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.

This manuscript is a follow up of a previous study published in Journal of Climate. Both studies used an idealized box-model to understand the stability and persistence of a centennial oscillation of the AMOC. Whereas the first study focused solely on the active role of salinity, the present work studies the effect of including the temperature as an active variable.

As mentioned in the previous round, this work tackles an interesting topic and is a nice and needed follow up of the authors' previous work.

The revised version partially answered some of my comments, but once again only a little of my concerns and the response led to changes/clarifications in the new version of the manuscript. For instance, the physical description remains elusive for a reader (that will not have access to the response of the authors). Also, the lack of consistency of the set of parameters (i.e., λ) throughout the study makes the paper hard to follow. In particular I do not understand why the parameters should not be fixed to allow for an easier comparison with Part I (LY22).

Overall, I feel that the study might be publishable at some point but there remained significant concerns that I would like the authors to address prior to publication.

Hence, I recommend this work for major revision.

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 these following aspects:

1) To be consistent with LY22, the key parameter λ is fixed at 14 Sv kg^{-1} m^3 throughout the paper, except for the stability analyses that require changing λ.

2) The thermal effects are generalized as two points: 1. shortening the oscillation period, and 2. stabilizing the system, which are more concise now. The role of subpolar temperature stratification is no longer our focus; thus Fig. 5 of the 2^{nd} revised manuscript is removed.

3) The strategy for deriving the equilibrium values is made clearer and more reasonable.

Major Comments:

I kept the numbering from the previous round of reviews.

1) Inconsistency between the theory and numerical simulations

Responses: Thank you very much for your valuable suggestions and patience.

I think the authors have a perfect analysis to make a nice paper, but once again it felt short because of the lack of consistency in the use of their parameters… The linear simulations (section 3 and Figs. 3 and 4) and sensitivity model (section 5 and Figs.8 and 9) are using λ *= 13 Sv kg*$^{-1}$ *m*^{3}*, whereas the nonlinear simulations are using* λ *= 14 Sv kg*$^{-1}$ *m*^{3} *(section 4 and Figs. 6 and 7). This is extremely confusing.*

It is obvious that the most interesting dynamics occur for λ *= 14 Sv kg*$^{-1}$ *m*^{3}*. This leads to TS3 being stable and TS4 being unstable and the existence of a possible window of oscillation. This would be consistent with the idea of enhanced mixing in the subpolar essentially switching between TS3 and TS4, allowing for a stable nonlinear oscillation. Here you would have generalized the nice result of LY22 to the presence of Temperature.*

Following your suggestions, we have changed λ to 14 Sv kg^{-1} m^{3} throughout the paper, except for stability analyses that require altering λ in Fig. 8 of the 3rd revised manuscript.

The temperature will not be without effect (section 3), since it moves the window of oscillation to higher λ*. The increase of* λ *essentially increases the salinity feedback to counteract the damping effect of the temperature restoring (when one compares the couple S3 and S4 with the couple TS3 and TS4).*

The temperature variation has effect on the system; however, its effect will be dampened by the surface temperature restoring. The thermal processes consist of the negative temperature advection feedback and the surface temperature restoring. The former stabilizes the system and dampens AMOC anomaly, the latter dampens the temperature anomaly thus limits the negative temperature advection feedback. Therefore, the net effect of surface temperature restoring is to destabilize the system instead of dampening the AMOC anomaly. Since the surface temperature restoring cannot overcome the negative temperature advection feedback, the overall effect of the thermal processes (negative temperature advection feedback + surface temperature restoring) is still to stabilize the system (dampen the AMOC anomaly).

Here you have everything you need to make a clear and nice paper. I will strongly suggest you to do so.

This would require you to

*a) Focus on a single standard λ = 14 Sv kg^{-1} m^3 (instead of 13 Sv kg^{-1} m^3), *throughout* the manuscript;*

We have changed λ to 14 Sv kg^{-1} m^3 throughout the paper, except for stability analyses in Fig. 8 of the revised manuscript.

b) Discuss the effect of temperature as a "displacement" of the oscillatory-window (where range λ could be discussed);

In the revised manuscript the thermal effects are summarized as: (1) shortening the oscillation period, and (2) stabilizing the system. The latter reveals that the stability criteria (λ_r) for the 3TS and 4TS models are shifted compared to the 3S and 4S models.

c) Add the critical value of TS3 to Figs.8 and 9, to highlight the oscillatory window as you did in LY22.

The essence of the self-sustained oscillation is a growing oscillation restrained by nonlinearity, which is also true in LY22. The presence of a "window" of stable linear 3-box model and unstable linear 4-box model is not the precondition for self-sustained oscillation. We did not consider the temperature processes three years ago when we worked on LY22, which led to a plausible but incomplete conclusion that there is a "window" for a self-sustained MCO of an S-only system. Now after three more years of in-depth study of the multicentennial oscillation (MCO) of AMOC, and more importantly, the linear oscillation, we think this conclusion should be improved to be "***a growing oscillation restrained by nonlinearity***" instead of the "***window***". Recalling that the λ_r (stability criterion under the parameters in the 3rd revised manuscript) for the 3S, 4S, 3TS, and 4TS models are 13.58, 12.10, 14.39, and 13.68 Sv kg^{-1} m^3, respectively.

Specifically, (1) in 4S model, when considering the nonlinear subpolar vertical mixing, there is a "window" of self-sustained oscillation from unstable 4S to stable 3S. The self-sustained oscillation can only occur in the window defined by λ, that is, λ_r(4S) (12.10) < λ < λ_r(3S) (13.58) (Fig. R1a). This is the case studied in LY22.

(2) in 4TS model, when considering the nonlinear subpolar vertical mixing, the self-sustained oscillation can occur even when both the 3TS and 4TS models are slightly unstable (Fig. R1b), as long as λ is not too large.

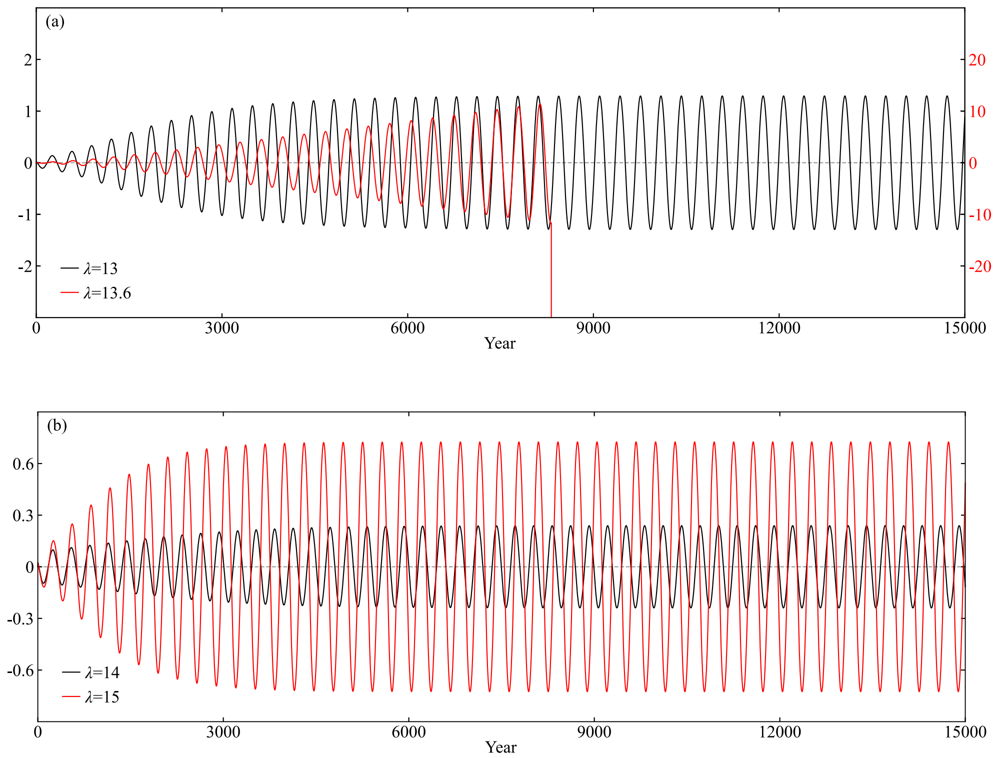

FIG. R1. Time series for q' (units: Sv) in different models. (a) Self-sustained oscillation under λ = 13 Sv kg^{-1} m^3 (black curve, left y-axis) and unsustainable growing oscillation under λ = 13.6 Sv kg^{-1} m^3 (red curve, right y-axis), in 4S model where the subpolar vertical mixing with κ = 10^{-4} m^{-3} s is included. (b) Self-sustained oscillations under λ = 14 Sv kg^{-1} m^3 (black curve) and 15 Sv kg^{-1} m^3 (red curve), in 4TS model where the subpolar vertical mixing with κ = 10^{-4} m^{-3} s is included.

Consequently, we did not add the stability criterion for 3TS in Figs. 7-9 of the 3rd revised manuscript, as the essential mechanism is not determined by the "window" anymore. We plot both the stability criteria for the 3TS and 4TS in Fig. R2, for the following discussions.

Since the role of relative stability between the 3TS and 4TS is not essential (the "window" will not determine the self-sustained oscillation), we did not focus on it in the 3rd revised manuscript. Also, as the role of subpolar temperature (T) stratification is less important now, we have removed it from the thermal effects, and Fig. 5 of the 2nd revised manuscript is thus removed.

The case with λ = 14 Sv kg^{-1} m^3 still corresponds to an unstable mode of 4TS. Including nonlinearity leads to a self-sustained oscillation, which is consistent between this study and LY22.

The previous discussions on subpolar T stratification have caused confusion, which is also seen in your following major comment 2b. Since it is essential for understanding that the "window" is not a precondition for self-sustained oscillation, we will try to explain the role of subpolar T stratification here (although it is no longer emphasized in the 3rd revised manuscript), which also explains your major comment 2b.

As discussed in the 1st and 2nd revised manuscripts, the subpolar S and T stratifications have destabilizing and stabilizing effects on the system, respectively. Compared to the 3S, the 4S has the destabilizing S stratification, thus the 4S is always less stable than 3S. However, the 4TS has both the destabilizing S stratification and the stabilizing T stratification, so whether the 4TS is less stable than the 3TS depends on the competition of S and T stratifications.

When the thermal effects are weak, the stabilization from T stratification is weaker than the destabilization from S stratification, and the 4TS will be less stable than the 3TS. This corresponds to a wide range of γ (Fig. 4b of the 3rd revised manuscript), including the standard value γ = (1 year)$^{-1}$, which leads to a less stable 4TS than 3TS in Fig. 2c of the 3rd revised manuscript.

When the thermal effects are strong, the stabilization from T stratification can overcome the destabilization from S stratification, thus the 4TS will be more stable than the 3TS. This corresponds to $\gamma \approx$ (0.2 year)$^{-1}$ (Fig. 4b of the 3rd revised manuscript). Also, it is seen in Figs. R2a and R2b that under certain model geometry settings (lower left) while keeping other parameters to standard values in Table 1, the 4TS can also be more stable than the 3TS. Therefore, although there is always a "window" of unstable ***4S*** and stable ***3S***, there is not always a "window" of unstable ***4TS*** and stable ***3TS***.

It has been shown in the 1st revised manuscript that, even if there is no "window" of unstable 4TS and stable 3TS under the old model parameters, there could be self-sustained oscillation if the linear growing oscillation is restrained by nonlinearity. Under the new model parameters in the 3rd revised manuscript, between 13.68 < λ < 14.39 Sv kg^{-1} m^3 there is unstable 4TS and stable 3TS. However, even under a higher λ where both the 3TS and 4TS are unstable (like λ = 15 Sv kg^{-1} m^3 in Fig. R1b), introducing nonlinearity can lead to self-sustained oscillation (Fig. R1b). Therefore, such "window" is not the essence and precondition of self-sustained oscillation in the 4TS model.

The essence of the self-sustained oscillation is a growing oscillation restrained by nonlinearity. This mechanism is also the essence of LY22, which is consistent with this current study. The problem is, LY22 didn't generalize the self-sustained oscillation mechanism well enough, which arose from our limited understanding of linear oscillation at that time. To address this, we improved the mechanism to be "a growing oscillation restrained by nonlinearity" in part II. Therefore, the significance of part II lies not only in revealing the role of temperature variation in AMOC MCO, but also in improving the self-sustained oscillation mechanism in LY22. The key point here is the ***improvement*** of the self-sustained oscillation mechanism, rather than the ***inconsistency***.

We hope that this can resolve the confusions persistent through the 3 rounds of review.

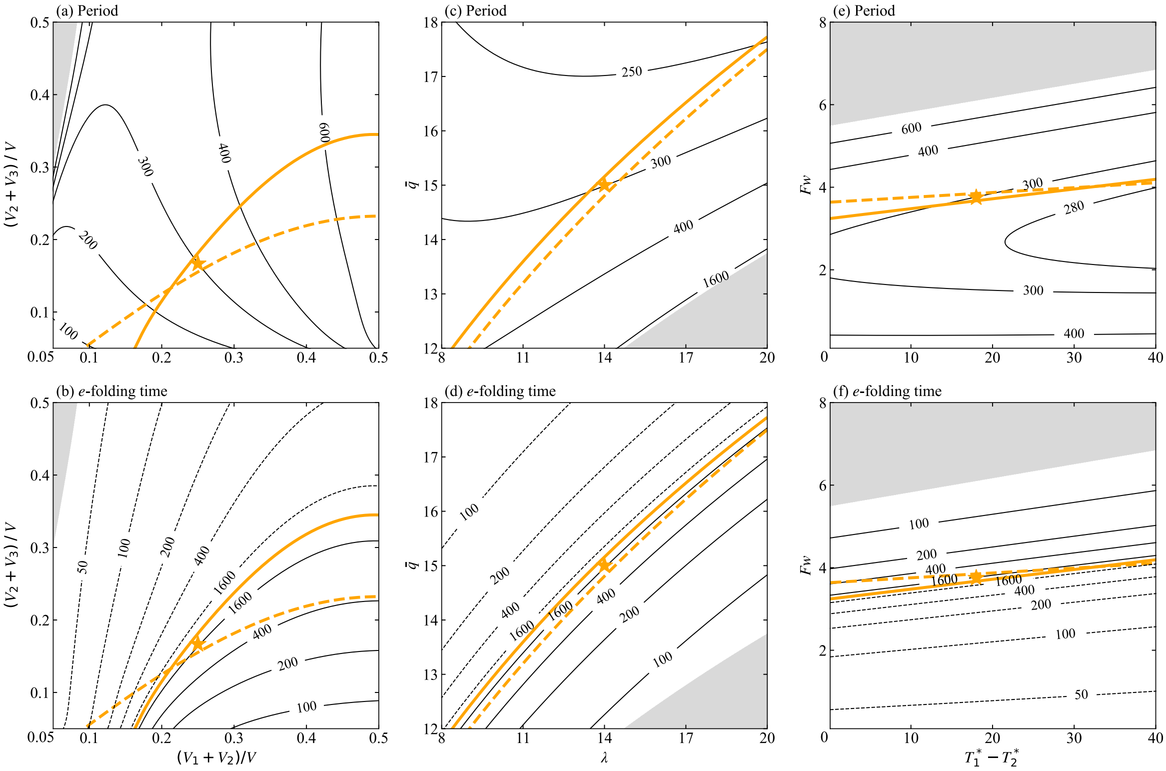

FIG. R2. Sensitivity of the conjugate eigenmode in the 4TS model to (a, b) model geometry, (c, d) the mean strength of AMOC and the AMOC's sensitivity to density perturbation, and (e, f) surface virtual salt flux and meridional difference of restoring temperature. The orange star denotes the standard parameters. The solid and dashed orange curves denote the stability thresholds for the 4TS and 3TS models, respectively.

2) Inconsistency between the LY22 when temperature anomaly is allowed.

Responses: Thank you very much for your valuable comments.

I do not understand the problem suggested by the authors in their response. Actually I was satisfied by Fig. 2c (which is essential to the study), but the authors seem to interpret it in a way that makes the study still inconsistent with Part I (LY22)… Also, the discussion in the response is not aligned with the figure…

We agree that Fig. 2c is essential to this study. From Fig. 2, especially Figs. 2a and 2c, we conclude the thermal effects as: 1. shortening the oscillation period, and 2. stabilizing the system. We choose not to focus on the stabilizing effect of the subpolar T stratification, which is no longer essential to the study.

2a) There is a clear window of oscillation equivalent to LY for λ = 14 Sv kg^{-1} m^3. This should be the value chosen in the rest of the study for consistency with LY22.

According to your suggestion, we have fixed λ to 14 Sv kg^{-1} m^3.

2b) Unlike stated in the response ("Therefore, there is always a window with unstable 4S and stable 3S. However, after including thermal effects, subpolar temperature stratification is included, which has stabilizing effect and tends to stabilize the 4TS model. This stabilizing effect is absent in 3TS."), now 4TS is always less stable than 3TS (equivalently to 4S and 3S in LY22)

The relative stability of the 4S and 3S is simple. The 4S has an additional destabilizing subpolar S stratification compared to the 3S, so the 4S is always less stable than the 3S, no thermal effects (stabilization from subpolar T stratification) need to be taken into consideration.

Turning to the TS models, the condition of Fig. 2c is only under γ = (1 year)$^{-1}$, that is, the strength of the thermal processes is fixed. As we have stated in the response to your major comment 1c, whether the 3TS model is more stable than the 4TS model depends on the relative strength of destabilization from subpolar S stratification and stabilization from subpolar T stratification. Under the standard strength of thermal effect at γ = (1 year)$^{-1}$, the stabilization from subpolar T stratification is weaker than the destabilization from subpolar S stratification, thus the 4TS is less stable than 3TS. Altering the strength of thermal effects through altering γ can affect the relative stability between the 4TS and 3TS. For example, upon switching to around γ = (0.2 year)$^{-1}$, the thermal effects are stronger and the stabilization from subpolar T stratification overcomes the destabilization from subpolar S stratification, and the 4TS is more stable than 3TS (Fig. 4b of the 3rd revised manuscript).

Note that in most cases, the 4TS is less stable than 3TS [including the standard condition γ = (1 year)$^{-1}$ of Fig. 2c]. Only when γ is around (0.2 year)$^{-1}$ (Fig. 4b of the 3rd revised manuscript), can the 4TS be more stable than 3TS, evidenced by the slightly larger Re(ω) of the 3TS than 4TS in Fig. 4b.

Altering the thermal effects leads to difference in relative stability between the 4TS and 3TS. This is due to the stabilization from subpolar T stratification, which ***reflects the role of thermal processes instead of the inconsistency*** between LY22 and this study. It is impossible that the behavior of the system remains completely unmodified after including T variation, or in other words, the thermal processes are completely effectless for the AMOC MCO. Although we choose not to focus on subpolar T stratification in the 3rd revised manuscript, explaining its effect is still essential for resolving your concern. We would like to emphasize again that, whether there is a "window" of unstable 4TS and stable 3TS, and whether the 4TS is less stable than the 3TS, are not the prerequisites for self-sustained oscillation. The essence for self-sustained oscillation in both LY22 and this study is a growing oscillation restrained by nonlinearity.

Finally, under the new standard value λ = 14 Sv kg^{-1} m^{3} according to your suggestion, the paper becomes easier to follow, and a self-sustained oscillation can be realized when including nonlinearity, which will not induce conflict with LY22 even from a reader's viewpoint. Additionally, we no longer focus on the subpolar T stratification and the relative stability between the 4TS and 3TS, which also reduces the risk of confusing the readers.

2c) The stabilizing effect of T is obvious from Fig. 2c, it pushes in both 3-box and 4-box the instability toward higher value of λ. So that for a given value of λ, TS is always more stable than S models (for both 3-box and 4-box).

Yes, that's true. This is the stabilizing effect of the thermal processes.

It is clear for me that the reference parameter should be λ = 14 Sv kg^{-1} m^{3}. Otherwise, the manuscript has no resemblance to LY22 and should not be a Part II, with little to no relation to the methodology or Part I.

3) Lack of information

After two rounds of review, I finally managed to understand what is actually done in the study. This leads me to have serious concern on the strategy.

You need to include a summary of your response to my major comment describing your potentially questionable strategy. This is essential for transparency and so the reader can judge the validity of the approach. Otherwise, the manuscript remains confusing and there is a lack of information.

I suggest adding a paragraph such as: "The equilibrium solution could be obtained by solving Eq. (1). However here we choose a slightly different strategy, for being able to explore the sensitivity of the variability to mean and parameter changes. Hence here we fix the model parameters as well as two mean variables ($\overline{q}$ and $\overline{S_1}$). The values of the remaining mean variables are then computed using Eq. (2)."

Beyond this lack of information, I have a serious concern regarding the chosen strategy. It is quite unphysical and problematic to compute F_w based on $\overline{q}$ and $\overline{S_{1,2}}$. F_w is set "externally" (Precipitation-Evaporation). It would be better to set F_w as a fix parameter and the mean variables $\overline{q}$ and $\overline{S_1}$. Then using mean variable $\overline{S_2}$ (as well as $\overline{S_{3,4}}$ could be recomputed using Eq. (2b). I highly recommend you to do so for the acceptance of the manuscript.

Responses: Thank you very much for these valuable suggestions. In this work, we regard F_w as a model parameter. And the equilibrium values for model variables are derived from Eq. (1). To obtain the exact equilibrium values, and to explore the sensitivity of model behaviors to changes in equilibrium values and parameters, we fix the model parameters as well as $\overline{q}$ and $\overline{S_1}$. Other equilibrium values $\overline{S_{2-4}}$ and $\overline{T_{1-4}}$ could be derived from Eq. (2) of the 3rd revised manuscript.

In lines 172-176 of the 3rd revised manuscript, we add details to explain our strategy for obtaining the equilibrium values: "*For being able to explore the sensitivity of the model's behavior to changes in mean states and model parameters, we fix the model parameters as well as mean states for two variables ($\overline{q}$ and $\overline{S_1}$), then the corresponding $\overline{S_{2-4}}$ and $\overline{T_{1-4}}$ can be calculated following Eq. (2). For simplicity, the model parameters and $\overline{q}$, $\overline{S_1}$ will be collectively regarded as model parameters.*"

Specific Comments:

1. *Line 114-116: This sentence suggesting that the AMOC oscillation of 170 years is irrelevant because it is slightly shorter than the period studied here is unacceptable.*

I am personally quite convinced that they are fundamentally the same. Note that in subsequent studies, using a slightly more complex idealized loop model, Sévellec and Fedorov (2014 and 2015) found a period of ~250 years.

Given your sentence in line 582 (which I agree with), the sentence line 114-116 needs to be revised to stress the consistency of the approach and results.

Responses: Thank you very much for your encouraging comments.

We have moved the reviews on loop model studies to lines 116-117 of the 3rd revised manuscript as: "*The self-sustained multicentennial oscillation (MCO) of AMOC has been found in loop models.*" Now, the logic of this part becomes: currently there is a need for theoretical studies; the AMOC MCO has been studied utilizing loop models; adopting a different approach, our part I LY22 studied the AMOC MCO using a box model.

From our understanding, the periods in Sévellec et al. (2006) and Sévellec and Fedorov (2014, 2015) lie in between centennial and multicentennial timescales. The loop model in Sévellec et al. (2006) considers both T and S variations and is integrated in its original form (with an explicit angle ϕ of the loop), while the model in Sévellec and Fedorov (2014, 2015) keeps T fixed and is integrated in a reduced form (it has undergone a Fourier decomposition and is reduced to a 3-variable waterwheel model). Therefore, the model in Sévellec et al. (2006) is more complex than Sévellec and Fedorov (2014, 2015). Our study shows that thermal processes shorten the oscillation period, implying that the shorter period in Sévellec et al. (2006) (170 years) than that in Sévellec and Fedorov (2014, 2015) (250 years) might be attributed to the inclusion of T variation. According to your comment, in the 3rd revised manuscript we have moved the reviews on the loop model studies and no longer underscore the 170-year period, which we hope to be a satisfactory revision.

References:

Sévellec, F., T. Huck, and M. Ben Jelloul, 2006: On the mechanism of centennial thermohaline oscillations. J. Mar. Res., 64, 355-392.

Sévellec, F. and A. V. Fedorov, 2014: Millennial variability in an idealized ocean model: predicting the AMOC regime shifts, J. Climate, 27, 3551-3564.

Sévellec, F. and A. V. Fedorov, 2015: Unstable AMOC during Glacial Intervals and millennial variability: the role of mean sea ice extent, Earth and Planetary Science Letters, 429, 60-68.

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.

I appreciate the authors' efforts in revising the manuscript and their responses to my previous comments. Most of my questions have been answered and I find this version more clear and well-organized. As the manuscript is now, I have some remaining comments or questions, and I would recommend this manuscript for minor revision. Detailed comments are listed below.

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 these following aspects:

1) To be consistent with LY22, the key parameter λ is fixed at 14 Sv kg^{-1} m^{3} throughout the paper, except for the stability analyses that require changing λ.
2) The thermal effects are generalized as two points: 1. shortening the oscillation period, and 2. stabilizing the system, which are more concise now. The role of subpolar temperature stratification is no longer our focus; thus Fig. 5 of the 2nd revised manuscript is removed.
3) The strategy for deriving the equilibrium values is made clearer and more reasonable.

1. Line 32–33: you should explain why you care about the so-called "self-sustained" oscillation. Is this because only self-sustained oscillations can be robustly observed? This is not necessarily true, as long as the decay time-scale is longer compared to the oscillatory time-scale.

Response: We focus on the oceanic self-sustained AMOC oscillation, that is, the AMOC oscillation is sustained by oceanic processes, instead of the stochastic or chaotic atmosphere (external forcing). Accordingly, in lines 29-30 of the 3rd revised manuscript we revised this sentence to be "*Introducing nonlinearity into the system can lead to self-sustained AMOC MCO that is controlled by ocean internal dynamics.*"

Additionally, although both the strongly damped and growing oscillations can exist in our theoretical model (i.e., linear oscillations with extremely short negative/positive e-folding timescales), they are unlikely to be observed in real ocean. Only the sustainable oscillations can be observed in real ocean. In our study, we focus on the ocean-dominated self-sustained oscillation, which is a subset of the sustainable oscillation. If the self-sustained oscillation with a stricter criterion (without the help of stochastic and chaotic forcing) can arise in theoretical model, the sustainable oscillations (either internally- or externally-driven) have a great possibility to exist in the real ocean, where the nonlinearity, stochastic forcing, and chaotic forcing are likely to present in some form.

2. *Line 33–34: change "as a combination of growing oscillation and nonlinear restraining effect" to "a growing oscillation restrained by nonlinearity".* Revised.
3. *Line 37: it is not clear what you mean by "flow properties". Also in line 137.*

Response: Thank you very much for this comment. We have replaced the "flow properties" with "mean strength of AMOC and the AMOC's sensitivity to density perturbation" throughout the 3rd revised manuscript.

4. *Line 49: does it have to be a "small" internal variability?*

Response: The climate variation occurred during the glacial-interglacial transition has been regarded as large-amplitude climate shift between different equilibria of the climate system. On the other hand, the stable Holocene climate was more related to small amplitude variation around a single equilibrium. Accordingly, we have revised the start-up sentence of the introduction and also this sentence to be: "*Compared to the glacial-interglacial climate shift…*" and "*It is thus reasonable to deduce that small-amplitude internal variability around a single equilibrium was crucial for climate variability during this period*" in lines 44-47 of the 3rd revised manuscript.

Moreover, the theory we proposed is a linear theory, which is usually adopted for explaining small-amplitude oscillation around a single equilibrium. For explaining the climate shift between multiple equilibria, the more complex nonlinear theories are usually adopted.

5. *Introduction: it will be helpful if you can give some example phenomena on the relevant time-scale before you discuss their mechanism.*

Response: Thank you very much for this suggestion. Paleoclimate evidences of the multicentennial climate variability have been reviewed in the first paragraph of Part I (LY22), thus we did not include it in Part II (the current study). According to your suggestion, we added some more recent references (Askjær et al. 2022; Li et al. 2023) on AMOC-related multicentennial climate variability revealed in proxy data, in line 48 of the 3rd revised manuscript.

References:

Askjær, T. G., and Coauthors, 2022: Multi-centennial Holocene climate variability in proxy records and transient model simulations. Quat. Sci. Rev., 296, 20.

Li, Y. W., and Coauthors, 2023: 550-year climate periodicity in the Yunnan-Guizhou Plateau during the mate mid-Holocene: Insights and implications. *Geophys. Res. Lett.*, **50**

6. *Line 65: change "bearing" to "governed by".* Revised.
7. *Line 74: if heat brought by NADW is accumulated at mid-depth, convection should warm the surface with an increase in salinity. I do not understand why it is "convective cooling".*

Response: The warmer mid-depth water will be exposed to the cold surface air through deep convection, thus is "convectively cooled" by the cold surface air. For the entire Weddell Sea from the surface to the deep layer, the overall density is increased since more heat is released at the surface. Accordingly, in lines 72-74 of the 3rd revised manuscript, we have revised this sentence to be "*When the heat accumulation becomes too extreme, deep convection in the Weddell Sea is triggered, connecting the warm mid-depth water to the cold surface air. This leads to heat loss of the Weddell Sea…*"

8. *Line 85–89: whether an increase in MOC will increase or decrease salinity in North Atlantic depends on the regime of the ocean in models and this is called MOC stability. It is not clear which is happening in the real ocean. You may want to make this comment here for this mechanism.*

Response: Thank you very much for this suggestion. In studies focusing on the current real ocean, the AMOC is predominantly in the thermal mode, where the mean circulation in the upper layer is northward and controlled by meridional temperature difference, so is the CESM1 simulation analyzed in LY22. To make this point clearer, in lines 85-87 of the 3rd revised manuscript we have stressed that the mean AMOC and positive AMOC anomaly are northward: "*Starting with a positive AMOC anomaly, the northward perturbation advection transports more low-latitude water with higher salinity to the subpolar region, enhancing the AMOC anomaly.*"

9. *Line 111-112: why it is a problem that the mechanism is dominated by diffusion?*

Response: Thank you very much for this comment. In lines 110-111 of the 3rd revised manuscript, we have revised the review of Roebber (1995) as: "*Roebber (1995) realized a 683-year AMOC oscillation in a 3-box ocean model coupled with a Lorenz atmospheric model. Yet, this oscillation is sustained by chaotic atmospheric processes*", to stress that the AMOC oscillation in this study is sustained by the chaotic atmosphere, instead of the oceanic processes that we focus on.

10. Line 239: you can give a physical explanation what this initial condition represents. A fresher subtropical ocean? How does this correspond to what is happening in the real world?

Response: Technically, this initial condition S_1' = -0.01 psu can represent a fresh perturbation of the tropical upper ocean in the real world. However, this initial perturbation (S_1' = -0.01 psu) is actually only one form of perturbation that enables the start of the oscillation in our theoretical model. Other perturbations like S_1' = +0.01 psu or S_2' = -0.01 psu also enable the oscillation. The sign and size of the initial perturbation does not matter, as long as it is small. This is also unrelated to what is happening in the real ocean.

11. Figure 2: I find that letter λ is used both for the parameter in the model, and eigenvalues is confusing.

Response: λ is used only as a model parameter, including in Fig. 2. The λ_r and λ_i denote the values of λ where Re(ω) and Im(ω) equal to 0, respectively.

12. Line 300–301: you mention that thermal process stabilizes the system, but you only show this to be true for a range of parameter λ. How is this robust in other parameter regimes? And I am a little confused that in table 4, your results show that stronger temperature restoring makes the system more unstable. Is this consistent with your conclusion that thermal process has a stabilizing effect?

Response: The thermal processes always stabilize the system. We would like to emphasize that the thermal processes consist of the temperature advection feedback and the surface temperature restoring. The former stabilizes the system while the latter destabilizes the system. Since the destabilizing effect of the temperature restoring is realized through its limit on the temperature advection, the restoring cannot overrun the temperature advection. Therefore, although the surface temperature restoring destabilizes the system, the overall effect of the thermal processes (temperature advection + surface temperature restoring) is to stabilize the system. This is evident from Fig. 2c that Re(ω) of 3S (4S) is always larger than 3TS (4TS), no matter the value of λ.

13. Line 421: replace "sustainability" with "being self-sustained". Revised.

14. Line 460: as you discussed in your response, you can mention here that other nonlinear forms of mixing can also enable self-sustained oscillation. This way your results are more robust.

Response: Thanks for this suggestion. Statements are added in lines 449-450 of the 3rd revised manuscript as: "*Other forms of nonlinearity can also enable self-sustained oscillation (like* $k_m = |q'|$*, figure not shown).*"

15. Line 486: remove the word "can". Revised.

16. Line 515: replace "oscillatory potential" with "oscillation". Revised.

17. Line 515–516: remove "because of negative Re(ω)" and "due to positive Re(ω)". Revised.

18. General comment: Another non-linearity can come from equation of state. Your equilibrium solutions of the salinity equations imply that all salinity values can change by an arbitrary constant, and it is possible that this may shift the whole system into another regime. This point is worth a comment, maybe in conclusion.

Response: Thanks for your comments. In lines 594-595 of the 3rd revised manuscript, we added "*For simplification, the nonlinearity in equation of state for seawater is not considered, which is idealized.*" Additionally, our theoretical model exclusively studied the thermal mode of the AMOC, while the multi-equilibrium and climate shift are not considered.