Replies to Reviewer #1:

We sincerely appreciate your valuable comments and the constructive feedbacks. Our manuscript has been revised in accordance with your recommendations. The subsequent sections include our detailed, point-by-point responses.

Summary: In this study, the authors identify the characteristics of the primary low-frequency mode of the AMOC MCO and analyze them in a GCM to propose a new mechanism for the AMOC MCO that is driven mostly by processes in the North Atlantic. Their analysis focuses mostly on salinity anomalies that both drive and respond to changes in the overturning circulation. Using lead-lag regressions, they identify the northward advection of salinities anomalies in the upper ocean by the perturbed circulation to be a key positive feedback enhancing a given phase of AMOC, and advection by the mean circulation of salinity anomalies of the opposite sign in the subtropic intermediate waters to be a key negative feedback contributing to a phase change of the AMOC.

Overall, this is an interesting study that proposes a new mechanism and analyzes it in great detail in their model, and build a solid 3-dimensional picture of the feedbacks at play in the AMOC MCO. My main comments pertain to the usefulness of the box model in providing theoretical support for their mechanism, clarity around discussion of their key figures and the subpolar-to-subtropical feedback direction of their mechanism, and the implication of this study for alternate proposed AMOC MOC mechanisms.

Responses: Thank you very much for your valuable suggestions, which help us improve the manuscript tremendously. Considering the comments from all the reviewers, we revised the manuscript primarily in the following aspects:

1) The rationale for employing the theoretical model is revised. Instead of using the coupled model to “validate” the theoretical model, now the theoretical model is used to perform sensitivity (or, denial) experiment to examine the indispensability of the anomalous and mean advections in this North Atlantic Ocean-originated AMOC multicentennial oscillation (MCO). Accordingly, section 5b and the appendix have undergone significant revisions.

2) We computed freshwater transport to quantify the contributions of different processes to the subpolar salinity anomaly. As salinity transport across a section cannot reflect the actual effect on salinity change/tendency of the subpolar region, it is no longer adopted. Accordingly, section 5a is rewritten.

3) The evolution of salinity anomalies, especially the subtropical intermediate salinity anomaly, is depicted and analyzed in more detail. This corresponds to section 4b and the schematic Fig. 13.

4) Difference between this North Atlantic Ocean-originated AMOC MCO and other AMOC MCOs, as well as the implication of this study on interpretating other AMOC MCOs, are delineated in section 6.

5) The necessary information regarding the introduction section and illustration on the theoretical model are complemented. Now the revised manuscript is more standalone and it is no longer mandatory for readers to refer to Li and Yang (2022) (LY22) for the basic information. This corresponds to section 1 and the appendix.

Major Comments:

1. *While I understand that the paper builds on previous work by the same authors using a box model, the paper should still be able to be read as a standalone one without having to refer to that paper. To that end, I think the authors need to add more description in the Introduction of how their box model (in previous work) identified the North Atlantic paradigm for the AMOC MCO, and whether their box model had the capability to test/identify the other proposed AMOC MCO mechanisms (Southern Ocean paradigm, Arctic Ocean paradigm, etc.).*

Responses: Thank you very much for this comment. In lines 128-136 of the revised manuscript, we added the information: "*The theoretical box model in LY22 is inspired by another theoretical box model proposed by Griffies and Tziperman (1995) (hereafter GT95), which focuses on AMOC multidecadal oscillation. Both theoretical models only include processes in the North Atlantic, and are therefore unable to exhibit AMOC oscillations related to other ocean basins. In the GT95 theoretical model, the depth of the upper boxes is set at only 300 m. In addition, the volume of the subpolar boxes is set to only 1/11 of the North Atlantic, which is too small as the latitude for deep water formation (DWF) spans from approximately 50°N to 75°N. LY22 adopted new model parameters including thicker upper boxes and larger subpolar boxes, and found that the model exhibits AMOC MCO.*" LY22 is based on GT95 but adopted more realistic model parameters and included a necessary nonlinear subpolar vertical mixing. Both LY22 and GT95 included processes in the North Atlantic only and hence focused on North Atlantic Ocean-originated AMOC oscillations, while LY22 is on multicentennial timescale and GT95 is on multidecadal timescale.

Under the new parameters and physics, a self-sustained AMOC MCO is realized in LY22, that is how we identified the North Atlantic Ocean-originated AMOC MCO in LY22.

As the LY22 model includes only North Atlantic processes, it cannot test or identify AMOC modes related to the Arctic Ocean, Southern Ocean, and other basins in its original form. However, processes related to other basins can be appended to the LY22 model to make it capable of analyzing other modes of the AMOC oscillations. For instance, Wei and Zhang (2022) included a negative feedback originated from the Arctic Ocean in the two-box model raised by Stommel (1961) (not the LY22 model) whose essence is also the advection in the subtropical-subpolar North Atlantic. This way, Wei and Zhang (2022) related their study to the Arctic Ocean-originated AMOC multidecadal and multicentennial oscillations.

Additionally, we have enriched our reviews on AMOC MCO studies in the introduction to make this manuscript more “standalone”. Sentences like “*Please refer to LY22 and YY23 and references therein…*” are deleted.

References:

Stommel, H., 1961: Thermohaline convection with two stable regimes of flow. *Tellus*, **13**, 224-230, https://doi.org/10.1111/j.2153-3490.1961.tb00079.x

Griffies, S. M., and E. Tziperman, 1995: A linear thermohaline oscillator driven by stochastic atmospheric forcing. *J. Climate*, **8**, 2440-2453, https://doi.org/10.1175/1520-0442(1995)008<2440:ALTODB>2.0.CO;2

Li, Y., and H. J. Yang, 2022: A theory for self-sustained multicentennial oscillation of the Atlantic meridional overturning circulation. *J. Climate*, **35**, 5883-5896, https://doi.org/10.1175/JCLI-D-21-0685.1

Wei, X., and R. Zhang, 2022: A simple conceptual model for the self-sustained multidecadal AMOC variability. *Geophys. Res. Lett.*, **49**, 11, https://doi.org/10.1029/2022GL099800

2. *The authors emphasize in the introduction the importance of using box models to provide a theoretical basis for studies using more complex/realistic models. However, it is not clear to me in this particular case what additional understanding the use of the box model adds to the study. The authors' analysis of the GCM in this study provides a plausible mechanistic understanding*

of AMOC MCO through careful analysis of lead-lag relationships between salinity anomalies and phases of the AMOC MCO and separation of contributions from mean and perturbed circulations. Simply showing that the results here are consistent with those from the box model does not seem to me to contribute to the 'theoretical' understanding of the AMOC MCO. The analysis in Fig. 12 seems to take inspiration from the box model (in plotting similar quantities to those that were analyzed in the box model), but it does not, in my view, add further "proof" that the AMOC MCO is in fact caused by the mechanism proposed; it merely shows that those quantities are in/out of phase with AMOC. Overall, I think the use of the box model needs to be further justified, specifically what understanding is gained from the box model that could not have been deduced from the GCM analysis alone. Alternatively, if the box model was merely useful in generating the North Atlantic paradigm theory, that is okay too, but the paper should acknowledge that.

Responses: Thank you very much for your quite insightful comments on the use of theoretical models in your major comments 2 and 3. We have revised the relation between theoretical model and coupled model results throughout the revised manuscript. Specifically, theoretical studies (not limited to LY22) provide theoretical perspectives for analysis on coupled model results, as we have stated in lines 71-75 of the revised manuscript: "*Under the linear framework, the single-equilibrium oscillation is perceived as an anomaly hovering around the unstable equilibrium. Positive and negative feedbacks enhance and weaken the anomaly, collectively leading to the anomaly's phase transition and therefore its cyclic evolution. This forms our foundational "theoretical interpretation" of linear oscillation, by which we are inspired to review the aforementioned coupled model studies on AMOC MCO.*" The existing coupled model studies on AMOC MCO have not adopted such theoretical view in their analyses, or more specifically, they have not regarded their AMOC MCO as an oscillatory system (a dynamic system/problem). For instance, Delworth and Zeng (2012), Jiang et al. (2021), and Meccia et al. (2023) all emphasized the mean advection as the process governing the entire evolution, although mean advection is a weakening process for the AMOC anomaly and there should be positive feedback remains to be found. This reflects a lack of theoretical/dynamic perspective in their analyses.

Additionally, we no longer attempt to "verify" the theory using coupled model results. On the contrary, we regarded the theory as a condensation of the coupled model mechanism. By conducting sensitivity/denial experiments utilizing the theoretical model, it can serve our interpretation on the modeled AMOC MCO and even real-world AMOC multicentennial variability (MCV). Therefore, we replaced the original Fig. 12 with a denial experiment: artificially deactivating the anomalous advection or mean advection in the upper AMOC branch in the theoretical model of LY22. The results suggest that, the self-sustained AMOC MCO in the LY22 model can only exist when both the anomalous and mean advections in the upper branch (the two paramount processes in the CESM1 AMOC MCO) are active. This further proves the significance of these two processes in the North Atlantic Ocean-originated AMOC MCO.

References:

Delworth, T. L., and F. R. Zeng, 2012: Multicentennial variability of the Atlantic meridional overturning circulation and its climatic influence in a 4000 year simulation of the GFDL CM2.1 climate model. *Geophys. Res. Lett.*, **39**, https://doi.org/10.1029/2012GL052107

Jiang, W. M., G. Gastineau, and F. Codron, 2021: Multicentennial variability driven by salinity exchanges between the Atlantic and the Arctic Ocean in a coupled climate model. J. Adv. Model. Earth Syst., 13, e2020MS002366, https://doi.org/10.1029/2020MS002366

Meccia, V. L., R. Fuentes-Franco, P. Davini, K. Bellomo, F. Fabiano, S. T. Yang, and J. von Hardenberg, 2023: Internal multi-centennial variability of the Atlantic meridional overturning circulation simulated by EC-Earth3. Climate Dyn., 18, https://doi.org/10.1007/s00382-022-06534-4

3. *Relatedly, the above question could be (partially) addressed in lines 121-128 if the authors could be more specific about what "theoretical support" they are looking for and explain how their box model would provide such support. The GCM studies (including this one) do offer mechanistic explanations, though they are perhaps not 'causal' because they cannot say that the AMOC MCO would be destroyed without them. For example, do you hope to show that the proposed mechanism is a necessary ingredient for the AMOC MCO by showing in simple models that you can't get an MCO without it? In my view, the main advantage of using a box model is in establishing causal relationships in the mechanisms proposed by performing mechanism denial experiments (for example, suppressing circulation perturbations in the upper*

two boxes but allowing all other quantities to change—do you still get an MCO?), but the authors don't use the box model in any such way here.

Responses: Thank you very much again for your comments on the use of theoretical model. The old lines 121-128 are removed. In the revised manuscript, we viewed the LY22 theoretical model as a condensation of the coupled model mechanism, and performed denial experiments using it to examine the indispensability of the anomalous and mean advections. We no longer try to verify the theory using the coupled model results. “Theoretical support” is a theoretical perspective for analyzing the coupled model AMOC MCO, that is, to regard it as a linear oscillation, and look for the positive feedback and negative feedback that enhance and weaken the anomalies, enable the phase transition and thus the full cycle. Such theoretical perspective is inspired by the enormous theoretical studies that regard the coupled model results or real-world phenomena as dynamic systems, rather than being solely inspired by the study of LY22 alone. However, now the LY22 model also directly supports our analysis of the coupled model results, through the denial experiments that further underscore the importance of anomalous and mean advections.

4. *Throughout section 4, the authors refer to the quantities in Figs. 4, 5, 6, etc as "salinity anomalies", using language like "the negative salinity anomaly evolves…". However, the shading in those figures do not show actual salinity anomalies, but regression coefficients between salinity anomalies and the AMOC LFC1. Of course, one can relate regression coefficients to the absolute values of the salinity anomalies themselves if one picks a specific phase of the AMOC MCO to focus on. For example, if we focus on the phase when the AMOC is at its peak strength at year 0, then indeed strong negative regression coefficients at lag -200 years can be interpreted as negative salinity anomalies preceding a strong AMOC. The authors need to make it clear that this is what they are doing in section 4, or else a lot of their language leads to confusion. For example, in line 309, the authors say "...positive salinity anomalies gradually develop… which mirrors the similar evolution of the AMOC." This sentence confuses the reader into thinking we are looking directly at salinity anomalies in Fig. 4, and that we should compare them to some depicted evolution of AMOC, when in fact Fig. 4 is precisely showing salinity anomalies and AMOC changing together because it depicts lead-lag regression coefficients. Similarly, in lines 293-294, "The salinity anomalies in the South Atlantic are weak, having negligible contributions to salinity variation in the North Atlantic" is*

misleading—there could be strong salinity anomalies in the SA, but what figure 4 shows us is that they have low correlation with AMOC LFC1. This language throughout all of section 4 needs to be modified for clarity.

Responses: Thank you very much for this comment. We have added clarification in lines 309-313 of the revised manuscript as: "*Positive and negative regression coefficients at lag n years represent that generally there are positive and negative salinity anomalies in the corresponding regions, respectively, when salinity anomalies lag the AMOC LFC1 by n years. For conciseness, positive/negative salinity anomaly is used to represent positive/negative regression coefficient, which can be rough to some extent,*" to indicate that what Fig. 4 (and also Figs. 5, 6, 8, and 10) depicts are lead/lag regression coefficients instead of actual salinity anomalies, and the use of "salinity anomaly" instead of "regression coefficient" in section 4 is for conciseness. Additionally, in line 329 of the revised manuscript, we rephrased the old sentence "*The salinity anomalies in the South Atlantic are weak*" to be "*In the South Atlantic, salinity anomalies do not reflect evolution synchronous with the AMOC*" to avoid misleading.

However, the actual salinity anomalies in the South Atlantic are still much weaker than those in the North Atlantic (Fig. R1). With the aforementioned refinements and clarifications, we suppose that the expressions in the revised manuscript will not lead to confusion and imprecision.

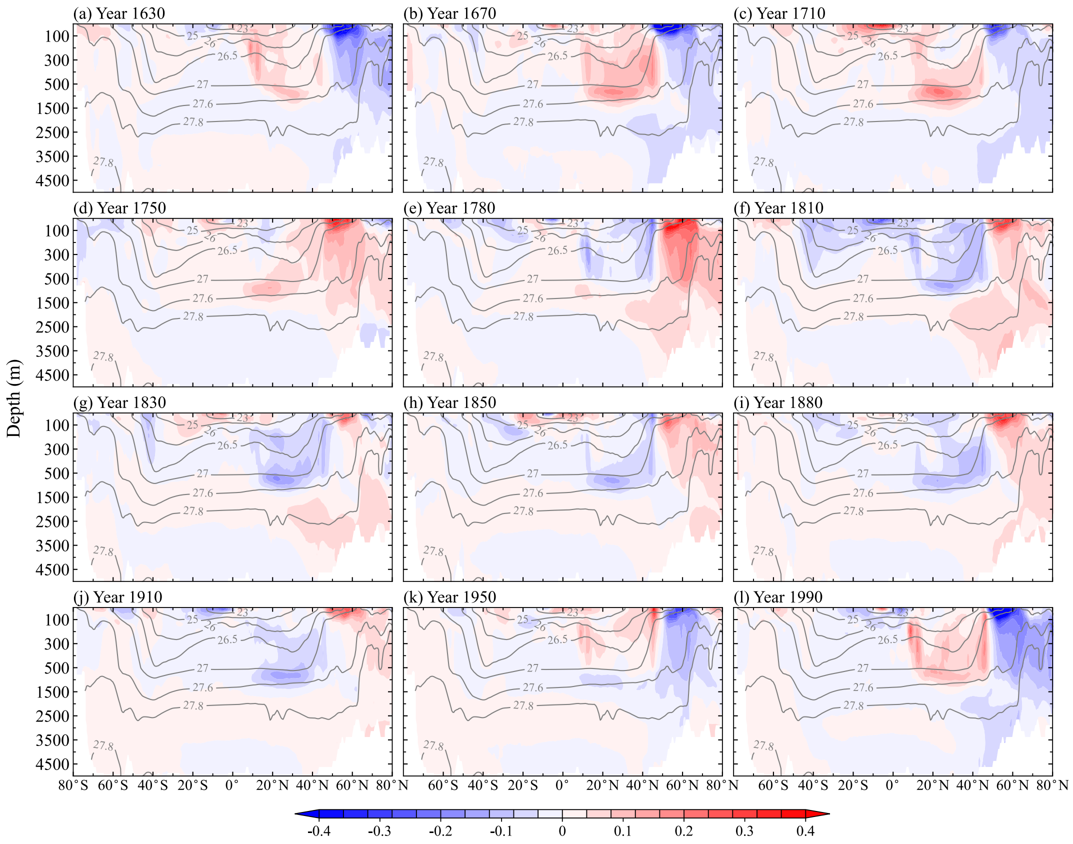

FIG. R1. Zonally averaged actual salinity anomalies in the Atlantic (shading; units: psu) during years 1630-1990. Contours show the zonally averaged climatological potential density σ_θ in the Atlantic (units: kg m^{-3}).

5. *The authors do a good job of explaining the mechanisms in the North Atlantic that enhance AMOC anomalies, i.e. the discussion of Figures 4-7, how the subtropics drive changes in the subpolar region through the upper branch of the AMOC. However, I feel the explanation of how changes in the North Atlantic/subpolar eventually cause changes in the subtropics was less clear. Specifically, what drives the phase change of the salinity anomalies in the subtropical intermediate waters? This seems key to explain the full feedback cycle that allows oscillations. During the discussion of Fig. 9, I thought this was because AMOC strength affects the subtropical circulation strength, which in turn changes the advection of the mean salinity into the subtropics and modifies the salinity there. However, in the summary paragraph (subsection c), I had the impression that the feedback from the North Atlantic to the subtropics was more through the recirculation of the sinking North Atlantic salinity anomaly by the mean DWBC, in other words that the fresh anomaly in the subpolar region associated with the weak AMOC gets advected downward and southward until it moves upward in the subtropics, causing a negative salinity anomaly there by the time AMOC has entered a strong phase. Overall, this part of the full AMOC MCO feedback cycle (connection from the deep subpolar to the intermediate subtropics) needs some clarification.*

Responses: Thank you very much for this comment. Salinity anomaly descended from the DWF region hardly has effect on the subtropical intermediate salinity anomaly. In lines 444-458 (section 4b), lines 478-501 (the summarizing section 4c), and the schematic Fig. 13 of the revised manuscript, we have explained the evolution mechanism for the subtropical intermediate salinity anomaly more in detail.

Instead of calculating salinity transport as in the original manuscript, in the revised manuscript we employed freshwater mass transport/flux (computation method is introduced in lines 196-213 of the revised manuscript), which is more physical as the salinity transport cannot represent the actual change of salinity in the study area. Take the theoretical model of LY22 for instance, freshwater transport into subpolar upper box 2 is proportional to $-q(S_1 - S_2)$, signifying the actual salinity tendency of the subpolar upper ocean. In contrast, salinity transport represents qS_1, which fails to represent the change of S_2. The velocity used for freshwater transport computation is the total velocity that consists of the Eulerian-mean and eddy-induced velocities.

We conducted freshwater budget analysis (Fig. R2) for the 10°-35°N, 0-1000 m subtropical North Atlantic where the subtropical intermediate salinity anomaly evolves. It is revealed that the major contributions come from liquid freshwater transports at 10°N (Fig. R2a, black curve) and 35°N (Fig. R2a, orange curve) boundaries. Surface freshwater flux (Fig. R2a, green curve) is weak; freshwater transport at the 1000 m bottom is nearly negligible (Fig. R2a, red curve). Therefore, it is the upper-intermediate ocean transports instead of the deep ocean or surface that affects salinity anomaly in the 10°-35°N, 0-1000 m subtropics.

At the AMOC minimum (Fig. R3a), the subtropical intermediate salinity anomaly is at its positive peak. Given the positive salinity anomalies at the 35°N boundary, the northward mean Gulf Stream carries these positive salinity anomalies out of the subtropics. Therefore, this mean advection process increases freshwater transport into the subtropics (Fig. R2b, dashed orange curve at lag -200 years) and weakens the subtropical positive salinity anomaly. Salinity anomalies at the 10°N boundary are positive as well. Because the mean current here is northward but slower, mean advection at the 10°N boundary decreases freshwater transport into the subtropics (Fig. R2b, dashed black curve at lag -200 years) and enhances the positive subtropical salinity anomaly, but is much weaker than that of 35°N. Mean salinity on the western flank of the 35°N boundary is close to the spatially averaged climatological salinity of the 10°-35°N, 0-1000 m subtropics (Fig. R4), rendering the effect of anomalous advection therein minimal. Mean salinity on the eastern side of the 35°N boundary is higher. The weak but southward anomalous advection therein (Fig. R4a) decreases freshwater transport into the subtropics (Fig. R2b, solid orange curve at lag -200 years), thereby enhances the positive subtropical salinity anomaly. At the 10°N boundary, the southeastward anomaly of the equatorial western boundary current (WBC; Fig. R4a) decreases the northward transport of equatorial freshwater (Fig. R2b, solid black curve at lag -200 years), becoming the strongest process enhancing the positive subtropical salinity anomaly. Now the weakening processes for the positive subtropical salinity anomaly are stronger than the enhancing processes, causing the positive subtropical salinity anomaly to weaken (Figs. R3a-d) then transition to negative (Fig. R3e).

Subsequently, mean advection at the 35°N boundary transports negative salinity anomalies northward away from the subtropics (Fig. R3e), reducing freshwater transport into the subtropics (Fig. R2b, dashed orange curve at lag -40 years). The effect of mean advection at the 10°N boundary (Fig. R2b, dashed black curve at lag -40 years) is contrary to and much weaker than that of the 35°N boundary. The anomalous advection on the eastern flank of the 35°N boundary shifts northward (Fig. R4e), hence increasing freshwater transport into the subtropics (Fig. R2b, solid orange curve at lag -40 years). Because the equatorial WBC now exhibits a northward anomaly (Fig. R4e), more equatorial freshwater is advected into the subtropics at the 10°N boundary (Fig.

R2b, solid black curve at lag -40 years). At this stage, the enhancing processes for the negative subtropical salinity anomaly overcome the weakening processes, and the negative subtropical salinity anomaly gradually marches toward the negative maximum (Figs. R3e-g). Subsequently, the weakening processes again outweigh the enhancing processes, and the cycle enters the opposite phase.

As the salinity anomaly in the deep water formation (DWF) region dominates this AMOC MCO, the subtropical salinity anomaly is a component of the DWF region salinity anomaly's cycle. Despite its importance, the subtropical salinity anomaly remains secondary to the DWF region salinity anomaly, which is why we focused more on the latter in the revised manuscript. Consequently, in the revised manuscript which already has 15 figures including the appendix, Fig. R2 is not included, and we mainly stressed the anomalous advection along the equatorial WBC (Fig. R2b, solid black curve) and the mean advection along the Gulf Stream (Fig. R2b, dashed orange curve). They are the two major contributors to the subtropical salinity anomaly, and are more impactful than the anomalous subtropical circulation.

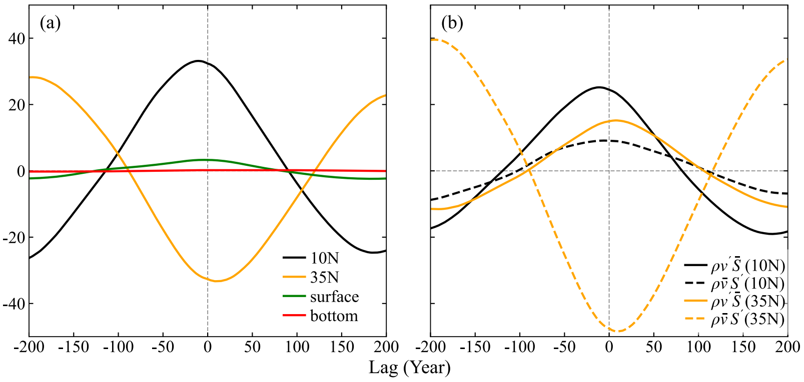

FIG. R2. Lead/lag regression coefficients of components constituting the freshwater mass budget of the 10°-35°N, 0-1000 m subtropical North Atlantic onto the AMOC LFC1 (units: 10^6 kg s^{-1}). (a) Total liquid freshwater mass transports into the given region, at 10°N (black curve) and 35°N (orange curve), respectively. Surface freshwater mass flux into the given region (green curve). Total liquid freshwater mass transport at the 1000 m bottom of the given region (red curve). (b) Liquid freshwater mass transports into the given region at the 10°N boundary (annotated in Fig. R3), induced by anomalous advection of mean salinity (solid black curve) and mean advection of salinity anomaly (dashed black curve). The orange curves are the same as the black curves, but for transports at the 35°N boundary. Negative lag means the AMOC LFC1 lags the freshwater terms (units: year).

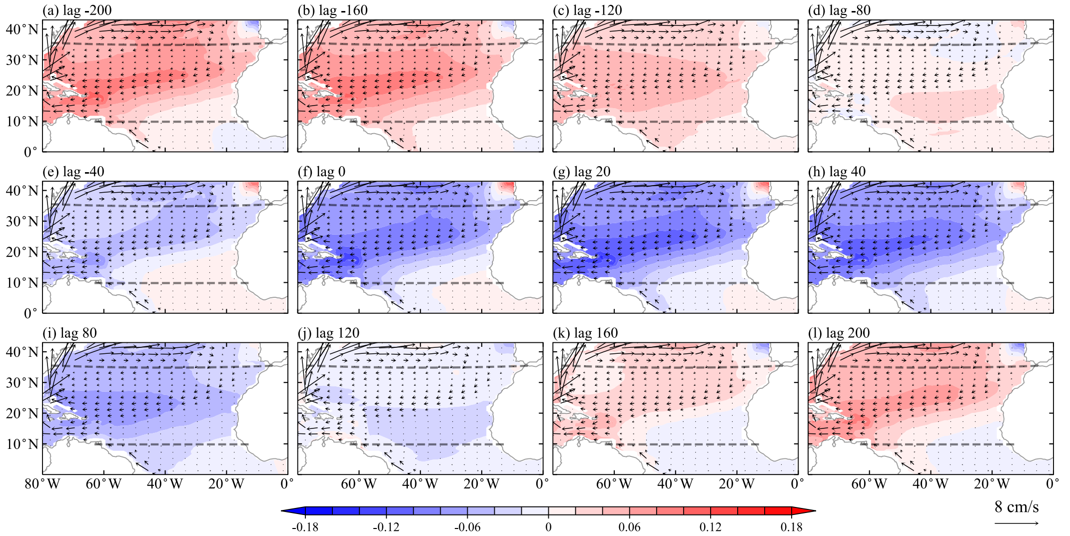

FIG. R3. Lead/lag regression coefficients of salinity anomalies averaged over 26.5-27.6 σ_θ on the AMOC LFC1 (units: psu), superimposed with climatological currents averaged over the same potential density range (vector; units: cm s^{-1}). Negative lag means the AMOC LFC1 lags the salinity anomalies (units: year). The 10°N and 35°N boundaries are annotated by the dashed gray curves.

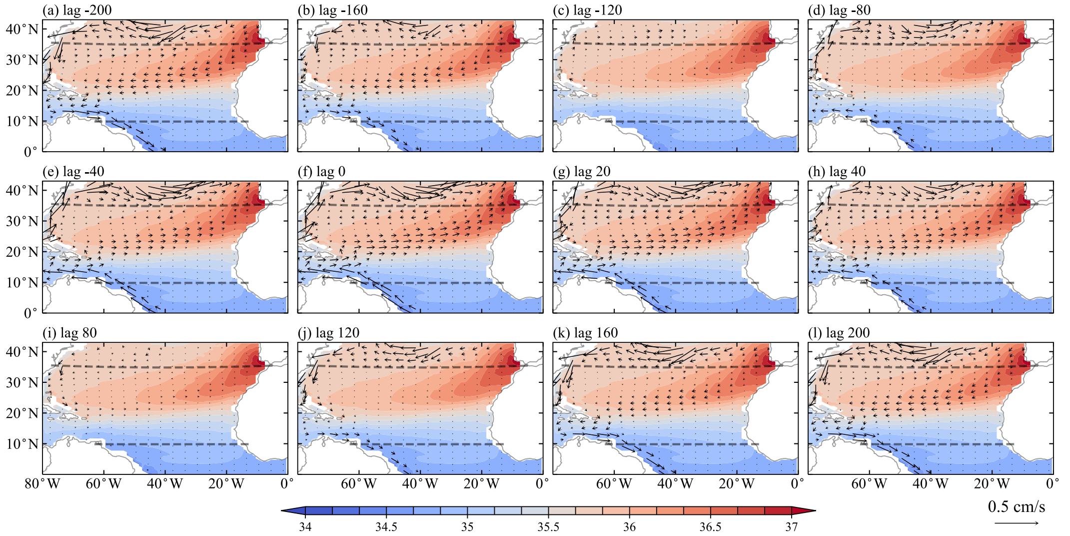

FIG. R4. Lead/lag regression coefficients of current anomalies averaged over 26.5-27.6 σ_θ on the AMOC LFC1 (units: cm s^{-1}), superimposed with climatological salinity averaged over the same potential density range (shading; units: psu). Negative lag means the AMOC LFC1 lags the current anomalies (units: year). This figure is the same as Fig. 9 of the revised manuscript, except that here the 10°N and 35°N boundaries are annotated by the dashed gray curves.

6. *This paper puts forward a convincing new theory for a mechanism driving the AMOC MCO. However, I think the impact of the paper would be much greater if they demonstrated what their findings imply for other theories of the AMOC MCO. For example, they mention that different mechanisms proposed by other studies are not necessarily contradictory, but could represent different modes of the AMOC MCO. Do you find evidence of those other modes/mechanisms here? If so, can you discuss how much of the low frequency behavior is explained by the North*

Atlantic paradigm and how much is explained by others? If not, why did other studies identify those modes—does it depend on the model used? How would one (hypothetically) test which of many proposed mechanisms is/are the "real" driver of the AMOC MCO? Could your box model or a more complicated one be used to do such hypothesis testing? Without some discussion on these points, this mechanism seems equally viable to others proposed for explaining the AMOC MCO.

Responses: Thank you very much for this comment. We have rewritten the section 6 (conclusion and discussion). In lines 653-667 of the revised manuscript, we discussed the difference between the North Atlantic Ocean-originated AMOC MCO and other AMOC MCOs as: "*The AMOC MCO analyzed in this study is North Atlantic Ocean-originated, differentiating it from the "flip-flop" AMOC MCO in Park and Latif (2008), the Southern Ocean-originated AMOC MCO in Delworth and Zeng (2012), and the Arctic Ocean-originated AMOC MCO in Jiang et al. (2021) and Meccia et al. (2023). The "flip-flop" AMOC MCO is a multi-equilibrium phenomenon that is markedly distinct from our study. The main difference between the Southern Ocean-originated AMOC MCO and our North Atlantic Ocean-originated AMOC MCO lies in the location of the salinity anomaly advected northward by mean advection. In their study, mean advection moves salinity anomaly in the upper Southern Ocean northward, whereas in ours, the northward salinity anomaly is located in the subtropical intermediate ocean. In Jiang et al. (2021) and Meccia et al. (2023), clear current and salinity anomalies are exhibited north of the subpolar North Atlantic, yet in our study the anomalies therein are rather weak, especially the current anomalies. Given the connection between salinity anomalies from the Arctic Ocean and the sea ice thermodynamics, distinctions between the Arctic Ocean-originated and North Atlantic Ocean-originated AMOC MCOs likely stem from the difference in both the ocean model and the sea ice model utilized.*"

Following, in lines 668-694 of the revised manuscript, we discussed the implication of this study on understanding other AMOC MCOs, as: "*Despite these differences, our study still aids the understanding of other AMOC MCOs. The positive salinity advection feedback in the subtropical-subpolar upper ocean is pivotal not only in this study, but also in the Southern Ocean-originated and Arctic Ocean-originated AMOC MCOs. In Fig. 5a of Delworth and Zeng (2012), the continuous northward salinity anomaly that is symbolic of mean advection reaches approximately 45 °N. North of 45 °N, salinity anomalies evolve nearly synchronously, mirroring the corresponding pattern in Fig. 5a of our current study. Therefore, this evolution of salinity anomaly is likely driven by the positive salinity advection feedback. Local salinity evolution north of 45 °N is also reflected in Fig. 5a in Jiang et al. (2021) and Fig. 5a in Meccia et al. (2023), suggesting that the positive salinity advection feedback is likely to be the enhancing process for AMOC anomaly in the Arctic*

Ocean-originated AMOC MCOs, which is not addressed by these two studies. The AMOC MCO in the intermediate-complexity model study of Mehling et al. (2023) has a similar mechanism to the Arctic Ocean-originated AMOC MCOs, and they employed a box model adapted from Stommel (1961) to explain their mechanism and emphasize the Arctic Ocean-originated salinity anomaly. Likewise, Wei and Zhang (2022) also utilized a revised Stommel's two-box model to account for the Arctic Ocean-originated AMOC multidecadal oscillation, whose essence is similar to that of the Arctic Ocean-originated AMOC MCOs. Both theoretical models employed in these two studies actually incorporate the positive salinity advection feedback and mean advection process in the subtropical-subpolar upper ocean, although their focus is salinity anomaly from the Arctic Ocean. Hence, the subtropical-subpolar positive salinity advection feedback likely serves as the essential enhancing process for AMOC anomaly in the North Atlantic Ocean-originated, Southern Ocean-originated, and Arctic Ocean-originated AMOC MCOs. The primary difference among them is perhaps the location of the salinity anomaly that is advected into the DWF region through mean advection. By incorporating additional boxes representing the South Atlantic/Southern Ocean, the theoretical model in LY22 can potentially account for the South Ocean-originated AMOC MCO through capturing salinity anomaly in the Southern Ocean." Figure 5a in the current study, in Delworth and Zeng (2012), in Jiang et al. (2021), and in Meccia et al. (2023) are presented in Fig. R5.

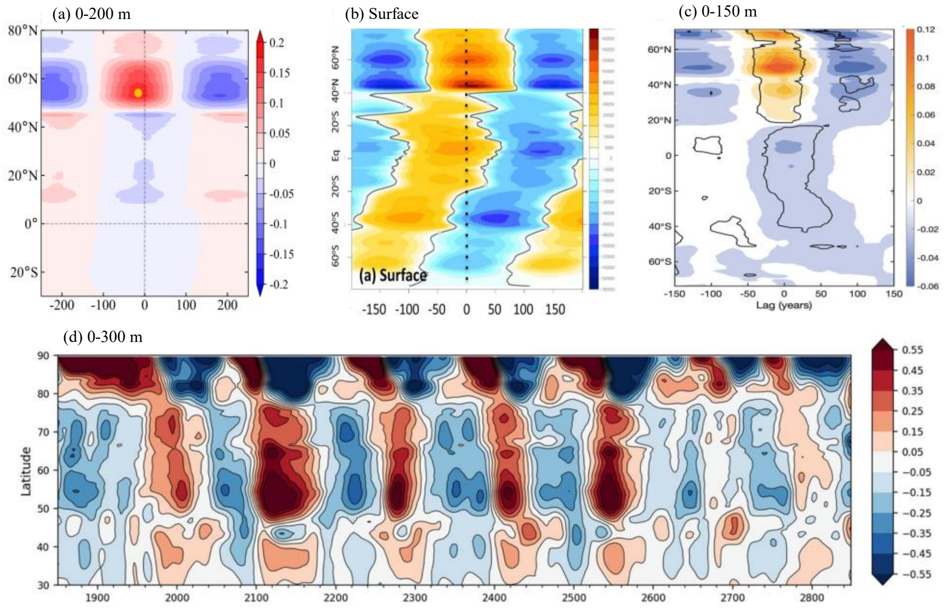

FIG. R5. Figure 5a in (a) the current study, (b) Delworth and Zeng (2012), (c) Jiang et al. (2021), and (d) Meccia et al. (2023), representing lead/lag regression coefficients of zonally and vertically averaged salinity anomalies in the Atlantic on the AMOC LFC1/AMOC index (units: psu/psu Sv^{-1}).

We have not found obvious evidence of other AMOC MCOs in the current study. As we have stated before, difference between these AMOC MCOs perhaps lies in the ocean and sea ice models utilized.

In reality, the variation of AMOC reflected by proxy data is better termed as multicentennial variability (MCV) instead of MCO, as variability is from an observation and statistical perspective but oscillation is a more physical term that is determined by a specific mode. Likewise, Liu (2012), Sutton et al. (2018), and Zhang et al. (2019) have also discussed such terminology on the Atlantic multidecadal variability. Compared to “AMO”, they suppose that it is better to use “AMV”, which might include several oscillations that are controlled by distinct modes and have their own unique periods. Because in each AMOC MCV coupled model study, the time series of the AMOC index exhibits a clear multicentennial period and can be interpretated as a linear oscillation controlled by positive and negative feedbacks, these modeled AMOC MCVs can be regarded as AMOC MCOs and might represent different modes of the real-world AMOC MCV.

The most convincible manner for testing whether the real-world AMOC MCV actually consists of several modes (AMOC MCOs), and whether the coupled models have simulated the modes (if exist in reality) rightfully, is through analysis on the direct observation instead of model results. Therefore, examining which AMOC MCO found in coupled models is true and contributes the most to the real-world AMOC MCV is obscured by the dearth of observation. We have added such discussion in lines 695-700 of the revised manuscript as: “*To the best of our knowledge, this study is perhaps the first to identify the North Atlantic Ocean-originated AMOC MCO in coupled models. Therefore, this mode will be more convincing if identified in other coupled models. However, the assessments of whether this North Atlantic Ocean-originated AMOC MCO and other previously proposed AMOC MCOs genuinely exist in the Earth’s climate system, as well as their relative contributions to the real-world AMOC MCV, are inhibited by the limited direct observations that are unfortunately unavailable in the foreseeable future.*”

However, our study still contributes to the understanding of other modeled AMOC MCOs, and the theoretical model in LY22 can be extended to account for the Southern Ocean-originated and Arctic Ocean-originated AMOC MCOs, through appending South Atlantic boxes or negative feedbacks from the Arctic Ocean. By including these additional elements, multiple types of AMOC MCOs can coexist in a theoretical model, and each of them might dominate under certain circumstance. Whether the hypothetical examination of their relative importance is true, could be highly dependent on the parameter setting and the parameterization of the physical processes.

References:

Stommel, H., 1961: Thermohaline convection with two stable regimes of flow. *Tellus*, **13**, 224-230, https://doi.org/10.1111/j.2153-3490.1961.tb00079.x

Delworth, T. L., and F. R. Zeng, 2012: Multicentennial variability of the Atlantic meridional overturning circulation and its climatic influence in a 4000 year simulation of the GFDL CM2.1 climate model. *Geophys. Res. Lett.*, **39**, https://doi.org/10.1029/2012GL052107

Liu, Z. Y., 2012: Dynamics of interdecadal climate variability: A historical perspective. *J. Climate*, **25**, 1963-1995, https://doi.org/10.1175/2011JCLI3980.1

Sutton, R. T., G. D. McCarthy, J. Robson, B. Sinha, A. T. Archibald, and L. J. Gray, 2018: Atlantic multidecadal variability and the U.K. ACSIS program. Bull. Amer. Meteor. Soc., 99, 415-425, https://doi.org/10.1175/BAMS-D-16-0266.1

Zhang, R., and Coauthors, 2019: A review of the role of the Atlantic meridional overturning circulation in Atlantic Multidecadal Variability and associated climate impacts. *Rev. Geophys.*, **57**, 316-375, https://doi.org/10.1029/2019RG000644

Jiang, W. M., G. Gastineau, and F. Codron, 2021: Multicentennial variability driven by salinity exchanges between the Atlantic and the Arctic Ocean in a coupled climate model. J. Adv. Model. Earth Syst., 13, e2020MS002366, https://doi.org/10.1029/2020MS002366

Wei, X., and R. Zhang, 2022: A simple conceptual model for the self-sustained multidecadal AMOC variability. *Geophys. Res. Lett.*, **49**, 11, https://doi.org/10.1029/2022GL099800

Meccia, V. L., R. Fuentes-Franco, P. Davini, K. Bellomo, F. Fabiano, S. T. Yang, and J. von Hardenberg, 2023: Internal multi-centennial variability of the Atlantic meridional overturning circulation simulated by EC-Earth3. Climate Dyn., 18, https://doi.org/10.1007/s00382-022-06534-4

7. *If the eddy-induced AMOC plays a role in the MCO (Fig. 1h), how do you expect your results to be affected by the fact that eddies are parameterized in such a model, and that the ocean model may be experiencing significant numerical diffusion? This may be a major caveat to the applicability of this study to understanding the real-world low frequency AMOC variability.*

Responses: Thank you very much for this comment. If the North Atlantic Ocean-originated AMOC MCO is a robust mode of the real-world AMOC MCV, it has a high possibility to exist even without the eddy activities, no matter parameterized or not. Figure 1h of the revised manuscript suggests that the eddy-induced parameterized AMOC always weakens the AMOC anomaly in this simulation. Analysis in LY22 suggested that, salinity tendency of the subpolar upper ocean induced by eddy-induced salinity mixing weakens the AMOC anomaly, and its magnitude is proportional to the product of (1) the salinity difference between the subpolar upper and deeper oceans and (2) the square of the AMOC anomaly. Therefore, the eddy-induced mixing is a nonlinear process. Appending this nonlinear process into the purely linear theoretical model, the system can exhibit self-sustained AMOC MCO, because the linear advection system now has the necessary nonlinearity. In short, the role of eddies in this North Atlantic Ocean-originated AMOC MCO is perhaps more of a source of nonlinearity. As nonlinearity should be present in the real world in various forms and not limited to the eddy-related processes, the sustainable North Atlantic Ocean-originated AMOC MCO is likely to exist in the real world even without considering the eddies.

8. *I would use different colormaps for the figures where you're showing lead-lag regression coefficients in colors (Figs. 4, 5, 6) and for the figures where you're showing climatological values in colors and regression coefficients as vectors (Figs. 7, 9), so that the reader can easily distinguish between these two different kinds of plots.*

Responses: Thank you very much for this suggestion. In the revised manuscript, we have changed the color scheme for climatological salinities in Figs. 3, 7, and 9, to distinguish them from the lead/lag regression coefficients in Figs. 4, 5, 6, 8, and 10.

Minor Comments:

1. *In the abstract line 34, it should be: "AMOC MCO more likely originates".*

Responses: Thank you very much for this comment. The abstract has been rewritten.

2. *Line 42: I don't believe Yang et al. (2023) actually appears in the reference list.*

Responses: Sorry for this careless mistake. Yang et al. (2023) is a manuscript of us submitted to *Journal of Climate*. As citing LY22 is also sufficient for the current study, we have deleted Yang et al. (2023) in the revised manuscript.

3. *Line 64: What does "theoretically" mean here?*

Responses: The introduction has been rewritten and this part no longer exists in the revised manuscript. However, we still regret any confusion the "theoretically" has caused, as we should have used "analytically" instead. The theoretical model in LY22 is a four-box model, and its eigenvalue can be computed numerically only. After simplifying the four-box model into a three-box model with less freedom, its eigenvalue can be calculated analytically, that is what we meant by "theoretically".

4. *Line 65: The term "perturbation advection" here and throughout the paper seems unusual and/or needs to be defined—maybe call it "advection by the perturbed circulation"?*

Responses: Thank you very much for this suggestion. We have rephrased "perturbation advection" as "anomalous advection" throughout the revised manuscript.

5. *Line 66: "dampened" should be "damped".*

Responses: Thank you very much for this suggestion. This part has been rewritten. Throughout the revised manuscript, we used "weaken" instead of "dampen" when stating that a process tends to attenuate the anomaly.

6. *Line 143: Is the ice sheet component of the model active?*

Responses: Thank you very much for this comment. We have stated in line 159 of the revised manuscript that the ice sheet component is not active.

7. *Line 187: I wouldn't say that it's an oscillation around an equilibrium as an "equilibrium" implies that AMOC would approach that strength and remain there if all forcing were held fixed.*

Responses: Thank you very much for this comment. In lines 219-220 of the revised manuscript, we revised this sentence to be "*The AMOC index shows a stable oscillation around its mean state*."

8. *Line 208: Should be "The power spectrum…"* Revised.

9. You mention that LFP1 explains 87.5% of the low frequency variability of AMOC—does LFP2 contribute significantly to the remaining unexplained variability, and if so, does it have a strong signal in the Southern Ocean?

Responses: Thank you very much for this comment. LFP2 of the AMOC (Fig. R6) explains 8.5% of the total low-frequency AMOC variance, that is, around 68% of the remaining low-frequency variability. However, the model output of the Atlantic meridional streamfunction is only valid north of around 30°S. Figure R7 shows the LFC2s and LFP2s of the global SSD, SSD anomaly induced by SSS anomaly, and SSD anomaly induced by SST anomaly. No obvious signal is observed in the Atlantic and Southern Ocean (Figs. R7b-d). Therefore, it is reasonable to focus on LFC1s of the AMOC and global buoyancy fields.

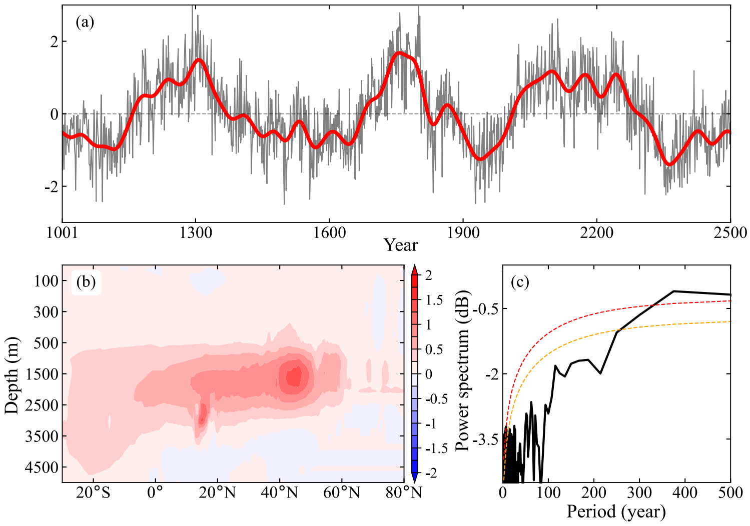

FIG. R6. (a) LFC2 (gray curve) and the low-pass-filtered LFC2 of the Atlantic meridional streamfunction. (b) LFP2 of the Atlantic meridional streamfunction (units: Sv). (c) Power spectrum (units: dB) of the unfiltered AMOC LFC2, with period as the abscissa. The dashed orange and red curves represent the best-fit first-order Markov red noise spectrum and the significance at 95% confidence level, respectively. Before applying the low-frequency component analysis (LFCA) method, the data is detrended and then weighted according to the square root of grid cell thicknesses. The Lanczos filter is used for the LFCA.

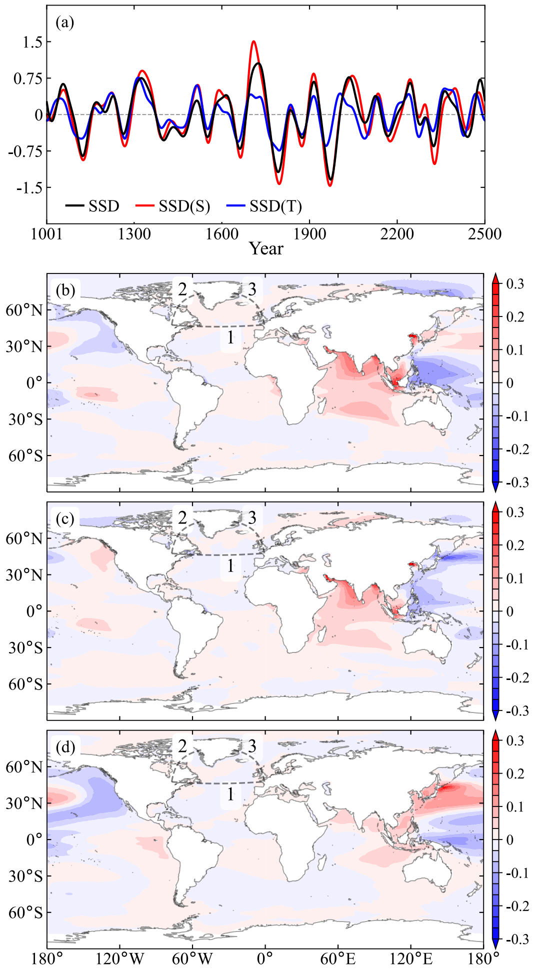

FIG. R7. (a) Filtered LFC1s (units: dimensionless) of the global SSD anomaly (black curve), SSD anomaly induced by SSS anomaly (red curve), and SSD anomaly induced by SST anomaly (blue curve). (b) LFP1 (units: kg m^{-3}) of the global SSD anomaly. (c) and (d) are the same as (b), but for the SSD anomalies induced by SSS and SST anomalies, respectively. Before the LFCA, the data is detrended and then weighted according to the square root of grid cell areas. The Lanczos filter is applied in (a); and the LFCA, in (b)-(d).

10. Line 252 (and line 585 later): Since you don't actually identify any low frequency modes in Earth's climate more generally (beyond AMOC), I would refrain from statements claiming that the MCO can act as a pacemaker for low-frequency climate variability. Related statements are deleted.

11. Line 314: At this point you have not shown any causal connections/mechanisms between the Arctic, deep Atlantic, and subpolar upper ocean, so I would refrain from saying one is "influenced" by the other.

Responses: Thank you very much for this comment. The related sentences are rephrased as "*Throughout the entire cycle, salinity anomalies in the Arctic Ocean and Atlantic deep ocean are largely synchronized with the upper DWF region. Signals in the South Atlantic are much weaker compared to those in the North Atlantic*" in lines 348-351 of the revised manuscript.

12. Line 362: Do you mean that advection by the mean circulation cannot explain the evolution of salinity regression coefficients with AMOC?

Responses: Thank you very much for this comment. In lines 395-397 of the revised manuscript, we have rephrased this sentence to be "*However, this mean advection of salinity anomaly is too weak and may not be enough to determine the weakening (Figs. 6a-c, g-i) and phase transition (Figs. 6c, i) of salinity anomaly in the DWF region*."

13. Line 398: I think you mean "positive to negative" here. Revised.

14. Line 400-402: This is a nice theory (one feedback overcoming the other to start the phase shift in AMOC), but I don't think any of your analyses support it directly. Ideally you would show a salinity budget for the subpolar upper ocean (if possible) and show how contributions from the subtropical upper ocean and intermediate ocean compete until one signal wins out over the other.

Responses: Thank you very much for this comment. This sentence has been rephrased to "*When weakening processes outweigh enhancing processes for salinity anomaly in the upper DWF region, its magnitude peaks and starts to decrease*" in lines 436-438 of the revised manuscript. The original version is not rigorous enough. As Fig. 11b of the revised manuscript reveals, the two paramount processes for salinity anomaly in the 0-1000 m DWF region are the anomalous and mean advections in the subtropical-subpolar North Atlantic. These two processes offset each other to a large extent but do not achieve exact compensation, as there are other processes enhancing and weakening salinity anomaly in the DWF region.

15. Line 631: Should this be "viability" rather than "liability"? This sentence has been rephrased.

Replies to Reviewer #2:

We sincerely appreciate your valuable comments and the constructive feedbacks. Our manuscript has been revised in accordance with your recommendations. The subsequent sections include our detailed, point-by-point responses.

The paper deciphers the mechanism of multicentennial variability of the AMOC in a realistic coupled model, in analogy with a previously published theoretical analysis of centennial oscillations in a box-model.

It presents a very comprehensive and convincing analysis. It is well written, logically organized, figures are well prepared and clearly labelled.

Overall, I think the paper deserves publication, except that one central reference is not provided and some (not so) major comments should be addressed.

Responses: Thank you very much for your valuable suggestions, which help us improve the manuscript tremendously. Considering the comments from all the reviewers, we revised the manuscript primarily in the following aspects:

1) The rationale for employing the theoretical model is revised. Instead of using the coupled model to “validate” the theoretical model, now the theoretical model is used to perform sensitivity (or, denial) experiment to examine the indispensability of the anomalous and mean advections in this North Atlantic Ocean-originated AMOC multicentennial oscillation (MCO). Accordingly, section 5b and the appendix have undergone significant revisions.

2) We computed freshwater transport to quantify the contributions of different processes to the subpolar salinity anomaly. As salinity transport across a section cannot reflect the actual effect on salinity change/tendency of the subpolar region, it is no longer adopted. Accordingly, section 5a is rewritten.

3) The evolution of salinity anomalies, especially the subtropical intermediate salinity anomaly, is depicted and analyzed in more detail. This corresponds to section 4b and the schematic Fig. 13.

4) Difference between this North Atlantic Ocean-originated AMOC MCO and other AMOC MCOs, as well as the implication of this study on interpretating other AMOC MCOs, are delineated in section 6.

5) The necessary information regarding the introduction section and illustration on the theoretical model are complemented. Now the revised manuscript is more standalone and it is no longer mandatory for readers to refer to Li and Yang (2022) (LY22) for the basic information. This corresponds to section 1 and the appendix.

Major Comments:

1. *Line 42: Yang et al. (2023) does not appear in the references. As this submitted manuscript is presented as a follow-up study of this cited paper, it is difficult to evaluate how much is new in the present manuscript. The full reference is therefore needed.*

   (I am sorry I could not find it by myself: In Journal of Climate, there are already 53 papers with author Yang published in 2023.)

Response: Sorry for this careless mistake. Yang et al. (2023) is a manuscript of us submitted to *Journal of Climate*. It is a theoretical study that extends the theoretical model in LY22, by including temperature variation in the salinity-only box model in LY22. As it is also sufficient to only cite LY22, in the revised manuscript we have deleted Yang et al. (2023). This will not hamper the interpretation of the revised manuscript.

2. *Page 30, lines 666-667: To allow the full understanding of the theoretical model within this manuscript (without a mandatory reading of the Li and Yang 2022 paper), I think some explanation on the parameterization of q' should be given here, or/and elsewhere in the manuscript.*

Response: Sorry for the oversight. In the revised manuscript, the necessary information for making this study a standalone paper is included. These pieces of information are related not only to the theoretical model but also to other sections, such as the introduction. More importantly, the rationale for utilizing the theoretical model is revised. Now, the LY22 box model is used for conducting sensitivity (or, denial) experiment to validate the indispensability of anomalous and mean advections in the upper AMOC branch for the North Atlantic Ocean-originated AMOC MCO. By artificially deactivating either the anomalous or mean advection in the upper AMOC branch in the theoretical model, the AMOC MCO cannot exist (Fig. 12 of the revised manuscript). This indicates that the anomalous and mean advections in the upper AMOC branch are paramount for this North Atlantic Ocean-originated AMOC MCO.

As the theoretical model is now used to test the coupled model mechanism, instead of using the coupled model to verify the theoretical model, the parameters and parameterization for q' in the theoretical model are now set according to the CESM1 simulation.

In Eq. (A5) of the revised manuscript we have explained how q' is parameterized. The parameterization is now based on Fig. A2 of the revised manuscript, that is, the AMOC anomaly is linearly proportional to the anomaly of meridional difference of potential density, with the former having a magnitude around 70 times that of the latter. Therefore, λ is given as 70 Sv kg^{-1} m^{3} in the theoretical model.

The choices of other parameters are illustrated in lines 733-743 of the revised manuscript. Accordingly, $\overline{q}$ in the theoretical model is set as 24 Sv, the mean AMOC strength in the CESM1 simulation.

We hope that with the refinements and clarifications in the revised manuscript, the utilization of the theoretical model becomes more reasonable, the parameters and parameterizations of the theoretical model become rightful and clear for the readers, and it is no longer mandatory for the readers to read LY22 for the information that should have been elaborated by us in this manuscript.

3. *Page 25, line 551: explanation should be given why the model maximum overturning is not used instead of this quite arbitrary 15 Sv. Is it the model overturning at the edge of the boxes?*

Response: Thank you very much for this comment. As we have stated in the response to your major comment 2, now $\overline{q}$ is set to 24 Sv, the climatological AMOC strength in the CESM1 simulation.

4. *Figure 12 and in section 5b: The parameterization of q' and its validity in the model could be tested as well.*

Response: Thank you very much for this comment. As we have stated in the response to your major comment 2, now the parameterization of q' is derived from the linear relation between AMOC anomaly and anomaly of meridional difference of potential density in the CESM1 simulation (Fig. A2 of the revised manuscript). Following Fig. A2, λ is given the value of 70 Sv kg^{-1} m^{3}.

5. *Overall, I think the relationship between the NAC and AMOC could be addressed in more details, it is vaguely discussed in different places of the paper and NAC current anomalies are*

shown in some figures. Are the NAC anomalies in geostrophic balance with the zonally averaged density (salinity) anomalies?

Response: Thank you very much for this comment. As the NAC is part of the AMOC's upper limb, a positive (negative) AMOC anomaly is always accompanied with a faster (slower) NAC. The NAC anomalies are in geostrophic balance with the zonally averaged salinity anomalies. For instance, when the zonally averaged salinity anomalies in the upper DWF region and subtropics are negative and positive (Figs. R8a, l), the NAC exhibits westward anomaly (Figs. R9a, l). Similarly, when the zonally averaged salinity anomalies in the upper DWF region and subtropics are positive and negative (Fig. R8f), the NAC exhibits eastward anomaly (Figs. R9f).

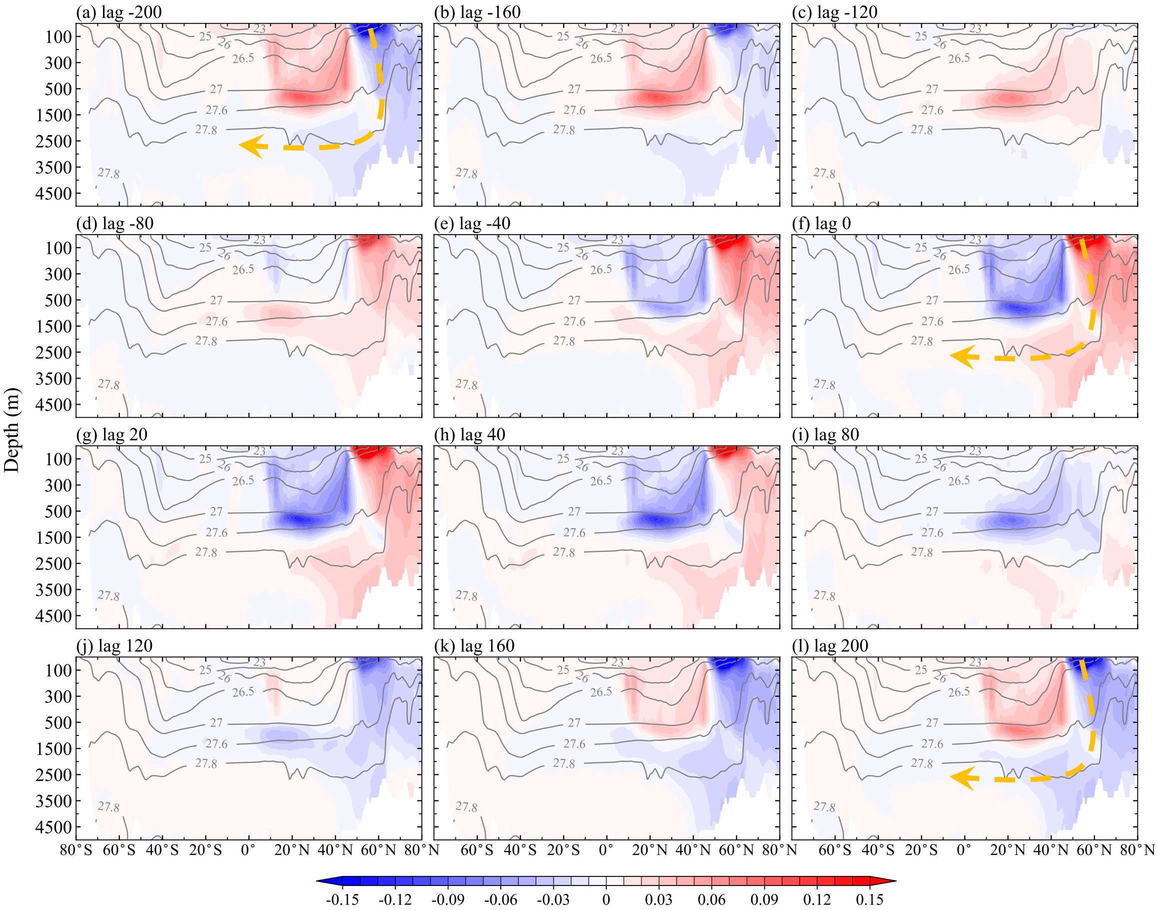

FIG. R8. Lead/lag regression coefficients of zonally averaged salinity anomalies in the Atlantic on the AMOC LFC1 (shading; units: psu). Negative lag means the AMOC LFC1 lags the salinity anomalies (units: year). Contours show the zonally averaged climatological potential density σ_θ in the Atlantic (units: kg m^{-3}). Orange arrows in (a), (f), and (l) show schematically the downward and southward movements of salinity anomalies. This figure is the same as Fig. 4 of the revised manuscript.

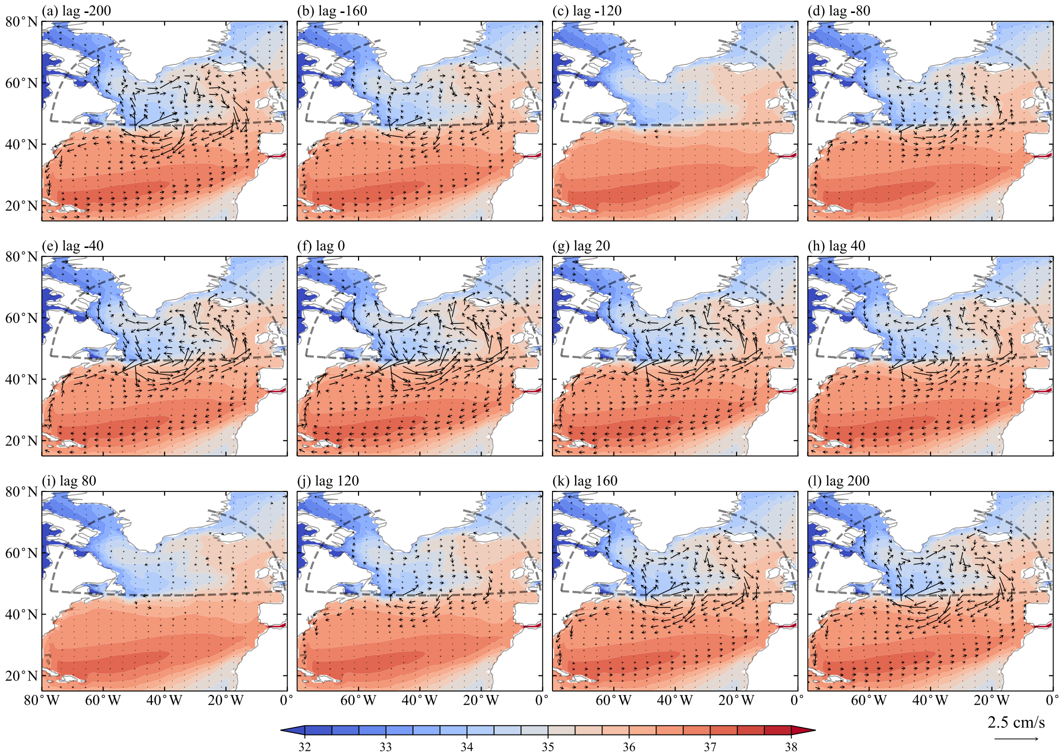

FIG. R9. Lead/lag regression coefficients of current anomalies averaged over 0-200 m on the AMOC LFC1 (units: cm s^{-1}), superimposed with climatological salinity averaged over the same depth range (shading; units: psu). Negative lag means the AMOC LFC1 lags the current anomalies (units: year). This figure is the same as Fig. 7 of the revised manuscript.

6. *Figure 13: The schematic is nice, but the most important processes are how the change of phase occurs, maybe intermediate figures could be shown to highlight this issue to provide the full story.*

Response: Thank you very much for this suggestion. In the revised manuscript, the schematic Fig. 13 incorporates four subpanels depicting the oceanic states when the AMOC: (1) is at the minimum, (2) has just turned slightly positive after phase transition, (3) reaches the maximum, and (4) becomes slightly negative after phase transition. More descriptions on the full cycle, including the growing, weakening, and phase transition of the AMOC anomaly, are presented in lines 614-637 of the revised manuscript.

7. *Page 27: The only way to prove that the mode is really internal would be to exhibit ocean-only simulations that reproduce it, in the same way as Gastineau et al. 2018 for instance.*

Reference:

Gastineau. G., J. Mignot, O. Arzel, and T. Huck, 2018: North Atlantic Ocean internal decadal variability: Role of the mean state and ocean–atmosphere coupling. J. Geophys. Res., 123, 5949–5970, https://doi.org/10.1029/2018JC014074.

Response: Thank you very much for this comment. This part has been deleted. Throughout the revised manuscript, we have stated that the AMOC MCO in our study is “primarily” instead of “completely” driven by internal oceanic processes. This AMOC MCO is still an internal variability of the Earth’s climate system, as the external natural forcing is fixed throughout the simulation.

Minor Comments:

1. *Fig. 2: it is not obvious from the figure to compare the relative amplitudes of the salinity and temperature induced SSD anomalies in the total. The have the same amplitude in panel a, and colorbars in panels c and d do not allow to determine easily their maximum amplitudes.*

Response: Thank you very much for this suggestion. The colorbar limits in Figs. 2b-d are made the same in the revised manuscript. Now it is obvious that the sign of the SSD signals follows that of the SSS-induced SSD signals, hence salinity anomaly has larger effect on the AMOC MCO than temperature anomaly does.

Additionally, we would like to acknowledge an oversight in our original manuscript: the sign of LFC1 of the SSD induced by SST anomaly (Fig. 2a, blue curve) should be reversed. This has been rectified in the revised manuscript.

2. *Page 22, line 468: typo at "structure".* Revised.

3. *Page 30, lines 666-667: I think some details about the parameterization of q' should be given here, since this is quite central to the mechanism.*

Response: Sorry again for the lack of necessary information in the original manuscript. Now the details on parameterization in the theoretical model and the parameter values are included in the revised manuscript. The whole manuscript, including the contents of the theoretical model, the introduction, as well as other parts, are made more of an independent paper. It is no longer mandatory for readers to refer to LY22 for the basic information that should be included in this manuscript.