Replies to Reviewer #1:

Thank you very much for these constructive comments. We have revised the manuscript carefully based on these suggestions. The followings are our point-to-point replies.

This manuscript presents a novel conceptual investigation into the potential for self-sustained, multicentennial-scale oscillations of the Atlantic Meridional Overturning Circulation (AMOC) within a two-hemisphere box model framework. Building upon earlier single-hemisphere box models, the authors introduce the thermohaline and wind-driven gyre circulation, and freshwater/heat inputs into a two-hemisphere box configuration. The results shed new light on the possibility of internal, low-frequency AMOC variability driven mainly by the salinity advection feedback in the North Atlantic. The work is highly relevant to theoretical climate dynamics and offers potential implications for understanding long-term Atlantic variability in both past and future climates. The model's simplicity allows the authors to isolate and interpret the mechanisms that are often obscured in more complex GCMs (2 examples are analyzed), while largely reproducing the oscillation modes from them. The manuscript is generally clearly written. To further enhance the manuscript's completeness and clarity, several minor issues regarding model assumptions, physical realism, and broader climate implications should be addressed.

Responses: Thank you very much for your encouraging comments. Combining suggestions from all reviewers, we have revised the manuscript substantially in the following areas:

(1) The section discussing the millennial mode has been removed.

(2) The section titled "Linear Oscillations Excited by Stochastic Forcing" has been removed.

(3) A new section, "Oscillatory Timescales of Multicentennial Oscillations", has been added. This material was previously part of the Discussion section.

(4) Substantial revisions have been made to both the Methods and Discussion sections. In the Methods, we now include a more detailed description of the box model, including the fundamental oceanic processes it represents and the treatment of wind-driven processes. In the Discussion, we have expanded the text to better emphasize the novelty of this study and to clearly acknowledge its limitations.

Specific issues

1. *Oversimplification of ocean geometry (in Table 1) and basin asymmetry. The model treats the Atlantic basin as a symmetric two-hemisphere system with simplified vertical and meridional box divisions. This symmetry ignores key features of the real climate system, such as the configuration of the Southern Ocean and the role of the Agulhas leakage which are known to influence AMOC strength and variability (e.g. Biastoch et al. 2008). The authors can discuss in more details how geometric asymmetries and inter-basin exchange processes (e.g., with the Indo-Pacific; Sun et al. 2021) might alter or inhibit the oscillatory behavior seen in the model.*

Responses: Thank you for your comments. We agree that the highly idealized geometry and dynamics in our model may omit some important influences on the AMOC multicentennial variability. In response, we have expanded the discussion to include the potential roles of the Southern Ocean, Agulhas leakage, and inter-basin exchange processes.

Our model does not simulate Southern Ocean-driven mechanisms seen in more complex coupled models. For instance, in the Kiel Climate Model and GFDL CM2.1 (Park and Latif 2008; Delworth and Zeng 2012), multicentennial AMOC oscillations are triggered by heat or salinity anomalies in the Weddell Sea that propagate to the North Atlantic, influencing North Atlantic Deep Water (NADW) formation. The absence of such signals in our simulation likely reflects the simplified model structure and boundary conditions.

While few studies have explored the influence of Agulhas leakage or Indo-Pacific exchange on AMOC multicentennial variability, they are recognized contributors to shorter-timescale variability. Biastoch et al. (2008), for example, demonstrate that dynamic signals originating in the Agulhas leakage region can modulate decadal MOC variability, with impacts in the North Atlantic comparable to those from the northern source. Sun et al. (2021) further suggest that during warm phases of the Atlantic multidecadal oscillation, changes in sea surface temperatures and atmospheric circulation over the Indo-Pacific may also feedback into the Atlantic system via inter-basin connections.

We now acknowledge these processes more fully in the revised manuscript, while noting that the extent to which they influence AMOC variability on multicentennial timescales remains an open research question. Nevertheless, we believe that despite omitting some key processes, the highly idealized model highlights the central role of oceanic dynamics, which we consider a strength of its simplified design.

2. *The absence of some potentially important atmosphere-ocean-sea-ice feedbacks/processes. While the model includes some key climate feedbacks/processes (such as the salt advection*

feedback, vertical mixing, wind stress SST gradient feedback), it omits other critical atmosphere-ocean-sea-ice couplings/processes. For instance, feedbacks involving sea ice (e.g., insulating effects on surface fluxes, brine rejection; Maccia et al. 2022) are not fully incorporated which can also be important at multicentennial timescales and can fundamentally alter the behavior of the oscillator system. These potential limitations and how they impact the oscillation mechanism should be acknowledged and discussed.

Responses: Thank you for your comments. We appreciate your helpful suggestion to expand the discussion on the limitations of our model and the potential influence of atmosphere-ocean-sea-ice feedbacks on AMOC multicentennial oscillations. These feedbacks are indeed present in the real climate system, and numerous studies have highlighted their importance. In response, we have revised the discussion section to explicitly acknowledge this limitation.

Our model does not include key atmosphere-sea-ice feedbacks identified in recent studies of Arctic-origin AMOC multicentennial oscillations. For example, Jiang et al. (2021) and Meccia et al. (2023) reported that sea ice melt during strong AMOC phases generates freshwater anomalies, which are advected southward and weaken subpolar deep convection. Although these oscillations are primarily governed by mean advection, the maintenance of anomalies likely involves coupled feedbacks not captured in our model. Similar limitations apply to Mehling et al. (2023) and Vellinga and Wu (2004), which emphasize simplified or air-sea-driven mechanisms.

However, we also regard this as a strength of our modeling framework. The exclusion of coupled atmosphere-ocean-sea-ice feedbacks enables us to isolate and demonstrate that AMOC multicentennial oscillations can emerge solely from internal oceanic dynamics. This finding implies the existence of a more fundamental, intrinsic mechanism underpinning such variability, operating independently of the full complexity of coupled climate interactions.

References:

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

Meccia, V. L., R. Fuentes-Franco, P. Davini, K. Bellomo, F. Fabiano, S. Yang, and J. Von Hardenberg, 2023: Internal multi-centennial variability of the Atlantic Meridional Overturning Circulation simulated by EC-Earth3. Climate Dyn., 60, 3695-3712.

Mehling, O., K. Bellomo, and J. von Hardenberg, 2024: Centennial-scale variability of the Atlantic meridional overturning circulation in CMIP6 models shaped by Arctic – North Atlantic interactions and sea ice biases. Geophys. Res. Lett., 51, e2024GL110791.

Vellinga, M., and P. Wu, 2004: Low-Latitude Freshwater Influence on Centennial Variability of the Atlantic Thermohaline Circulation. J. Climate, 17, 4498-4511.

3. *Validation against paleoclimate periodicities. The authors note that the centennial-scale oscillations produced by the model are not clearly observed in some proxy records (e.g., Stocker and Mysak, 1992) and propose several possible explanations for this discrepancy. I encourage the authors to elaborate more fully on these potential factors qualitatively where appropriate, and consider whether some of them could be explored or tested within the framework of the modified model. This would help clarify the model's applicability to real-world climate variability and enhance its relevance to paleoclimate interpretations.*

Responses: Thank you for your comments. We agree that it would be valuable to explore more potential factors such as surface freshwater fluxes in our idealized model framework, as this could help clarify their roles in modulating oceanic oscillations. Such extensions would indeed enhance the model’s relevance to paleoclimate variability.

However, we have chosen not to include additional experiments in the current paper for several reasons. First, the scope of this study is already extensive, and adding new sensitivity experiments would go beyond the paper’s current focus. We plan to investigate these additional factors in future work. We have also expanded the discussion section to address this point more thoroughly.

Second, we believe that idealized models are best suited for isolating and understanding the fundamental dynamics of individual components, rather than reproducing the full complexity of the real climate system. For the latter, fully GCMs are more appropriate tools.

Finally, our main objective in this study is to understand the internal ocean dynamics that give rise to multicentennial oscillations. We therefore focus on how basin geometry and AMOC strength affect the oscillation timescale in the new Section 4. These factors alone can account for most of the variability in the modeled periods.

Editorial comment: L35 The abstract is not complete.

Responses: Thank you for your comments. We have completed the abstract in the revised version.

Ref:

Biastoch et al. (2008). Agulhas leakage dynamics affects decadal variability in Atlantic overturning circulation. Nature, 456(7221), 489-492.

Sun et al. (2021). The importance of inter‐basin atmospheric teleconnection in the SST footprint of Atlantic multidecadal oscillation over western Pacific. Climate Dynamics, 57, 239-252.

Replies to Reviewer #2:

Thank you very much for these constructive comments. We have revised the manuscript carefully based on these suggestions. The followings are our point-to-point replies.

The authors presented a self-sustained oscillatory AMOC in multi-century timescale using a two-hemisphere 6-box model. Their study showed that the AMOC is largely driven by salinity. The AMOC become weak/damped when wind-driven circulation is introduced. The study is very helpful in understanding the AMOC, and can be published after major/minor revision.

Responses: Thank you very much for your encouraging comments. Combining suggestions from all reviewers, we have revised the manuscript substantially in the following areas:

(1) The section discussing the millennial mode has been removed.

(2) The section titled "Linear Oscillations Excited by Stochastic Forcing" has been removed.

(3) A new section, "Oscillatory Timescales of Multicentennial Oscillations", has been added. This material was previously part of the Discussion section.

(4) Substantial revisions have been made to both the Methods and Discussion sections. In the Methods, we now include a more detailed description of the box model, including the fundamental oceanic processes it represents and the treatment of wind-driven processes. In the Discussion, we have expanded the text to better emphasize the novelty of this study and to clearly acknowledge its limitations.

Major issues

My major concerns are (1) Clarity of the parameter definitions and their physical meaning, or how these values are derived based on some kind observations if possible,

Responses: Thank you for your comments. The parameters in our model are selected based on observational data or outputs from climate models. For example, the boundary conditions, including freshwater fluxes and restoring temperatures, are set to values that produce equilibrium temperature and salinity fields closely matching CESM1 output (Yang et al., 2015). Standard physical constants follow widely accepted values, and wind-driven advection coefficients are chosen to reproduce realistic subtropical cell transports. The κ is set to produce the desired amplitude of the self-sustained oscillation. The total basin volume is adjusted to reflect realistic ocean conditions, and the basin geometry parameters are tuned to capture the targeted oscillatory behavior, with their sensitivity discussed in Section 3b. The mean AMOC strength is also based on CESM1 output, and its influence on the oscillation period is examined in Section 4. The linear closure parameter is chosen to obtain a relatively robust oscillation period, with details provided in the manuscript. Hence, all equilibrium values are selected to be consistent with results from climate models. Details on parameter selection are provided at their first mention in the manuscript. A clarifying note has also been added to the caption of the parameter table.

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.

(2) the difference between the current and previous work may need being discussed or clarified,

This study enhances our understanding of the AMOC multicentennial variability by extending earlier one-hemisphere box model analyses to a more physically realistic two-hemisphere configuration. While the fundamental mechanisms governing AMOC oscillations are broadly consistent across both models, the two-hemisphere framework offers several important advances. These include the influence of basin geometry and vertical ocean structure on the period and stability of oscillations, an improved parameterization of AMOC anomalies based on inter-hemispheric density differences, and a theoretical expression linking system properties to oscillation timescales. The study also demonstrates that thermohaline circulation is both necessary and sufficient for generating multicentennial variability, and provides the first analytical explanation for the absence of such variability in the presence of wind-driven meridional overturning circulation (WD-MOC) alone.

Beyond our previous work, some studies have reported multicentennial AMOC oscillations in two-hemisphere three-box models (e.g., Lucarini and Stone, 2005a, 2005b; Scott et al., 1999), but their analyses focus mainly on eigenmode computations and time series outputs, lacking a detailed account of the underlying feedback mechanisms. Moreover, the oscillations identified in these studies are typically damped or unstable, whereas the present work demonstrates a sustained multicentennial oscillation. Some spatially resolved, though still highly idealized, models also exhibit AMOC multicentennial variability. Mysak et al. (1993) identified 200-300-year oscillations triggered by stochastic freshwater forcing, but their analysis focuses solely on the positive salinity-advection feedback and lacked a complete feedback loop explanation. Similarly, Wolfe and Cessi (2015) report ~380-year oscillations characterized by large-scale sea surface salinity (SSS) anomalies of opposite sign in the two hemispheres. However, their analysis primarily focuses on the positive salinity-advection feedback and does not consider the full suite of feedback mechanisms required to sustain such oscillations.

Overall, in contrast to these previous studies, our work provides a systematic and physically interpretable framework for understanding sustained AMOC multicentennial oscillations in a two-hemisphere box model. We explicitly derive the oscillation period, clarify the necessary and sufficient role of thermohaline circulation, and explain why wind-driven circulation alone cannot sustain such variability. This work contributes to bridging conceptual box models with more complex climate models, offering insights into AMOC dynamics during the Holocene and their potential relevance to long-term societal change.

We have added a comparison between this work and previous studies, including results obtained from spatially resolved yet highly idealized models, in the Discussion in the revised manuscript.

References:

Lucarini, V., and P. H. Stone, 2005a: Thermohaline circulation stability: A box model study. Part I: Uncoupled model. J. Climate, 18, 501-513.

——, 2005b: Thermohaline circulation stability: a box model study. Part II: coupled atmosphere – ocean model. J. Climate, 18, 514-5291.

Scott, J. R., J. Marotzke, and P. H. Stone, 1999: Interhemispheric thermohaline circulation in a coupled box model. J. Phys. Oceanogr., 29, 351-365.

(3) I have difficulty understand the solid and dashed qs and qn in Figure 1b, which does not seem to be necessary in the associated equations,

In this study, we assume that the wind-driven circulation consists of two components: a poleward branch and an equatorward branch. The poleward transport in the tropics is primarily driven by Ekman transport (solid arrows in Fig. 1b), while the equatorward return flow is geostrophic (dashed arrows in Fig. 1b). These two components are approximately balanced in the tropical upper ocean, ensuring mass conservation. Although the total mass is conserved, integration over the entire upper ocean shows that the net salinity and heat transports are poleward. Therefore, qn and qs represent the total mass transports of the wind-driven circulation in the upper ocean in the Northern and Southern Hemispheres, respectively. The dashed arrows are retained to emphasize mass conservation in the model. Further explanation has been added to the revised manuscript.

and (4) Implication of this study to the observed AMOC change should be welcomed if possible.

We agree that it is helpful to relate our results to the observed changes in the AMOC. Our study shows that multicentennial AMOC oscillations can occur due to internal ocean dynamics alone, even without complex feedbacks from the atmosphere or sea ice. This suggests that some of the long-term AMOC variability seen in observations may come from natural ocean processes, rather than external climate forcing. Although our model is simplified, it provides a clear framework for understanding the possible timescale of these oscillations. However, we recognize that real-world climate processes are much more complex and may weaken or mask such oscillations, which could explain why they are hard to detect or irregular in proxy records. We have added a short discussion to explain how our results may be relevant to observed AMOC changes.

Specific issues

1. *L28, sensitivity => sensitive*

Responses: Thank you for your comments. We have revised this.

2. *L54, arose => arisen?*

Responses: Thank you for your comments. We have revised this.

3. *L81-83, It might be helpful to summarize the difference two-hemisphere model in the current study and previous ones (Lucarini and Stone 2005a, 2005b; Scott et al. 1999). What is new in this study?*

Responses: Thank you for your comments. The previous studies primarily computed the eigenmodes and presented the AMOC time series without providing a clear explanation of the underlying feedback mechanisms. In addition, the multicentennial oscillations identified in their work are either damped or unstable, whereas our study presents a sustained oscillation. We have revised the manuscript to clarify these points.

4. *L101, 6S model, Authors must have a reason of using this name, but it might be helpful to readers if the name could be explain briefly.*

Responses: Thank you for your comments. We have revised the phrase "only salinity equations" to "six salinity equations" for clarity. This change will help readers better understand the abbreviation.

5. *L124-126, a little confused about qn (qs) northward (southward) in solid arrows and poleward (equatorward) arrows. First, qn and qs is in green solid; second, it is not clearly described why we have two wind-driven transports at the same time. Are they driven by westerlies in the mid-latitude and easterlies in the tropical oceans, respectively? The connections between theoretical box model and physical reason behind should be described.*

Consider: "and in green arrows represent transports by the wind-driven circulation. The solid and dashed arrows represent poleward and equatorward wind-driven transports, respectively."

Responses: Thank you for your comments. In this study, we assume that the wind-driven circulation consists of two components: a poleward branch and an equatorward branch. The poleward transport in the tropics is primarily driven by Ekman transport (represented by solid arrows), while the equatorward return flow is geostrophic (represented by dashed arrows). These two components are approximately balanced in the tropical upper ocean, ensuring mass conservation. Although the total mass is conserved, when integrated over the entire upper ocean, the net salinity and heat transport remains poleward. Therefore, qn and qs represent the total mass transports of the wind-driven circulation in the upper ocean, in the Northern and Southern Hemispheres, respectively.

This explanation was not clearly provided in the original manuscript, and the meaning of the dashed arrows was also missing. For clarity, we have now added this explanation both in the figure caption and in the relevant section discussing the wind-driven circulation.

6. *Table 1: Not clear how some of these numbers were determined: T1*, T2*, T3*, lamda, K, qn bar, qs bar etc.*

Responses: Thank you for your comments. We agree that parameter selection is a critical issue. Most of the parameters used in our model, including geometric configurations, boundary conditions, and equilibrium values, are chosen to be consistent with observations, realistic assumptions, or results from coupled climate models. For example, the boundary conditions, including freshwater fluxes and restoring temperatures, are set to values that produce equilibrium temperature and salinity fields closely matching CESM1 output (Yang et al., 2015), as described in the main text and Appendix A. The relationship between the boundary conditions and the equilibrium fields is provided in Eqs. (8). Other physical constants, such as the thermal expansion coefficient, saline contraction coefficient and reference seawater density are based on commonly accepted values. The wind-driven advection coefficients for the North and South Atlantic are selected to produce realistic mean mass transports associated with subtropical cells. These values are consistent with estimates from climate models, as mentioned in the manuscript. The total basin volume is also adjusted to reflect realistic oceanic conditions, and the parameters related to basin geometry are tuned to better capture the targeted oscillatory behavior. Since these parameters can influence the oscillation, their sensitivity is discussed in Section 3b. Although the mean AMOC strength used in this study (24 Sv) is relatively strong compared to values used in some previous studies, it is tuned to match the CESM1 result. Besides, the influence of the mean AMOC strength on the oscillation period is also discussed in the Section 4 in the manuscript. The linear closure parameter is chosen to obtain a relatively robust oscillation period, with details provided in the manuscript. We have added a clarifying note in the caption of the parameter table to reflect these choices.

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.

7. *L168-169, what is the unit of omega or just an eigenvalue?*

Responses: Thank you for your comments. The unit of omega is 10^{-10} s^{-1}. We have added this information to the context.

8. *L171, how is the e-folding time is estimated or derived?*

Responses: Thank you for your comments. The e-folding time is calculated from the imaginary part of the eigenvalue and represents the time interval over which an exponentially growing quantity decreases by a factor of e. This explanation has been incorporated into the context.

9. *Table 2, 6TS_THC and 6TS_WDC have not been introduced. What is "In 10–10 s–1" in frequency and "In Year" in period? Why are there multi-values in the last row in Damped mode?*

Responses: Thank you for your comments. We have addressed Table 2 in the revised manuscript from the following three aspects:

(1) We added a brief explanation of the 6TS models in the table caption. Since these models are introduced later in Section 5, some distance from where Table 2 appears, we also included a sentence to help readers make the connection.

(2) "In 10–10 s–1" (for frequency) and "In Year" (for period) are the units. The real part of each eigenvalue is used to calculate the e-folding time, and the imaginary part is used to compute the period of the related oscillatory mode. We have added this explanation to the revised manuscript.

(3) The 6TS models in the last row have multiple damped modes because they have more degrees of freedom. As this is a secondary point, we chose not to include it in the main text.

10. *L178, weakly, do you have a criterion defining the "weakly" or "strongly"?*

Responses: Thank you for your comments. For unstable oscillations, "weakly" means the e-folding time is longer than the oscillation period, so the amplitude increases slowly. In contrast, "strongly" means the e-folding time is shorter, leading to faster growth. For decaying oscillations, where the e-folding time is negative, "weakly" refers to a smaller (more negative) e-folding time and slower decay, while "strongly" indicates a larger (less negative) e-folding time and faster decay. This explanation has been incorporated into the context.

11. *L179-180, "The real part of the eigenvalue [Re(omega)] increases with lamda", how about those after fork when >50?*

Responses: Thank you for your comments. When lambda>50, the multicentennial eigenmode disappears. We have added this information to the context.

12. *L180, maximum, also minimum when Im(omega) <0?*

Responses: Thank you for your comments. It's correct that Im(omega) reaches its minimum when considering the sign. We have revised this point in the context.

13. *L187, ordinate => y-ordinate?*

Responses: Thank you for your comments. The term "ordinate" refers to the y-coordinate, so its use here is correct.

14. *L190, "-40" => 40? How are periods of 1800 and e-folding time of -40 year estimated? "-40" mean the model is unstable as later discussions show.*

Responses: Thank you for your comments. Based on all the reviewers' suggestions, we have removed the section about the millennial mode. However, we would like to address your question here. The period and e-folding time are derived from the eigenvalues of the box model by converting the units. In this context, the millennial mode has an e-folding time of -40 years, which indicates that it takes 40 years to decrease to 1/e of the original value, signifying stability. I believe it is important to retain the sign to clarify whether the system is stable or unstable.

15. *L191, unchanged, this should be cautious because the Im(w) changes signs.*

Responses: Thank you for your comments. Based on the reviewers' suggestions, we have removed the section about the millennial mode. However, we will be more cautious regarding the signs of the eigenvalues.

16. *L192, smaller, do you mean the value including the sign? note the negative sign in Re(w). I suggest authors use the values such as about -10 unit, +/-5 unit as presented in the figure. This will make readers easy to follow.*

Responses: Thank you for your comments. Following the suggestions of all three reviewers, we have removed the section on the millennial mode. However, we address the related question here: the sign is considered when comparing the real and imaginary parts of the eigenvalues. We appreciate the suggestion to use specific values and units to describe growth/decay and maxima/minima. This has been revised accordingly in the remaining parts of the paper.

17. *L191-193, for a given of millennial mode of 1800 years with e-folding time of 40 years, I am afraid that the mode may not be real in the real-world, because the mode will be damped and disappear within a small fraction of the mode period. Since the millennial model is discussed here very briefly, I suggest deleting it.*

Responses: Thank you for your comments. We agree with your suggestion, and based on the reviewers' feedback, we have removed the section about the millennial mode.

18. *L209, again, how is k determined? A reference is needed if this is determined based on previous studies.*

Responses: Thank you for your comments. The value of κ affects the amplitude of the oscillation; in this study, it is set to produce the desired amplitude of the self-sustained oscillation. We follow the value used in our previous work (LY22), whereas YYL24 adopted a value an order of magnitude smaller. This indicates that even weak subpolar vertical mixing can transform a growing oscillation into a self-sustained one. Therefore, the specific value of κ is relatively flexible, as its primary role is to introduce a nonlinear mechanism that limits the unbounded growth of the oscillation. We have added this statement and the corresponding reference in the revised manuscript.

19. *L267, global, which might be misleading, "total Atlantic" might be more reasonable.*

Responses: Thank you for your comments. We have revised this.

20. *L269, "multi-centennial mode is analyzed", I suppose, at this point, we don't know the timescale of the mode. The statement may need revising.*

Responses: Thank you for your comments. We have revised "multi-centennial mode" to "least damped oscillatory mode".

21. *L271, 5S, may need explanation such as "5-box salinity-driven model", which is also true to 6S model.*

Responses: Thank you for your comments. We fully agree with your suggestion to explain the model abbreviations to help readers better understand the text, and we have updated the relevant parts accordingly. Following all the reviewers' suggestions, we decided that introducing the 5S model at this stage was unnecessary and might cause confusion, so we have removed it. Instead, it is now introduced in the new Section 4. For the 6S model, we have already provided an explanation, as mentioned in our response to point 4.

22. *L406-406, Small Re(w) should result in a smaller damping, therefore the thermal process should make the damping weaker. Correct me if this not right.*

Responses: Thank you for your comments. We believe there was a misunderstanding. When comparing Re(ω) (the real part of the eigenvalues, in units of s^{-1}), both the sign and the value are considered. A smaller Re(ω) means the oscillation is more stable, with stronger damping. Therefore, the thermal process should increase the damping effect. We understand that the term "damping rate" might be confusing, so we have changed all such expressions to "the imaginary part". We have also replaced all uses of "frequency" with "real par" to make the meaning clearer.

23. *L426, Should 6TS be 6TS_THC or 6TS_THC+WDC, or both?*

Responses: Thank you for your comments. The 6TS models include both the 6TS_THC and 6TS_THC+WDC models. We have revised "6TS model" to "both 6TS_THC and 6TS_THC+WDC models" for greater clarity.

24. *L479-480, How about T2'? Should T2' increase, decrease, or it is assumed being no change?*

25. *L485, "Southward wind-driven advection", is it in the South (qs solid in Fig. 1) or North Atlantic (qs dashed)? How about the northward wind-driven advection, which is not explicitly discussed?*

Responses: Thank you for your comments. The original expression may lead to confusion. In this context, we specifically refer to wind-driven advection in the South Atlantic, while that in the North Atlantic has already been addressed earlier in the manuscript. For clarification, the wind-driven circulation is represented by a poleward Ekman transport (solid arrows) and an equatorward geostrophic return flow (dashed arrows), which are approximately balanced in the tropical upper ocean to conserve mass. Although mass is conserved, the net salinity and heat transport remains poleward. Accordingly, qn and qs denote the total upper-ocean mass transport associated with wind-driven circulation in the Northern and Southern Hemispheres, respectively. This explanation, along with the meaning of the dashed arrows, has been clarified in the revised figure caption and manuscript text.

26. *L546-547, "damped oscillatory mode", I am confused, since the results presented in 6TS_THC and 6TS_THC+WDC (Figs. 6a, 7a) are oscillatory except for 6TS_WDC in Fig. 6a.*

Responses: Thank you for your comments. Based on the reviewers' suggestions, we have removed this section as it does not offer substantial new insights. However, we would still like to address your question here. The term “damped oscillatory mode” refers to the fact that the eigenvalues under these conditions are complex with negative real parts, indicating damped oscillations. When sustained stochastic forcing is applied, these oscillations can be maintained over time.

27. *L551-552, it may be true if I compare THC first and then THC+WDC. However, it might be different, if I compare WDC first and then THC+WDC, the THC appears a damped effect, at least between 1k-2k years and 4k-5k years.*

Responses: Thank you for your comments. In response to the reviewers' suggestions, we have removed this section from the manuscript, as it does not provide substantial new insights. However, we would still like to address your question here. It is appropriate to compare the salinity amplitudes between the THC and THC+WDC models, as both exhibit oscillatory behavior. In contrast, such a comparison between the WDC and THC+WDC models is not appropriate, as the WDC models do not exhibit oscillations. Instead, the salinity variations in the WDC models follow an irregular, stochastic pattern, which may occasionally produce larger amplitude values.

28. *L561, Fig. 9d y-axis label may be simply "N".*

Responses: Thank you for your comments. Based on the reviewers' suggestions, we have removed this section as it does not provide substantial new insights; accordingly, the associated figure has also been removed.

29. *L570 "box model", better to specify as "6-box model".*

Responses: Thank you for your comments. We have revised this.

30. *L569-579, what is new in this study should be emphasized? Does LY22 and YYL24 have the multi-centennial oscillation?*

Responses: Thank you for your comments. Although multicentennial oscillations of the AMOC have also been verified in the one-hemisphere box model in our previous work (LY22 and YYL24), and the core physical mechanisms behind these oscillations are broadly similar between the one- and two-hemisphere models, this study introduces several new elements not addressed in earlier work:

(1) Clarifying the Mechanism Consistency:

The statement that "the core physical mechanisms of AMOC multicentennial oscillations are similar between the one-hemisphere and two-hemisphere box models" is itself a significant point requiring explicit clarification. Only by confirming this similarity can we confidently apply insights from the one-hemisphere model to the more realistic two-hemisphere case. We find that the period of multicentennial oscillations are less sensitive to some factors such as basin geometry and the structure of the ocean layers. The two-hemisphere model, with its increased number of boxes, introduces more degrees of freedom and therefore exhibits different oscillatory characteristics from the one-hemisphere version. These distinctions help bridge the gap between idealized and coupled models.

(2) Relevance to Holocene Variability and Human Civilization:

As discussed in the introduction, the overarching goal of our box model research is to understand centennial-to-millennial natural climate variability during the Holocene—a period critical to the development of human civilization. Historical documents, particularly those from ancient China, suggest the presence of multicentennial variations in human activity. Understanding the corresponding variability in the climate system, especially the AMOC, may therefore provide insights into the trajectory of human history. A one-hemisphere model alone is insufficient to explain the variability observed in proxy records and coupled model simulations, in which the

AMOC necessarily operates across both hemispheres. For this reason, the two-hemisphere box model is essential for our analysis.

(3) Physical Differences in Parameterization:

While the mathematics of the one- and two-hemisphere models are similar and produce broadly comparable outcomes, key physical differences remain. The two-hemisphere model more closely resembles coupled models by incorporating the AMOC's inter-hemispheric nature. Additionally, the models differ in how they parameterize the AMOC anomaly: the one-hemisphere model assumes it is proportional to the density gradient anomaly between the equator and the pole, while the two-hemisphere model uses the gradient between the two poles. This latter formulation is more physically reasonable and is supported by diagnostic analyses from two coupled models (see Appendix A).

(4) New Theoretical Insight into Oscillation Period:

By further simplifying the two-hemisphere box model, we derived a theoretical expression for the oscillation period. This allows us to explicitly assess how factors such as mean AMOC strength, basin volume, and ocean layer structure influence the period of oscillation. To our knowledge, such an analytical formulation has not been previously reported.

(5) Role of Wind-Driven Circulation (WD-MOC):

This study offers the first systematic exploration of the impact of wind-driven meridional overturning circulation (WD-MOC) on multicentennial variability. The two-hemisphere box model is the simplest configuration capable of simultaneously representing thermohaline and wind-driven circulations. For the first time, we derive a theoretical result showing that WD-MOC alone cannot generate oscillatory behavior in this framework. We consider this a meaningful contribution to understanding the role of WD-MOC in climate variability across different timescales. (It is worth noting that horizontal wind-driven gyres can exhibit oscillatory behavior due to wave dynamics, which is distinct from the mechanisms considered here.)

(6) Necessary and Sufficient Role of Thermohaline Circulation:

For the first time, we demonstrate that thermohaline circulation is both a necessary and sufficient condition for generating multicentennial oscillations in this simplified model. This conclusion cannot be obtained from either observational data or complex coupled models alone.

Finally, this work remains a critical part of our ongoing research series on AMOC multicentennial variability across a hierarchy of models with increasing complexity. The two-hemisphere box model serves as a key intermediate step between simple conceptual models and more complex tools such as 2D and 3D ocean models, ocean general circulation models (OGCMs), and fully coupled climate models.

We have also revised the discussion section of the manuscript to highlight and address the new insights presented in this study.

Replies to Reviewer #3:

Thank you very much for these constructive comments. We have revised the manuscript carefully based on these suggestions. The followings are our point-to-point replies.

In this manuscript, the authors expand their previously proposed box models for multi-centennial AMOC oscillations to include the Southern hemisphere as well as the wind-driven circulation. This is an interesting step towards more a realistic process representation and yields some new insights, for example about the role of the wind-driven vs. thermohaline circulation in sustaining AMOC oscillations. While the manuscript is quite long, the modelling choices and especially parameter values are nevertheless not sufficiently motivated. However, I think the manuscript should be publishable after addressing these points (and some additional ones listed below).

Responses: Thank you very much for your encouraging comments. Combining suggestions from all reviewers, we have revised the manuscript substantially in the following areas:

(1) The section discussing the millennial mode has been removed.

(2) The section titled "Linear Oscillations Excited by Stochastic Forcing" has been removed.

(3) A new section, "Oscillatory Timescales of Multicentennial Oscillations", has been added. This material was previously part of the Discussion section.

(4) Substantial revisions have been made to both the Methods and Discussion sections. In the Methods, we now include a more detailed description of the box model, including the fundamental oceanic processes it represents and the treatment of wind-driven processes. In the Discussion, we have expanded the text to better emphasize the novelty of this study and to clearly acknowledge its limitations.

Major Comments

1. *The box model design, and especially the parameter values, are not motivated very well. However, the reader should at least somewhat understand which processes are included and which assumptions/simplifications go into the box model equations. This does not need to be exhaustive, but some points to consider are, e.g.:*

Responses: Thank you for your comments. We agree that additional background on the model is necessary. In the revised manuscript, we have incorporated further explanations, addressing the questions listed below.

- How are the transports terms expressed? (proportional to density difference? etc.)

We assume a positive linear correlation between transport strength and the north-south density difference. While this assumption is now stated explicitly in the paper, the transport formula is not presented directly, as our focus is on the linearized equations.

- Is the global salt content conserved?

Yes. Eq. (2e) represents the conservation of salt content within the model.

- Linear equation of state? How are the delta_rho terms obtained?

The AMOC anomaly q' is parameterized as a linear function of the density anomaly difference between the northern and southern boxes. The density anomaly is determined by the salinity anomaly and is vertically integrated.

- Why does the vertical mixing term have this form? etc. Probably some of these points are addressed in the previous papers (LY22, YYL24), but this paper should be able to stand on its own.

The form of vertical mixing is a reasonable outcome derived from the coupled model (LY22) and represents one of several possible mechanisms that can generate the self-sustained oscillation (YYL24).

2. *A bigger concern, which should however be easy to address, is that it is currently unclear how the parameter values are obtained. It is only very briefly mentioned in L144 that "the model is tuned so that its equilibria nearly agree with the results of the two coupled models", but it is not detailed what is tuned exactly. Given that Eqs. 2a-e only have three degrees of freedom, it should be mentioned more explicitly which variables were fitted to the model(s) (only CESM or both?) and which weren't. Then, it would be good to check whether the remaining variables have reasonable values compared to the model(s), or whether they deviate. (This wouldn't be a big problem, but it should be transparently reported.)*

Responses: Thank you for your comments. We agree that parameter selection is a critical issue. Most of the parameters used in our model, including geometric configurations, boundary conditions, and equilibrium values, are chosen to be consistent with observations, realistic assumptions, or results from coupled climate models. For example, the boundary conditions, including freshwater fluxes and restoring temperatures, are set to values that produce equilibrium temperature and salinity fields closely matching CESM1 output (Yang et al., 2015), as described in the main text and Appendix A. The relationship between the boundary conditions and the equilibrium fields is provided in Eqs. (8). Other physical constants, such as the thermal expansion coefficient, saline contraction coefficient and reference seawater density are based on commonly accepted values. The wind-driven advection coefficients for the North and South Atlantic are selected to produce realistic mean mass transports associated with subtropical cells. These values are consistent with estimates from climate models, as mentioned in the manuscript. The κ is set to produce the desired amplitude of the self-sustained oscillation. The total basin volume is also adjusted to reflect realistic oceanic conditions, and the parameters related to basin geometry are tuned to better capture the targeted oscillatory behavior. Since these parameters can influence the oscillation, their sensitivity is discussed in Section 3b. Although the mean AMOC strength used in this study (24 Sv) is relatively strong compared to values used in some previous studies, it is tuned to match the CESM1 result. Besides, the influence of the mean AMOC strength on the oscillation period is also discussed in the Section 4 in the manuscript. The linear closure parameter is chosen to obtain a relatively robust oscillation period, with details provided in the manuscript. We have added a clarifying note in the caption of the parameter table to reflect these choices.

It would also be good to state whether parameters were at any point tuned with the goal to obtain a similar period of the oscillations as in Li & Yang (2022) or more generally in the CESM.

This is an insightful question. Most of the parameters used in our model, including boundary conditions, and equilibrium values, were not intentionally selected to reproduce the same oscillation period as reported in LY22. Instead, they were chosen based on observational constraints, realistic assumptions, or outputs from coupled climate models, as previously discussed. However, the basin geometry was adjusted to help achieve this outcome, as it is a particularly sensitive factor, although it remains within a reasonable range. Furthermore, we have examined the sensitivity of the oscillation period to basin geometry, and under realistic parameter settings, the period is very likely to fall within the multicentennial range. This statement has been incorporated into the Discussion section of the revised manuscript.

The problem of unclear parameter choices of course persists for the more complex box models introduced later in the manuscript, which probably have more degrees of freedom. Again, it should be reported which values were fitted/tuned from the CMIP models, which ones were chosen "by hand" or from theory, and which ones simply result from the choice of the other parameter values. For the most important "hand-chosen" parameters (e.g., kappa in L209, I would guess?), it would be good to physically motivate at least the order of magnitude.

All box models, regardless of their complexity, are configured with different boundary conditions to achieve a consistent equilibrium state. The closure parameters vary between models, while all other parameters are held constant and are listed in Table 1. Additional clarification on these choices has been provided in the revised manuscript.

The value of κ affects the amplitude of the oscillation; in this study, it is set to produce the desired amplitude of the self-sustained oscillation. We follow the value used in our previous work (LY22), whereas YYL24 adopted a value an order of magnitude smaller. This indicates that even weak subpolar vertical mixing can transform a growing oscillation into a self-sustained one. Therefore, the specific value of κ is relatively flexible, as its primary role is to introduce a nonlinear mechanism that limits the unbounded growth of the oscillation.

One particularly important parameter is the closure coefficient lamba, on which the stability of the solutions depends sensitively (e.g., Fig. 2). It would be good to explain in a bit more detail how this parameter is obtained, and what could be a reasonable range from different climate models (at least from the two in Appendix A, if it can be obtained by fitting these data).

Although the oscillation period is sensitive to the closure parameter lambda over a broad range, its variation becomes limited near the minimum period, which corresponds to the maximum imaginary part. Therefore, we use this minimum period as a representative estimate. It is difficult to compare the values of lambda between the box model and climate models, or to define a reasonable range for lambda, given the simplified nature of the box model. In our climate model results, lambda can exceed 30, as shown in the Appendix, and may reach values above 80 depending on the choice of box regions (not shown). Therefore, we consider it more appropriate to select lambda based on its behavior in the sensitivity curves.

Around L428, different values of this parameter are considered. How are these values chosen?

As mentioned earlier, in different models, we select the ‘standard’ value of lambda as the value at which the imaginary part reaches its maximum (corresponding to the minimum period), since the values are relatively stable in its vicinity.

3. *While the box model analysis undoubtedly yields some interesting results, the Discussion currently does not link or contrast them with the results from previous box model studies beyond the authors' own previous research. The only other study that is discussed briefly regarding mechanisms is that of Sévellec et al. (2006) (L636). Following this example, and given that quite a lot of other previous studies are cited in the Introduction, it would be very helpful to*

include some of them in the discussion. In addition, it improve the Discussion to link the results obtained here to studies with spatially resolved but still highly idealized approaches such as Mysak et al. (1993) and Wolfe & Cessi (2015).

Responses: Thank you for your comments. Several conceptual studies, including our previous work, have identified the low-frequency AMOC oscillations in one-hemisphere models (e.g., Griffies and Tziperman, 1995; Rivin and Tziperman, 1997; Wei and Zhang, 2022). However, systematic investigations of AMOC multicentennial variability in two-hemisphere configurations remain limited—an important gap that this study aims to address.

Although some studies have reported multicentennial AMOC oscillations in two-hemisphere three-box models (e.g., Lucarini and Stone, 2005a, 2005b; Scott et al., 1999), their analyses primarily involve eigenmode computations and time series outputs, without offering a detailed explanation of the underlying feedback mechanisms. Moreover, the oscillations identified in these studies are typically damped or unstable, whereas the present work demonstrates a sustained multicentennial oscillation.

Sévellec et al. (2006) also investigate centennial-scale oscillations using a two-dimensional latitude-depth model under mixed boundary conditions, complemented by a simplified Malkus loop oscillator to explore the oscillation mechanism. Their linear stability analysis incorporates both temperature and salinity, making it analytically complex. The oscillations reported—on the order of 100-200 years in the one-hemisphere configuration and over 500 years in the two-hemisphere case—are either strongly damped or highly unstable, with growth rates comparable to or exceeding the oscillation periods. For example, their Table 2 reports a 170-year period with a 454-year growth rate (unstable), and a 605-year period with a −908-year growth rate (damped), suggesting that these modes may not be physically robust.

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	∞

Some spatially resolved, though still highly idealized, models also exhibit AMOC multicentennial variability. Mysak et al. (1993) identified 200-300-year oscillations triggered by stochastic freshwater forcing, but their analysis focuses solely on the positive salinity-advection feedback and lacked a complete feedback loop explanation. Similarly, Wolfe and Cessi (2015) report ~380-year oscillations characterized by large-scale sea surface salinity (SSS) anomalies of opposite sign in the two hemispheres. However, their analysis primarily focuses on the positive salinity-advection feedback and does not consider the full suite of feedback mechanisms required to sustain such oscillations.

Overall, in contrast to these previous studies, our work provides a systematic and physically interpretable framework for understanding sustained AMOC multicentennial oscillations in a two-hemisphere box model. We explicitly derive the oscillation period, clarify the necessary and sufficient role of thermohaline circulation, and explain why wind-driven circulation alone cannot sustain such variability. This work contributes to bridging conceptual box models with more complex climate models, offering insights into AMOC dynamics during the Holocene and their potential relevance to long-term societal change.

We have added a comparison between this work and previous studies, including results obtained from spatially resolved yet highly idealized models, in the Discussion in the revised manuscript.

Specific comments

1. *L30: "the AMOC itself is a necessary and sufficient condition (⋯)": This sounds a bit odd. Of course the AMOC cannot exhibit oscillations if it doesn't exist. I think what the authors mean is that the shallow wind-driven overturning cannot exhibit multi-centennial oscillations without the thermohaline component (in this model)? Please clarify.*

Responses: Thank you for your comments. We completely agree with your idea that this statement seems not rigorous. We have revised this expression.

2. *L33-35: This sentence is incomplete (understanding what?). Besides, I am unsure whether this millennial mode, which is only mentioned briefly in the text and the authors are "not sure whether this mode is physically meaningful" deserves to be highlighted in the abstract. In my view, it would be better to end the abstract with a conclusion/take-home message.*

Responses: Thank you for your comments. Following the suggestions of all reviewers, we have removed the discussion of the millennial mode from both the main text and the abstract in the revised manuscript. In addition, as suggested, we have added a concluding sentence at the end of the abstract to summarize the main findings.

3. *L49: The recent paper by Mehling et al. (2024) investigated (multi-)centennial AMOC variability in a multi-model ensemble and it seems appropriate to cite it here.*

Responses: Thank you for your comments. We have added it to the citation.

4. *L56: Jiang et al. and Meccia et al. (already mentioned in the previous citation bracket) also showed the importance of salinity anomalies in the deep-water formation regions.*

Responses: Thank you for your comments. We have added them to the citation.

5. *L58: "generated" -> "amplified" (by definition of a feedback)*

Responses: Thank you for your comments. We have revised this.

6. *L106: I don't think "prove" is the correct word here - the existence of the eigenmode is already demonstrated by the analysis of the deterministic equations in Section 2b.*

Responses: Thank you for your comments. Section 5 has been removed as suggested in point 23, along with the related sentence.

7. *L129: Please mention here that q is not yet defined, and the closure will only be introduced later (for the linearized system).*

Responses: Thank you for your comments. We have revised this.

8. *L141: It seems rather unintuitive that the upper-ocean salinity in the subpolar North Atlantic (S_1) should be equal to the deep-ocean salinities (Eq. 2d). Is this an artifact of the simplifications in this model or actually supported by the two CMIP models shown in Appendix A? Unfortunately, only the upper-ocean salinities are reported there.*

Responses: Thank you for your comments. This is a very thoughtful consideration. We assume that vertical mixing in the NADW formation region is sufficiently strong and that the combined effects of NADW advection and deep-ocean diffusion are adequate to mix surface and deep waters. We have also examined this in the coupled model simulations. Fig. R2 shows the equilibrium salinity section in the Atlantic basin, where the salinity in the NADW formation region (around 60°N) is clearly similar to that of the deep ocean, particularly between 2000 and 4000 meters. Therefore, we believe that our choice of equilibrium salinity values is reasonable.

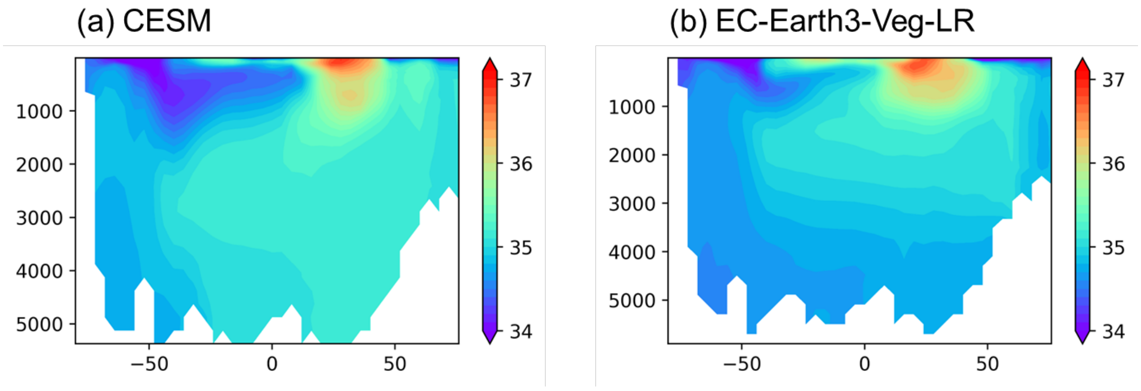

Fig. R2. Equilibrium salinity section in the Atlantic basin from CESM1 and EC-Earth3-Veg-LR simulations.

9. *L145: "nearly agree": in which variables? Surely this is not the case for all?*

Responses: Thank you for your comments. The equilibrium salinities and mean AMOC strength have been tuned to match the CESM1 values. We have also revised the corresponding statement in the manuscript accordingly.

10. *Table 1: An equilibrium AMOC strength of 24 Sv appears rather high (the observed value, albeit not strictly in equilibrium, is around 17 Sv). Is this tuned to the CESM1 value?*

Responses: Thank you for your comments. An equilibrium AMOC strength of 24 Sv is tune to our CESM1 control run. As the equilibria values follows the results in our model, we choose the mean AMOC strength to keep consistence.

11. *L167: "obtained numerically": with which software/method?*

Responses: Thank you for your comments. The eigenvalues were computed using Python's NumPy library (numpy.linalg.eig), which employs LAPACK's QR algorithm for general matrices. We have included the relevant information in the paper.

12. *L219: "gradually enhancing amplitude": It could be mentioned that it grows exponentially*

Responses: Thank you for your comments. We have revised this.

13. *L261: "minor role of the Southern Ocean": Is this consistent with the CMIP models shown in the Appendix?*

Responses: Thank you for your comments. We acknowledge an error in the manuscript: "Southern Ocean" should be corrected to "South Atlantic." In addition, we have examined the variability of salinity anomalies in the North and South Atlantic using CESM1, as illustrated in Fig. R3. The amplitude of salinity anomalies in the South Atlantic is significantly smaller than that in the North Atlantic. This result from the climate model is consistent with the findings from our box model.

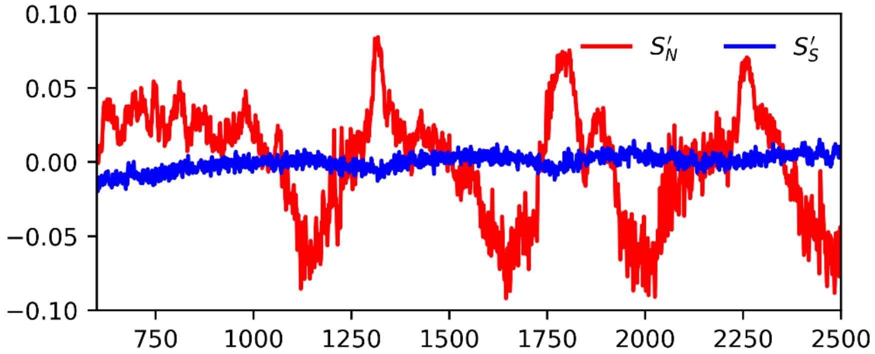

Fig. R3. Salinity anomalies in the North (red) and South (blue) Atlantic, integrated over the upper 4000 meters. The averaging regions correspond to the boxes defined in the Appendix.

14. *L271: It seems unnecessary to introduce an additional model ("5S") here that is never used afterwards, and which doesn't seem too relevant for these results. Please consider removing it for more clarity.*

Responses: Thank you for your comments. We agree that the introduction of the 5S model at this point was unnecessary and potentially confusing, so we have removed it. Instead, it will now be introduced in the new Section 4.

15. *L274: "regardless of the presence of enhanced vertical mixing": Is this for a fixed value of the mixing coefficient \kappa or for varying values?*

Responses: Thank you for your comments. Region three has been removed from the manuscript in response to the suggestions of all reviewers. However, we provide an explanation here for clarity. The original region three represented a regime where the oscillation remained unstable even under extremely strong mixing. This implies that the mixing coefficient, \kappa, must take variable values and could become very large in this regime.

16. *L292: Varying the depth of the upper-ocean boxes could be better motivated by linking the scale depth of the AMOC to its mean strength (Nayak et al., 2024).*

Responses: Thank you for your comments. We agree with the general view that the scale depth of the AMOC is linked to its mean strength. However, we have chosen not to include this relationship in our model or add a corresponding figure for the following reasons:

(1) This study focuses on small amplitude multicentennial oscillations of the AMOC. In this context, variations in scale depth associated with oscillation strength may be negligible. While the mean AMOC strength, which is connected to the broader climate state, could indeed relate to scale depth, this idea is acknowledged but not the central focus of our paper.

(2) Exploring the relationship between mean AMOC strength and scale depth provides limited insight in our model because the closure parameter is fixed. Our previous work (LLY24) has already demonstrated that the AMOC multicentennial mode is sensitive to both the mean AMOC strength and the closure parameter. However, simultaneously changing all three factors, namely mean AMOC strength, scale depth, and closure parameter, in the 6S model would introduce unnecessary complexity without offering additional explanatory value.

(3) Instead, we consider the 3S model to be a useful tool for investigating this question and have examined the relationship between scale depth and mean strength within that framework. As shown in Eq. (X) in the main text, the period of the AMOC multicentennial oscillation is influenced by the closure parameter and a prescribed mean state, both of which are treated as independent parameters in our model. Although the oscillatory period is sensitive to the closure parameter over a broad range, its variation becomes small near the minimum value. Therefore, we use this minimum period as an estimate. This approach links the closure parameter to the climate mean state, allowing us to reduce the number of independent variables. Fig. RX shows the theoretically estimated oscillation period under the changed mean AMOC strength and depth. With both the increase of AMOC strength and depth, the e-folding time and period increase slowly.

(3) However, we have chosen not to include these results in the main text, as introducing this new relationship would significantly increase the complexity of the model and deviate from the paper's primary focus.

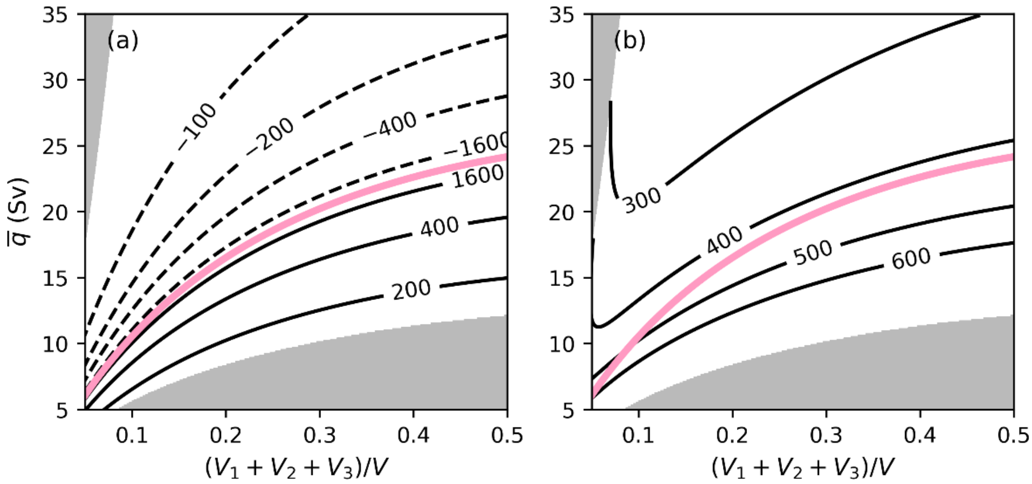

Fig. R4.

17. L313: This dependence on the temperature gradient is only motivated 2 pages later, please mention this here to avoid confusion (e.g., "which will be motivated later in Eq. 11").

Responses: Thank you for your comments. We have revised this.

18. L445: It would be good to provide a short take-home message at the end of this paragraph. For example, is the addition of the (somewhat artificial) additional vertical mixing in the subpolar North Atlantic still necessary to obtain self-sustained oscillations? If yes, what does the inclusion of the wind-driven component change qualitatively?

Responses: Thank you for your comments. In our box model, the inclusion of wind-driven circulation does not introduce a nonlinear mechanism; therefore, enhanced vertical mixing in the subpolar North Atlantic boxes remains essential. Rather, the wind-driven circulation contributes by dampening the multicentennial oscillation through a reduction in its amplitude. We have clarified this point in the revised manuscript.

19. L484: Are these feedbacks realistic? Could you try diagnosing some of them from any of the two CMIP models?

The sheared Ekman component AMOCE has almost no low-frequency variability

The compensation between thermohaline and wind-driven circulations has been verified in the Fast Ocean-Atmosphere Model (FOAM) (Yang et al. 2013)

20. L485-497: This paragraph is quite long for essentially saying (from my understanding, I might have misread!) "the feedbacks in the Southern hemisphere are the same as in the Northern hemisphere, but with a weaker magnitude". It might be possible to shorten this paragraph considerably.

Responses: Thank you for your comments. We fully agree with your viewpoint and have revised the paragraph.

21. L527, L602 and elsewhere (especially Appendix B): By "theoretical", you mean "analytical"?

Responses: Thank you for your comments. We clarified that “theoretical solution” refers to “analytical solution”. Since “analytical solution” is a more precise term, we have replaced “theoretical solution” with “analytical solution”.

22. *L533: "sufficient and necessary" seems like a too strong statement, since other mechanisms could be at play (as mentioned in the Introduction). At least add "in this model" to this sentence.*

Responses: Thank you for your comments. We have revised this section and made similar changes elsewhere.

23. *L535-L567: I would suggest removing Section 5 entirely. As far as I can see, it does not really yield new insights compared to Li & Yang (2022), adding relatively little to an already very long article. In addition, the approach of adding red noise directly to the temperature and salinity equations is not really consistent with the literature, where white noise is usually just added to the surface freshwater flux (e.g., Cessi, 1994; Mikolajewicz & Maier-Reimer, 1990) or surface temperature where no salinity equations exist (e.g., Griffies & Tziperman, 1995).*

Responses: Thank you for your comments. We have removed this section as it does not offer substantial new insights.

24. *L576: Is it a coincidence that both the AMOC strength and total ocean volume are twice as large here as in LY22? Does this have to do with the tuning choices taken here (maybe to match a given period)?*

Responses: Thank you for your comments. This is an insightful question. It is indeed a coincidence that both the AMOC strength and total ocean volume in our current setup are approximately twice as large as those in LY22. The mean AMOC strength is tuned to match the results from coupled models, while the total ocean volume is adjusted to reflect more realistic values. Although the mean AMOC strength and total ocean volume are among the most important factors in determining the oscillation period, we believe that the matched period between the one-hemisphere and two-hemisphere models is partly coincidental, since other model parameters such as basin geometry and freshwater forcing also influence the period. Nonetheless, this agreement shows that it is possible to reproduce similar timescales in both models, which further supports the consistency of the theoretical framework.

25. *L596-L607: It is a bit unusual to introduce new results with figures, and even a new box model, in the Discussion. I would suggest moving the results of the 3-box model to a dedicated section in the Results, if they are deemed relevant for the overall message of the article.*

Responses: Thank you for your comments. We have added a new section presenting the results of the 3-box model, with a focus on providing more detail about the timescale of the multicentennial mode.

26. *L615: The Okazaki et al. refence discusses (relatively short-lived) deep-water formation in the North Pacific 15 thousand, not 15 million years ago. Maybe the review by Ferreira et al. (2018) would be more appropriate to cite here?*

Responses: Thank you for your comments. We agree that the review by Ferreira et al. (2018) is more appropriate to cite in this context, and we have updated the citation accordingly in the revised manuscript.

27. *L619: I don't want to open a can of worms, but from my understanding D-O events in the paleoclimatic record are not described as periodic oscillations but rather as a jump process with exponentially distributed waiting times (Ditlevsen, 1999). However, some climate models do show millennial-scale "D-O-like" oscillations which are often quite regular (Malmierca-Vallet et al., 2023). Maybe these model simulations would provide a better motivation to study millennial-scale oscillations from a conceptual viewpoint.*

Responses: Thank you for your comments. As the conceptual model alone is not sufficient to provide a convincing new insight and this topic is only a secondary focus of the paper, we have removed the discussion of the millennial mode from the manuscript.

28. *L633: "highly likely": this cannot be concluded from a simple box model, please reword or remove this statement.*

Responses: Thank you for your comments. We have removed this statement.

29. *L637: "believed" -> "showed"? (in their model)*

Responses: Thank you for your comments. We have revised this.

30. *L638-641: This discussion of multiple equilibria is a bit confusing. First, even very simple models like the Stommel (1961) box model can have multiple equilibria under fixed freshwater forcing. From my understanding, the box model in this study cannot have (physically meaningful) multiple equilibria by construction, since the model is first linearized around an equilibrium before introducing the closure for q. That being said, unless the authors see a need to clarify their idea, this part of the discussion could also simply be omitted.*

Responses: Thank you for your comments. We agree that the discussion of multiple equilibria is not appropriate in the context of our linearized model, which focuses on small-amplitude oscillations around a single equilibrium. Accordingly, we have removed this discussion in the revised manuscript.

31. *L644: I don't understand why there is "interaction between the surface freshwater flux and [the] closure parameter". Surely this can only be a one-way interaction (i.e., the freshwater flux influencing the closure parameter and not the other way round), since the freshwater flux is a prescribed forcing that doesn't change? After clarifying this, what is the relevance of this interaction?*

Responses: Thank you for your comments. We greatly appreciate your careful reading and for pointing out the incorrect expression, which has been very helpful to us. As shown in Eq. (X) in the main text, the period of the AMOC multicentennial oscillation is influenced by both the surface freshwater flux and the closure parameter, which are treated as independent parameters in our model. Although the oscillatory period is sensitive to the closure parameter over a broad range, its variation becomes small near the minimum value. Therefore, we use this minimum period as an estimate. In this context, the freshwater flux influences the value of the closure parameter required to achieve the minimum period, rather than the closure parameter affecting the freshwater flux. This is because the freshwater flux is prescribed as an external forcing in our model. This interpretation aligns with your comment and underscores the important role of surface freshwater flux in determining the oscillatory period. We have revised the corresponding expression in the manuscript accordingly.

32. *L652: It is unclear what is meant by "its sustainability in the real world is a serious concern".*

Responses: Thank you for your comments. We have revised this statement to emphasize that the multicentennial oscillation is influenced by numerous climatic factors beyond the thermohaline circulation itself.

33. *L661-L669: This outlook reads almost like the abstract of another paper. It would be good to shorten it a bit.*

Responses: Thank you for your comments. This part about the millennial mode has been removed, as suggested by all reviewers. We have also made major revisions to the discussion. A summary of these changes is provided at the beginning of our response letter.

34. *L677: To facilitate replication, it would be great to make the scripts used for the numerical solutions available.*

Responses: Thank you for your comments. We have revised the 'Data Availability Statement'. Anyone interested in reproducing the results is welcome to contact us to obtain the codes.

35. *L692-694: Why specifically this choice for the northern and southern boxes? They seem quite small and their limits are somewhat arbitrary (especially in the South). Does the linear density scaling still hold for differently defined boxes?*

Responses: Thank you for your comments. We selected the box boundaries based on the regression pattern of AMOC onto density anomalies (Fig. R5b). The northern box corresponds to a region with a strong positive regression and the southern box corresponds to a region with a negative regression. However, we acknowledge that the choice of box boundaries is somewhat arbitrary. To test the robustness of our results, we explored alternative box definitions and found that the main relationship is not significantly affected by the specific box ranges. For example, using CESM model output, we tested larger northern and southern boxes: the North Atlantic box covers 60°W-20°E, 45°-75°N, and the South Atlantic box covers 60°W-20°E, 30°-70°S. As shown in Fig. R5c and R5d, the linear relationship still holds under these alternative definitions. We also tested other regions with positive regression in the North Atlantic, and the results were very similar; therefore, they are not shown here.

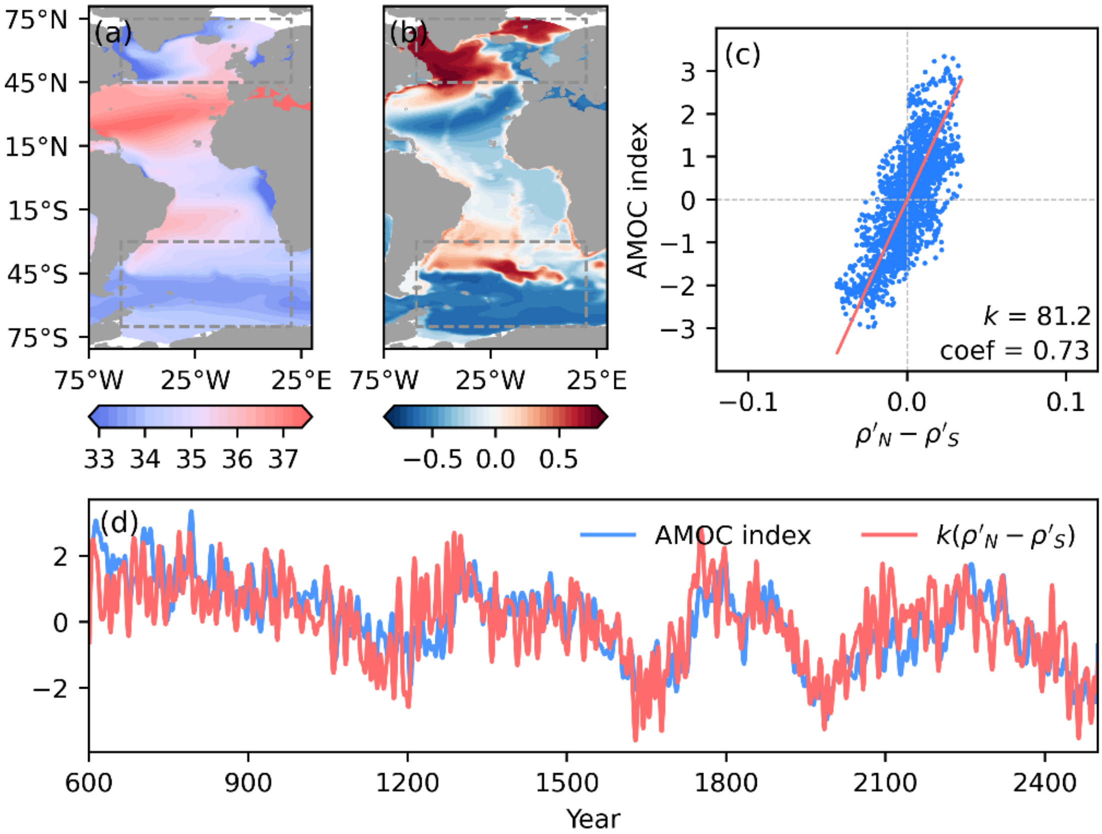

Fig. R5. (a) Climatology of sea-surface salinity (units: psu) in CESM 1.0. Dashed boxes outline the subpolar North, tropical, and subpolar South Atlantic boxes, respectively. (b) Regression of AMOC anomaly (units: Sv) on density anomaly integrated above 4000-m depth (units: kg m^{-3}). (c) Scatter plot of AMOC anomaly (ordinate) versus the difference of density anomaly (abscissa) averaged between the two regions in subpolar North and South Atlantic oceans, respectively. The red line represents the reduced major axis regression with a coefficient of 0.73 and a slope of 81.2 Sv kg^{-1} m^3. (d) Time series of AMOC anomaly (blue curve) and its estimation (red curve) from the reduced major axis regression. In (c) and (d), the anomalies of AMOC index and density are lowpass-filtered with a cutoff period of 10 years.

36. L725: It is unclear why a model with only a 500-year control simulation is chosen here when a >1200-year control run from the same model family (EC-Earth3) is also available on the ESGF.

Responses: Thank you for your comments. We have now included a control simulation using the EC-Earth3-LR model, which spans 2000 years. The relevant text in the manuscript has also been revised accordingly.

Technical comments

1. *L168/169: units for \omega missing* Revised.

2. *Table 2: "Year" should be "Year ^ -1"* No, we have converted the value to have the unit 'year.'

3. *L182: Remove "under"* Since lambda does not represent the eigenvalue, we have replaced “under” with “for” instead of removing it.

4. *L305: "with" missing after "consistent"* Revised.

5. *Fig. 2: Please put the units directly on the y-axis labels instead of only in the figure caption.* Revised. We have also revised Fig. 5 and Fig. B2.

6. *L537: "killed" is maybe not the most scientific wording here* Revised in L527.

7. *L638: Remove "largely"* Revised.

8. *L658: "become the" -> "shift towards"* Revised.

9. *L693 and elsewhere: subtropic -> subtropical* Revised.

10. *L711: scattering -> scatter* Revised.

Additional references

11. *Cessi, P. (1994). A Simple Box Model of Stochastically Forced Thermohaline Flow. Journal of Physical Oceanography, 24(9), 1911-1920. https://doi.org/10.1175/1520-0485(1994)024<1911:ASBMOS>2.0.CO;2*

12. *Ditlevsen, P. D. (1999). Observation of* α*-stable noise induced millennial climate changes from an ice-core record. Geophysical Research Letters, 26(10), 1441-1444. https://doi.org/10.1029/1999GL900252*

13. *Ferreira, D., Cessi, P., Coxall, H. K., de Boer, A., Dijkstra, H. A., Drijfhout, S. S., et al. (2018). Atlantic-Pacific Asymmetry in Deep Water Formation. Annual Review of Earth and Planetary Sciences, 46(1), 327-352. https://doi.org/10.1146/annurev-earth-082517-010045*

14. *Malmierca-Vallet, I., Sime, L. C., & the D-O community members. (2023). Dansgaard-Oeschger events in climate models: review and baseline Marine Isotope Stage 3 (MIS3) protocol. Climate of the Past, 19(5), 915-942. https://doi.org/10.5194/cp-19-915-2023*

15. *Mehling, O., Bellomo, K., & von Hardenberg, J. (2024). Centennial-Scale Variability of the Atlantic Meridional Overturning Circulation in CMIP6 Models Shaped by Arctic-North Atlantic Interactions and Sea Ice Biases. Geophysical Research Letters, 51(20), e2024GL110791. https://doi.org/10.1029/2024GL110791*

16. *Mysak, L. A., Stocker, T. F., & Huang, F. (1993). Century-scale variability in a randomly forced, two-dimensional thermohaline ocean circulation model. Climate Dynamics, 8(3), 103-116. https://doi.org/10.1007/BF00208091*

17. *Nayak, M. S., Bonan, D. B., Newsom, E. R., & Thompson, A. F. (2024). Controls on the Strength and Structure of the Atlantic Meridional Overturning Circulation in Climate Models. Geophysical Research Letters, 51(11), e2024GL109055. https://doi.org/10.1029/2024GL109055*

18. *Stommel, H. (1961). Thermohaline Convection with Two Stable Regimes of Flow. Tellus, 13(2), 224-230. https://doi.org/10.1111/j.2153-3490.1961.tb00079.x*

19. *Wolfe, C. L., & Cessi, P. (2015). Multiple Regimes and Low-Frequency Variability in the Quasi-Adiabatic Overturning Circulation. Journal of Physical Oceanography, 45(6), 1690-1708. https://doi.org/10.1175/JPO-D-14-0095.1*