Replies to Reviewer #1:

We sincerely appreciate your valuable comments and the constructive feedbacks. The manuscript has been carefully revised according to your minor comments.

The authors have made impressive and substantial revisions that address my concerns and I believe substantially strengthen the paper. In particular, the use of their earlier box model (to perform mechanism denial experiments and provide an analytical framework) is now much more effective, and they have situated their MCO mechanism in the context of other modes of variability found in the literature much better. The summary of the mechanism in conjunction with Fig. 13 is also substantially clearer. I have a few minor comments about points of confusion/incomplete logic, but otherwise I recommend this paper for publication.

1. *I appreciate the effort to clarify the language around salinity anomalies vs. lead-lag regression coefficients (with respect to Figures 4,5,6, etc.), however the authors are still missing the key language/concept that would connect lead-lag regression coefficients to actually salinity anomalies: that if they define a certain phase of the AMOC LFC1 to be centered at lag = 0, then the lead-lag coefficients can be interpreted as salinity anomalies. In their case specifically, I would recommend that they add language to lines 310-314 saying that, for conciseness they consider the case where peak AMOC strength occurs at lag = 0, and that therefore negative regression coefficients at lag = -200 can be thought of as negative salinity anomalies preceding a strong AMOC.*

Responses: Thank you very much for this suggestion. “Lag -200 and 0 years can be regarded as the negative and positive peaks of the AMOC” has been added in lines 320-321 of the revised manuscript.

2. *The mechanism denial experiments are a great addition, however, the authors should state what they actually do when they "deactivate" a given term in the circulation. Do they remove the term entirely from the model, or do they set that term fixed to some mean/initial value? The latter is what would make more sense, as that would prevent the given term (anomalous advection of mean salinity or mean advection of anomalous salinity) from feeding back on the AMOC.*

Responses: Thank you very much for this comment. As the box model is a linearized model, the mean values for $q'(\overline{S_1} - \overline{S_2})$ and $\overline{q}(S_1' - S_2')$ are both 0. Therefore, both approaches lead to the same results.

3. *I understand now that the freshening of the intermediate subpolar waters comes from the anomalously northward WBC advecting fresh water from the equatorial region. However, why does a strong AMOC lead to this northward WBC anomaly? The authors should add a sentence explaining why a WBC anomaly develops from a strengthened AMOC in order for this to be a "feedback" (i.e., so far they have shown that intermediate subtropical salinity affects AMOC strength once it affects the subpolar surface/intermediate salinities, but they should then explain why AMOC strength in turn affects intermediate salinity through WBC anomalies for this to be a complete feedback cycle).*

Responses: Thank you very much for this comment. The equatorial WBC is part of the AMOC upper branch. This statement has been added in line 455 of the revised manuscript. A stronger (weaker) than usual AMOC is always accompanied with a northward (southward) WBC anomaly, rather than an AMOC anomaly leads to a WBC anomaly through certain feedbacks.

4. *Line 314 it is not clear what is meant by "which can be rough to some extent".* This part of the sentence has been deleted.

5. *Line 456: do you mean "moves the negative salinity anomaly northward" instead of "removes negative salinity anomaly northward"?* Revised.

Replies to Reviewer #2:

We sincerely appreciate your valuable comments and the constructive feedbacks. The manuscript has been carefully revised according to your minor comments.

I have very much appreciated the tuning on the box model on the full ocean model.

1. *I have a single remaining minor issue with the equations of the box model given in the Appendix where (lines 732 and 733) some arguments on the parameterization of* k_m *as* $\kappa q'^2$ *should be given: from a physical point of view, I would expect that* k_m *only depends on* S_2 *and* S_3*, whereas* q' *depends on all salinities. That would deserve some explanations.*

Responses: Thank you very much for this comment. The parameterization of k_m is adopted from LY22. This statement has been added in line 746 of the revised manuscript. LY22 reveals that, the subpolar salinity mixing is in cubic relation ($F_{mixing} \sim -\Delta S'^3$) with the subpolar vertical salinity difference (Fig. R1a). Additionally, the subpolar vertical salinity difference (Fig. R2a) is positively correlated to the AMOC anomaly (Fig. R3a). Therefore, LY22 adopted the parameterization of $F_{mixing} \sim -\Delta S'^3 \sim -q'^2 \Delta S' \sim -k_m \Delta S'$.

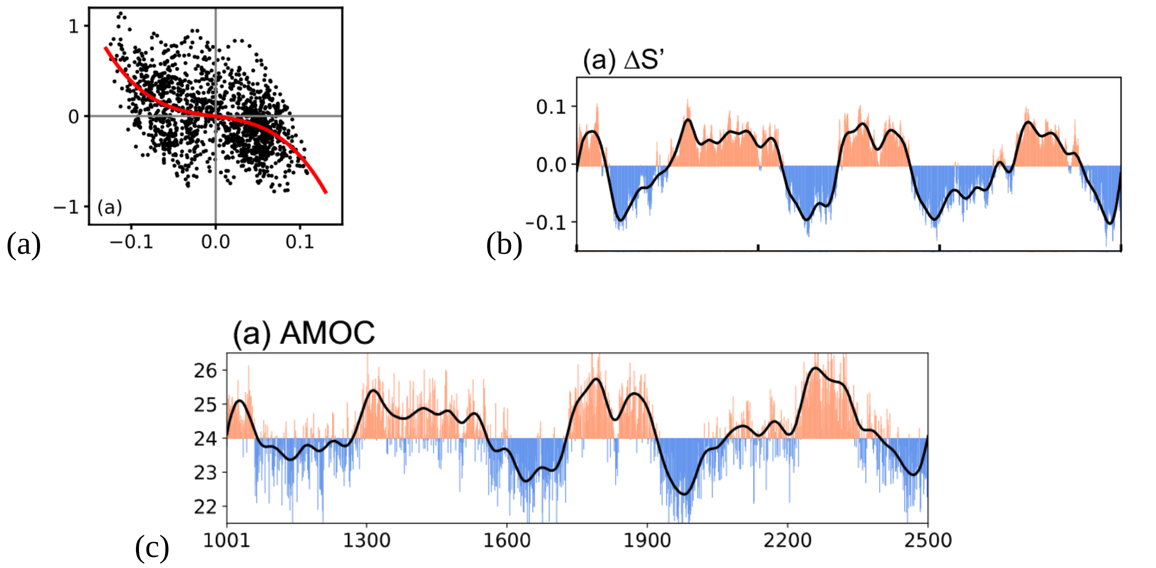

FIG. R1. (a) Figure 3a of LY22, the y-axis represents salinity mixing (units: psu Sv^{-1}), the x-axis represents salinity difference between subpolar 0-1000 m upper ocean and ocean below 1000 m (units: psu). (b) Figure 2a of LY22, the y-axis represents salinity difference between subpolar 0-1000 m upper ocean and ocean below 1000 m. (c) Figure 1a of LY22, representing the AMOC index (units: Sv). This simulation analyzed is the same as the one in the manuscript.

2. *Regarding the figures, the different panel organization between the 4x3 panels of Fig. 4 and the 3x4 panels of the following ones (6 7 8 9 10) could be avoided, unless there are some reasons to do so.*

Responses: Thank you very much for this comment. As the subplots of Fig. 4 are much wider than those of Figs. 6-10, we would like to choose the 4×3 layout for Fig. 4. Figure 4 is for meridional-depth sections from 80°S to 80°N, while Figs. 6-10 are for horizontal patterns with narrow zonal scale. Figure R2 is the 3×4 layout version of Fig. 4, which does not look as clear as the 4×3 layout.

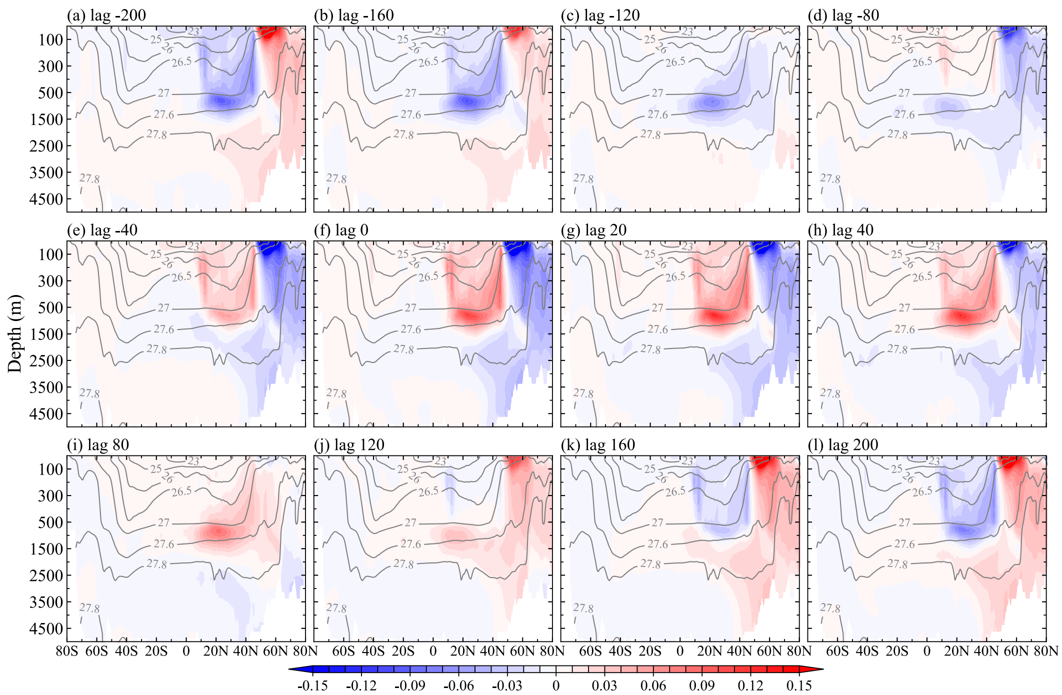

FIG. R2. A 3×4 layout version of Fig. 4.

3. *As a minor comment, some care should be given to the writing of Eqs. (1, 2) and the following paragraph to slightly separate the ρ u (v) dz and dy (dx).* A dot has been added in between.