Replies to Reviewer #1:

Thank you very much for all of your constructive comments. We have carefully revised our manuscript based on the advice by you and other reviewers. The following are our point-by-point replies.

This manuscript is a follow up of two previous studies published in Journal of Climate. As its predecessors, this study uses an idealized box-model to understand the centennial variability of the AMOC. The difference here is the addition of a Southern Hemisphere and subtropical wind.

This work, in principle, tackles an interesting topic potentially suitable for publication in Journal of Climate. However, there is basically no significant changes in the results from previous analyses, making this incremental study not needed. Beyond that there is many fundamental issues and misunderstanding as described more specifically below.

Hence, I recommend this work to be rejected.

Responses: Thank you for your comments. We fully understand your concerns, and also agree with you that, scientifically, the two-hemisphere box model is lack of fundamental advancement to the one-hemisphere box model. However, we believe this work is still an essential part of our ongoing research series on the natural centennial-millennial variability of the AMOC. Additionally, this work does introduce new contents that have not addressed in our previous studies.

(1) As stated in the introduction of our manuscript, the ultimate goal of our box model studies is to understand the centennial-millennial natural variability during the Holocene period—a time when human civilization underwent significant development. Historical records, particularly those from ancient Chinese documents, suggest that human civilization also experienced multicentennial variations during this period. Therefore, understanding the multicentennial variability of the climate system, specifically the AMOC, may offer insights into the evolution of human civilization. If we were to rely solely on a one-hemisphere box model theory, it would be challenging to comprehend the centennial-millennial variability observed in proxy data and coupled models, where the AMOC necessarily exhibits a two-hemispheric structure. For this reason, we believe that the study of the two-hemisphere box model is essential.

(2) Mathematically, the two-hemisphere box model is nearly identical to the one-hemisphere box model, resulting in quite similar outcomes. However, there are subtle but important physical differences between the two models. Firstly, we have to admit that the two-hemisphere model is more similar to the coupled models as it incorporates the inter-hemispheric structure of the AMOC. Secondly, the parameterization of the AMOC anomaly differs: in the one-hemisphere model, it is assumed to be linearly proportional to the anomaly of the density gradient between the equatorial

and polar basins, whereas in the two-hemisphere model, it is proportional to the anomaly between the two polar basins. The latter assumption is more reasonable and is supported by diagnostic analyses of results from two coupled models (Appendix A). To date, we have not encountered coupled model results that support the former assumption.

(3) In this work, we thoroughly investigate the effect of wind-driven meridional overturning circulation (WD-MOC) on multicentennial climate variability. The two-hemispheric model is, to date, the simplest model that can incorporate both thermohaline circulation and wind-driven circulation. For the first time, we have derived a theoretical solution that explicitly explains why there is no oscillatory behavior when only WD-MOC is present. We believe this represents an important scientific advancement in understanding the role of WD-MOC in natural climate variability across various timescales. It is important to note that wind-driven horizontal gyres do exhibit oscillatory behavior, primarily due to wave dynamics. We will address the reviewer's concern regarding mass conservation in the presence of WD-MOC in the subsequent Q&A section.

(4) For the first time, we demonstrate that the presence of thermohaline circulation is both a sufficient and necessary condition for multicentennial oscillations in this simple model. This conclusion cannot be derived from observations or complex models.

(5) By further simplifying the two-hemisphere box model, we derived a theoretical solution for the oscillation period (Appendix B), allowing us to clearly see how factors such as the mean AMOC strength, ocean basin volume, and ocean structure can influence the period. To our knowledge, this has not been accomplished in previous studies.

Major Comments:

1. *The paper is only slightly incremental when compared to previous paper by the authors. Two additions are done: the inclusion of two boxes to represent the subpolar South Atlantic and the (wrong) consideration of the wind. In principle, I am not against an incremental study if the results from the former, simpler study is altered. But here the conclusions are that the results are not changed. Since it is a theoretical paper, this only reinforce value of the previous paper that already demonstrated the key aspects of the oscillation. My conclusion is that LY22 (and to some extent YYL23) did a good job in their simplification and extract the essence of the centennial oscillation. There is no need for more complex theoretical model, since it does not change significantly their conclusions. Hence the current paper does not seem to be important.*

Responses: Thank you for your comments. We agree that the core physical mechanisms driving AMOC multicentennial oscillations are very similar between the one-hemisphere and two-hemisphere box models. In the preceding content, we have provided several reasons why it remains meaningful to use a two-hemisphere box model to study these oscillations.

In fact, the statement that "the core physical mechanisms of AMOC multicentennial oscillations are similar between the one-hemisphere and two-hemisphere box models" is itself an important issue that requires clarification. Only after this verification can we confidently assert that the results from the one-hemisphere model can be reliably applied to the two-hemisphere model.

We have found that the period and stability of multicentennial oscillations are strongly influenced by the basin geometry and the division of the upper ocean, lower ocean, and ocean box. The two-hemisphere model, which includes more boxes, provides a greater number of degrees of freedom, leading to different oscillatory behaviors of the AMOC from one-hemisphere models. These differences also can help us understand the discrepancies between idealized models and coupled models.

This work is part of our ongoing research series on AMOC multicentennial oscillations across a hierarchy of models with varying levels of complexity. The study of the two-hemisphere box model is a crucial step towards understanding more complex models, such as 2-D models, 3-D models, ocean general circulation models, and coupled models.

2. *The manuscript strongly oversell the novelty of LY22 and YYL23 (starting with the first sentence of the abstract) ignoring the previous literature (despite being cited and acknowledged in LY22 and YYL23). Namely the growth of the mode is through the well-established salinity feedbacks (e.g., Stommel, 1961; Marotzke, 1996), the oscillatory behavior has been already theoretically established (e.g., Sévellec et al, 2006). LY22 only explain how this mode can be sustained by the damping of convection mitigating the salinity feedback growth. It is interesting but it is far from "establishing a theory". It is especially wrong given that previous elements of the theory are 20, 30 and 60 years old… To some extent they are "common knowledge". This wrong claim of novelty is highly problematic.*

Responses: Thank you for your comments. We are fully aware that the theories on thermohaline circulation were “well established 20, 30, and 60 years ago,” and that some of these theories are considered “common knowledge.” These theories primarily focus on the **stability**, **bifurcation**, and **regime shifts** of thermohaline circulation, which we are very familiar with.

However, as we have emphasized several times in LY22 and YYL23, our studies address a different issue—namely, the ***very low-frequency internal variability of the AMOC (q'), under a highly stable mean AMOC climate ($\overline{q}$)***. This low-frequency variability has the following characteristics: (1) The variability has a very limited amplitude, i.e., $q' \ll \overline{q}$, allowing the system to be linearized. Linear stability analysis can thus be used to solve this low-frequency variability problem and identify the eigenmodes; (2) External forcings, such as heat flux, freshwater (virtual salinity) flux, and surface wind forcing, are assumed to affect only $\overline{q}$,, meaning that q' is entirely driven by internal salinity-advection and temperature-advection feedbacks; and (3) Our focus is on the centennial-millennial timescale, making this low-frequency variability of the AMOC relevant for understanding internal climate variability since the mid-Holocene, during which the Earth's climate has been relatively stable. A theory specifically addressing climate variability during the Holocene has not yet been well established.

In his classic paper, Stommel (1961) studied the multi-equilibrium state of the thermohaline circulation driven by heat and salt fluxes. However, he did not explore the internal linear oscillations of the thermohaline circulation around a stable climate.

In Marotzke (1996), a 4-box ocean-atmosphere model (consisting of 2 ocean boxes and 2 atmosphere boxes) in one hemisphere is used to examine the feedbacks between atmospheric meridional transports and the thermohaline circulation. This model includes both temperature and salinity, with surface flux conditions provided. However, this work also does not address the issue of low-frequency oscillations of the thermohaline circulation.

In Sévellec et al. (2006), a 2-D latitude-depth model under mixed boundary conditions is initially used to study the centennial oscillations of the thermohaline circulation. Then, the Howard-Malkus loop oscillator is further employed to understand the oscillatory and growth mechanisms. Linear stability analysis is applied to both models. My comments on their work are as follows:

(1) Their models include both temperature and salinity, along with boundary conditions, which makes their linear stability analysis very complex and difficult to understand.

(2) The period of oscillations is roughly 100-200 years in their one-hemisphere model and over 500 years in their two-hemisphere model (see their Table 2 below). The problem is that the growth rate (in years) of these oscillations is roughly comparable to, or even shorter than, the period, which makes the oscillatory mode either strongly damped or too unstable, and also irregular (see Figures 1-2 in their paper). I don’t believe this type of oscillation is physically meaningful. In their Figures 8 and 15, the Malkus oscillator shows a period of 170 years for the one-hemisphere basin and 605 years for the two-hemisphere basin. The corresponding growth rates (in years) are 454 and -908 years, indicating a strongly unstable mode and a damped mode, respectively. Neither oscillation would likely survive in the real world.

Table 2. Period and growth time scale obtained from time integration (nonlinear) and predicted by the linear stability analysis (linear) of the centennial modes. These values are done for all experiments (one and two hemispheres, with and without convection/ACC) for the 2-D latitude-depth model and the Howard-Malkus loop.

One	2D with conv.		2D with no conv.		Howard-Malkus loop	
Hemisphere	nonlinear	linear	nonlinear	linear	nonlinear	linear
Period (yr)	171	424	171	162	170	167
Growth (yr)	−507	−208	201	186	454	476

Two	2D model (linear)		Howard-Malkus loop	
Hemispheres	without ACC	with ACC	nonlinear	linear
Period (yr)	733	750	605	502
Growth (yr)	−67	−129	−908	∞

We are deeply inspired by these studies, and their insights and ideas have played a crucial role in our progress. However, I would like to emphasize that the questions we are investigating are quite different from theirs. We are focused on using the simplest model with straightforward, easy-to-be-understood physics to answer the following questions: Does the thermohaline circulation have an internal oscillation (eigenmode) with a centennial-to-millennial timescale? And if so, what is the underlying mechanism?

3. *Inclusion of the subtropical wind - this part is simply wrong for three fundamental reasons.*

1) *When including the Southern Hemisphere in an AMOC study, this is usual to incorporate the wind over the Southern Ocean (not the subtropical one) which is known to be key in the dynamics of the AMOC (Toggweiler and Samuels, 1995; Nikurashin and Vallis, 2012). A lack of state-of-the-art knowledge of the impact of the wind on the AMOC is worrying.*

Responses: Thank you for your comments. We are fully aware that winds over the Southern Ocean play an important role in affecting the AMOC. However, in this simple box model, it is challenging to parameterize the effect of perturbation winds over the Southern Ocean on the AMOC.

We would like to emphasize once again that we are using a linearized equation to identify the ***eigenmodes*** on a centennial-millennial timescale. This approach means that the linearized system has excluded external forcings. However, the mean background of the system, particularly **the mean AMOC ($\overline{q}$)**, can be considered to **have included the effects of mean wind forcing** over the Southern Ocean. It is well recognized that both buoyancy forcing in the North Atlantic and wind forcing over the Southern Ocean are critical to maintaining the climatological AMOC, and their perturbations are important to AMOC anomalies. However, for the purposes of this study to find the ***eigenmode*** of the system, both the anomalies of buoyancy forcing in the North Atlantic and wind forcing over the Southern Ocean are not considered.

2) *The way the wind is included is not compatible with mass conservation. In any zonally average models, including box models, the flow must be compensated at depth. Meaning that the wind-induced flow from the subtropical to the subpolar surface boxes should be exactly compensated by a flow from the subpolar to the subtropical deep boxes. Alternatively, and more realistically, the flow is compensated within the surface boxes, which suggests its absence, in this from, in a box model. Not conserving mass in such a simple model is extremely worrying.*

Responses: Thank you for your comments. Mass conservation must be satisfied when including wind-driven circulation in this study. We have been mindful of this requirement since the beginning of our research. The mass is indeed conserved in our model system. However, we acknowledge an error in Fig. 1b of the original manuscript, where the wind-driven return flow was not marked. In the revised Fig. 1b, the wind-driven poleward flow and equatorward flow are now depicted as green solid and dashed arrows, respectively, both of which are assumed to occur in the upper ocean—the region influenced by wind forcing. In this study, we assume the poleward wind-driven transport in the tropics is mainly the Ekman transport, while the equatorward return flow is the geostrophic flow, they are roughly balanced with each other in the tropical upper ocean.

The consideration of mass conservation due to wind forcing is reflected in Eq. (7). Taking temperature as an example, for the subpolar boxes 1 and 3, the wind-driven poleward heat transport is expressed as $+q_n(T_2 - T_1)$ in Eq. 7a1 and $+q_s(T_2 - T_3)$ in Eq. 7a3, respectively. For the equatorial box 2, the equatorward heat transport due to the return flow is expressed as $[-q_n(T_2 - T_1) - q_s(T_2 - T_3)]$ in Eq. 7a2. For the upper oceans, the total mass is conserved within each box, and the total heat transport due to wind-driven circulation is also conserved, which can be easily verified by summing Eqs. 7a1 to 7a3.

In the revised manuscript, we explicitly state that both mass transport and buoyancy transport due to wind-driven circulation are conserved. Additionally, we clarify that it is the wind-driven meridional overturning circulation, rather than horizontal gyres, that is being considered.

3) *The discussion on the effect of temperature on the wind is absolutely incorrect. First of all equation (11) is not built for this purpose. More importantly the "thermal wind relation" does not relate ocean temperature with atmospheric wind but ocean temperature with ocean flow. This is a fundamental misunderstanding. Also \tau^x is not a wind amplitude but a wind stress, which is further away from the "thermal wind relation". This confusion impacts all the results of the paper.*

Since this is at the core of this incremental paper (when compared at LY22 and YYL23). This on its own make me recommend this work to be rejected.

Responses: Thank you for your comments. We think it is correct that the wind-driven meridional mass transport can be appropriately scaled to be proportional to the meridional temperature gradient.

In the tropical-subtropical ocean, the wind-driven meridional volume transport can be roughly scaled as the Ekman transport, which is proportional to the zonal surface wind stress, i.e., $V_E = -\tau^x/\rho f$. If we further assume that, the surface zonal wind stress is mainly caused by the meridional gradient of the sea level pressure, which is, in turn, caused by the meridional surface air temperature based on thermal wind relation, we will be able to connect the wind-driven meridional volume transport to the meridional temperature gradient in the upper ocean. That is, $V_E \sim \tau^x \sim \frac{\partial}{\partial_y}(T) \sim (T_2 - T_1)$. Eq. (11) in the original manuscript is revised, and in fact the whole paragraph has been carefully rewritten in the revised manuscript. The detailed derivation is as below.

The wind-driven volume transports can be roughly scaled as the Ekman transport, which is proportional to the zonal surface wind stress, that is,

$$V_E = -\frac{\tau^x}{\rho_o f} \tag{1}$$

where τ^x is the zonal surface wind stress, ρ_o is ocean density and f is the Coriolis parameter.

In the atmosphere, the thermal wind balance describes how changes in temperature with latitude affect wind patterns with height. In the zonal direction, the thermal wind equation is given by (Vallis, 2017):

$$\frac{\partial u}{\partial z} = -\frac{g}{fT_0}\frac{\partial T}{\partial y} \tag{2}$$

Where, u is the zonal wind velocity, z is the height, g is the gravity, T_0 is a reference temperature and $\frac{\partial T}{\partial y}$ is the meridional temperature gradient.

Apply (2) near the surface ($z = 0$), the geostrophic balance approximates the zonal wind (u) at the lower atmospheric boundary layer to be related to the temperature gradient:

$$u \approx -\frac{gh}{fT_0}\frac{\partial T}{\partial y} \tag{3}$$

where h could represent the depth of the atmospheric boundary layer. This expression indicates that the meridional temperature gradient drives changes in the zonal wind speed.

In the ocean, the surface currents are driven by wind stress through Ekman dynamics. The zonal wind stress τ^x at the ocean surface can be expressed as:

$$\tau^x = \rho_a C_D |u| u \tag{4}$$

where ρ_a is the air density, C_D is the drag coefficient.

If we further assume that the surface air temperature and sea surface temperature are similar, the connection between the zonal wind and meridional gradient of SST can be thus established:

$$\tau_x \approx -\frac{\rho_a g h C_D |u|}{T_0 f}\frac{\partial T}{\partial y} \sim -\alpha \frac{\partial T}{\partial y} \tag{5}$$

Overall, we emphasize that the two-hemisphere box model is essential for identifying the physical mechanism of the AMOC multicentennial oscillation, even though it may yield results similar to the one-hemisphere box model. Moreover, the two-hemisphere model allows us to address issues that cannot be explained by the one-hemisphere model. The effect of wind-driven meridional overturning circulation can only be discussed within the two-hemisphere box model, a topic that has not been explored in previous studies. In fact, the results related to wind-driven circulation are very interesting and non-trivial. We even derived a theoretical solution for the effect of wind-driven circulation.

We have rewritten the entire paper carefully to eliminate any potential misunderstandings. We greatly appreciate the reviewer's comments, which have significantly helped us improve the manuscript.

References:

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

Vallis, G. K., and R. Farneti, 2009: Meridional energy transport in the coupled atmosphere-ocean system: scaling and numerical experiments. Quart. J. Roy. Meteor. Soc., 135, 1643-1660.

Vallis, G. K., 2017: Atmospheric and oceanic fluid dynamics. Cambridge University Press.

Specific Comments:

1. *L.23-24: Most of the ingredients for the self-sustained oscillation already existed in the literature. To the best of my knowledge, from a theoretical point-of-view, salinity feedback is well-established (e.g., Stommel, 1961; Marotzke, 1996), the centennial oscillation and the role of the salinity feedback is also well-established (e.g., Sévellec et al, 2006), the role of convection as a "damping" mechanisms (by limiting density anomaly is also well-established. Hence I find the first sentence of the abstract quite surprising, to say the least. What the authors previously established is the possibility of maintaining the oscillation between a growing and a damping regime. Hence the authors gave a new insights in the theory of self-sustained centennial oscillation, but they are far from "establishing a theory" on their own. I remind the authors that the papers I cited are 60, 30 and 20 years old.... It is absolutely unacceptable to read on a first sentence of the abstract*

Responses: Thank you for your comments. Please refer to our replies to Major comments #1 and #2.

We acknowledge that some previous studies have contributed to the theory of self-sustained centennial oscillation. In our series of studies, we aim to establish a theory that is as simple as possible to explain centennial-millennial variability during the Holocene, which has not been achieved adequately in previous research. Additionally, for the multicentennial timescale we are concerned with, emphasizing the role of convection as a damping mechanism does not seem to be a good idea.

We have also carefully rewritten the entire paper, including the abstract, to eliminate any potential misunderstandings.

2. *that something already established by others 20 to 60 year ago has "been recently established" by the authors. A more decent sentence should be: "Previous studies have [...]" and citation of the vast literature on the topic should be properly done and acknowledged.*

Responses: Thank you for this comment. We have rewritten the abstract and introduction. More previous studies are introduced with more details.

3. *L.28-29: This is the same as Sévellec et al (2006), which also tested two-hemisphere circulation (in a range of models, including a theoretical derivation in an idealized model).*

Responses: Please refer to our reply to Major comment #2. Sévellec et al. (2006) explored the centennial oscillation of the AMOC in a 2D model and a loop model. However, there are some differences between their study and ours.

First, our box model is significantly simpler than the model used in Sévellec et al. (2006), making it possible to obtain a theoretical solution. Second, the centennial mode in Sévellec et al. (2006) is either strongly damped or unstable, leading to very irregular centennial oscillations. This instability necessitates external forcing, either to suppress the unstable mode (via freshwater flux) or to excite the damped mode (through stochastic forcing). However, in our series of studies, we explicitly identified a multicentennial mode with a very weak growth (or damping) rate, meaning the growth or damping timescale is much longer than the oscillation timescale. Only this type of eigenmode is physically meaningful and can survive in the real world.

In the revised manuscript, we have rewritten the abstract to emphasize these differences. Additionally, we have included a comparison of the theoretical solutions between our research and that of Sévellec et al. (2006) in the discussion section.

4. *L.61-63: There is no debate. It is well established that it is a salinity feedback (Marotzke, 1996). In this context the salinity is deemed to grow and the potentially oscillation to occur. There is no need to know "where it come from" but to know "what create it" (i.e., the salinity feedback). This sentence suggests a misunderstanding of the state-of-the art knowledge on ocean dynamics (including box-model dynamics).*

Responses: Thank you for your comments. Please refer to our reply to Major comment #2. The question we are addressing is very different from that in Marotzke (1996), although the box models used are fundamentally similar.

5. *L.74-75: Once again you are ignoring the vast theoretical literature on the topics.*

Responses: Thank you for your comments. We are fully aware of vast previous theoretical studies, but only the most relevant studies are cited in this work.

6. *L.76-86: Even if recent, these authors were not the first to study this. Even in LY22, it is acknowledged that the criteria for oscillation was already established 20 years ago by Sévellec et al (2006).*

Responses: Thank you for this comment. We fully acknowledge the contribution of Sévellec et al. (2006) and recognize that their research is of high quality. However, as stated in our response to Major Comment #2, their model incorporates both temperature and salinity, which results in a solution that is not sufficiently clean and clear. This issue actually motivated us to initiate our own series of studies.

7. *L.88-89: This is indeed annoyingly incremental. I feel that there was a missed opportunity to incorporate some of the 2-hemisphere result in the previous papers. But on its own, this new study does not grant a new publication, especially after two previous papers on extremely close topic, and the lack of fundamentally new insight with the addition of the southern hemisphere and subtropical wind. Overall, the oscillation remained the oscillation: so that the 1-hemisphere study was enough to grasp the key elements, which is what we asked from a theory.*

Responses: Thank you for the comments. It is true that the core physical mechanisms of the AMOC multicentennial oscillations in both the one-hemisphere and two-hemisphere box models are very similar. However, applying a two-hemisphere box model to study multicentennial oscillations is still meaningful for the following reasons:

First, the similarity in the core physical mechanisms between the one-hemisphere and two-hemisphere box models is itself an important result that requires verification.

Second, we have found that the period and stability of multicentennial oscillations are strongly influenced by basin geometry and the division of the upper ocean, lower ocean, and ocean box. The two-hemisphere model, which includes more boxes, offers a greater number of degrees of freedom, leading to different oscillatory behaviors of the AMOC compared to one-hemisphere models. These differences can also help us understand discrepancies between idealized models and coupled models.

Additionally, the effect of wind-driven meridional overturning circulation can only be explored within the two-hemisphere box model—a topic that has not been addressed in previous studies. In fact, our results related to wind-driven circulation are very interesting and non-trivial. We derived a theoretical solution for the effect of wind-driven circulation, which we believe provides new insights into the role of wind-driven meridional overturning circulation in very low-frequency climate variability.

8. *L.95-96: The use of box model for studying the role of the wind is extremely questionable... How to properly constrain the balance of eddy-turbulent-flow mean-Ekman-circulation key for the role of the wind for the AMOC? Is box model the right tool to do that. Historically, this is not by box models that this was theoretically tackled (Toggweiler and Samuels, 1995; Nikurashin and Vallis, 2012)*

Responses: Thank you for your comments. Please refer to our reply to your comment on the 1st page.

We agree that traditionally, the box model is used to study thermohaline circulation. However, we have found that the wind-driven meridional overturning circulation (WD-MOC), including the mean Ekman flow, can also be parameterized within the box model. On the other hand, eddy-turbulent flow is too complex to be considered in a box model. We have rewritten the relevant paragraph to emphasize that in this paper, we focus solely on discussing the large-scale WD-MOC, particularly the subtropical cell.

9. *L.101-104: This is hard to accept that a box model is significantly more realistic than another simpler, box model that roughly show the same results.*

Responses: Thank you for the comment. We have deleted the related expression.

10. *L.130-138, Eqs.1: You need an equation for q as well. This is key. It represents the momentum equation. Such a system without the momentum is ill-posed. Also salinity feedback cannot occur without S acting on q.*

Responses: Thank you for the suggestion. $\overline{q}$ is given as the constant and q' is expressed in Eqs. (4)-(5).

11. *L.145: This makes absolutely no sense. Do not set q solve q!* Please refer to reply above.

12. *Tab.1: Parameter values and equilibrated values should be clearly separated. For instance in such box model it is traditional to solve q, but here it is named as parameter (i.e., "mean" and not "equilibrium")... All these is unclear.* Revised as suggested.

13. *Eq.4: Why having a momentum for anomaly but not for the equilibrium. This is not "classical" and nowhere explains.*

Responses: Thank you for your comment. Please refer to our reply to Major comment #1 and #2.

We are addressing a small perturbation problem within a very stable thermohaline circulation, rather than focusing on issues of stability, bifurcation, or regime shifts in the thermohaline circulation itself. Consequently, we are dealing with an anomalous momentum and a prescribed equilibrium. This is indeed not a "classical" problem related to thermohaline circulation.

14. *L.174: Not at all, this is the conservation of salinity. You do not seem to fully understand what is going one on such simple system. This degenerated eigenvalue allows for an overall shift of the mean salinity without changing the dynamics.*

Responses: Thank you for your comment. We must avoid overinterpreting the physics of these two modes. It is correct to state that $\omega = 0$ indicates no oscillatory behavior in the system and implies an equilibrium state. However, more information would be required to assert that $\omega = 0$ represents "an overall shift of the mean salinity without changing the dynamics." Indeed, in this system, the conservation of salinity must be maintained for any ω, not just for $\omega = 0$.

15. *Fig.3: I am not convinced at all that the growth has stopped. Even if weaker that k=0, S1' seems to be growing and other indices (b-d) suggest the same. You need to make simulations at least 10 times longer to establish statistical equilibrium.*

Responses: Thank you for your comments. Figure 3 has been replotted in the revised manuscript, extending the time series to 8,000 years to more clearly observe the equilibrium state. The primary reason we initially plotted only 4,000 years in the original Figure 3 was to better illustrate the phase differences in salinity and the related terms in different boxes.

16. *L.297-299: Ok, so no need for this study what is the gain of a slightly more complex theoretical model. The simpler are by definition better theory.*

Responses: Thank you for your comments. We would never fully understand the role of the South Atlantic in AMOC multicentennial oscillations without using a two-hemisphere model. We believe this is a new insight. Only after studying the two-hemisphere box model can we confidently state that the one-hemisphere model is sufficient for studying the AMOC. However, this does not imply that the simpler one-hemisphere model is a better choice.

17. *Section.4: Representing the action of the wind in the context of the AMOC solely as a subtropical cells and completely missing the Southern Ocean component is not acceptable.*

Responses: Thank you for your comments. In this simple box model, we can currently only parameterize the subtropical cell (i.e., WD-MOC). We do not yet know how to parameterize wind forcing over the Southern Ocean. We are actively working on this issue using a 3-D planetary wave ocean model and hope to gain insight into the role of the Southern Ocean.

Please also refer to our reply to Major comment #3.1.

18. *L.303-305: ??? Not at all ! Ocean temperature gradient can induced atmospheric temperature gradient that can then lead to wind. But this is an indirect effect/feedback. The main characteristic of the wind is that it is a direct forcing of the atmosphere on the ocean.*

Responses: Thank you very much for your comment. Please also refer to our reply to Major comment #3.3.

19. *L.308-311: The more important wind in the context of the AMOC in the Southern Ocean. I suggest the authors to read description or "theory" of the AMOC. This is well-established knowledge.*

Responses: Thank you for your comments. We fully recognize the importance of the wind over the Southern Ocean. However, it is just not simply to parameterize the Southern Ocean wind effect in the box model. Please also refer to our reply to Major comment #3.1.

20. *Eqs.7: Here the return flow of the subtropical cell is not acknowledged, so that qn and qs does not conserve mass (i.e., they are not a baroclinic transport). This is not consistent with box-model equations. This makes the study inconsistent.*

Responses: Thank you for your comments. Please refer to our reply to Major comment #3.2.

21. *L.372-384: This does not make any sense. In particular the last part on the thermal-wind relation demonstrates a lack of understanding of what it is (which is basic ocean physics knowledge). Wind is not proportional to ocean meridional temperature gradient because of thermal-wind relation !!! This is not acceptable in a manuscript submitted to Journal of Climate.*

Responses: Thank you for your comments. Please refer to our reply to Major comment #3.3.

22. *Section.4b: There is not point to that. On 1st order wind is constant (so anomaly will only passively move around the wind-driven circulation) and here your mass flow is not conserved !!!*

Responses: Thank you for your comments. We would like to emphasize again that the mass is conserved in our model. Please also refer to our reply to Major comment #3.2.

Anomalous wind leads to anomalous circulation, and the two are positively correlated. This relationship can be considered as the physical mechanism explaining why there are no oscillatory behaviors when only the wind-driven subtropical cell is considered. Eqs. 14-17 in Section 4b mathematically prove this point. We believe this is a significant new insight.

23. *Section.5: I do not see the point of using a stochastic forcing in the case of sustained variability and theoretical analysis. The hypothesis that is tested here is never discussed. I do not see any value in this section.*

Responses: Thank you for your comment. We believe the value of this section lies in that, demonstrating the effect of stochastic forcing can help identify the eigenmode of the system: the resonant peak corresponds to the position of the eigenmode. This approach is commonly used in many studies (e.g., Griffies and Tziperman, 1995).

The eigenmodes of the box model can be obtained either theoretically (e.g., in LY22) or numerically (e.g., in this work). However, in the real world, the eigenmodes in a linear system cannot be directly observed unless they are excited by additional processes. Stochastic forcing provides an approach to examine whether an eigenmode can be observed and whether it is physically meaningful and practically useful.

We want to emphasize that the eigenmodes are either damped, unstable, or neutral, meaning that they cannot exhibit sustained oscillations on their own. (A neutral mode is not a self-sustained mode.) To achieve sustained oscillation, either external forcing or additional internal processes (e.g., weak nonlinear processes related to mixing) are required. Usually, stochastic forcing can turn a weakly-damped mode into a sustained oscillation, while internal processes, such as nonlinear mixing, can transform a weakly-unstable mode into a self-sustained oscillation. It's important to note the distinction: "self-sustained" oscillation is driven by internal processes, while oscillation excited by external forcing can only be termed "sustained."

In general, we would like to emphasize that in studies of linear stability problems, stochastic forcing is commonly used to examine the robustness of the eigenmode.

References: *Griffies, S. M., and E. Tziperman, 1995: A linear thermohaline oscillator driven by stochastic atmospheric forcing. J. Climate, 8, 2440-2453.*

24. *L.546-549: It is well established (see previous publications including yours) that the period is controlled by the mean circulation (i.e., the renewal of water: circulation*volume of the ocean).*

Without circulation there is no "natural" period, so the spectrum does not show a peak. Once again, this result is trivial and does not grant publication.

Responses: Thank you for your comments. It is true that many previous studies have specified that the period of the AMOC multicentennial oscillation is controlled by the mean strength of the AMOC, and that there is no natural oscillation without the mean circulation. Qualitatively, the period is similar to the turnover time of an ocean basin. However, there has been no analytical solution to the period until now. In this study, we derived an analytical solution for the oscillation period and found that the period is not simply determined by the ocean volume, but is also significantly influenced by how the ocean boxes are defined.

The oscillation period was obtained considering only thermohaline circulation. No oscillation occurs in the presence of wind-driven circulation alone. The AMOC multicentennial oscillation is a result of thermohaline circulation. We tested a scenario where the mean AMOC persists, but there is no anomalous AMOC (green curves in Fig. R1). In this case, no peak appears in the spectrum.

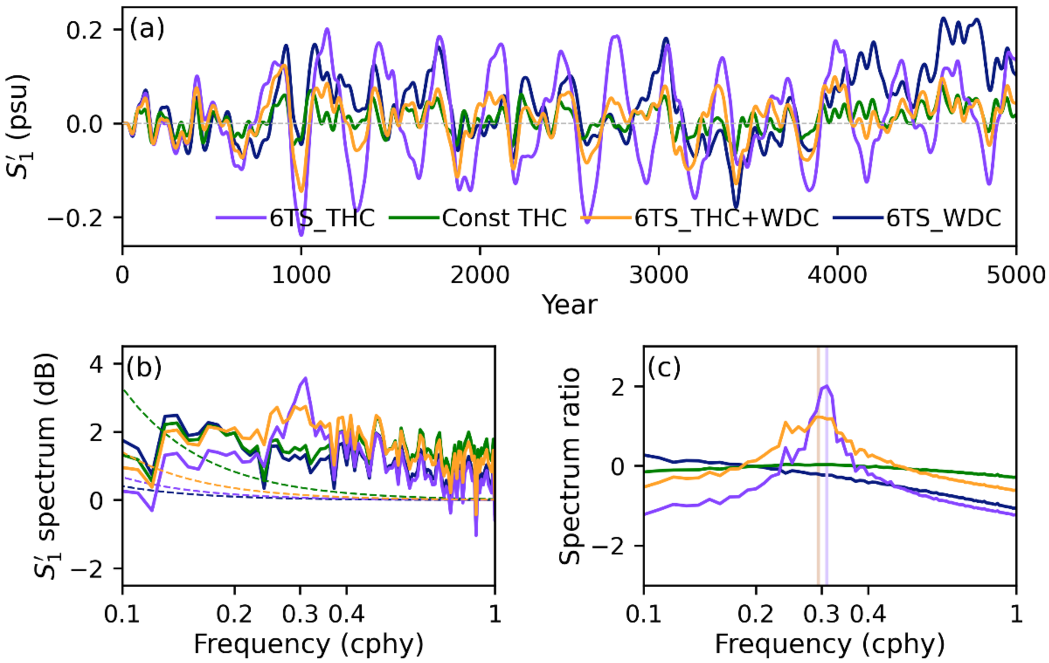

Fig. R1. Time series of (a) S_1' in the 6TS_THC model, 6TS_Const_THC, the 6TS_THC+WDC model, and the 6TS_WDC model, forced by stochastic freshwater and heat flux. λ is set to 21.0 and 23.0 Sv kg^{-1} m^3 for the 6TS_THC and 6TS_THC+WDC modes, respectively; and damped oscillatory modes are obtained in the presence of the thermohaline circulation. Other parameters take the values in Table 1. (b) The power spectra of S_1' for three cases with the confidence level 95%. (c) The ratios of S_1' spectrum to the noise spectrum (units: dB), with peaks around 0.31 and 0.29 cycles per a hundred year (cphy) (320 and 340 years) for the 6TS_THC and 6TS_THC+WDC modes, respectively. Colored curves are noted in panel (a).

25. *L.606-609: The shape of DO-events is not an harmonic oscillation (as your MCO is). Please read the literature regarding this point before making such claim.*

Responses: Thank you for your comments. It is true that, in most proxy data and references, the Dansgaard-Oeschger (DO) event is considered non-harmonic (and in some literatures, it is referred to as the “DO cycle”). However, the timescale of the DO event suggests that there may be a connection between the millennial mode in our study and the DO cycle, despite their differences. At this stage, we can only speculate that the millennial mode could provide a background for DO events. Additionally, some studies, such as Ganopolski et al. (2001), suggest that the DO event resembles a harmonic oscillation more closely in the Southern Hemisphere. Therefore, we believe it is valuable to explore millennial modes as a potential means to help explain the DO event.

References: Ganopolski, Andrey, and Stefan Rahmstorf, 2001: Rapid changes of glacial climate simulated in a coupled climate model. Nature, 409(6817):153-8.

References:

1. *Marotzke, J., 1996: Analysis of thermohaline feedbacks, in Decadal Climate Variability: Dynamics and Predictability, D. L. T. Anderson and J. Willebrand, eds., NATO ASI Series, 44, 333-378.*
2. *Nikurashin, M., and G. Vallis, A Theory of the Interhemispheric Meridional Overturning Circulation and Associated Stratification, J. Phys. Oceanogr., 42, 1652-1667.*
3. *Sévellec, F., T. Huck, and M. Ben Jelloul, 2006: On the mechanism of centennial thermohaline oscillations, J. Mar. Res., 64, 355-392.*
4. *Toggweiler, J. R., and B. Samuels, 1995: Effect of Drake passage on the global thermohaline circulation, Deep Sea Research, 42, 477-500.*
5. *Stommel, H., 1961: Thermohaline convection with stable regimes flow, Tellus, 13, 224-230.*

Replies to Reviewer #2:

Thank you very much for all of your constructive comments. We have carefully revised our manuscript based on the advice by you and other reviewers. The following are our point-by-point replies.

Due to the close relationship between changes in deep ocean circulation and climate change, such as the multi-century variations in the Atlantic meridional overturning circulation and multi-century climate changes, this is a highly researched and prominent topic in physical oceanography. The concept of thermohaline circulation is deeply rooted in physical oceanography and continues to dominate the study of deep ocean circulation, as is the case in this article, which still follows the traditional view of thermohaline circulation. However, the dynamic mechanisms of thermohaline circulation are a topic of debate, and nowadays, the meridional mass stream function on vertical planes is commonly used to describe deep ocean circulation, known as the meridional overturning circulation, referred to as the oceanic mass circulation by Wunsch (2002). This approach avoids the challenging question of what drives the circulation.

The authors attempt to explain the multi-century oscillations of the Atlantic meridional overturning circulation using a box model, and while the mathematical derivations are sound good, some of the physical explanations are not very clear and require further clarification. Specifically, these include:

Responses: Thank you very much for your comment. It is right that in this study we avoided the challenging question of what drives the thermohaline circulation, and address a different question: what causes the multicentennial-scale internal variability **($\boldsymbol{q'}$),** of the thermohaline circulation, under a highly stable mean climate **($\boldsymbol{\overline{q}}$)**. $q' \ll \overline{q}$, allowing the system to be linearized. Linear stability analysis can thus be used to solve this low-frequency variability problem and identify the eigenmodes. This question is relevant for understanding internal climate variability since the mid-Holocene, during which the Earth's climate has been very stable. A theory specifically addressing climate variability during the Holocene has not yet been well established.

Major issues:

1. *Stommel (1961) published a groundbreaking seminal article, postulating that deep ocean currents are driven by temperature and salinity differences between high and low latitudes. Because the density of seawater is determined by its temperature and salinity, density gradients in seawater propel the flow of deep ocean currents. This article still builds upon Stommel's theoretical framework, as Carl Wunsch (2002) has noted the commonalities in such research:*

"When Stommel (1961) first introduced the term 'thermohaline circulation' in a box model, he explicitly provided a source of mechanical energy in the form of mixing devices. These devices disappeared in subsequent discussions and extensions of this influential model." "Returning the downwelling mass flux upward across the stable stratification requires a finite amount of work, manifested as the turbulent mixing carrying dense fluid across the density gradient. The only possible sources of this work are tidal stirring and the wind field (Munk and Wunsch, 1998; Paparella and Young, 2002). And Paparella and Young (2002) have shown that a convective mode of motion cannot generate the turbulence required to carry the meridional overturning circulation across the stable stratification." Then, Wunsch believed that the ocean's mass flux is sustained primarily by the wind, and secondarily by tidal forcing. Both in models and the real ocean, surface buoyancy boundary conditions strongly influence the transport of heat and salt, because the fluid must become dense enough to sink, but these boundary conditions do not actually drive the circulation. Numerous studies have pointed out the shortcomings and deficiencies in Stommel's theory, for example: De Boer et al. (2009) indeed showed that the meridional density gradients do not control the Atlantic meridional overturning circulation by scaling analysis and numerical modeling. How do the authors explain this?

Responses: Thank you for your comments. We are fully aware that wind forcings, particularly wind-driven Ekman pumping in the Southern Ocean, play important roles in driving thermohaline circulation. While Stommel's pioneering theory has its deficiencies, it has inspired the search for more fundamental mechanisms that control or drive thermohaline circulation.

We agree with De Boer et al. (2009) that the meridional density gradient does not control the AMOC. Our understanding is that, for the mean AMOC, both buoyancy forcing in the North Atlantic and wind forcing over the Southern Ocean are critical. This view is widely accepted within the physical oceanography community. Of course, changes in the AMOC can be driven by various factors, including remote forcings (wind + buoyancy) and local forcings (wind + buoyancy). However, it is also generally agreed that perturbations in the meridional density gradient can lead to significant changes in the AMOC, which is a key assumption in our work.

We would like to emphasize once again that we are not studying what controls or drives the thermohaline circulation, and what factors can lead to the **instability**, **bifurcation**, and **regime shifts** of thermohaline circulation. Instead, we focus on what internal processes can cause the oscillation of thermohaline circulation at centennial-millennial timescale. Therefore, we use a linearized equation to identify the ***eigenmodes*** on a centennial-millennial timescale. This approach means that the linearized system has excluded external forcings. However, the mean background of the system, particularly **the mean thermohaline circulation ($\overline{q}$)**, can be considered as to **have**

included the effects of mean wind forcing over the Southern Ocean. However, for the purposes of this study to find the ***eigenmode*** of the system, both the anomalies of buoyancy flux in the North Atlantic and wind forcing over the Southern Ocean are not considered.

Additionally, our findings indicate that wind-driven circulation, specifically the wind-driven subtropical cell, does not favor the multicentennial oscillation of the AMOC.

In general, while results from our highly idealized box model are valuable in helping us understand certain aspects of reality, they do not capture the full complexity of the real world.

2. *In the fourth section, the description of the wind-driven circulation does not distinguish between horizontal and vertical components, as authors conflate the two. The horizontal component indicates wind-driven gyre, Vallis and Farneti (2009) study of meridional energy transport, which encompasses the vertical wind-driven meridional overturning circulation, as described by Wunsch (2002) in terms of the oceanic mass circulation for energy transport. Clearly, the wind acts as an external forcing on the ocean, which contradicts the notion of self-sustainability. Can the author provide clarification?*

Responses: Thank you very much for your comments. In this work, the wind-driven circulation specifically refers to the subtropical cells (STCs), and we examine the meridional overturning component rather than the horizontal gyre of the STCs. The strength of the STCs is positively correlated with the strength of the wind forcing, and therefore with the meridional gradient of oceanic temperature. This correlation is the physical mechanism explaining why there is no self-sustained oscillation when only wind-driven circulation is present. The theoretical solution to the box model with only wind-driven circulation, as presented in Section 4b, mathematically proves this point.

In the revised manuscript, Section 4 has been carefully rewritten for clearer description and understanding.

3. *In the paragraph starting at Line 622, authors mentioned the issue of freshwater, which is not the focus of this study but is worth noting. Bryan (1986) presented three modes of the Atlantic meridional overturning circulation through numerical simulations, showing that only changes in salinity can switch the circulation state, while temperature has no effect. The impact of salinity or freshwater, especially in numerical simulations, is now a consensus, with the two being like two sides of a coin. However, the Stommel's box model cannot explain why only changes in salinity, and not temperature, can alter the state of the Meridional Overturning*

Circulation. This limitation arises from the buoyancy assumption of the Stommel box model. Guan and Huang (2008) found that changes in salinity can alter the state of the Meridional Overturning Circulation when driven by energy provided by winds, while temperature cannot. This finding also supports the viewpoint that external mechanical energy drives the Meridional Overturning Circulation, as indicated in Question 1. Linear stability analysis shows that the buoyancy assumption of the Stommel theory considers temperature to have stable and unstable states, while salinity only has one stable state, which is inconsistent with Bryan's numerical simulations. Under the mechanical energy driving assumption, temperature has only one state, while salinity has stable and unstable states. When salinity changes, the Meridional Overturning Circulation can undergo transitions between states, providing a good interpretation of Bryan's numerical simulation results. What considerations do authors have regarding these findings?

Responses: Thank you very much for raising these interesting questions. As a discussion, we would like to try to present our thoughts here.

First of all, we do not think that the Stommel's theory is inconsistent with Bryan's numerical results. The AMOC in the present-day is in its thermal regime, i.e., the current mean AMOC is determined mainly by the mean meridional temperature distribution in the Atlantic and strong Ekman pumping in the Southern Ocean, since the mean meridional salinity distribution does not favor the deep-water formation in the North Atlantic. We think this is one of key points in Stommel model.

Second, in terms of changes in the AMOC, in the low temperature regime (such like in the subpolar ocean), change in subpolar salinity is much more efficient than change in the subpolar temperature, to alter the AMOC. Particularly, when hosing the freshwater into the subpolar ocean, the AMOC can be changed easily from its thermal regime to so-called "saline regime". This is idea that Byran's work want to communicate, that is, "only changes in salinity can switch the circulation state, while temperature has no effect". We also think, at this point, the Stommel model can explain that only changes in salinity, and not temperature, can alter the state of the AMOC.

Third, in terms of mechanical energy provide by winds, we do not think the fundamental idea in Guan and Huang (2008) is different from Byran (1986)'s. The wind can change the ocean surface evaporation, so to change the sea surface salinity (SSS). The wind can also change the ocean latent and sensible heat fluxes, so to change the SST. In the subpolar ocean, once again, it comes to that change in SSS is much more efficient than change in the SST, to alter the AMOC. The only difference is energy source of the SSS change: mechanical energy by winds or heating by the Sun.

Stommel-like model shows clearly that there can be the multiple equilibria of the AMOC. Boundary conditions (restoring, flux or mixed), external forcings, model parameters, and the sensitivity of the AMOC to meridional buoyancy gradient and etc., all could lead to the regime shift of the AMOC. For certain degree, we understand why there is a major difference between the Stommel-like model and the energy-constraint model of Guan and Huang (2008), i.e., the difference in bifurcation structure. Stommel-like model has two thermal modes (one stable and another one unstable) and one stable saline mode, whereas the energy-constraint model has one stable thermal mode and two saline modes (one stable and another one unstable). The major physics and the major forcings are quite different in these two models, the ways to perturbing the system are also very different, how can we anticipate the same results? We think the fundamental reason is that the temperature and salinity have opposite (or compensation) effect on the density, which brings big uncertainty in the AMOC regime.

Although Stommel-like model has been explored extensively, there remains lots of interesting questions related to the AMOC to be investigated further. This study examines the internal variability of the AMOC under a stable mean climate, and we hope it will deepen our understanding of the possible behaviors of the AMOC and its possible impact on the global climate.

Minors:

1. *Line 941-941: This reference is incomplete.*

Responses: Thank you for pointing out this problem. We have revised it.

2. *Line 950, first name should be GAO and YU.*

Responses: Thank you for pointing out this problem. We have revised it.

References:

1. *Bryan, F., 1986: High-latitude salinity effects and interhemispheric thermohaline circulations. Nature, 323, 301–304.*
2. *de Boer, AM, et al., 2010：Meridional Density Gradients Do Not Control the Atlantic Overturning Circulation. J. Phys. Oceanogr., 40, 368–380.*
3. *Guan, Y. P., and R. X. Huang, 2008: Stommel's box model of thermohaline circulation revisited—The role of mechanical energy supporting mixing and the wind-driven gyration. J. Phys. Oceanogr., 38, 909-917.*

4. *Munk, W. H., and C. Wunsch, 1998: Abyssal recipes II: Energetics of tidal and wind mixing. Deep-Sea Res. I, 45, 1977–2010.*

5. *Paparella, F., and W. R. Young, 2002: Horizontal convection is non-turbulent. J. Fluid Mech., 466, 205–214.*

6. *Stommel, H., 1961: Thermohaline convection with two stable regimes of flow. Tellus, 13, 224–230.*

7. *Vallis, G. K., and R. Farneti, 2009: Meridional energy transport in the coupled atmosphere-ocean system: scaling and numerical experiments. Quart. J. Roy. Meteor. Soc., 135, 1643-1660.*

8. *Wunsch, C., 2002: What is the thermohaline circulation? Science, 298, 1179–1181.*

Replies to Reviewer #3:

Thank you very much for all of your constructive comments. We have carefully revised our manuscript based on the advice by you and other reviewers. The following are our point-by-point replies.

This manuscript is the third part of a study on the self-sustained multicentennial oscillation of AMOC. In the previous parts, the authors used a single-hemisphere model that includes the temperature effects or not, and found that enhanced high-latitude mixing, or nonlinearity are essential for self-sustained multicentennial oscillation to occur. In this work, the authors proceed to use a more realistic two-hemisphere model with or without wind- driven circulation, and study how different parameters, such as model geometry, and contribution from wind driven circulation alter the behavior of this oscillation. This work is an interesting and meaningful continuation of earlier papers, and makes the study on AMOC multicentennial oscillation more complete. However, I do have some questions and concerns for the manuscript as it is now, and would recommend a major revision. Please see comments below.

Major Comments

1. *This model for AMOC, seems to assume that the overturning circulation is being pushed rather than pulled (for example, in Appendix A authors studies how North Atlantic Ocean density is correlated with the strength of AMOC). This is fine for the earlier single-hemisphere models. But with a double-hemisphere model, the Southern Ocean dynamics need to be discussed as well (e.g., Wolfe and Cessi, 2010; Nikurashin and Vallis, 2011, 2012). It seems to me that your 3S model in Appendix B is more relevant to the perspective that AMOC is being pulled from the Southern Ocean, only that q in your model is not determined by Southern Ocean dynamics.*

Responses: Thank you very much for your comments. Actually, in this study, we deliberately avoided addressing the challenging question of what primarily drives the thermohaline circulation—whether the AMOC is more pushed by deep-water formation in the subpolar North Atlantic or more pulled by wind-driven Ekman pumping in the Southern Ocean. We fully recognize that both the pulling mechanism due to Ekman pumping in the Southern Ocean and the pushing mechanism due to deep-water formation in the subpolar North Atlantic are critical to the AMOC. This is also a widely accepted understanding within the physical oceanography community.

In this study, we address a different question: what causes the multicentennial-scale internal variability $(\boldsymbol{q'})$, of the AMOC, under a highly stable mean AMOC $(\overline{\boldsymbol{q}})$. $q' \ll \overline{q}$, allowing the system to be linearized. Linear stability analysis can thus be used to solve this low-frequency variability problem and identify the eigenmodes. However, the mean background climate, particularly **the mean AMOC ($\overline{q}$)**, can be considered as to **have included the effects of mean wind forcing** over the Southern Ocean. For the purposes of this study to find the ***eigenmode*** of the linearized system, external forcings, i.e., both the anomalies of buoyancy flux in the North Atlantic and wind forcing over the Southern Ocean, are not considered. This kind of consideration applies to both one-hemisphere model and two-hemisphere model.

2. *For the parameters used in the model (lines 145–146 and table 1), how are they chosen? If they are based on some datasets, what are these datasets? If they are chosen to tune the model, are they compared to any datasets (observations, data assimilation, etc.)? For example, 24 Sv seems too strong for an averaged strength of AMOC (e.g., Lumpkin and Speer, 2007). If the purpose of this paper is to make the mode more "realistic", such comparison will be important.*

Responses: Thank you very much for your comments.

Some parameters, such as equilibrium temperature, salinity, and mean AMOC, are based on the results of CESM1 (Yang et al., 2015), as mentioned in the main text and Appendix 1. The restoring temperatures and surface virtual salt fluxes are calculated from the equilibrium temperature and salinity using Eqs. (8). Certain parameters, such as the thermal expansion coefficient, saline contraction coefficient, and reference seawater density, are commonly used values. The wind-driven advection coefficients for the North Atlantic and South Atlantic are selected to produce reasonable mean mass transports by the wind-driven subtropical cells, a point also mentioned in the text. Parameters related to basin geometry were chosen to tune the model, which might affect the oscillation; thus, the sensitivity of these parameters is discussed in Section 3.b.

The mean AMOC value (24 Sv) is relatively strong compared to those used in other studies. Therefore, we conducted sensitivity experiments to test how the mean AMOC strength influences the oscillation period.

References: Yang, H., Q. Li, K. Wang, Y. Sun, and D. Sun, 2015: Decomposing the meridional heat transport in the climate system. Climate Dyn., 44, 2751-2768.

3. *For the sensitivity tests (like Fig. 4), what is the physical meaning of changing these volumes? An expansion of NADW, or upper ocean thermocline?*

Responses: These volumes significantly affect the oscillation period of the AMOC. Physically, changing $V_1 + V_4$ alters the area of deep-water formation, while changing $V_1 + V_2 + V_3$ modifies the influencing region of the upper branch of the AMOC, both of which are critical to the mass balance of the AMOC, and thus to a certain degree determine the turnover time of water in the Atlantic.

Does it really make sense to change these values in such a large range that maximum volumes may be as five time as large as the minimum?

We believe it makes sense to test the sensitivity over such a large range. The timescale of AMOC variability can shift from multicentennial to decadal as the volume changes from very large to very small. In studies by Griffies and Tziperman (1995) and Rivin and Tziperman (1997), a similar box model was used, and the volume of the subpolar box was chosen to be very small, resulting in AMOC variability on a decadal to multidecadal timescale. The volume of the deep-water formation is indeed a critical factor.

We also think that dramatic changes in the volume of deep-water formation could have occurred in Earth's history, given the significant changes in Earth's climate.

Do you keep other parameters unchanged for this, like when you change V1 + V2 + V3, what happened to V4 + V5 + V6?

In our study, the total volume of the ocean basin is fixed. Therefore, changing $V_1 + V_2 + V_3$ is equivalent to changing $V_4 + V_5 + V_6$.

To me, it is more interesting and important to test sensitivity to other parameters whose changes have a clear physical meaning (such as q, freshwater fluxes, or restoring temperature).

The theoretical solution of the multicentennial oscillation indicates that the volumes are key factors in influencing both the stability and the period of the oscillation. Additionally, our focus is primarily on the eigenmode itself rather than on the influence of external factors such as freshwater fluxes and restoring temperature.

Also, in caption of Fig. 4 you have (a) period and (b) e−folding time, but otherwise in titles for each panel. Maybe this is a typo?

Thank you very much for pointing out this typo. It should be (a) e-folding time and (b) period. The figure caption is revised.

4. *For the model with wind-driven circulation, you claim that the strength of qn and qs is determined by meridional temperature gradient. However, the strength of wind-driven circulation is determined by wind- stress in traditional western boundary current models*

(Stommel, 1948; Munk, 1950). And in equation (11) and in lines 383 it is stated that wind-driven circulation is proportional to τx, which is correct. But the claim that τx is determined by meridional temperature gradient is not correct. If τx is to be connected to some temperature gradient, it should be that of the atmosphere but not that of the ocean. In lines 576–579, the mechanism for oscillation in this model involves a weakened wind-driven circulation as a result of reduced oceanic meridional temperature gradient, also seems wrong.

Responses: Thank you for your comments. We think it is correct that the wind-driven meridional mass transport can be appropriately scaled to be proportional to the meridional temperature gradient.

In the tropical-subtropical ocean, the wind-driven meridional volume transport can be roughly scaled as the Ekman transport, which is proportional to the zonal surface wind stress, i.e., $V_E = \tau^x/\rho f$. If we further assume that, the surface zonal wind stress is mainly caused by the meridional gradient of the sea surface pressure, which is, in turn, caused by the meridional surface air temperature based on thermal wind relation, we will be able to connect the wind-driven meridional volume transport to the meridional temperature gradient in the upper ocean. That is, $V_E \sim \tau^x \sim \frac{\partial}{\partial_y}(T) \sim (T_2 - T_1)$. Eq. (11) in the original manuscript is revised, and in fact the whole paragraph has been carefully rewritten in the revised manuscript. The detailed derivation is as below.

The wind-driven volume transports can be roughly scaled as the Ekman transport, which is proportional to the zonal surface wind stress, that is,

$$V_E = \frac{\tau^x}{\rho_o f} \quad (1)$$

where τ^x is the zonal surface wind stress, ρ_o is ocean density and f is the Coriolis parameter.

In the atmosphere, the thermal wind balance describes how changes in temperature with latitude affect wind patterns with height. In the zonal direction, the thermal wind equation is given by (Vallis, 2017):

$$\frac{\partial u}{\partial z} = -\frac{g}{f T_0}\frac{\partial T}{\partial y} \quad (2)$$

Where, u is the zonal wind velocity, z is the height, g is the gravity, T_0 is a reference temperature and $\frac{\partial T}{\partial y}$ is the meridional temperature gradient.

Apply (2) near the surface ($z = 0$), the geostrophic balance approximates the zonal wind (u) at the lower atmospheric boundary layer to be related to the temperature gradient:

$$u \approx -\frac{gh}{fT_0}\frac{\partial T}{\partial y} \tag{3}$$

where h could represent the depth of the atmospheric boundary layer. This expression indicates that the meridional temperature gradient drives changes in the zonal wind speed.

In the ocean, the surface currents are driven by wind stress through Ekman dynamics. The zonal wind stress τ^x at the ocean surface can be expressed as:

$$\tau^x = \rho_a C_D |u| u \tag{4}$$

where ρ_a is the air density, C_D is the drag coefficient.

If we further assume that the surface air temperature and sea surface temperature are similar, the connection between the zonal wind and meridional gradient of SST can be thus established:

$$\tau_x \approx -\frac{\rho_a g h C_D |u|}{T_0 f}\frac{\partial T}{\partial y} \sim -\alpha\frac{\partial T}{\partial y} \tag{5}$$

We have rewritten the entire paper carefully to eliminate any potential misunderstandings. We greatly appreciate the reviewer's comments, which have significantly helped us improve the manuscript.

References:

Vallis, G. K., 2017: Atmospheric and oceanic fluid dynamics. Cambridge University Press.

Also, in the models without wind- driven circulation, mass conservation works fine. But for the model in Fig. 1b, where do qn and qs go? Mass conservation seems to be violated here.

Mass conservation must be satisfied when including wind-driven circulation. We have been mindful of this requirement since the beginning of our research. The mass is indeed conserved in our model system. However, we acknowledge an error in Fig. 1b of the original manuscript, where the wind-driven return flow was not marked. In the revised Fig. 1b, the wind-driven poleward flow and equatorward flow are now depicted as green solid and dashed arrows, respectively, both of which are assumed to occur in the upper ocean—the region influenced by wind forcing. In this study, we assume the poleward wind-driven transport in the tropics is mainly the Ekman transport, while the equatorward return flow is the geostrophic flow, they are roughly balanced with each other in the tropical upper ocean.

The consideration of mass conservation due to wind forcing is reflected in Eq. (7). Taking temperature as an example, for the subpolar boxes 1 and 3, the wind-driven poleward heat transport is expressed as $+q_n(T_2 - T_1)$ in Eq. 7a1 and $+q_s(T_2 - T_3)$ in Eq. 7a3, respectively. For the equatorial box 2, the equatorward heat transport due to the return flow is expressed as $[-q_n(T_2 - T_1) - q_s(T_2 - T_3)]$ in Eq. 7a2. For the upper oceans, the total mass is conserved within each box, and the total heat transport due to wind-driven circulation is also conserved, which can be easily verified by summing Eqs. 7a1 to 7a3.

In the revised manuscript, we explicitly state that both mass transport and buoyancy transport due to wind-driven circulation are conserved. Additionally, we clarify that it is the wind-driven meridional overturning circulation, rather than horizontal gyres, that is being considered.

5. *I do not understand why it is necessary to look for self-sustained oscillation in a system without thermohaline circulation at all (section. 4b). As implied by the title, the oscillation you are interested in is a behavior of AMOC. In a system without AMOC, even if any multicentennial oscillation is found, it would be irrelevant.*

Responses: Thank you very much for your comments. We aim to address the following question: Does the AMOC serve as a sufficient and necessary condition for the self-sustained multicentennial oscillation? When the AMOC is present, self-sustained oscillation occurs, indicating that the existence of the AMOC is a sufficient condition for the oscillation. Conversely, when the AMOC disappears, the oscillation ceases, suggesting that the AMOC acts as a necessary condition for the oscillation. It is essential to examine both aspects to fully understand this relationship.

Minor comments

1. *The abbreviation "MCO" can be confused with "MOC" too easily, and sometimes it distracts the reader too much to distinguish them.*

Responses: We have replaced all "MCO" with its original form "multicentennial oscillation".

2. *In several places, "AABW" is mentioned and discussed. But this is a model for the upper cell of AMOC (that of NADW), AABW is not included in this model, even implicitly. For example, the density in box 3 and box 6 does not correspond to "AABW", but the upwelling branch of NADW.*

Responses: Thank you very much for your invaluable suggestions. We have revised the related expression in this article.

3. *Line 285 "... has a higher probability to occur". It is not clear what two models or scenarios are being compared here.*

Responses: Here, the sensitivity of the 6S model is compared with that in the 4S model (one-hemisphere box model). We have revised the expression.

References:

1. *Lumpkin, R. and Speer, K. (2007). Global ocean meridional overturning. Journal of Physical Oceanography, 37(10):2550–2562.*
2. *Munk, W. H. (1950). On the wind-driven ocean circulation. Journal of meteorology, 7(2):80–93.*
3. *Nikurashin, M. and Vallis, G. (2011). A theory of deep stratification and overturning circulation in the ocean. J. Phys. Oceanogr., 41(3):485–502.*
4. *Nikurashin, M. and Vallis, G. (2012). A theory of the interhemispheric meridional overturning circulation and associated stratification. J. Phys. Oceanogr., 42:1652–1667.*
5. *Stommel, H. (1948). The westward intensification of wind-driven ocean currents. Transactions, American Geo- physical Union, 29(2):202–206.*
6. *Wolfe, C. L. and Cessi, P. (2010). What Sets the Strength of the Middepth Stratification and Overturning Circu- lation in Eddying Ocean Models? Journal of Physical Oceanography, 40:1520–1538.*