Review 1:

We thank the reviewer for the time he took to make a careful reading of our manuscript and for all the suggestions that are very useful to improve it. In the following the comments of the reviewer are in black, while our responses are in blue.

The manuscript by Swingedouw and co-authors describes an experimental protocol to systematically analyse AMOC sensitivity under external forcing from both freshwater input and climate change. The authors propose four interesting experiments, which will primarily be performed with Earth System Models (CMIP6+), and which are well suited to address the main scientific questions. The overarching aim is to obtain a better understanding of AMOC sensitivity and, in doing so, to assess AMOC-related impacts that are relevant for society.

I strongly recommend this manuscript for publication in Geoscientific Model Development (GMD). The experimental design is well motivated and clearly presented, with supporting figures and tables. This TIPMIP-OCEAN protocol will provide valuable information for the AMOC community and for assessing the risk of AMOC tipping under future climate change.

That being said, I have a few moderate and minor comments that should be addressed prior to publication.

We thank the reviewer for those encouraging words as well as for the useful comments to which we reply below.

Moderate comments and suggestions:

1. General aim of Experiments A, and specifically the reversibility component (Lines 191–192): I find the proposed experiments useful for studying AMOC weakening due to the joint contributions of global warming and freshwater input. The imposed forcing leads to a “major AMOC weakening” (Line 548; could this be quantified?) during the first 150 model years. Thereafter, forcing conditions are ramped down over 100 years and continued for an additional 100 years under no external forcing.

   Given the typical multi-centennial adjustment timescale of the AMOC (Bonan et al., 2022, already cited) and results from existing pulse-forcing experiments (Jackson et al., 2023, already cited; van Westen et al., 2025, https://doi.org/10.1029/2025JC022651), it is unlikely that a coherent picture of AMOC reversibility can be obtained from the outlined experimental design (Fig. 2). Addressing reversibility would require substantially longer simulations for experiment “Stabilized (Ad)”, which are not mandatory under Tier 1 (Table 1). The 250-year experiment “Stabilized (Ab)” is likely sufficiently long in duration, but because it includes a modified background forcing, it cannot be used to assess reversibility.

My main point is that some aspects of AMOC stability cannot be addressed using simulations that are shorter than the AMOC adjustment timescale. This also applies to Experiments C and D, where the minimum required “stabilization” periods are 50 and 100 years, respectively. Only if modelling groups are willing to extend the stabilization (i.e. no external forcing) phases beyond 100 years can these questions be meaningfully addressed.

I would therefore suggest using more careful language and tempering expectations regarding AMOC reversibility, which is one of the central scientific questions in the TIPMIP-OCEAN protocol (Line 551).

We fully agree with this comment. This question of timescale of forcing changes and associated response is actually key for what concerns tipping dynamics. Therefore, we have adjusted the manuscript to better insist on this aspect. Experiments that we propose are clearly too short for assessing steady state reversibility, which would require at least a millennium. What we were targeting here was more about reversibility on a time frame relevant for society (a few generations).

We have clarified this in the introduction section of the paper (lines 167-169, 195-196, 206).

The reference provided with a doi (van Westen et al. 2025) has been added in the manuscript. Concerning the question “The imposed forcing leads to a “major AMOC weakening” (Line 548; could this be quantified?)”, we think that the experimental design is aiming to lead to a massive weakening (cf. Jackson et al. 2023), but it is possible that it won’t be the case in some models. This will constitute a result from the protocol, but we believe that in this paper, it is not possible to specify any quantification of such a weakening. We have therefore not modified this part of the text.

2. The proposed experiments involve relatively rapid changes in forcing conditions (typically 100 years), which are short compared to the multi-centennial adjustment timescale of the AMOC. As a result, the simulated AMOC responses are likely to be dominated by the imposed forcing rather than by intrinsic AMOC dynamics. While the forcing may still trigger destabilising feedbacks (e.g., salt-advection), the experimental design appears less suitable for investigating true “tipping dynamics of the AMOC”, as suggested by the manuscript’s title. I therefore suggest reconsidering the title, for example:“Tipping sensitivity of the AMOC.”

   This comment is tightly linked with former ones, and we agree that since forcing and response timescales are similar, it is more specific to change the title according to the proposition made by the reviewer. We have modified the manuscript accordingly.

3. I note that the manuscript appears to miss a substantial portion of relevant AMOC literature, despite the authors being highly active in this research area and publishing influential work. While the AMOC literature is rapidly evolving and extensive, I have included a number of relevant (non-exhaustive) recent studies in the minor comments below to help complement the discussion. I do not expect the authors to include all of the listed references, but I encourage them to consider the most relevant ones.

   We thank the reviewer for pointing towards important papers we might have forgotten to mention.

Minor comments and suggestions:

- Lines 68 – 69: I also recommend including the more recent studies by Armstrong McKay et al. (2022) and Wunderling et al. (2024), which are already cited in the manuscript.

  Modified accordingly.

- Line 74: The term "exact thresholds" is somewhat vague and would benefit from clarification; please specify what is meant here. In addition, it may be helpful to explicitly identify the relevant ocean tipping elements in this context, particularly the two major ones: the AMOC and the North Atlantic subpolar gyre.

  We have improved this sentence into "*exact thresholds beyond which some systems can tip are poorly known*" which is hopefully clearer. Concerning the suggestion to specify AMOC and SPG systems here, we argue that this is too early in the manuscript, since this is only the 7th line of the introduction and AMOC and SPG has not been defined yet: we are in this paragraph providing very general insights on tipping points to the reader.

- Line 77: Please write out AMOC here; this is the first time the acronym is mentioned in the main text.

  Corrected, thanks.

- Lines 77 – 78: Either quantify "warm and salty" and "cold and fresh" or add 'relatively' here.

  The word "relatively" has been added, thank you for the suggestion.

- Line 80: after "Labrador, Irminger, and Nordic seas". This should have a reference, e.g., Katsman et al. (2018, https://doi.org/10.1029/2017JC013329).

  This is actually useful to support this sentence. However, we think that it is more appropriate to use a reference pointing towards observations and not only to models, as it is the case in the suggested reference. This is why we have added Lozier et al. (2019) instead.

- Lines 81 – 82: "The Atlantic nortward upper limb of the AMOC", the term "Atlantic" is redudant here and could be removed.

  Corrected, thanks.

- Line 84: I would also add the study by Orihuela-Pinto et al. (2022, https://doi.org/10.1038/s41558-022-01380-y) here, which nicely provides an overview of the relevant global climatic responses (their Figure 6).

  Good idea, we have added this reference.

- Line 85: The term "climatological equator" is unclear in this context. Do you mean that the ITCZ and/or the maximum zonally-averaged temperature is located north of the geographic equator? Please clarify.

  This reference was actually unclear. We were referring to the ITCZ location, but we believe this was too specific as compared to what was said before: there was no clear reason here to put a focus on the ITCZ while we are just saying in this sentence that the climatic impacts are widespread. We therefore remove this part of the sentence. Thank you for spotting this unclear statement.

- Lines 85 – 89: These climate impacts require substantially more comprehensive and appropriate referencing, as a single citation is clearly insufficient. Moreover, the study by Ben-Yami et al. (2024) focuses primarily on ITCZ shifts under different AMOC strengths and does not cover the broader range of climate impacts discussed here. A number of recent modelling studies have addressed AMOC-related climate impacts more comprehensively, including the mentioned effects here.
  For clarity, I provide below a (non-exhaustive) overview of relevant and recent studies:

  General impacts:

  - Orihuela-Pinto et al. (2022, https://doi.org/10.1038/s41558-022-01380-y)
  - Bellomo et al. (2023, https://doi.org/10.1007/s00382-023-06754-2)
  - van Westen et al. (2024, already cited)
  - Bellomo & Mehling (2024, https://doi.org/10.1029/2023GL107624)

  Temperature impacts and winter storms:
  - Meccia et al. (2024, https://doi.org/10.1088/1748-9326/ad14b0)
  - Meccia et al. (2025, https://doi.org/10.1088/1748-9326/ada3e7)
  - van Westen & Baatsen (2025, already cited)

  Hydroclimate:

- Saini et al. (2025, https://doi.org/10.1029/2024PA004967)
- van Westen et al. (2025, https://doi.org/10.5194/hess-29-6607-2025)

Sea level:

- Volkov et al. (2023, https://doi.org/10.1038/s41467-023-40848-z)
- Howard et al. (2024, https://doi.org/10.1088/2515-7620/ad3368)
- van Westen et al. (2026, https://doi.org/10.5194/os-22-1353-2026)

We thank the reviewer for providing such an extensive list of references to better support our claims. We have added some of them (one for each impact to avoid making the list unbalanced and uneasy to read) in the manuscript in order to be more specific for supporting the list of impacts cited.

- Line 91 – 92: A recent study by Nian et al. (2026, https://www.nature.com/articles/s43247-026-03427-w) might be worth including/discussing here.

  We have added a short discussion on it line 103-104

- Line 98: Perhaps you could roughly quantify the "large freshwater outburst" for the general reader, for example by expressing it in Sverdrups and/or as a multiple of the present-day Greenland Ice Sheet (GrIS) melt rate. I do acknowledge that this is being discussed in greater detail in lines 505 – 518.

  Indeed, this is stated and discussed later on. We do not believe it is worth providing such specific information already here in the introduction, and prefer to enter into numbers later in the manuscript.

- Line 108: I can't find the (Dijkstra, 2024) entry in the reference list, please include.

  This has been corrected.

- Lines 117–119: There is indeed a *significant* weakening trend across all SSP scenarios (Weijer et al., 2020, already cited; Dijkstra & van Westen, 2026;https://doi.org/10.1146/annurev-marine-040324-024822). However, more recent studies provide additional nuance. For instance, Bonan et al. (2025, https://doi.org/10.1038/s41561-025-01709-0) suggest that the decline is more coherent across CMIP6 models, while Portmann et al. (2026, https://www.science.org/doi/10.1126/sciadv.adx4298) report even larger declines when accounting for model biases. It may be worthwhile to also discuss these more recent findings here.

  Thank you for this suggestion. We believe that the discussion of observational constraint results is a bit too complex to be added in just a few words here. Indeed, the fact that the two references cited provide different results require an assessment of their approaches, to try to explain their differences, in order to avoid providing the feeling that they are inconsistent. However, such a (necessary) assessment is out of the scope of the present paper.

- Lines 119 – 120: The 21st century climate change-induced AMOC weakening is primarily driven by changing surface heat fluxes, with a limited contribution by surface freshwater fluxes. See the surface buoyancy flux analyses in van Westen et al. (2025, https://doi.org/10.1029/2025JC022651) and Drijfhout et al. (2025, already cited).

  We respectfully disagree with this statement. The analysis by Drijfhout et al. (2025) relies on water mass transformation and formation rates over the convection sites to argue that surface freshwater fluxes play a secondary role. However, salinity changes resulting from freshwater forcing anomalies elsewhere can also strongly influence the preconditioning of deep convection sites. Furthermore, deep convection brings relatively warm and saline deep waters to the surface, which are subsequently cooled and densified during winter. Therefore, a substantial part of the water mass transformation is driven by surface heat loss, while salinity can play a critical role in determining whether convection is initiated or suppressed. Importantly, the salinity evolution at convection sites is not controlled solely by local freshwater fluxes, but also by freshwater redistribution and transport from remote regions.

  We acknowledge that this line of reasoning can be debated, and the reviewer may not find these arguments entirely convincing. However, we do not consider that introducing such a detailed discussion on the relative roles of heat flux and freshwater forcing in driving AMOC changes is essential in the context of the present manuscript. We therefore prefer not to expand this discussion further here.

- Line 124: The 'substantial biases in CMIP-class model' needs some references:
  - van Westen & Dijkstra (2024, already cited)
  - Dijkstra & van Westen (2024, https://doi.org/10.16993/tellusa.3246)
  - Portmann et al. (2026, https://www.science.org/doi/10.1126/sciadv.adx4298)

  These references are actually necessary and have been added.

- Line 126 – 127: 'with indications that a tipping point may have been crossed', this has been suggested in the analysis by van Westen et al. (2025, https://doi.org/10.1029/2025JC022651).

  This reference has been added.

- Line 128 – 130: A reference to Vanderborght et al. (2025, https://doi.org/10.1175/JCLI-D-24-0570.1) is also appropriate here, as they quantify the 'key feedbacks governing its stability and tipping behaviour'.

This reference has been added.

- Line 131: entrance -> southern entrance

  Corrected.

- Line 132: Atlantic -> North Atlantic; as the salt-advection feedback convergences freshwater anomalies over the northern North Atlantic

  Corrected.

- Line 134: ‘presence of an AMOC tipping point’, the review study by Dijkstra et al. (2026, https://doi.org/10.1002/wcc.70049) points to the strong evidence that the present-day AMOC is indeed in a multi-stable regime (i.e., the presence of an AMOC tipping point).

  This reference has been added.

- Line 137: The statement “may not behave” is somewhat vague. This likely reflects that (quasi-)equilibrium conditions are not satisfied. Only under such conditions is an appropriate stability indicator, as it can then be meaningfully linked to the strength of the salt-advection feedback (Rahmstorf, 1996, already cited; Vanderborght et al., 2025, https://doi.org/10.1175/JCLI-D-24-0570.1). Please clarify this point.

  We have added some text (lines 148-151) and included a reference.

- Lines 141 – 142: A reference to Hankel (2025, https://doi.org/10.1073/pnas.2411357121) is also appropriate here.

  This reference has been added.

- Lines 145 – 147: What about the results from Weijer et al. (2012, https://doi.org/10.1029/2012GL051611), where a comparison is made between a high-resolution and low-resolution stand-alone ocean model simulation.

  This reference has been added.

- Lines 150–153: I do not fully agree with this statement. The pulse-forcing experiments in Jackson et al. (2023) (+0.3 Sv for 100 years) are relatively short in duration (mostly 200 years in total), which may be insufficient to assess AMOC stability. For example, van Westen et al. (2025,https://doi.org/10.1029/2025JC022651) show in their pulse-forcing experiments (their Fig. 6a, blue curve) that a 200-year integration is too short to robustly assess AMOC stability based on AMOC strength alone. This limitation should be acknowledged or discussed more carefully.

  This limitation has been acknowledged lines 167-169

- Lines 157–161: The statement regarding an overshoot does not appear to be supported by all relevant studies. For instance, the CO_2 hysteresis experiment in Willeit & Ganopolski (2024,https://doi.org/10.5194/esd-15-1417-2024) does not exhibit an overshoot, nor do the results presented by Gérard & Crucifix (2024, https://doi.org/10.5194/esd-15-293-2024).

  Thank you for this comment and for pointing us towards those interesting papers. However, we do not think they contradict our sentence which was: "often with an overshoot in AMOC strength (where the AMOC recovers to a temporarily stronger state than initially) "

  We were referring here to the transient response over time, as illustrated by the use of the term *temporarily*, rather than to the steady-state response represented in a hysteresis diagram. Furthermore, the use of the word *often* explicitly indicates that this statement does not necessarily apply to all models. We agree that the occurrence of overshoots in hysteresis diagrams is also of great interest. However, as the reviewer has pointed out several times, our experimental design does not allow us to draw fully robust conclusions regarding hysteresis behaviour because of computational limitations.

- Lines 233 & 469: The compensation indeed prevents model drift; however, more importantly, it is required to obtain equilibrium solutions (i.e., to ensure an autonomous system).

  Thank you for this comment, we have added a sentence for this argument (line 251).

- Line 235: Maybe you can already state here that you are proposing a volume compensation.

  We think this will be a bit premature to do so, for the sake of the clarity of the paper.

- Line 361: 'Impact the GMSAT', maybe good to add a reference here? For example, Drijfhout (2015, https://www.nature.com/articles/srep14877) or van Westen et al. (2024, already cited).

  This is a good idea, we have added this citation.

- Line 371: The more recent study by Pontes & Menviel (2024, https://doi.org/10.1038/s41561-024-01568-1) might be worth adding here.

  This reference has been added.

- Line 447: This claim needs some references, for example Dijkstra (2024, https://doi.org/10.1016/j.physd.2023.133984).

  This reference has been added.

- Line 451 (and elsewhere): It may be helpful to refer to this forcing protocol as a "pulse forcing" experiment, following the terminology used in Dijkstra & van Westen (2026,https://doi.org/10.1146/annurev-marine-040324-024822). Suggestion: "… constant freshwater flux for a fixed time, a so-called pulse forcing (Dijkstra & van Westen, 2026)."

  We have added that it is known as a pulse forcing experiment here.

- Line 454: "there is an indication that a threshold has been crossed", this is quite vague. Please clarify which threshold.

  This has now been rewritten to be more specific

- Line 482: There is a consensus on the sign (i.e., weakening) in AMOC projections (Weijer et al., 2020, already cited; Dijkstra & van Westen, 2026, https://doi.org/10.1146/annurev-marine-040324-024822), but there is much more uncertainty on the magnitude of AMOC weakening (Bonan et al., 2025, https://doi.org/10.1038/s41561-025-01709-0; Portmann et al., 2026, https://www.science.org/doi/10.1126/sciadv.adx4298). Maybe rewrite this sentence.

  Agreed, this is now rewritten.