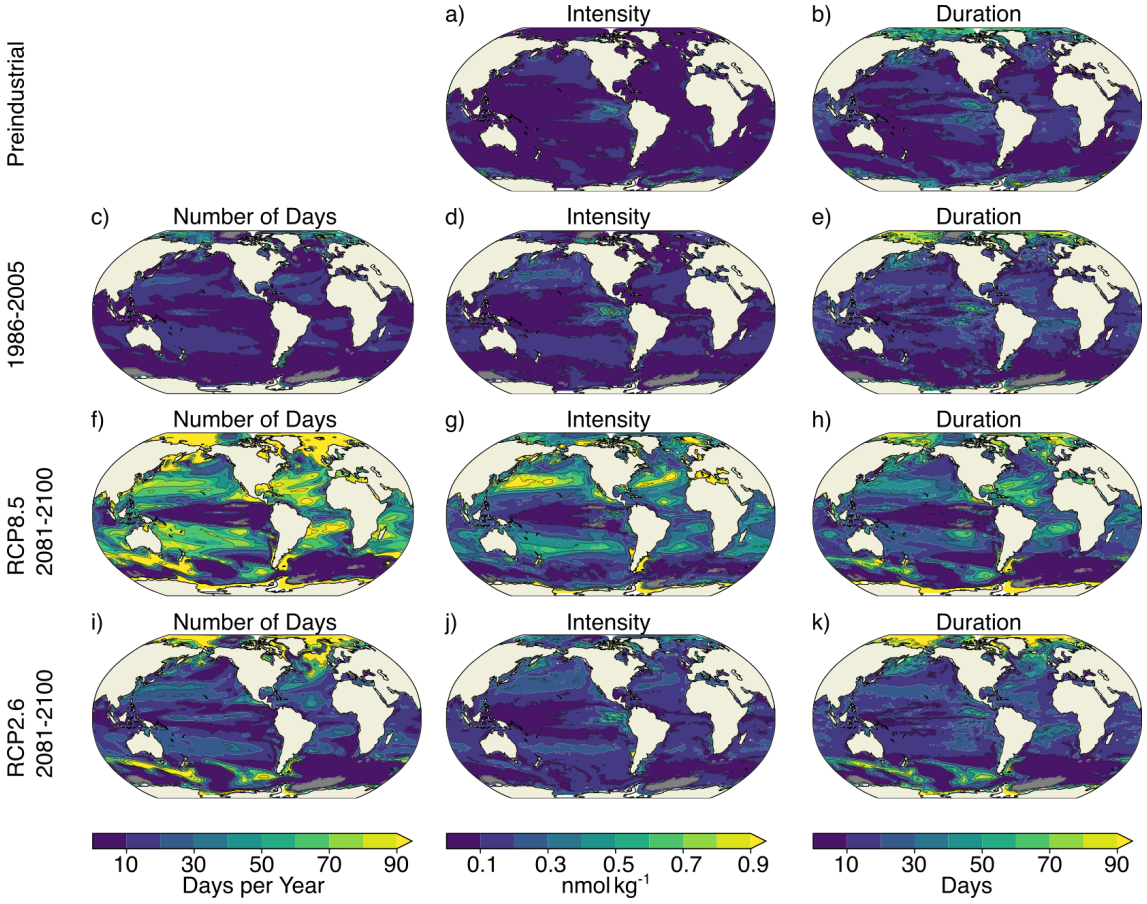


Figure A2. Simulated characteristics of surface [H^+] extreme variability events for preindustrial (a,b), ~~present-day~~ 1986-2005 ensemble mean (~~1986-2005,~~ c-e), RCP8.5 ~~end of century~~ 2081-2100 ensemble mean (~~2081-2100,~~ f-h), and RCP2.6 ~~end of century~~ 2081-2100 ensemble mean (~~2081-2100,~~ i-k). Grey colors represent regions where no ensemble member simulates variability extremes. The black contours highlight the pattern structures.

References

Anderson, J. L., Balaji., V., Broccoli, A. J., Cooke, W. F., Delworth, T. L., Dixon, K. W., Donner, L. J., Dunne, K. A., Freidenreich, S. M., Garner, S. T., Gudgel, R. G., Gordon, C. T., Held, I. M., Hemler, R. S., Horowitz, L. W., Klein, S. A., Knutson, T. R., Kushner, P. J., Langenhost, A. R., Cheung, L. N., Liang, Z., Malyshev, S. L., Milly, P. C. D., Nath, M. J., Ploshay, J. J., Ramaswamy, V., Schwarzkopf, M. D., Shevliakova, E., Sirutis, J. J., Soden, B. J., Stern, W. F., Thompson, L. A., Wilson, R. J., Wittenberg, A. T., and Wyman, B. L.: The New GFDL Global Atmosphere and Land Model AM2–LM2: Evaluation with Prescribed SST Simulations, J. Clim., 17, 4641–4673, https://doi.org/10.1175/JCLI-3223.1, 2004.

Andrade, J. and Estévez-Pérez, M.: Statistical comparison of the slopes of two regression lines: A tutorial, Anal. Chim. Acta, 838, 1 – 12, https://doi.org/https://doi.org/10.1016/j.aca.2014.04.057, 2014.

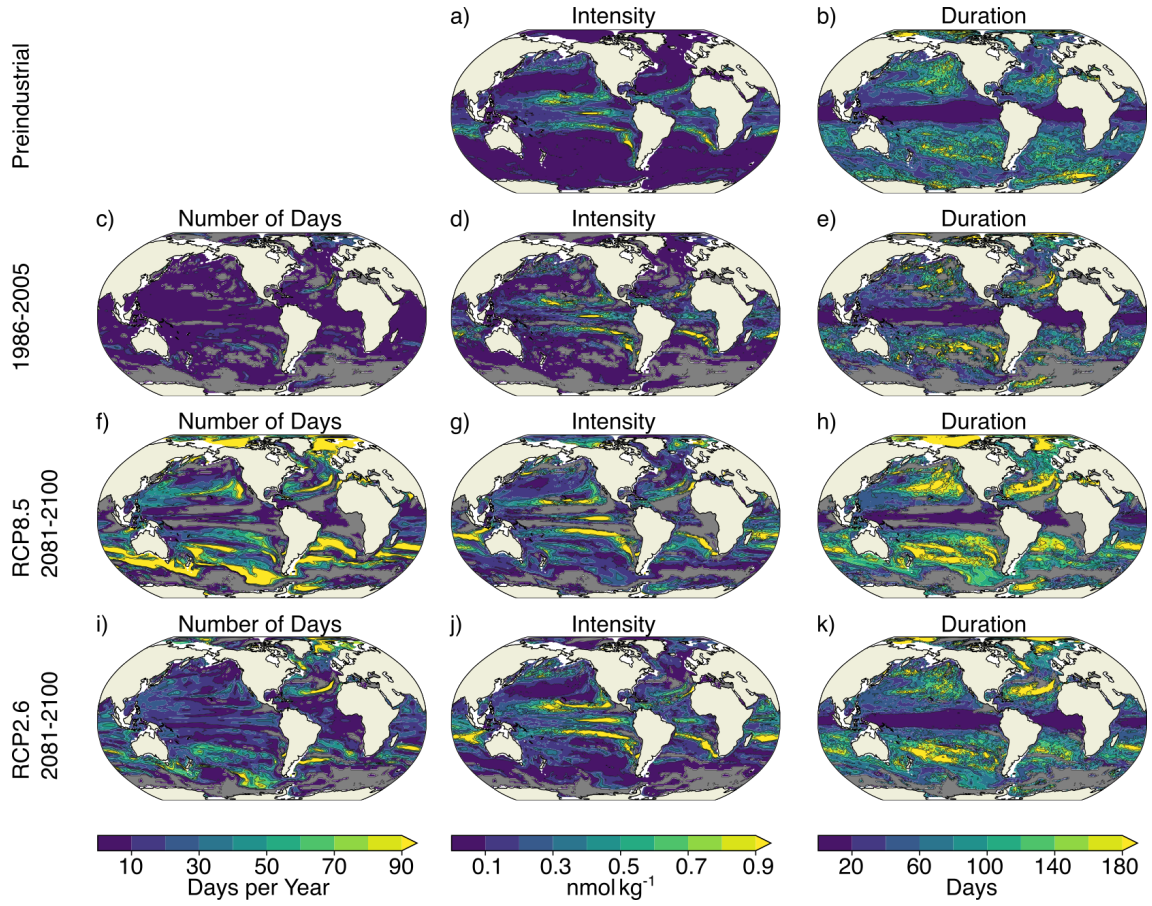


Figure A3. The same as Figure A2, but for 200 m~~depth~~. The color scale for the duration plots changed with respect to that in Figure A2.

Bakker, D. C. E., Pfeil, B., O'Brien, K. M., Currie, K. I., Jones, S. D., Landa, C. S., Lauvset, S. K., Metzl, N., Munro, D. R., Nakaoka, S.-I., Olsen, A., Pierrot, D., Saito, S., Smith, K., Sweeney, C., Takahashi, T., Wada, C., Wanninkhof, R., Alin, S. R., Becker, M., Bellerby, R. G. J., Borges, A. V., Boutin, J., Bozec, Y., Burger, E., Cai, W.-J., Castle, R. D., Cosca, C. E., DeGrandpre, M. D., Donnelly, M., Eischeid, G., Feely, R. A., Gkritzalis, T., González-Dávila, M., Goyet, C., Guillot, A., Hardman-Mountford, N. J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Ibánhez, J. S. P., Ichikawa, T., Ishii, M., Juranek, L. W., Kitidis, V., Körtzinger, A., Koffi, U. K., Kozyr, A., Kuwata, A., Lefèvre, N., Lo Monaco, C., Manke, A., Marrec, P., Mathis, J. T., Millero, F. J., Monacci, N., Monteiro, P. M. S., Murata, A., Newberger, T., Nojiri, Y., Nonaka, I., Omar, A. M., Ono, T., Padín, X. A., Rehder, G., Rutgersson, A., Sabine, C. L., Salisbury, J., Santana-Casiano, J. M., Sasano, D., Schuster, U., Sieger, R., Skjelvan, I., Steinhoff, T., Sullivan, K., Sutherland, S. C., Sutton, A., Tadokoro, K., Telszewski, M., Thomas, H., Tilbrook, B., van Heuven, S., Vandemark, D., Wallace, D. W., and Woosley, R.: Surface Ocean CO_2 Atlas (SOCAT) V4, https://doi.org/10.1594/PANGAEA.866856, 2016.

Bednaršek, N., Tarling, G. A., Bakker, D. C. E., Fielding, S., Jones, E. M., Venables, H. J., Ward, P., Kuzirian, A., Lézé, B., Feely, R. A., and Murphy, E. J.: Extensive dissolution of live pteropods in the Southern Ocean, Nat. Geosci., 5, 881–885, https://doi.org/10.1038/ngeo1635, 2012.

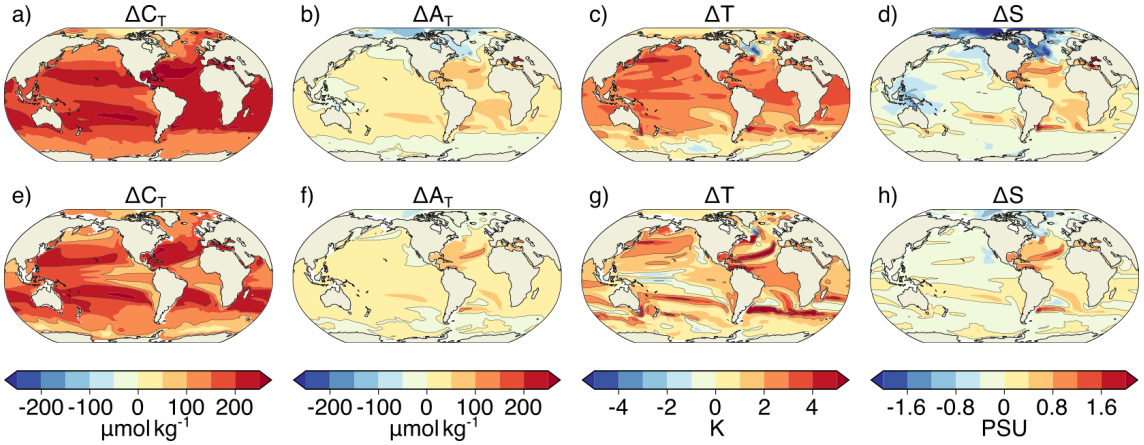


Figure A4. Simulated ensemble mean changes in C_T (a,e), A_T (b,f), T (c,g), and S (d,h) from preindustrial to 2081-2100 following the RCP8.5 scenario. Shown are changes for (a-d) the surface and (e-h) at 200 m ~~depth~~. The black contours highlight the pattern structures.

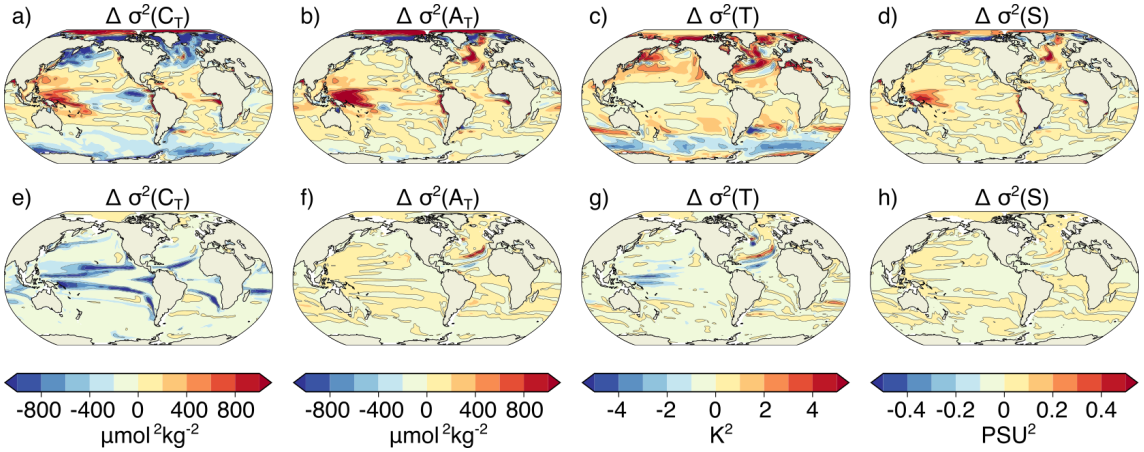


Figure A5. Simulated ensemble mean changes in the ~~standard deviations~~ variances of C_T (a,e), A_T (b,f), T (c,g), and S (d,h) from preindustrial to 2081-2100 under the RCP8.5 scenario. Shown are changes for (a-d) the surface and (e-h) at 200 m ~~depth~~. The black contours highlight the pattern structures.

Bednaršek, N., Feely, R. A., Reum, J. C. P., Peterson, B., Menkel, J., Alin, S. R., and Hales, B.: *Limacina helicina* shell dissolution as an indicator of declining habitat suitability owing to ocean acidification in the California Current Ecosystem, Proc. R. Soc. B, 281, 20140 123, https://doi.org/10.1098/rspb.2014.0123, 2014.

Bindoff, N., Cheung, W., Kairo, J., Aristegui, J., Guinder, V., Hallberg, R., Hilmi, N., Jiao, N., Karim, M., Levin, L., O'Donoghue, S., Purca Cuicapusa, S., Rinkevich, B., Suga, T., Tagliabue, A., and Williamson, P.: Chapter 5: Changing Ocean, Marine Ecosystems, and Dependent Communities, IPCC Special Report on the Ocean and Cryosphere (SROCC), 2019.

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and Vichi, M.: Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models, Biogeosciences, 10, https://doi.org/10.5194/bg-10-6225-2013, 2013.

Britton, D., Cornwall, C. E., Revill, A. T., Hurd, C. L., and Johnson, C. R.: Ocean acidification reverses the positive effects of seawater pH fluctuations on growth and photosynthesis of the habitat-forming kelp, Ecklonia radiata, Sci. Rep., 6, 26 036, https://doi.org/10.1038/srep26036, 2016.

Caldeira, K. and Wickett, M. E.: Anthropogenic carbon and ocean pH, Nature, 425, 365–365, https://doi.org/10.1038/425365a, 2003.

Carter, B. R., Frölicher, T. L., Dunne, J. P., Rodgers, K. B., Slater, R. D., and Sarmiento, J. L.: When can ocean acidification impacts be detected from decadal alkalinity measurements?, Global Biogeochem. Cycles, 30, 595–612, https://doi.org/10.1002/2015GB005308, 2016.

Carter, B. R., Feely, R. A., Williams, N. L., Dickson, A. G., Fong, M. B., and Takeshita, Y.: Updated methods for global locally interpolated estimation of alkalinity, pH, and nitrate, Limnol. Oceanogr. Methods, 16, 119–131, https://doi.org/10.1002/lom3.10232, 2018.

Chatfield, C.: The analysis of time series - an introduction, Chapman & Hall/CRC, 5th edn., 1996.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer-Verlag London, 2001.

Collins, M., Sutherland, M., Bouwer, L., Cheong, S.-M., Frölicher, T., Des Combes, H., Koll, M., Losada, I., McInnes, K., Ratter, B., Rivera-Arriga, E., Susanto, R., Swingedouw, D., and Tibig, L.: Chapter 6: Extremes, Abrupt Changes and Managing Risks, IPCC Special Report on the Ocean and Cryosphere (SROCC), 2019.

Cornwall, C. E., Comeau, S., DeCarlo, T. M., Larcombe, E., Moore, B., Giltrow, K., Puerzer, F., D'Alexis, Q., and McCulloch, M. T.: A coralline alga gains tolerance to ocean acidification over multiple generations of exposure, Nat. Clim. Change, 10, 143–146, https://doi.org/10.1038/s41558-019-0681-8, 2020.

Dickson, A. and Millero, F.: A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media, DEEP-SEA RES PT A, 34, 1733 – 1743, https://doi.org/https://doi.org/10.1016/0198-0149(87)90021-5, 1987.

Dickson, A. and Riley, J.: The effect of analytical error on the evaluation of the components of the aquatic carbon-dioxide system, Mar. Chem., 6, 77 – 85, https://doi.org/https://doi.org/10.1016/0304-4203(78)90008-7, http://www.sciencedirect.com/science/article/pii/0304420378900087, 1978.

Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.: Ocean Acidification: The Other CO_2 Problem, Annual Review of Marine Science, 1, 169–192, https://doi.org/10.1146/annurev.marine.010908.163834, 2009.

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W., Dunne, K. A., Harrison, M. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Phillipps, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A. T., and Zadeh, N.: GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics, J. Clim., 25, 6646–6665, https://doi.org/10.1175/JCLI-D-11-00560.1, 2012.

Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Sentman, L. T., Adcroft, A. J., Cooke, W., Dunne, K. A., Griffies, S. M., Hallberg, R. W., Harrison, M. J., Levy, H., Wittenberg, A. T., Phillips, P. J., and Zadeh, N.: GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation Characteristics, J. Clim., 26, 2247–2267, https://doi.org/10.1175/JCLI-D-12-00150.1, 2013.

Fassbender, A. J., Rodgers, K. B., Palevsky, H. I., and Sabine, C. L.: Seasonal Asymmetry in the Evolution of Surface Ocean pCO2 and pH Thermodynamic Drivers and the Influence on Sea-Air CO2 Flux, Global Biogeochem. Cycles, 32, 1476–1497, https://doi.org/10.1029/2017GB005855, 2018.

Feely, R., Chris, S., Hernandez-Ayon, J., Ianson, D., and Hales, B.: Evidence for Upwelling of Corrosive " Acidified" Water onto the Continental Shelf, Science, 320, 1490–1492, https://doi.org/10.1126/science.1155676, 2008.

Form, A. U. and Riebesell, U.: Acclimation to ocean acidification during long-term CO_2 exposure in the cold-water coral Lophelia pertusa, Glob. Change Biol., 18, 843–853, https://doi.org/10.1111/j.1365-2486.2011.02583.x, 2012.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, Earth Syst. Sci. Data, 11, 1783–1838, https://doi.org/10.5194/essd-11-1783-2019, 2019.

Frölicher, T. L., Rodgers, K. B., Stock, C. A., and Cheung, W. W. L.: Sources of uncertainties in 21st century projections of potential ocean ecosystem stressors, Global Biogeoch Cycles, 30, 1224–1243, https://doi.org/10.1002/2015GB005338, 2016.

Frölicher, T. L., Fischer, E. M., and Gruber, N.: Marine heatwaves under global warming, Nature, 560, 360–364, https://doi.org/10.1038/s41586-018-0383-9, 2018.

Frölicher, T. L., Ramseyer, L., Raible, C. C., Rodgers, K. B., and Dunne, J.: Potential predictability of marine ecosystem drivers, Biogeosciences, 17, 2061–2083, https://doi.org/10.5194/bg-17-2061-2020, 2020.

Frölicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Krasting, J. P., and Winton, M.: Dominance of the Southern Ocean in Anthropogenic Carbon and Heat Uptake in CMIP5 Models, J. Clim., 28, 862–886, https://doi.org/10.1175/JCLI-D-14-00117.1, 2015.

Gallego, M. A., Timmermann, A., Friedrich, T., and Zeebe, R. E.: Drivers of future seasonal cycle changes in oceanic pCO_2, Biogeosciences, 15, 5315–5327, https://doi.org/10.5194/bg-15-5315-2018, 2018.

Gattuso, J.-P. and Buddemeier, R. W.: Calcification and CO_2, Nature, 407, 311–313, https://doi.org/10.1038/35030280, 2000.

Gehlen, M., Gruber, N., Gangstro, R., Bopp, L., and Oschlies, A.: Chapter 12: Biogeochemical Consequences of Ocean Acidification and Feedback to the Earth System, in: Ocean Acidification, edited by Gattuso, J.-P. and Hansson, L., pp. 230–248, Oxford University Press, 2012.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res. Oceans, 118, 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.

Gray, A. R., Johnson, K. S., Bushinsky, S. M., Riser, S. C., Russell, J. L., Talley, L. D., Wanninkhof, R., Williams, N. L., and Sarmiento, J. L.: Autonomous Biogeochemical Floats Detect Significant Carbon Dioxide Outgassing in the High-Latitude Southern Ocean, Geophys. Res. Lett., 45, 9049–9057, https://doi.org/10.1029/2018GL078013, 2018.

Griffies, S.: ELEMENTS OF MOM4p1, GFDL ocean group technical report no.6, NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University Forrestal Campus, 201 Forrestal Road, Princeton, NJ 08540-6649, 2009.

Gruber, N.: Warming up, turning sour, losing breath: ocean biogeochemistry under global change, Philos. Trans. R. Soc. A, 369, 1980–1996, https://doi.org/10.1098/rsta.2011.0003, 2011.

Hall-Spencer, J. M., Rodolfo-Metalpa, R., Martin, S., Ransome, E., Fine, M., Turner, S. M., Rowley, S. J., Tedesco, D., and Buia, M.-C.: Volcanic carbon dioxide vents show ecosystem effects of ocean acidification, Nature, 454, 96–99, https://doi.org/10.1038/nature07051, 2008.

Hartmann, D., Klein Tank, A., Rusticucci, M., Alexander, L., Brönnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., and Zhai, P.: Observations: Atmosphere and Surface Supplementary Material, Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 2013.

Hauck, J. and Völker, C.: Rising atmospheric CO_2 leads to large impact of biology on Southern Ocean CO_2 uptake via changes of the Revelle factor, Geophys. Res. Lett., 42, 1459–1464, https://doi.org/10.1002/2015GL063070, 2015.

Hauri, C., Gruber, N., McDonnell, A. M. P., and Vogt, M.: The intensity, duration, and severity of low aragonite saturation state events on the California continental shelf, Geophys. Res. Lett., 40, 3424–3428, https://doi.org/10.1002/grl.50618, 2013.

Hofmann, G. E., Smith, J. E., Johnson, K. S., Send, U., Levin, L. A., Micheli, F., Paytan, A., Price, N. N., Peterson, B., Takeshita, Y., Matson, P. G., Crook, E. D., Kroeker, K. J., Gambi, M. C., Rivest, E. B., Frieder, C. A., Yu, P. C., and Martz, T. R.: High-Frequency Dynamics of Ocean pH: A Multi-Ecosystem Comparison, PloS One, 6, 1–11, https://doi.org/10.1371/journal.pone.0028983, 2011.

Joint, I., Doney, S. C., and Karl, D. M.: Will ocean acidification affect marine microbes?, ISME J., 5, 1–7, https://doi.org/10.1038/ismej.2010.79, 2011.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., Jung, M., Kawamiya, M., Koven, C., Pongratz, J., Raddatz, T., Randerson, J. T., and Zaehle, S.: C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6, Geosci. Model Dev., 9, 2853–2880, https://doi.org/10.5194/gmd-9-2853-2016, 2016.

Keller, K. M., Joos, F., and Raible, C. C.: Time of emergence of trends in ocean biogeochemistry, Biogeosciences, 11, 3647–3659, https://doi.org/10.5194/bg-11-3647-2014, 2014.

Kroeker, K. J., Micheli, F., Gambi, M. C., and Martz, T. R.: Divergent ecosystem responses within a benthic marine community to ocean acidification, Proc. Natl. Acad. Sci. U.S.A., 108, 14 515–14 520, https://doi.org/10.1073/pnas.1107789108, 2011.

Kroeker, K. J., Kordas, R. L., Crim, R., Hendriks, I. E., Ramajo, L., Singh, G. S., Duarte, C. M., and Gattuso, J.-P.: Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming, Glob. Change Biol., 19, 1884–1896, https://doi.org/10.1111/gcb.12179, 2013.

Kroeker, K. J., Bell, L. E., Donham, E. M., Hoshijima, U., Lummis, S., Toy, J. A., and Willis-Norton, E.: Ecological change in dynamic environments: Accounting for temporal environmental variability in studies of ocean change biology, Glob. Change Biol., 26, 54–67, https://doi.org/10.1111/gcb.14868, 2020.

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

Kwiatkowski, L., Gaylord, B., Hill, T., Hosfelt, J., Kroeker, K. J., Nebuchina, Y., Ninokawa, A., Russell, A. D., Rivest, E. B., Sesboüé, M., and Caldeira, K.: Nighttime dissolution in a temperate coastal ocean ecosystem increases under acidification, Sci. Rep., 6, 22 984, https://doi.org/10.1038/srep22984, 2016.

Landschützer, P., Gruber, N., and Bakker, D. C. E.: Decadal variations and trends of the global ocean carbon sink, Global Biogeochem. Cycles, 30, 1396–1417, https://doi.org/10.1002/2015GB005359, 2016.

Landschützer, P., Gruber, N., Bakker, D. C. E., Stemmler, I., and Six, K. D.: Strengthening seasonal marine CO_2 variations due to increasing atmospheric CO_2, Nat. Clim. Change, 8, 146–150, https://doi.org/10.1038/s41558-017-0057-x, 2018.

Lauvset, S. K., Carter, B. R., Perez, F. F., Jiang, L.-Q., Feely, R. A., Velo, A., and Olsen, A.: Processes Driving Global Interior Ocean pH Distribution, Global Biogeochem. Cycles, 34, e2019GB006 229, https://doi.org/10.1029/2019GB006229, 2020.

Leinweber, A. and Gruber, N.: Variability and trends of ocean acidification in the Southern California Current System: A time series from Santa Monica Bay, J. Geophys. Res. Oceans, 118, 3622–3633, https://doi.org/10.1002/jgrc.20259, 2013.

McNeil, B. I. and Matear, R. J.: Southern Ocean acidification: A tipping point at 450-ppm atmospheric CO_2, Proc. Natl. Acad. Sci. U.S.A., 105, 18 860–18 864, https://doi.org/10.1073/pnas.0806318105, 2008.

McNeil, B. I. and Sasse, T. P.: Future ocean hypercapnia driven by anthropogenic amplification of the natural CO_2 cycle, Nature, 529, 383–386, https://doi.org/10.1038/nature16156, 2016.

Mehrbach, C., Culberson, C. H., Hawley, J. E., and Pytkowicx, R. M.: MEASUREMENT OF THE APPARENT DISSOCIATION CONSTANTS OF CARBONIC ACID IN SEAWATER AT ATMOSPHERIC PRESSURE 1, Limnol. Oceanogr., 18, 897–907, https://doi.org/10.4319/lo.1973.18.6.0897, 1973.

Najjar, R. and Orr, J.: Design of OCMIP-2 simulations of chlorofluorocarbons, the solubility pump and common biogeochemistry, internal ocmip report, LSCE/CEA Saclay, Gif-sur-Yvette, France, ocmip5.ipsl.jussieu.fr/OCMIP/phase2/simulations/design.ps, 1998.

Oppenheimer, M., Glavovic, B., Hinkel, J., Van de Wal, R., Magnan, A., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., Deconto, R., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., and Sebesvari, Z.: Chapter 4: Sea Level Rise and Implications for Low Lying Islands, Coasts and Communities, IPCC Special Report on the Ocean and Cryosphere (SROCC), 2019.

Orr, J. C. and Epitalon, J.-M.: Improved routines to model the ocean carbonate system: mocsy 2.0, Geosci. Model Dev., 8, 485–499, https://doi.org/10.5194/gmd-8-485-2015, 2015.

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y., and Yool, A.: Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, Nature, 437, 681–686, https://doi.org/10.1038/nature04095, 2005.

Orr, J. C., Epitalon, J.-M., Dickson, A. G., and Gattuso, J.-P.: Routine uncertainty propagation for the marine carbon dioxide system, Mar. Chem., 207, 84 – 107, https://doi.org/https://doi.org/10.1016/j.marchem.2018.10.006, http://www.sciencedirect.com/science/article/pii/S030442031830149X, 2018.

Palter, J. B., Frölicher, T. L., Paynter, D., and John, J. G.: Climate, ocean circulation, and sea level changes under stabilization and overshoot pathways to 1.5 K warming, Earth Syst. Dyn., 9, 817–828, https://doi.org/10.5194/esd-9-817-2018, 2018.

Resplandy, L., Séférian, R., and Bopp, L.: Natural variability of CO_2 and O_2 fluxes: What can we learn from centuries-long climate models simulations?, J. Geophys. Res. Oceans, 120, 384–404, https://doi.org/10.1002/2014JC010463, 2015.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., and Rafaj, P.: RCP8.5 — A scenario of comparatively high greenhouse gas emissions, Clim. Change, 109, 33, https://doi.org/10.1007/s10584-011-0149-y, 2011.

Riebesell, U., Zondervan, I., Rost, B., Tortell, P. D., Zeebe, R. E., and Morel, F. M. M.: Reduced calcification of marine plankton in response to increased atmospheric CO_2, Nature, 407, 364–367, https://doi.org/10.1038/35030078, 2000.

Rivest, E. B., Comeau, S., and Cornwall, C. E.: The Role of Natural Variability in Shaping the Response of Coral Reef Organisms to Climate Change, Curr. Clim. Change Rep., 3, 271–281, https://doi.org/10.1007/s40641-017-0082-x, 2017.

Rodgers, K. B., Sarmiento, J. L., Aumont, O., Crevoisier, C., de Boyer Montégut, C., and Metzl, N.: A wintertime uptake window for anthropogenic CO_2 in the North Pacific, Global Biogeochem. Cycles, 22, 1–16, https://doi.org/10.1029/2006GB002920, 2008.

Sarmiento, J. and Gruber, N.: Ocean Biogeochemical Dynamics, Princeton University Press, 2006.

Sentman, L. T., Shevliakova, E., Stouffer, R. J., and Malyshev, S.: Time Scales of Terrestrial Carbon Response Related to Land-Use Application: Implications for Initializing an Earth System Model, Earth Interact., 15, 1–16, https://doi.org/10.1175/2011EI401.1, 2011.

Shevliakova, E., Pacala, S. W., Malyshev, S., Hurtt, G. C., Milly, P. C. D., Caspersen, J. P., Sentman, L. T., Fisk, J. P., Wirth, C., and Crevoisier, C.: Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink, Global Biogeochem. Cycles, 23, 1–16, https://doi.org/10.1029/2007GB003176, 2009.

Smale, D. A., Wernberg, T., Oliver, E. C. J., Thomsen, M., Harvey, B. P., Straub, S. C., Burrows, M. T., Alexander, L. V., Benthuysen, J. A., Donat, M. G., Feng, M., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Sen Gupta, A., Payne, B. L., and Moore, P. J.: Marine heatwaves threaten global biodiversity and the provision of ecosystem services, Nat. Clim. Change, 9, 306–312, https://doi.org/10.1038/s41558-019-0412-1, 2019.

Turi, G., Alexander, M., Lovenduski, N. S., Capotondi, A., Scott, J., Stock, C., Dunne, J., John, J., and Jacox, M.: Response of O_2 and pH to ENSO in the California Current System in a high-resolution global climate model, Ocean Sci., 14, 69–86, https://doi.org/10.5194/os-14-69-2018, 2018.

van Heuven, S., Pierrot, D., Rae, J., Lewis, E., and Wallace, D.: MATLAB Program Developed for CO_2 System Calculations, http://gts.sourceforge.net/, 2011.

van Vuuren, D. P., Stehfest, E., den Elzen, M. G. J., Kram, T., van Vliet, J., Deetman, S., Isaac, M., Klein Goldewijk, K., Hof, A., Mendoza Beltran, A., Oostenrijk, R., and van Ruijven, B.: RCP2.6: exploring the possibility to keep global mean temperature increase below 2°C, Clim. Change, 109, 95, https://doi.org/10.1007/s10584-011-0152-3, 2011.

Weiss, R.: Carbon dioxide in water and seawater: the solubility of a non-ideal gas, Mar. Chem., 2, 203 – 215, https://doi.org/https://doi.org/10.1016/0304-4203(74)90015-2, 1974.

Wernberg, T., Bennett, S., Babcock, R. C., de Bettignies, T., Cure, K., Depczynski, M., Dufois, F., Fromont, J., Fulton, C. J., Hovey, R. K., Harvey, E. S., Holmes, T. H., Kendrick, G. A., Radford, B., Santana-Garcon, J., Saunders, B. J., Smale, D. A., Thomsen, M. S., Tuckett, C. A., Tuya, F., Vanderklift, M. A., and Wilson, S.: Climate-driven regime shift of a temperate marine ecosystem, Science, 353, 169–172, https://doi.org/10.1126/science.aad8745, 2016.

Winton, M.: A Reformulated Three-Layer Sea Ice Model, J. Atmos. Ocean. Technol., 17, 525–531, https://doi.org/10.1175/1520-0426(2000)017<0525:ARTLSI>2.0.CO;2, 2000.

Wittenberg, A. T., Rosati, A., Delworth, T. L., Vecchi, G. A., and Zeng, F.: ENSO Modulation: Is It Decadally Predictable?, J. Clim., 27, 2667–2681, https://doi.org/10.1175/JCLI-D-13-00577.1, 2014.