Reviewer 1

This is an interesting, thorough and well-written study that investigates the effects of how global salinity is compensated for in hosing experiments of larger climate models. In particular, they investigate the differences between surface and volume compensation, which shows a lower AMOC collapse threshold for volume compensation compared to surface compensation. They further find that the AMOC threshold for a collapse is similar for volume compensation and no compensation at all, concluding that volume compensation is probably providing more trustworthy estimates of AMOC stability compared to surface compensation. They then tease out the origin of these differences, concluding that surface compensating over the Pacific can be a pragmatic compromise that limits artificial stabilization of the AMOC, if volume compensation cannot be used for technical or scientific reasons. In all cases, the study shows the importance of any study clearly stating how salinity is compensated for, and that the quantitative outputs from studies should be interpreted in that light.

The study is an important contribution to better understanding of model outputs from large and intermediate complexity climate models, and is a substantial step forward for providing more robust estimates of the risk of an AMOC collapse.

We thank the reviewer for the careful reading of the manuscript and for the positive assessment!

Minor comments

Should Delta H be Delta F_H in the caption of figure 1? In Figure B2 it is called Delta H.

Thank you for spotting this, this should indeed be Delta F_H. We will correct the labels in Figs. 1 and B2.

Figure B4: a, b, c, d indications are missing in the figure.

The missing labels will be added.

Reviewer 2

This study presents a detailed assessment of the impact of different freshwater compensation methods that are routinely applied in climate models when investigating the stability of the AMOC to freshwater hosing in the North Atlantic. It investigates differences in the critical freshwater forcing leading to a collapse of the AMOC originating from compensating for the freshwater hosing flux either globally at the surface or over the global ocean volume. By using a stand-alone ocean model and a climate model of intermediate complexity they show that the results are robust. The main conclusion is that the volume compensation approach is more suitable to quantify AMOC tipping points, while global surface compensation overestimates the stability of the AMOC due to the applied surface salinity flux over areas of the Atlantic outside of the hosing region.

I find the results to be relevant for the debate about the stability of the AMOC in climate models. The paper is very well written, the analyses are very detailed and the results are clearly presented. I therefore recommend publication after the few minor comments below have been addressed.

We thank the reviewer for the positive assessment and for the comments that have helped improve the manuscript.

Minor comments

L.11: This is true only for CLIMBER-X, but not for POP, for which the AMOC recovery is more sensitive to the compensation choice than the AMOC shutdown (Fig. 1a).

This is a good point. What we wanted to highlight here is that there is no qualitative difference in the recovery as in Jackson et al. (2017). Therefore, we will change this sentence to

> In contrast *to an earlier study*, the compensation method *does not introduce qualitative differences* in AMOC recovery when tracing the full hysteresis loop.

In the abstract it could be worthwhile making more explicit which compensation method is the 'most appropriate', as a conclusion of the analyses presented in the paper.

We appreciate this suggestion and will add the following sentence to the abstract:

> *In light of our results, volume compensation appears to provide the best trade-off between global salinity conservation and similarity to the effects of a physical freshwater flux.*

Section 3.2, Fig. 3 and Fig. 4: it is unclear for which model this is shown. Please clarify.

We will state explicitly that these plots as well as Section 3.2 only deal with POP. We reference the corresponding supplementary figures for Climber-X where available.

Fig. 3: There are many overlapping lines in this figure. I would suggest to consider splitting this into two panels, S-comp and V-comp. Or alternatively play with different colors and line styles. Please also consider using color-blind friendly color schemes.

We intend to keep both S-comp and V-comp in the same panel to demonstrate that the S-comp reconstruction matches that of V-comp well after subtracting the compensation tracer. However we happily take up the reviewer's suggestion to make the lines more distinguishable using different line styles. We will also check all figures with a color-blind simulator and choose different line colors where appropriate.

L.207-208: And what about mixed layer/convection?

Yes, to our knowledge this also includes mixing and convection. We therefore adapt the following, more general wording:

> The passive tracer follows *the same model physics* as the prognostic salinity.

Given the results in Fig. 7, I'm wondering to what extent plotting the AMOC hysteresis curves as a function of net surface freshwater flux into the whole Atlantic and Arctic (considering both the hosing and the compensation fluxes), instead of FH, would make the S-comp and V-comp curves overlap.

Thank you for this very interesting suggestion. Indeed S-comp and V-comp match very well in POP (and at least become closer in Climber-X) when the AMOC strength is plotted against this "effective F_H" taken at the surface over the entire Atlantic and Arctic basins. This figure (Fig. R1) will be included in the appendix of the revised manuscript along with a short discussion in the main text.

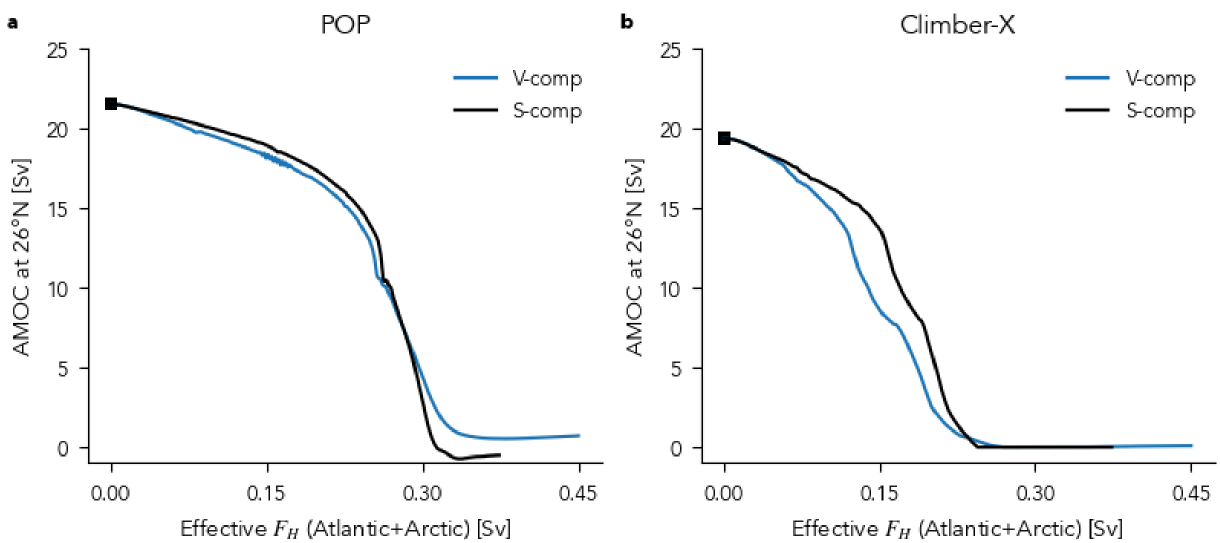

Fig. R1: Forward branch of the hysteresis curves against net freshwater forcing over the entire Atlantic and Arctic basins ("effective F_H").

The 'plateau' in Fig. B2 resembles the weak AMOC state described in Willeit & Ganopolski (2024), which they describe as a stable state in CLIMBER-X (see e.g. their Fig. 4b).

Thank you for bringing this to our attention. Our initial assessment was based on a ramp-down simulation from F_H=0.15 Sv, but we now performed equilibration simulations at different levels of F_H and found that this is indeed a stable state for some values of constant forcing (e.g., F_H=0.175 Sv). Therefore, we will change the caption of Fig. B2 to

> *The "plateau" in the V-comp case appears to be an intermediate stable state resembling the stable "weak AMOC" state of Willeit & Ganopolski (2024).*