AGU Advances

Peer Review History of

Column-Compound Extremes in the Global Ocean

Joel Wong[1], Matthias Münnich[1], and Nicolas Gruber[1]

[1] Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zurich, Switzerland

Files Uploaded Separately

Original Version of Manuscript (2023AV001059)
First Revision of Manuscript (2023AV001059R)
Second Revision of Manuscript [Accepted] (2023AV001059RR)
Author Response to Peer Review Comments

Peer Review Comments on 2023AV001059

Reviewer #1

The authors use CESM hindcast simulation over the last six decades to identify single extremes and compound extremes in the global upper ocean water column. Requiring individual events to cover at least 50m oder the upper 300m, they analyze high temperature (MHW), high [H+] (OAX), and low-oxygen (LOX) extremes, as well as compound MHW-OAX, MHW-LOX, OAX-LOX, and MHW-OAX-LOX extremes (column single extremes (CSX) and column compound extremes (CCX), respectively). The authors use both relative and absolute thresholds for defining extremes. They analyze occurrence relative to a fixed baseline (1950s conditions) as well as relative to a moving baseline, where the thresholds evolve are shifting to account for the trends in T, [H+], and [O2]. After evaluating the model simulation against observation-based surface T and [H+], they analyze changes in CSX and CCX occurrence under a fixed baseline. Relying on a moving baseline, they analyze difference in characteristics of the CCX and their co-occurrence with ENSO variability, as well as spatial patterns of occurrence and event metrics. Finally, the depth structure of CCX is analyzed regionally using a k-means clustering approach. This is done by defining six 50m depth bins and summing the intensity indices of single extremes during the column compound events in each bin over the whole time series.

In my point of view, the study is very comprehensive, an impressive piece of work, and

represents an important contribution both to the research on oceanic compound events and also on subsurface extremes in the ocean. As such it is suitable for publication in AGU Advances, given my points from above can be addressed.

Main comments:
* Is habitat compression a good name, given that impacts can not be directly inferred from the identified extremes, as pointed out in the discussion? It sounds at places a bit overstated, e.g. in L460-461. I would propose going with something more neutral like "affected ocean volume" or so, and then go into the potential ecosystem impacts in the discussion.
* I am unsure about the fact that MHW volume fraction is above 5% during the first years (Figure 5a). The average volume covered by T > 95th percentile should be 5% by definition (assuming the fixed baseline is representative for the first years). CSX pose the additional constraint that at least 50m of volume need to be extreme at the same time - so I think these should have a smaller volume fraction than 5%. I am likely missing something, but some clarification would be appreciated. In a wider sense I am surprised how close all volume fractions in the first years in Fig. 5a are to the 5% that would be expected without constraining at least 50m to be affected - does this mean the 50m constraint does not exclude many extreme events in single grid cells?
* Is there a simple reason why CCX including [H+] do not become more frequent over time under under a moving baseline (Figure 6), as it may be expected from increases in [H+] variability (e.g. Orr and Kwiatkowski, 2018, Burger et al., 2020)?
* I am somewhat confused with the different size and location metrics (Table 2). I suppose the "volume fraction" in Figs. 5 and 6 refer to how much ocean volume is under CCX or CSX conditions at any point in time, averaged over the 60y duration. But how does that compare to the "vertical fraction" panels in Figs 8 and 9 that go all the way to 100%? I expect this metric refers to how much water column is extreme during the CCX, and not in general. I would propose to add more details to the Extreme Event Metrics section and / or Table 2. Furthermore, I don't see yet the advantage of the metric "contiguous habitable space" compared to just the "vertical fraction". Some more motivation what we learn from this metric would be good.
* How well does the model simulate mean oxygen concentrations at subsurface? The skill in doing so is necessary to have meaningful extremes relative to the absolute threshold 150 muM. Mean oxygen content at subsurface may be something that can be evaluated based on observational data.
* For the clustering, I am not sure whether the conclusions are a bit too strong. As far as I understand, the depth structures represented by the cluster centroids could also arise as a result of different types of CCX occurring in the same region, particularly relevant when a bimodal structure occurs. See my comments on lines 592-599 and 663-679.

Specific comments:
* L65: What is meant exactly with "when stressors reinforce each other"? that one stressor causes another stressor, or that the impact from multiple stressors at the same time on an organism is larger than if they had occurred independently?
* L81-L86: I see the difficulty of being comprehensive in this part of the text, but these numbers are somewhat dependent on the extreme event definition and there may be ways of hinting towards the threshold definitions, like "Based on approach X, Y showed that the frequency of MHWs increased by Z".
* L109-L111: How about adding Amaya et al., 2023 (https://doi.org/10.1038/s41467-023-36567-

0) here?
* L260: Please refer here to Supporting Information Text S1, otherwise the definition of the seasonally varying thresholds is unclear (e.g., the application of a 11d rolling window)
* L270: The definition of the fixed baseline appears somewhat unusual relative to standard fixed baseline approaches where percentiles are just defined over some baseline period, e.g. 1961-1990. How is the choice here motivated?
* How is the moving baseline defined for [O2] with the additional criterion [O2] < 150 muM on top of the relative thresholds?
* L308: How is the intensity index for a single grid cell aggregated over the water column for CSX? The average over all affected grid cells in the CSX, or the maximum over these?
* Equation 4: I think the case N=Nr != 0 is not covered, although I would expect CP to be zero then. This could be probably solved by using >= or <= in one of the first two cases.
* L334-335: CP is not really proportional to N, is it? that would be CP = constant * N. Please clarify.
* What are differences (advances / disadvantages) of taking CP vs LMF? For example, an advantage of CP seems to always have the same range of values, irrespective of threshold definition. On the other hand LMF seems to be advantageous, as a LMF of, say, 2 has direct meaning, implying events to be twice as likely as by chance. It may be worth to add a sentence motivating the choice.
* L357ff: why are CESM MHW evaluated against monthly SST data, given availability on daily resolution? Is it because the hindcast simulation was forced with monthly atmospheric data? Connected: For the evaluation of MHW, were CESM MHW recalculated based on monthly-mean resolution? (L364ff)
* L393: I would add here the reference Ma et al., 2023 (10.1029/2023GB007765), since therein supp info table 6, the low non-seasonal variability in OceanSODA (for pCO2) is nicely shown
* L454-455: triple compound events have become 6 times more intense -> This sounds like a lot, but I think it is important to note that the CCX intensity, as defined by the Euclidian norm over the 3 intensities, is here basically controlled by [H+] intensity. Some care must be taken in interpreting these numbers, as it does not mean that the triple compound event has become a lot more intense in all dimensions.
* L477-480: I think the Spearman correlation coefficient may not be a very good choice when comparing a numerical variable (volume fraction of double CCX -> I suppose the total volume covered by either of the three CCX?) with a categorical variable (|ONI|>0.5 -> why do you use the absolute value here? wouldn't this also include La Nina?). Is the correlation coefficient similarly high when comparing directly against ONI?
* L477-488: are the numbers provided in this section all for the tropical Pacific?
* L481f: What is meant with "an asymmetric relationship with CCX volume fraction as compared with El Niño"?
* L489-494: OAX-LOX events affect most volumes because they have a large tendency to fall together? and longest duration because they tend to occur at subsurface where events tend to last longer? Some discussion would be interesting here.
* L497f: It should be "maximum intensity index" I think
* L490: I am confused by the correspondence between volume fraction and absolute numbers in km3. It says 735km3 would be equal 0.73% of the volume in the upper 300m, if I understand correctly. However, the global ocean has a surface area of around 360e6 km2, multiplied by 0.3km (for 300m) yields a total volume of around 100e6km3, such that 735km3 should be something like 1/1000 of a percent rather than almost a percent. This point concerns all numbers in section 4.2. Please clarify. Hopefully, this means that the 12 km3 for the triple CCX

is not correct, this number is really small given that a single water column in a 1{degree sign} grid would have a volume of about 100km*100km*0.3km=3000km3.
L495-498 and Figure 6: Why is the intensity index for MHW-OAX larger than for the other two double CCX? some more discussion would be interesting
* L508-509; Figure 7a,b: How much can the model results be trusted close to the Antarctic continent? given the relatively coarse model resolution there and the lack of high-quality observational data to compare against.
* L511: "typically less than a week" seems to propose that MHWOAX frequency is very similar to the subtropical and southern high latitude ocean, where "one week" is under these conditions. Some less rounded numbers would be helpful. See also the comment below about colorbars in Figure 7.
* L514-515: Am I right in assuming this is because [O2] does not fall below 150 muM in these regions? I think this should be brought up. The same aspect comes back in L535, where it is discussed that CCX including oxygen mostly occur in the tropics.
L516: "Last typically less than 7 days" - is this referring to number of days per year or duration?
* Figure 7: I have the impression increasing the number of levels in the colorbar for panels a,c,e,g would help highlighting the spatial patterns and discussing them in the text
* Figure 8: panels c) and d) are already shown in Figure 7, correct? If so, I would propose to mention that in the caption. Same holds for Figure 9c,d
* L542-543: Not sure if I understand correctly. The intensity at the threshold during a CSX would be 1, the corresponding CCX intensity for this case would be thus sqrt(2). One could also write: "2 to 4 times the intensity of the threshold for a single event".
* L544-546: From Fig 8b, I can not really read off that MHW-OAX have low intensity throughout the tropics. Maybe be more specific.
* L546: should it be "while" instead of "when"?
* Figures 8 and 9, panels b: intensity can not be below 1, can it? In this case the colorbar should start from 1 and not from 0.
* L563: "for the two other CCX event types"
* L592-599: I am not sure whether I understood it correctly. This text reads like any individual CCX must have this bimodal depth structure for grid cells that fall under cluster 1. However, with my understanding of the clustering approach, it may as well be that there are separate surface CCX and to a lesser extend subsurface CCX that occur in these regions, and that the MHW-OAX signal in individual events thus not need to be transported from surface to subsurface or vice versa.
* L627-628: "Cluster 2 events are located higher in the water column" -> this does not seem the case, from Fig. 10 it just looks the other way round
* L634-636: The conclusion that extremes occur because of variations in the extend of OMZs seems only one option. Extremes could also occur in a water that is permanent below 150 muM, just because the additional 5th pctl threshold is surpassed.
* L649-651: I would write that MHW-OAX-LOX cluster 5 and MHW-LOX cluster 4 are rather similar than identical. The numbers provided in this sentence mix the characteristics from both clusters, which is confusing, e.g. duration is from MHW-LOX, while intensity is from the triple CCX. Likewise, I would point out everywhere in the paragraph the similarity between MHW-LOX and MHW-OAX-LOX clusters rather than stating that they are identical.
* L660-663: Am I right that this is a bit speculative as the actual drivers can not be identified in the study? If this is true, please reformulate a bit more cautiously. For example, the role of biology can not be identified, or can it?
* L663-679: The discussion here is very interesting. However, it would be interesting to me

whether the individual triple compound events in this region generally have this bimodal temperature structure seen in Fig. 10 (as also shown in the supporting movie) or whether one could get this type of cluster also from different types of triple compound events occurring in the same region, one with surface warming and one with subsurface warming (related to the point on L592).
* L684ff: the discussion of cluster 4 - "there is not any known deep water mass of higher temperature in these areas". How about the role of seasonality in general here? During which seasons do the tripple compound events occur in the Bering Sea? During winter time, it may well be possible to have higher subsurface temperatures. Such a more general discussion may be necessary, since cluster 4 occurs not only after the blob, but throughout the time series (Fig. S11)
* L708: "up to 10 days per year" I would change the colorbar in Figure 11 to a finer scale - otherwise increases by up to 10 days per year are hard to see.
* L710-712: I can not see how Fig 11a is 20 days larger than Fig11b in the subtropics and tropics - particularly since there is quite a distinct increase in MHW-OAX during La Nina in the western Pacific. Do I understand the statement correctly?
* Figure 11: El Nino and La Nina phases are defined as in Fig. 6 or differently?
* L763: "the entire UPPER OCEAN water column", same in L863
* L767: Rather Section 4.4 than Section 4?
* L783: and remineralization also enriches the water with carbon, driving OAX. This may be worth adding.
* L798-800: I struggle to see how MHW-OAX events in the subtropics are driven by ENSO when looking into Fig. 11a,b - connected to my point above in Section 4.5.
* Figure 12: Values for the fixed baseline: are these also averages over the whole period 1961-2020? The colors for MHW-OAX-LOX 4 and 5 are relatively hard to distinguish on the map, maybe it would be worth to increase the brightness difference between the two.
* L813-814: how can you tell that the association to ENSO is low due to neglection of lag times, and not that the cluster may be also strongly driven by local processes or other large scale modes of variability?
* L816: "strong sub-cluster enhancements and suppression": The relatively little amount of CCX data during El Nino / La Nina phases surely also adds considerable noise to the results, making the shifts in CCX occurrence from ENSO potentially hard to detect. This may be worth adding.
* L828: is harbinger the right word here?
For the vertically migrating organisms, is it also relevant how much of vertical displacement the anomalies in T, [O2], and [H+] correspond to during the extreme events? For example, when an extreme event in [H+] would only imply conditions in [H+] that are usually prevalent only 10m deeper in the water column, I could image the impacts to be not so severly, compared to if they would correspond to a much larger displacement. [H+] may be a particularly good example, given that global [H+] increases by about 5 mumol kg-1 from surface to 300m depth. If an extreme in [H+] would imply a departure of [H+] by 0.5 mumol kg-1, that would mean something on the order of 30m of vertical displacement. I think this can definitely be significant - just thinking loudly here because it seems like a relevant aspect to me (the line of thought is a bit similar to the "Thermal displacement by marine heatwaves" paper by Jacox et al., just for the vertical).
* L887-888: This sounds a bit like the biases would only occur in upwelling regions, which is not the case (Fig 4d,e). It may be also worth noting here that the biases can be read both ways - it may also be the data product that underestimates variability, specifically for OceanSODA.
* L892: is the co-occurrence propensity really only a product of large scale processes? This

sounds a bit too much simplified to me.
* L925-929: A nice recent reference that points towards the necessity of compound extreme conditions for impacts may be Le Grix et al., 2023 (https://doi.org/10.1111/gcb.16968)
* L936-937: as pointed out before, the occurrence of all CCX including oxygen mainly in the tropics and the north Pacific seems a direct consequence of the assumed fixed threshold of 150 muM for oxygen. I think the manuscript would benefit from pointing this out
* L941-943: this seems a bit contradictory to the discussion on the ENSO related MHW-LOX cluster 4 and MHW-OAX-LOX cluster 5, where the bimodal structure for temperature is pointed out.

Small comments:
L356: remove dot before 'covering'
L527: "understand"
L660: ". ." Typo
L805: "El Ni no"
L871: "occur" instead of "occurs"
Supporting information, L8: please adjust "SX", see also L31
Supporting information Figs. S4 and S6: The depth threshold = 100m plots seem to show almost everywhere extremes (non-white area), that are not present for 75m depth thresholds. Is there a plotting problem?

Reviewer #2

Review of Column-Compound Extremes in the Global Ocean by Wong et al.

Summary
The authors present a study of single and multiple extremes found in the water column throughout the global ocean over the past 60 years. To do so, the authors utilize a single hindcast simulation from the Community Earth System Model with embedded biogeochemistry model to explore where and when marine extremes of high temperature (heatwaves), low oxygen, and acidity occur, both in isolation and together, with impacts on marine life still an open question. Furthermore, the authors analyze various metrics, including event duration, intensity, and frequency and evaluate how these metrics have changed over the study time period (largely, increased). Lastly, the authors present a complex but informative summation and categorization of their findings through a k-means clustering scheme in order to highlight particular hotspots in the global ocean where the water column is especially vulnerable to compound extremes in the water column, driven by complementary or competing factors, likely owing to both local (e.g., sea ice) and large-scale (e.g., ENSO) variability.

Appropriate for AGU Advances?
While studies until very recently have primarily focused on investigating single surface extreme events, especially marine heatwaves (MHWs), the presented study dives deeper into the water column and across types of extremes in order to focus on and include the conditions most relevant to marine ecosystems, and their management in the future. While this not the first study to use a hindcast simulation to investigate global temperature, oxygen, and pH extremes in the water column (e.g., Gruber et al. 2021), this study presents a new metric or identification scheme by which to identify multiple extremes in the water column, coined Column-Compound Extremes (CCXs); this could either turn out to be a one-off method or the basis of a future standardized scheme. While the majority of my comments (below) are minor in nature, there are a few items that the authors would need to consider addressing prior to acceptance and final publication.

Top comments:

The majority of the figures are illustrative and effective, while the tables present clear and useful information. Before final publication, however, the manuscript text as a whole requires additional attention concerning grammar and punctuation errors, to improve clarity and readability.

While the authors state their level of confidence in the model's representation of MHWs and OAXs, I am not as confident in the latter. I expand on some of the specifics behind my hesitation in the Specific Comments section below, but in short, the authors fail to provide the necessary quantitative information to be able to effectively draw conclusions on the utility of this particular model in representing OAXs (and completely fail to attempt any validation of oxygen/LOXs entirely). Instead, the authors present side-by-side visual (qualitative) comparisons, between the simulation and an observation-based product, which demands sole reliance on a subjective eye and full acceptance of the (seemingly inflated) conclusions presented by the authors. Therefore, I strongly recommend that the authors pay considerable

attention to this component of their analysis while in revision. A great deal of the interesting and substantial findings of this study involve OAXs (e.g., the most pronounced trends reported include an OAX component), making it all the more important that it's handled with care. Additionally, I suggest they include at least a short statement in the validation section as to why they cannot or will not attempt to validate LOXs (I believe the final section of the paper includes a first and only reference to a lack of data). See also specific comments further down.

I describe a significant source of perplexity as follows and I hope the authors can clear things up. The authors employ a single model simulation (hindcast) throughout their analysis but do not choose to use it to further identify the concurrent behavior and trends of supporting variables (upwelling, pH, MLD changes, etc.) to help explain and inform their results on extremes. Instead, it appears that they point to previous and independent studies to draw conclusions. While the mechanisms they employ to contextualize their extremes results are scientifically reasonable, without actual quantification with the same modeling tool, there's an obvious sense of unfinished business apparent throughout this study. The authors may be personally confident in the model's representation of MHWs and OAXs but this does not automatically translate to confidence in other atmospheric or physical and biogeochemical ocean processes that could influence these extremes.

General comments:

The authors define ocean acidity and low oxygen extremes as OAX and LOX, respectively. Throughout the manuscript, the in-/exclusion of an additional usage of the word extremes is inconsistent. For example, one sentence may read, "For OAX, trends increase … ." and another may read, "For OAX extremes, trends increase … ." I recommend that the authors choose one and make it consistent throughout.

The authors have isolated the regions of the World Ocean where and when multiple extremes tend to occur, highlighting the most prominent in severity. While the authors discuss possible drivers and general impacts on pelagic species, I feel there's also a missed opportunity to really hit home this key finding with just a couple of additional sentences that briefly highlight the significance of these hotspots with respect to well-known areas of the world, like the OAX-LOX and MHW-OAX CCXs near the Great Barrier Reef and the OAX-LOX CCXs throughout the major fisheries within the California Current Large Marine Ecosystem, for example. And if the authors were to pull on that thread even more, they could use these regional highlights as an opportunity to further discuss the longer-term trends at these sites specifically (e.g., how much have MHW-OAX CCXs increased in/near the Great Barrier Reef region over the study time period, highlighting how continued increases in extremes into the future would be devastating to this unique and important ecosystem).

Heads up, Supplementary Figures S3, S4, S5, and S6 are not explicitly cited/referenced in the main text (at least through a command-F quick search). This could be an error or it could suggest that they are not necessary. Also, the data repository the authors point to doesn't appear to have consistent figure counts/naming convention (e.g., I don't see a Figure 13 in the main manuscript but there are 'fig13' files in the repository).

Specific comments:

Key Point #1
The sentence structure of this point is confusing. Perhaps reword to read, "Column-compound extremes (CCX) occur when 50 m of the top 300 m is extreme in multiple parameters, reducing habitable space by up to 75%."

L26-29 Combining or rearranging these two sentences might improve readability and flow (and avoid beginning a sentence with 'they'). For example, "Removing this effect with a moving baseline permits us to better understand the key characteristics of CCXs, revealing a typical duration of 10-30 days and a predominant occurrence in the tropics and high latitudes, regions of high potential biological vulnerability."

L56-63 Rearranging this sentence might improve readability and flow. For example, "Receiving increased attention are extremes in ocean acidity (OAX) [citations] and low oxygen (LOX) [citations], with particular emerging concern about compound marine extremes, when conditions are extreme in more than one stressor [citations].

L64-66 The use of 'that is' here is clunky. Could it be removed?

L68 I believe the semicolon here is meant to be a comma.

L64-77 This paragraph reads as two versions of the same sentiment (L64-72 and then again L72-77). Is there a way to achieve conciseness through editing?

L78 What is meant by 'strong' here? Large/steep slope? Unwavering? Please clarify or present the info differently. (Further down in this paragraph, I see that the meaning of strong can be inferred but only in the case of OAX.)

L84 The authors should consider avoiding words like 'going' when describing quantitative results. Instead, 'For OAX, the trends are even stronger, increasing from … ."

L93 The authors might consider including a time reference to the "Blob" event.

L94 Remove comma after "They speculated … ."

L89-91 For some reason, this sentence doesn't sit right with me. Perhaps it's too strong of a statement or evokes too much certainty with the use of the word 'must.' Perhaps an alternative such as, "As a consequence of these trends in single-stressor extremes, increasing compound extremes naturally follow."

L91 To avoid too many mid-sentence breaks in this paragraph, the authors should consider beginning the sentence with 'For example, Gruber … ." instead.

L129-132 Sentence structure quite clunky. Perhaps rearranging will help with flow and efficiency, "As for generating and maintaining marine extremes at depth, other biological-physical mechanisms (e.g., mesoscale eddies; [citations]) are critical.

L145-147 This sentence jumps around unnecessarily, making it difficult to follow the thread. Instead, "… but they affect many remote regions through connected changes in large-scale

ocean and atmospheric circulations (aka teleconnections)."

L155 'also' unnecessary here

L159-177 The last 2 paragraphs of this section would benefit from editing, for all the reasons indicated so far.

****For the sake of time and to avoid copy editing the paper, I realize by the end of this first section that I must refrain from continuing to explicitly note the lines that I feel warrant editing for grammar, clarity, readability, flow, conciseness, etc. As such, I recommend that the authors address the rest of the paper themselves before final publication. Nevertheless, the comments so far serve as some appropriate examples.

Section 2
While editing is needed, this is a valuable section and the figure is effective.

L253 The authors discard the first 3 years of the available hindcast simulation. Is this simply because an even 60 years is personally desired or because there still exists some residual spin up? Either way, please include a reason in the manuscript.

L260-262 Does this mean the authors first computed the 5th percentile across all O_2 concentrations or only on times and depths when $O_2 < 3.5$ ml/L? Just want to be clear on the order of operations for reproducibility. The use of the word 'simultaneously' blurs this a bit.

L267 Out of curiosity, what is the shortest duration event found in the hindcast simulation?

L275-277 The authors admit that the choice of a 50 m minimum extension is subjective (a few lines down). The 300 m depth analysis range also seems subjective, as vertical migrations occur throughout the twilight zone (can extend down to ~1000 m). I'm not advocating for an alternative depth cut-off for this study, just for additional explanation surrounding the choice of 300 m; I looked through the 3 citations provided and 300 m didn't necessarily jump out as an obvious maximum depth, so the authors must have put additional thought into this.

L288-290 This last sentence seems to be placed prematurely and is perhaps better suited under the subsequent subsection on such 'metrics.'

L325-327 Does 'high' and 'low' here mean positive and negative, respectively?

L331-332 Does the hindcast simulate leap years, where DY could also = 366?

Section 3.5
Are there previous studies that have used the OceanSODA-ETHZ observational product and have drawn their own conclusions on its quantitative representation of their region of interest? If so, please include this/these for the reader. If this confidence can be achieved from the original Gregor & Gruber, 2021 paper alone, please indicate this somehow (such as, "specific details on the data set's representation of the observed ocean can be found in ...). This is particularly relevant in L378-380, where an appropriate citation would be helpful.

Section 3.5
While low oxygen is one of the three extreme events analyzed in this study, there is no mention of O_2 in this section on model evaluation (not even to say there isn't sufficient observational O_2 data to assess, if that's the case). Yet, the authors conclude (L407-409) 'that the hindcast model is able to capture not only the mean state of the ocean's physical and biogeochemical state, but also its variability, which is a critical requirement to investigate extremes.' Since the authors only assess the model's representation of acidity ([H+]), their conclusion, as stated, is too strong, without sufficient discussion on the quantitative assessment of acidity and the same for O_2. As I'm sure the authors know, a model could capture characteristics of one biogeochemical variable to a certain level of confidence, but not necessarily for another.

L360-363 While well-known, the El-Nino Southern Oscillation should first be spelled out before shortening to ENSO.

Figure 3 (L364-376)
The authors' description and graphical representation of the comparison between the hindcast simulation and the observational product is more qualitative than quantitative. For example, the authors mention the model's simulation of the distribution of the mean duration of detected MHWs as 'correct,' but do not show a histogram/PDF (*unless the authors meant spatial distribution here, unclear). By eye (Figure 3b,c), the model-product difference is at least by 1 month, on average, across most of the global ocean. Again, this is just a result of one person's qualitative assessment of the model-product differences shown ("by eye"). I suggest the authors consider updating this figure to include more quantitative information, such as that computed from common metrics (bias, root mean square error, etc.). This will allow the reader to see for themselves all that the authors claim ('great fidelity', 'correctly,' 'very well', 'overestimate') in this accompanying paragraph. I appreciate that this model-product comparison is not the focus of this study, but since the authors have already chosen to include a figure, it could be slightly tweaked to highlight the most useful and quantitative information that will also help to minimize subjective evaluation.

Figure 3a/L365-366
Since the authors do not include an additional contour representing a chosen ENSO index for the reader to confirm the statement on 'closely coupled fluctuations,' perhaps the authors should consider including additional text pointing out the sharp peaks around the large El Nino events of ~1998 and ~2015, as guidance.

L371 The inclusion of the phrase, 'while there are clearly issues with the models,' reads clunky and out of place. What is meant by this? Perhaps this is just a restatement of the first part of the sentence (unnecessary), but if not, the authors should consider expanding on what they mean by such a statement in the appropriate context.

L394-396 The grammar muddles the meaning of this statement ('especially not with regard to'). Please clear up.

Page 12/Figure 4
Can a statement on trends in OAXs or their CCXs be made confidently when we're unsure of the model's representation of their duration and intensity? I suggest the authors explore additional ways that might help either (1) increase our confidence level in these reported findings or (2) to

consciously and explicitly overstate the inherent uncertainties throughout (and not just in the caveats section at the end).

Figure 4 (L377-397)
As for my comments on Figure 3 and its accompanying description above, Figure 4 and its description would also benefit from an update from qualitative to quantitative, especially given the model's poorer representation of H+ extremes, relative to SST extremes.

L417-419 I see value in reporting these underlying trend values. For example, what percent change in acidity did the global ocean experience over the 60 year period?

L419-421 If you mean that these results are consistent with the previous studies cited, please edit to be more explicit. Currently, it could be interpreted as the column results simply mimic/are tied to the surface within your own analysis. Also, could the authors expand on this point? How so? For example, the authors describe the OAX-CSX change as 'rapid,' but do these previous studies show an equivalent rapid change around the same time period as this study?

L424-426 Regarding the described 'smaller trends in oxygen,' please provide a reference value/magnitude. Also, is it truly just a small negative trend in underlying global oxygen concentration or is it a result of heterogeneity in global O2 trends that tend to cancel/dampen the resulting signal? If the former, please elaborate as to why the additional 150 uM constraint would act to further 'mute' the trend in LOX-CSX.

Figure 5a
Since confidence isn't necessarily strong in the hindcast's ability to represent OAX across the globe, the authors might consider including a discussion on specific model behavior that could be influencing the underlying trend in global ocean acidity. For example, what role does model drift, if any, play, especially at depth? And on that note, what proportion of the water column appears to drive the OAX-CSX volume trend (full 300 m, more of the surface, more of the subsurface, etc.)?

Figure 5b (L439-442)
Are the positive trends in double CCXs having OAX (MHW-OAX and OAX-LOX) driven purely by the trend in OAX-CSX (Figure 4a) or are there fundamental differences in the spatial distribution globally of these double events vs. single OAX events?

L440 Another example of editing needs: the use of 'jumped up' is too colloquial.

Section 4.2
An accompanying table would be beneficial for the quantitative information reported in this section, either for the reported ENSO correlations or the stressor type characteristics (duration, volume, etc.).

L477-488 Please be explicit about significance tests employed and where/when significance is achieved alongside the reported correlation values.

L514-515 This kind of result is the perfect opportunity for the authors to expand upon/include a discussion on the 'why' behind these spatial patterns. For example, what about these basins

would preclude OAX-LOX CCX events? Is 300 m too shallow to see a signal here? Is the N. Atlantic well-oxygenated relative to other basins? Please provide more interpretation and discussion.

Figure 7
This figure made me realize that an additional or supplementary figure containing maps of the global average temperature, oxygen concentration, and acidity would be helpful, perhaps shown at the surface and again at 300 m; especially helpful to those that can't immediately recall which basins are relatively low in [O2], for example. A reference/baseline of the average model parameters of interest like this would allow for better interpretation of Figure 7.

Figure 7 caption
It would also be helpful to include in the caption a short refresher on what +/- propensity values indicate so the reader doesn't have to go back and reference the Methods section.

L588-590 It's unclear whether the authors have actually confirmed/documented increased stratification, reduced nutrient supply, and suppressed biological productivity in the hindcast or if these are hypotheses.

L600-623 I appreciate the inclusion of discussion and mechanistic understanding in this paragraph (and the rest of this subsection), but just want to point out that the authors have also chosen to have an entire main section dedicated to discussion (5); this paragraph sticks out because up until now, any expansion or discussion has been non-existent or muted. The answer is subjective, but perhaps the authors should reflect on the overall structure of their manuscript and make final adjustments accordingly. ... Have the authors confirmed any of these underlying mechanisms in the hindcast? It reads as if the mechanisms cited are hypotheses based on current understanding. For example (L617-619), have the westerlies over the ACC strengthened to drive enhanced upwelling of CDW here? If the authors have not, what is their reasoning? Regardless, I recommend these analyses, which would be consistent within the model framework (and readily available).

Figure 10
The colors corresponding to the violin plots are not explicitly stated. Please include either as a key within the figure or noted in the caption.

L624-631 I don't see a consistent match between what's presented in the figure (10, top middle, violin plots) and the descriptive text. For example, "Cluster 2 events are located higher in the water column, between 50 and 200m ..." yet the violin plots show both events as lower in the water column (centered ~200 m) relative to Cluster 1. Please double-check the validity of all OAX-LOX clusters shown/described here.

L678-679 Thanks for sharing the video; illustrative to see the authors' description come to life.

L675-678 Figure 12 is invoked before Figure 11. Please reorder to make it sequential.

Figure 12 With such little area occupied by the MHW-OAX-LOX 4 and 5 clusters combined with the very little shading difference between the two dark/black colors chosen, it's very difficult to discern which is which on the map itself. Please make the black shades more distinct from one

another (which may mean the authors have to lighten the color of the continents further). Also, the caption states that the values from the fixed baseline are in brackets but appear in parentheses in the figure; please update. It would also be helpful to again include the single-double-triple event coloring triangle key from Figure 1b for ease/reference.

L819-823 The authors state, "These CCXs have been found to be driven by … " and go on to cite previous studies. The way this is worded suggests that previous studies have indeed investigated CCXs before and that this is not the first study to define and analyze them in this way.

L893-897 The authors could consider employing an observational product, like the recent GOBAI-O_2 product from Jon Sharp, to include in their model validation efforts, especially as there is currently no comparison to observed O_2 whatsoever in the study. Sharp study here: doi.org/10.5194/essd-15-4481-2023.