We would like to thank the reviewers for their constructive comments. Below, we respond to each comment in turn; our replies are preceded by “>”.

REVIEWER 1

General comments

We thank the authors for their response and for incorporating most of the feedback. We believe the manuscript has been substantially improved. However, we still have some comments on the structure of the results section and the general writing. With regards to the writing, the authors are advised to carefully check the grammar and for missing verbs in sentences, and to rephrase when necessary. Please see detailed comments about the content below. Line numbers refer to the manuscript version (not to the tracked changes version).

> We checked the grammar and added missing verbs or other words, often resulting from regrouping earlier text fragments. We highlight the small textual changes in the diff document provided with this reply.

L335 add before “The” at the end: A second mechanism causing the freshening is the convective feedback by which reduced mixing leads to fresher surface waters that further inhibit the vertical mixing (Kuhlbrodt et al., 2001).

Main comments

The structure of the results section could be improved by placing the section about the warming thresholds (CDFs) in a more suitable place. The methodology is already used in Figure 3 and Figure 10, and should therefore be described before being used. As suggested before, the CDF methodology should be placed in the Methods section, but at least before the CDFs are used for the first time in the Results section. Since the warming thresholds for each system are already displayed in Figure 3, it could be an option to integrate the results of section 3.10 in this first part of the results. This would place the CDFs in the appropriate context and a clearer connection between figures 3 and 17 & 18 can be made.

> We moved figures 3 to section 3.10, and included a table at the start of section 3. Similarly moved figures 10 and 11 to section 4. We agree that this ordering makes for a more logical narrative. We do make a forward reference from section 3 to the new version of figure 3 (now figure 16), since new table 1 does not contain the spatial information of the categories of SNS.

In the response to the reviewers, the authors responded that they made changes to the conclusion. However, this change was not implemented in the conclusion. The proposed changes by the authors were only placed earlier in the discussion. Please rephrase the final conclusion to appropriately reflect the conclusion of the assessment.

> We thank the reviewer for spotting this error and have replaced the relevant lines with: The frequency of SNS events increases steeply at global warming levels below 2 C and remains relatively high between 2–4 C. While the frequency of SNS rises further between 4–6C this is based on a rapidly decreasing number of models (dropping from ~54 to ~10 in this interval), which correspondingly decreases the robustness of the inferred frequency

Specific comments

Line 136-139. The last sentence of this paragraph does not add much to the methodology and could be removed; the computational complexity argument is not very relevant here. If the authors want to keep this argument, please rephrase it because the sentence is somewhat unclear.

> We agree that this sentence did not add substantially to the methodological description and have removed it.

Lines 180 – 184. The specific discussion about the monsoons is not necessary here, since it is already stated that this catalogue does not focus on atmosphere variables. The first part and the last sentence of this paragraph are however relevant.

> We have deleted lines 180–184, starting from 'likely'.

Figure 10 is only referenced once in the text, in the last sentence of section 3.4, and once in the caption of figure 11. Could the authors reference this figure earlier in the section and put the figure into the context of the section?

> As mentioned above, we moved these figures to Section 4. Discussion as it builds on figs 17 and 18, together with the last par. of section 3.4. An introduction of this section is added:

As discussed in section 3.2.2 based on the findings in Drijfhout et al. (2025) a collapse of deep mixing in the Labrador, Irminger and Nordic Seas is found to precede a shutdown of the AMOC related to the formation of North Atlantic Deep Water in CMIP6 models. This is further illustrated by Figure 10 where we see the Northern Hemisphere mixed-layer collapse to occur at 0.3 – 0.5°C earlier which corresponds to ca 20-30 years earlier. For this reason, we argue this is a clear example of a tipping point cascade. In addition, this cascade also occurs earlier than previously estimated …

With Figure 11, the authors added an explicit comparison with McKay, which is a valuable addition. However, we suggest that the temperature thresholds should be displayed differently. By taking the 0 and 100 percentiles of the CDFs of maximal change, you obtain a very large spread in expected thresholds. This comparison raises two issues that should be addressed:

- First, in McKay, the minimum and maximum are estimated through literature/expert assessment, meaning they have an error and should likely be taken as an estimated interval where a collapse is somewhat likely. By taking the 0 and 100 percentiles of the CDF , the authors effectively disregard this error on the McKay estimation. To address this, we recommend using e.g. a boxplot, violinplot or scatterplot for more nuanced comparison.

- Second, the authors analyze the full CDF of maximal change, while McKay looks at the likelihood range of a threshold. These are fundamentally different statistics: the uncertainty of the temperature threshold location within each simulation versus the uncertainty of an AMOC collapse occurring at a given temperature. Thus, it would maybe be more appropriate to use e.g. the median/mean/maximum slope of the CDFs as the basis for Figure 11. In any case, it requires an improved argumentation.

- Since the authors added a valuable comparison between this paper and McKay for the AMOC, the authors are encouraged to extend Table 1 by adding this comparison for all the systems (for those that were also assessed by McKay). The authors should choose a suitable error on their estimate depending on how they address the above two points.

> We extended the discussion to clarify that the two approaches use intrinsically different metrics. We have also clarified that our estimates are not thresholds, but warming levels associated with maximum detected change. Following the reviewer's suggestion, we included violin plots for the two mechanisms discussed in this section. We did not extend Table 1, because this comparison is not discussed for all categories in that part of the paper.

Paragraph at line 395. What do the authors mean with the first sentence ("… urging for adjusting the burning embers in these reviews.")? The authors claim here that there is a much earlier onset of collapse based on the analysis in figure 11. For this, see the comments above. Because of taking the full range of the CDF , this methodology will tend to estimate wider ranges than the other reviews. This is not because the collapse actually has to occur at much lower temperatures.

> We deleted the urging part of this sentence. The “earlier” aspect of this section was meant to relate to the median of old figure 11 (now figure 18). We believe the suggestion of the reviewer to convert the results from this work to a violin plot helps making this clear.

The causality mentioned at the end of the paragraph (lines 397-398) is not fully supported. In the mixed layer section (section 3.2 & 3.3), the order of events was reversed, where it is argued that freshening of the subpolar gyre is brought about by declining AMOC (line 334). In this paragraph (lines 397-398) it is said that a SPG collapse triggers an AMOC collapse which cannot be proven through the results shown. It is not given that a SPG collapse automatically leads to an AMOC collapse (models can show a stop of convection in the Labrador sea while still having an active AMOC). This point should also be addressed more carefully in the discussion at line 584.

> We agree that the causal sequence required clearer explanation. We therefore added/changed the following:

After being triggered by atmospheric forcing (e.g. warming), further restratification of the water column becomes more determined by freshening brought about by a declining AMOC and even more by convective feedbacks resulting from reduced vertical mixing amplifying the reduction further (Drijfhout et al., 2025). This process also develops faster than the full AMOC shutdown and thus the SPG mixing collapse precedes the AMOC shutdown. However, for the AMOC collapse a mixed-layer collapse in all deep convective regions (including Nordic Seas) is needed.

We changed line 573 to: Associated with a mixed-layer collapse in all deep convective regions in the North Atlantic occurring earlier in CMIP6 than in CMIP5, also …. (and replace the , with :)

On Line 584, we reversed the sentence:

We also show that a shutdown of the AMOC is preceded by a collapse of deep mixing in the SPG (including the Nordic Seas, Drijfhout et al. 2025). This interaction between collapsing deep mixing and a collapsing AMOC is a clear example …

In the discussion, it is stated that the change in method is unlikely to explain the increased number of SNS (line 562). Could the authors add what the alternative/more likely explanation is (CMIP5/CMIP6 model differences/updated scenarios?)?

> The increased number of SNS in these cases are due to a greater sensitivity of the AMOC to SSP scenarios than previously to RCP scenarios. Part of this may be explained to the larger climate sensitivity in some CMIP6 climate models (Zelinka et al https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019GL085782), although, even corrected for this the AMOC remains more sensitive (Mecking and Drijfhout, 2023). Another reason is the possibly too large sensitivity to the (removal of) aerosol forcing (Menary et al., 2020), in a large suite of models, although this then also should apply to the mixed-layer collapse in the deep convection regions. We are nevertheless hesitant to argue that in CMIP6 AMOC sensitivity to changing forcing is too high. First no extra meltwater from a decreasing Greenland Ice Sheet is accounted for in CMIP6 models. In addition, models with less bias, in particular in the South Atlantic salinity structure, show stronger AMOC decline (Portmann et al., 2026). Note that these SNS events preferably show in model simulations extended to year 2300 or beyond, and that in CMIP5 fewer model runs were extended beyond year 2100. The larger number of SNS events in sea-ice collapse is also partly associated with the larger number of extended runs, but also to the decreased bias in sea-ice thickness over the historical period in CMIP6 (Notz et al., 2020).

The mechanisms stated in the paragraph at lines 605 to 612 need references.

> We added the following citations:

Eastward extension Ross gyre: Holland et al., 2019; Naughten and al. 2023.

Low-frequency changes in sea-ice and deep mixing: Rheinlænder et al., 2021; Mchedlishvili et al., 2022.

The reference to Portmann (… Line 905) is inaccessible since no preprint is available. Please remove this reference unless it is available at the time of publishing.

> It is now available and full citation is included https://www.science.org/doi/10.1126/sciadv.adx4298

Technical corrections

- Line 261: missing verb, what happens to the mixed layer depth in this model?

- Line 343: reference to figure 7 incorrectly formatted

- Line 544: substitute “appears” with “is” since the statement is a fact.

- Last two sentences of section 2.1 should be swapped.
- The colours of the contourlines in Figure A2 are too similar, making it hard to distinguish them.

> We have implemented all listed technical corrections.

REVIEWER 3

Suggestions for revision or reasons for rejection

(visible to the public if the article is accepted and published)

The manuscript "Catalogue of Strong Nonlinear Surprises in ocean, sea-ice, and atmospheric variables in CMIP6" by Angevaare and Drijfhout. I think the paper is one of the most comprehensive studies o f strong nonlinear surprises (SNS) in CMIP6 data (although biosphere dynamics are left out from this work). I think the paper has much improved from the last round of revisions. Particular merits are that the paper discusses carefully which models show SNS under which conditions, whether they are counted as state transitions or abrupt shift. The paper also tries to disentangle (a bit lengthy though) how these definitions might be different to a tipping point elaborated in a careful discussion at the beginning of the article.

Overall, the paper can be recommended for publication in Earth System Dynamics with the following minor revisions in my view:

Minor comments:

1. L 29: At several occurrences in the paper quotation marks are used for stressing terms such as the "tipping points controversy", "abrupt shifts". I think these emphases should rather be explained (if they are needed at all) as not every reader will know what the tipping points controversy, etc. are.

> We removed the quotation marks for the three cases where this occurs. We also included the Kopp et al reference after the phrase tipping point controversy.

2. L 56 (and further): I am not sure I agree that I would call state changes that are observed in the CMIP6 datasets are really "surprising" as the authors state. Instead, I would expect such abrupt changes or state transitions as temperatures increase. However, I agree that SNS is an acceptable name.

> We have reworded these sentences as we can agree with the viewpoint that certain state transitions that go beyond a linear response to the forcing are not a real surprise if we know they are governed by nonlinear dynamics and can show tipping behavior.

We changed L53:

... and to detect large changes in climate variables that go beyond a linear response to the forcing and are governed by nonlinear dynamics associated with strong internal feedbacks. We deleted " With the word truly we mean that state".

3. L 143: I did not understand why the authors used "linearly spaced between 8x and 2.5x in 61 steps". Has this been explained in the appendix? If not, please add half a sentence.

> The threshold sequence is analogous to the example in Appendix A1. We added a short explanatory sentence and included a footnote link to the code documentation where the method is illustrated.

4. L 154: “Ab2: Abrupt change at the end of the dataset” – Can the authors explain how important this category is, i.e. how often abrupt changes occur at the end of the dataset (therefore, briefly outlining why it is important to include this category) as opposed to abrupt changes that do not occur at the end of the dataset.
> This is important because substantial changes may occur at the end of the dataset, which would otherwise not be found by criterion Ab1. Just one example is the panel Ab2 in figure 2. This category mainly detects changes that are not found by criterion Ab1, as explained in section A2.2. This category is a compensation mechanism for those cases that would otherwise be missed in category Ab1. There are relatively few cases that are identified by category Ab2.

5. Figure 2: Add panel labels (a, b, c, ...) – in particular because later on in the article it is referred to such panels.
> We agree, and added this to the figure.

6. Figures 3: Can the authors briefly explain why they don’t show values below 1.0°C of global warming and above 4.5°C? I assume it is because not many models include these temperatures. Might be good to briefly state in the caption
> The colour scale was chosen to resolve the range where most SNS occur, while retaining triangular extensions for out-of-range values below 1,${}^\circ$C and above 4.5,${}^\circ$C. The caption now states:
The triangular extensions at the ends of the colour scale indicate out-of-range values: $<1{}^\circ{}\text{C}$ at the lower end and $>4.5{}^\circ{}\text{C}$ at the upper end.

7. L 508: “panel iii” – this panel does not exist in Figure 2
> We implemented this fix.

8. L 574: “SSP126 scenario that is in agreement with the Paris Agreements” – scenario that fulfills the targets set in the Paris Agreement.
> We changed the wording as suggested by the reviewer.