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 answers all my comments. I am mainly happy with the new version of the manuscript except for a minor, but mandatory, comment described below.

Hence, I recommend this work for publication after a minor, but mandatory, modification.

I trust the authors and the editor to make sure that this recommendation is applied in the manuscript, I do not need to see the manuscript again.

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) Stability thresholds for both the 3TS and 4TS models are included in Figs. 7, 8, and 9 of the revised manuscript. A relevant discussion on difference in relative stability between LY22 and this study is provided in the discussion. Now the full story is clear for the readers.
2) We have refined some wordings for greater accuracy and provided more detailed explanations where needed.

Minor, but mandatory, comments:

I kept the numbering from the previous rounds of reviews.

1) The existence of a window of oscillation equivalent to LY22 and consistency with LY22.

I would like the figure R2 to be included in the manuscript. Current figures 7, 8, and 9 are not displaying the full story of the analysis. It is important to show the existence of the window of oscillation.

I understand that the results are more complex than the ones suggested initially by LY22. I also understand that the sustained oscillation is a growing oscillatory instability that saturates to a given amplitude because of nonlinearities. Note that it is always the case - even for LY22! I also understand that the window of instability is more complex than in LY22 with 4 regions now: the 3 regions suggested by LY22 (stability for both TS3 and TS4, instability for both TS3 and TS4, and the window of oscillation), as well as one new region (where there is an inversion of stability between TS3 and TS4).

However, since the manuscript claimed to be a part II, it is essential to give the reader the full perspective of the consistency. The authors are then free to stress the more complex behavior they now understand (as they suggest in their response and did in the manuscript). However, it is unacceptable (in my opinion) to not give the full information of the consistency and let the reader decide.

Hence, I would like both the stability of TS3 and TS4 to be depicted on figures 7, 8, and 9 (as it is done in Fig. R2 of the previous round of responses by the authors). A discussion on this point needs also to be included (i.e., consistency and inconsistency with LY22).

Responses: Thank you very much for your suggestions and patience. The stability thresholds for both the 3TS and 4TS models are included in Figs. 7, 8, and 9 of the revised manuscript, as is done in Fig. R2 of the last reply. In lines 566-577 of the revised manuscript, we have added a relevant discussion on this. We believe that now the difference of relative stabilities of 3, 4-box models between LY22 and the current study, and our explanation of this problem, are more clearly presented for the readers. Furthermore, we would like to express our profound gratitude for your instructions and patience throughout the four rounds of reviews.

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' responses to my previous comments, and the revision to improve the manuscript. However, some wordings and details of the manuscript are still confusing. I would recommend a minor revision for this manuscript. 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) Stability thresholds for both the 3TS and 4TS models are included in Figs. 7, 8, and 9 of the revised manuscript. A relevant discussion on difference in relative stability between LY22 and this study is provided in the discussion. Now the full story is clear for the readers.
2) We have refined some wordings for greater accuracy and provided more detailed explanations where needed.

1. *To your response to my first comment that your goal is to understand an ocean-dominated oscillation instead of those driven by the atmosphere. I appreciate this clarification, but the word "self-sustain" is still confusing to me. Do you use this word to emphasize the oceanic dominance (does the part "self" refers to the ocean itself?), or the fact that this oscillation does not grow or decay? Seems to me that you emphasize the second part throughout the manuscript, yet the focus on the oceanic dominance is not well conveyed. As I mentioned before, an oscillation does not have to be strictly "self-sustained" to be relevant to actual observed oceanic phenomenon, as long as it decays more slowly compared to its oscillation period. Do you choose to focus only on the strictly "self-sustained" modes out of mathematical convenience, or you choose to only study this subset of oscillations?*

Responses: Thank you very much for this comment. Instead of the specific "ocean", "self-sustained" refers to the oscillatory system from a perspective of dynamic system, that is, the oscillation is sustained by processes intrinsic to the system rather than external forcing like stochastic/chaotic atmospheric processes. Self-sustained oscillation is a growing oscillation that saturates to a given amplitude because of certain a nonlinearity. It is different from neutral mode that the oscillation does not grow or decay with time.

In this work, our aim is to focus on the "self-sustained" subset of oscillation. Self-sustained oscillation is much harder to be identified in both simple theory, complex models and observations than damped, growing and neutral oscillations. We assume that if theory can give rise to a ***self-sustained*** multicentennial oscillation (which implies that the theory is more robust than those theories without self-sustained oscillation), then this multicentennial oscillation might be also more robust in complex models and observations, because in complex models and observations, the multicentennial oscillation does not have to be self-sustained: it can be due to coupled processes or simply external stochastic forcings.

In this work, we emphasize that the AMOC multicentennial oscillation can be self-sustained without stochastic/chaotic forcing, and can be controlled only by oceanic processes. We do not state that the real-world AMOC oscillation is necessarily self-sustained or controlled by oceanic processes. Our theory is one possibility of the AMOC multicentennial oscillation, while the stochastically/chaotically-forced oscillations are also possible.

2. *The abbreviation MCO may be confusing with MOC, and it seems unnecessary since it is not mentioned that many times in the manuscript.* This abbreviation has been deleted.
3. *Lines 104–105: The connection between all the references you list and your goal to stress that this oscillation is dominated by ocean processes is not well established. There are also many other works showing that these oscillations are driven by air-sea interaction or stochastic forcing. Why are you certain that oceanic processes are indeed dominant? Are there any observations or modeling works to support this? Or are you just proposing a possible ocean-only explanation in this manuscript?*

Responses: Thank you very much for this comment. These lines are rephrased. In lines 107-115 of the revised manuscript, we stated: "*Generally, the low-frequency AMOC oscillation can be externally-forced or self-sustained. Griffies and Tziperman (1995) (hereafter GT95) realized a stochastically-forced AMOC multidecadal oscillation in their 4-box model. Roebber (1995) realized a 683-year AMOC oscillation in a 3-box ocean model forced by chaotic atmosphere. Rivin and Tziperman (1997) (hereafter RT97) and Wei and Zhang (2022) realized self-sustained AMOC oscillations in their box models, but the timescales of these two studies are multidecadal. The self-sustained multicentennial oscillation of AMOC has been found in loop models (Sévellec et al. 2006; Sévellec and Fedorov 2014, 2015). Adopting a hemispheric 4-box model, our first publication of*

LY22 also identified the self-sustained AMOC multicentennial oscillation, and delved into its mechanism” to convey that, there are other theories like stochastically/chaotically-forced oscillations proposed, LY22 and the current study chose to propose a theory where the low-frequency AMOC multicentennial oscillation can be self-sustained without the aid of stochastic/chaotic forcing, and can be controlled by oceanic processes only. This is a possible and ocean-only explanation, and we do not suppose that the real-world low-frequency AMOC oscillation is necessarily self-sustained and only controlled by oceanic processes.

4. *Lines 120–121: These references are all modeling works. However, this point is also true in the real ocean. Due to the nonlinearity of EOS, salinity plays an important role in determining density where temperature is low.*

Responses: Thank you very much for this comment. The references therein are deleted, as this point does not need modeling works as backup.

5. *Line 131 “more realistic parameters”: How are these parameters more realistic, compared to the 4S model?*

Responses: Thank you very much for this comment. According to your suggestions in the 1st and 2nd rounds of reviews, we have changed the upper boxes depth and mean AMOC strength from 500 m and 10 Sv in LY22 to 1000 m and 15 Sv in this study, which are more representative of the real-world AMOC. This is reflected in lines 172-175 of the revised manuscript as: “*To make the model more representative of the real ocean, we change the depth for the upper (deeper) ocean box from 500 m (3500 m) in LY22 to 1000 m (3000 m), corresponding to the actual thickness of the upper (lower) branch of the AMOC.* $\overline{q}$ *is set to a larger value of 15 Sv with a northward direction.*”

6. *Table 2: Maybe it is more straightforward and physical to show timescales instead of eigenvalues.* *e*-folding times and periods are added in Table 2.

7. *Lines 373–374: “without being self-sustained” is unnecessary, since “damped, neutral, or growing” is the definition of no self-sustained. Back to my first point, since your definition of “self-sustained” is neither decaying nor growing, your emphasis on oceanic dominance is not well delivered. Also, you should clarify why you choose to only focus on these self-sustained modes.*

Responses: Thank you very much for this comment. "Without being self-sustained" is deleted in the revised manuscript. As we have stated in the last round of response, 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 greater possibility to exist in the real ocean, where the nonlinearity, stochastic forcing, and chaotic forcing are likely to present in some form. We choose to offer a possible explanation that the low-frequency AMOC oscillation can be sustained by processes intrinsic to the dynamic system without the aid of stochastic/chaotic forcing, and can be controlled by ocean-only processes. We do not suppose that the real-world low-frequency AMOC oscillation is necessarily self-sustained and completely controlled by ocean processes.

8. *Line 423: Why do you call it "linear advection system" instead of something like linear 4TS model? This phrase is a little strange to me. Or more specifically, this section is about a nonlinear response of MOC to changes in meridional density gradient.*

Responses: Thank you very much for this comment. The 4TS, 3TS, 4S, and 3S models are all linear advection-dominated systems. As the essence for a self-sustained oscillation is a linear and growing oscillation restrained by nonlinearity, in section 4 we want to stress that no matter the nonlinearity is outside (nonlinear subpolar vertical mixing) or inside (nonlinear relation between AMOC anomaly and meridional density difference) of the linear advection equations, a self-sustained oscillation can exhibit. For being fundamental enough, we used "linear advection system" instead of "linear 4TS model", which is merely one kind of linear advection-dominated system.

9. *Line 428: It is not accurate to term it as "nonlinear advection". The advection term itself is still linear in form. Nonlinearity is implicit in the definition of MOC perturbation term.*

Responses: Thank you very much for this comment. We have rephrased this phrase to be "*nonlinearity in the anomalous advection*" in lines 423-424 of the revised manuscript.