Response to the second round of review of Willeit and Ganopolski's "Generalized stability landscape of the Atlantic Meridional Overturning Circulation" by Yvan Romé, University of Leeds.

We would like to thank the Reviewer again and provide the response to his comments in blue below.

General comments

Firstly, I would like to thank the authors for their fast and detailed response to the previous review. The authors addressed the comments in detail, and the manuscript was modified in depth to answer them. The most recent version of the manuscript improves on the previous one and satisfactorily answers the main concerns raised during the previous round of reviews. In particular, the text is clearer, more precise, and more nuanced.

Regarding the previous comments:

• The authors' comment about the scope of the abstract is convincing, and the new version of the abstract is a good summary of the work conducted in this study.

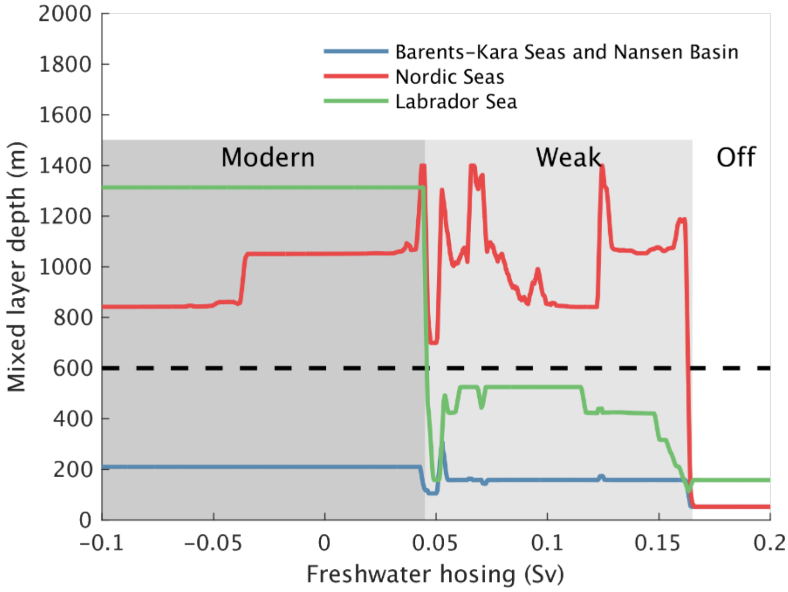
• The introduction was significantly improved. It answered thoroughly my previous comments, and I find the new version succinct and impactful. I understand the authors' response regarding the freshwater forcing and millennial-scale literature. Shifting the narrative to the limitations of previous hysteresis studies is an elegant way to avoid going through the extensive references on freshwater hosing simulations, which is, I agree, not entirely relevant to the paper.

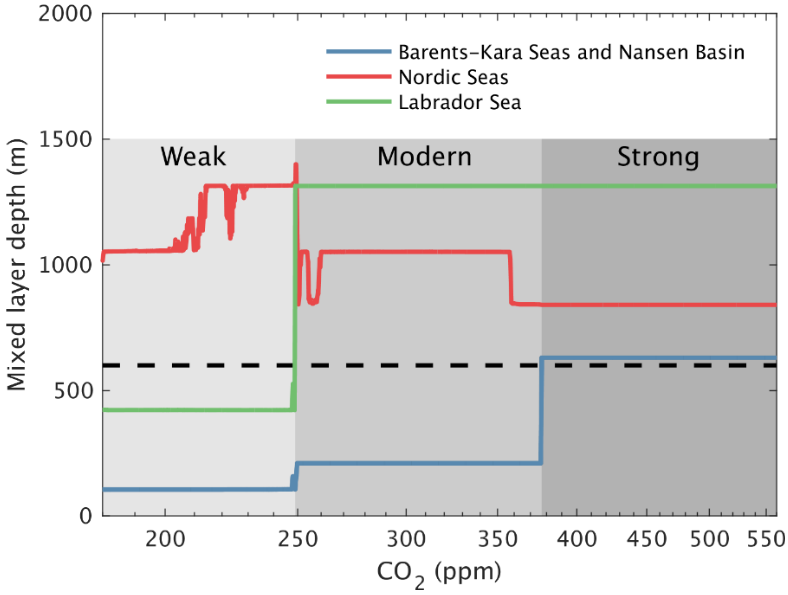
• The new version of the paper does an excellent job of highlighting the novelty of the FWF x CO2 AMOC stability diagram. The distinction between the previous studies and this method is also more transparent.

• I appreciate that the authors discussed the potential occurrence of millennial-scale variability in Figure 1 around 0.05 Sv. I have additional comments about the interpretation of this hysteresis cycle, but they can be addressed as specific comments.

• The definition of modes is better defined in the manuscript and is rigorous. I believe, however, that this should be introduced at the same time as the modes in the manuscript. In Sections 2 and 3, and Figures 1 and 2, it sounds like the AMOC modes are only defined by the strength of the AMOC, which makes the identification of the modes vague when it is actually done precisely. This is particularly relevant in Figure 2 where the transition between the modern and strong AMOC modes is not evident on the AMOC index alone. As argued in my previous review, an additional Figure/panel with MLD time series in the different regions for the hysteresis simulations could be a good way to illustrate this method. This could be in Supplementary information.

We added two additional figures (see below) to the Appendix showing the maximum mixed layer depth in the three different regions used to categorize the four different AMOC states for two selected curves of the FW and CO2 hysteresis experiments shown in Fig. 1 and 2 in the paper. We refer to these two new figures in sections 2 and 3, when discussing the FW and CO2 freshwater hysteresis experiments and introducing the different AMOC states.





• The additional sentences about the discussion of the freshwater fluxes make the argument more precise and nuanced.

• The stages of the construction of the stability landscape construction are more apparent. However, I think this may still be missing a short justification for the chosen design. For instance, why map the OFF mode only by going from high FWF to low FWF, when both directions matter in a hysteresis cycle? This makes sense when referring back to Figures 1 and 2, but a couple of sentences about this would improve the fluidity of the reading.

We have added a few sentences to the main text to justify and clarify our approach:
'The standard approach to tracing the AMOC stability diagram is to slowly change one of the control parameters (usually FWF) first in one direction and then in the opposite direction. However, such a method often fails to trace all equilibria, especially when more than two equilibrium states coexist at the same point in the phase space. Therefore, we combined the traditional approach, which works in our case for the Strong and Off modes, with a more sophisticated procedure where we alternate changes in FWF and CO2 space to trace the Modern and Weak states.'

We have also added a few sentences to the Appendix:
'The first step was to run experiments with increasing and decreasing FWF, starting from -0.5 Sv and +0.5 Sv, respectively, for all CO2 levels. These simulations are sufficient to trace the stability of the Off and Strong AMOC states (Fig. A2a,d), because (i) for large positive FWF the AMOC collapses under any CO2 concentration and (ii) for large negative FWF the AMOC always transitions to the Strong state. The stability analysis of the Modern and Weak AMOC states uses the pre-industrial state (CO2 of 280 ppm and zero FWF) as initial condition, but then requires a more sophisticated procedure to trace their stability through the 2D phase space (Fig. A2b,c).'

Although I do not have any major modifications to suggest, there are three more general revision points that I would like to raise:

• I think the manuscript is missing a clear definition of what the authors mean by stability. This is particularly important in the interpretation of the hysteresis figures. For instance, line 74 ("The AMOC in the model is monostable under pre-industrial conditions"): One could argue that this is not the case because both the overshoot and the modern AMOC branch exist at this point. Stability becomes more transparent in the slow FWF variation plot (0.005Sv/kyr), and it is not clear why the authors did not choose this experiment.

In the paper we mostly use the term "AMOC stability" in a colloquial but traditional sense as a shortened form of "tracing the AMOC stability diagram in a phase space, i.e. studying the phase portrait of the system". When we use the term "stable equilibrium state", we understand "stability" in the normal mathematical sense, i.e. small perturbation does not cause large changes. But, of course, with a numerical model we can only find stable equilibrium states. So "stable equilibrium state" = "equilibrium state" in our case. However, because to track 'true' equilibrium one would need to change the boundary conditions infinitively slowly, what we are presenting is an approximation of these states. In this context, the overshoot is clearly a result of the slow but transient nature of our experiments. We added the following sentence to highlight this:
'The overshoots are a result of the transient nature of our experiments and become less prominent with slower rates of FWF changes (Fig. B1).'
We have chosen to show the experiments with a 0.02 Sv/kyr rate of FWF change to be consistent with the rate used to perform the experiments in the 2D CO2-FWF space, for which using the even slower rate of 0.005 Sv/kyr would be computationally too expensive.

• I would also avoid the use of more or less stable, as it is not clear what this means (more or less potential modes accessible for a point in the phase space? Less likely transitions between multiple modes? Or even more or less stochastic variability"). For example in line 241 ("a CO2 increase drives the AMOC towards a stronger and more stable AMOC"): visually, this is true in Figure 2, but Figure 5 indicates that a CO2 increase from PI conditions means travelling from a monostable mode to two bistable modes, which, one could argue, is less stable. This confusion is also apparent in L219 ("Our AMOC stability landscape demonstrates why warm climates are generally stable and cold are unstable"): There again, different interpretations can arise from the reading of Figure 2 and Figure 5.

This is a good point and generally we tried to be clear on this in the paper. However, in a few places we overlooked this and have fixed it now:
Line 241: *'CO2 increase drives the AMOC towards a stronger and more stable AMOC' -> 'CO2 increase drives the AMOC towards a stronger state'*
Line 219: 'Our AMOC stability landscape demonstrates why warm climates are generally stable and cold are unstable' -> *'Our AMOC stability landscape demonstrates that interglacial climates of the Quaternary are generally stable, because of the mono-stability of the AMOC under pre-industrial-like*

conditions. The fact that the AMOC is monostable…. The existence of the Weak AMOC state has been shown by Willeit et al. (2024) to be related to Dansgaard-Oeschger events in the model, explaining the large AMOC variability observed during glacial times.'

• Figure 5 is a great tool to map the theoretical states accessible at a point of the phase space. However, because these modes also depend on the system's history, they can contradict the result of the hysteresis experiments from Figures 1 and 2 and create confusion in the text. In addition to the two examples introduced in the previous point (L241 and L219), the argument in L182 ("an Off AMOC state can not be achieved by varying CO2 alone, but only through a large enough FWF") is in contradiction with Figure 5. While these assertions are not wrong, they should be presented with more caution and nuance to capture the complexity of the tools presented here.

The AMOC modes shown in Fig. 5 don't depend on the history. "History" is always related to time, whereas stability diagram is not. The problem we face is how to trace all possible AMOC equilibrium states in 2D space using a numerical model. As mentioned in the response to a comment above, the standard method does not work for all AMOC states, and for modern and weak states we have to use a rather tricky approach. But this is purely a technical issue, because in our case any existing stable state can be "accessed" from any initial point in phase space - one just has to find the right trajectory. Of course, it is still theoretically possible that there are some equilibria that we have not been able to access with our approach. It is true that Fig. 1 and 2 only capture some of the states shown in Fig. 5, but that is actually one of the reasons why we performed experiments in the 2D phase space. This is mentioned already in the text: '…*but our results indicate that the standard method of tracing hysteresis in the FWF space may not be enough to find all possible AMOC modes*.' Generally, in the real world we can only find equilibrium states by doing transient experiments. However, by using such a slow rate of change, we are confident that we are really close to the equilibrium states, at least much closer than any previous studies.
The argument in L182 is correct, in the sense that starting from any of the 'on' AMOC states it is not possible to reach the 'Off' state by just (slowly) changing CO2.

The suggested revisions remain secondary, and I recommend publishing the manuscript after minor revisions without a new round of review.

Specific comments

Abstract

/

Introduction

/

Results

• Figure 1 – In the same way as the dip around 0.05 Sv was discussed, I believe the overshoot should also be mentioned in the text.

We have added the following sentence to describe this:
'When FWF is then slowly decreased again the AMOC recovers from the Off state with an overshoot at ~0.01 Sv, independently from the latitude at which the FWF is applied.'

• L100 – I understand the authors' point in my previous review, but I am still unsure if wrong is the right word to use here. Otherwise, one could argue that the 0.02 Sv/kyr experiment also gives a wrong illustration of the hysteresis cycle, as it produces different transitions from the slow experiment (0.005 Sv/kyr).

We are here specifically referring to whether a given rate of change captures or not the AMOC stability under pre-industrial conditions. For that it is correct to say that the faster rate of change experiments do not reflect the true (quasi)equilibrium states. In general it is of course true that also a rate of 0.02 Sv/kyr can be too fast to capture the true equilibrium states accurately, but specifically for the pre-industrial conditions the default rate of change of 0.02 Sv/kyr is slow enough to capture the correct stability landscape (monostable), while the 10x faster rate of change is not.

• L115 ("The warmer the climate the further north do the sites of deepwater formation shift, following the northward retreat of sea ice") - I find this sentence in contradiction with some of the following points of this paper, which is not what this figure shows. "A stronger AMOC is generally associated with a northward shift of the sites of deepwater formation shift, following the northward retreat of sea ice"?

Has been reformulated as suggested.

Discussion

• L214 ("for CO_2 below ∼250 ppm or FWF in the latitudinal belt 50–70◦N above ∼0.05 Sv") and L215 ("for CO_2 above ∼350 ppm or FWF in the latitudinal belt 50–70◦N below ∼-0.2 Sv") - I think this is oversimplifying the complex landscape presented in Figure 5. I would be more precise or remove these two assertions.

We have removed these two assertions to keep the statement more general. The details can be seen in Fig. 5.

• L219 (". The fact that the AMOC is monostable under pre-industrial-like conditions explains why the AMOC always recovered at the end of glacial terminations, after temporary shutdowns induced by the freshwater input (∼0.1 Sv) from rapidly melting ice sheets") - I think this sentence needs more nuance: what do the authors mean by pre-industrial-like conditions, there needs to be a reference about CO_2 concentrations during the previous interglacial, and I do not think that this explains the recovery of the AMOC during glacial termination given the transient nature of this phenomenon. Nonetheless, this is indeed a convincing argument in favour of it, and the discussion about the Pliocene is a good addition to the paper.

We have modified the sentence to:
'The fact that the AMOC is monostable for CO_2 concentrations around 280 ppm, a typical value for interglacials, in the absence of FWF also explains why the AMOC always recovered at the end of glacial terminations, after temporary shutdowns induced by the freshwater input (~0.1 Sv) from rapidly melting ice sheets.'
Interglacials typically last for ~5-10 kyrs, which is long time enough for the AMOC to equilibrate.

SI

/

Technical corrections

The technical corrections provided in the new version are satisfactory. I only have three minor points to raise:

• L51-53 – I am not sure I understand this sentence. Does this mean changes in CO2 and changes in surface ocean freshwater balance due to modifications in climate and land ice volume? Please ignore if this is not what was implied.

No, it means:
- changes in climate (which will affect temperature AND net surface freshwater flux)
- changes in surface freshwater flux due to land ice volume changes

• L103 (and Appendix A2) - It may be clearer to write 20ppm/kyr instead of 2%/kyr (0.002%/yr)? Or do is it increasing the CO2 concentration by 0.002 % every year (CO2(n+1) = CO2(n)*1.002) ?

It is actually an exponential CO2 increasing rate so that we get a roughly linear global temperature increase with time. The rate of change is 0.002%/yr, corresponding to CO2(n+1) = CO2(n)*1.00002, with a time step of 1 year. We have added the following to the Appendix A2 in order to further clarify this:
'We have chosen an exponential \chem{CO_2} change rate in order to get a roughly linear global temperature response with time, considering the logarithmic dependence of the \chem{CO_2} radiative forcing.'

• L200 ("with a cooling of ~15◦C") - cooling by up to ...

Has been fixed.