Authors' response to reviewers' comments on "Past, Present, and Future Variability of Atlantic Meridional Overturning Circulation in CMIP6 Ensembles"

We sincerely thank the reviewers for their insightful comments and constructive suggestions, which have enabled us to clarify key aspects and enhance the overall quality of our manuscript in this second revised version. Below, we provide detailed responses (in blue) to each of the reviewers' comments (in black), along with a revised version of the manuscript.

Arthur Coquereau, Florian Sévellec, Thierry Huck, Joël J.-M. Hirschi, and Quentin Jamet

Report #1

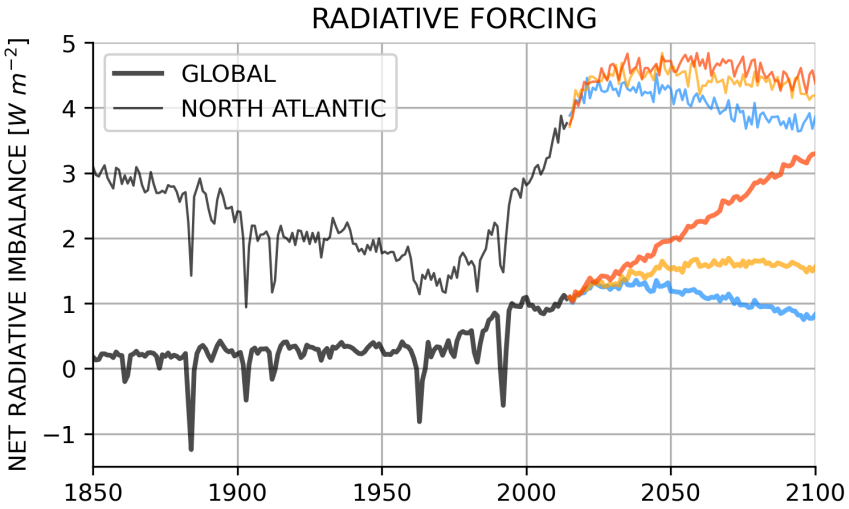
The authors have done a great job in addressing most points raised by reviewers and editor and the paper has accordingly improved.
I still disagree, however, with their response on my points 8 and 10. I recommend acceptance with minor revision, urging them to give a more balanced discussion here.

Main point (3.2.2):
In my view, after 1980-2000 and up to ca. 2050 aerosol removal and increasing greenhouse forcing apear similarly important in driving AMOC decline, with ssp126, ssp245 and ssp585 showing rather similar trends in the early half of the 21st century that only diverge after 2050.
The authors counteract this by noting that the globally averaged radiative forcing of the two is very dissimilar and my argument is not consistent with the evolution of the radiative forcing imbalance diverging quickly after 2015 in the ssp scenarios. I do not dispute this fact, but the authors seem to overlook that the two have a very different spatial pattern in terms of radiative forcing, with aerosols inducing a much more heterogeneous pattern with a strong maximum over the Atlantic NH sector, thus affecting deep mixing in the SPG much more effectively than global warming for the same globally averaged radiative imbalance caused by these two forcings.
They can find ample evidence for this in the two Menary and the Robson papers cited in this ms and maybe could give a look to the Deser et al. paper as well:

https://journals.ametsoc.org/view/journals/clim/33/18/jcliD200123.xml

We agree that local aerosol forcing could also play a role. However, we previously examined this hypothesis by analyzing the net radiative imbalance over the North Atlantic region (10°-60°N, 90°W-0°E) in the three large ensembles (see figure below), and we did not find comparable inter-scenario behavior up to 2050, as the scenarios begin to diverge much earlier. Nevertheless, we agree that our conclusion should be softened, given the possibility of alternative explanations and the fact that a comprehensive investigation of all potential mechanisms lies beyond the scope of the present study. The paragraph in the conclusion has therefore been updated accordingly:

"*The AMOC intensity decline associated with the transient increase of time variability, could suggest that this evolution is not forced by scenarios nor synchronous to anthropogenic emissions, because of the absence of clear scenario separation before 2050. In this respect, AMOC behavior differs from global CO2 concentration, anthropogenic radiative forcing, and temperature change, that already start diverging one or two decades earlier (O'Neill et al., 2016). Although this trend could be influenced by regional aerosol forcing (e.g., Deser et al., 2020; Menary et al., 2020; Robson et al., 2022), analyses of the local radiative imbalance over the North Atlantic (10°-60°N, 90°W-0°E) also show an early and rapid separation among scenarios (not shown).*" (l. 543-549)



Caption: Evolution of the net radiative imbalance averaged across the three large-ensemble models, shown at the global scale (thick lines) and over the North Atlantic region (10°–60° N, 90° W–0° E; thin lines). The historical period is shown in black, and the future scenarios are shown in blue (SSP1-2.6), orange (SSP2-4.5), and red (SSP5-8.5).

Minor point

I didn't understand the last sentence of the abstract. It reads as if internal variability is proportional with emission scenario intensity, but it also decreases proportional to AMOC intensity. The latter, however, is more inversely proportional to emission scenario intensity (I

think the authors mean radiative forcing), thus I think IV also is inversely proportional to this (decreasing simultaneously with AMOC strength) and the decrease in IV is proportional to the emission scenario intensity.

We agree with the reviewer's suggestion, we should have specified "inversely proportional". This has been updated in the abstract:

"*A key finding of this work is the evidence that internal variability decreases simultaneously with AMOC intensity and seems inversely proportional to emission-scenario intensity.*" (l. 14-15)

Report #2

This manuscript proposes an analysis of variance of the AMOC in the CMIP6 ensemble of historical and future simulations. The analysis is based on the ANOVA decomposition, previously used with 3 dimensions and here analysed in 4 dimensions. The ANOVA method has already been applied to explain the AMOC spread in terms of intermodal, intermember and inter-scenarios. The main novelty of the paper is to introduce the time dimension as an additional one, so as to assess the evolution of time variability along future scenarios and to compare uncertainty due internal variability as compared to the other sources of spread. Furthermore, one of the strength of ANOVA is to quantify the interactions among the various dimensions
The paper proposes an interesting view on an already well described result which is the spread of AMOC in climate simulations. The approach allows to formalize some assessment and sounds thus promising. Nevertheless, I propose that this paper is strongly revised before being published. My arguments are not so much on the science itself, but more on the way it is presented, argued and related to previous work.

1. the manuscript generally does not relate clearly to what was already shown/known in the literature. Some major papers assessing the spread and uncertainty in AMOC simulation are not cited (Buckley et al 2016 https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015RG000493, Baker et al 2023 https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2023GL103381, Jackson et al 2023 https://royalsocietypublishing.org/doi/epdf/10.1098/rsta.2022.0187, …, citing here only the main review ones). The fact that the effect of the scenarios on AMOC emerges only some decade after the end of the historical period was for example already discussed in previous studies.

Thank you for pointing us to these additional references on AMOC variability and on inter-model differences in North Atlantic state, AMOC intensity and variability, as well as the sources of related biases. We have incorporated these references into the revised manuscript.

"*In this regard, a recent review provided a comprehensive description of the scales involved in AMOC variability (Hirschi et al., 2020). It shows that the variability unfolds over a large collection of temporal and spatial scales, driven by diverse physical processes—from daily to interannual and even decadal fluctuations. These include fast synoptic weather systems and near-inertial gravity waves, mesoscale eddies and large-scale baroclinic waves, as well as slower climate modes such as the North Atlantic Oscillation or the Atlantic Multi-decadal Oscillation. Accordingly, a common simplified distinction is made between intra-annual variability, primarily linked to Ekman wind forcing, and interannual-to-decadal variability, which is more strongly associated with geostrophic processes and large-scale density contrasts (see Buckley and Marshall, 2016, for a review).*" (l. 26-33)

"*It should be noted, however, that despite continuous progress, models still exhibit significant biases in their representation of the AMOC. This stems in part from the fact that the AMOC arises from a complex interplay of processes, many of which involve small-scale dynamics—such as mesoscale and submesoscale eddies or narrow boundary currents—that remain unresolved in most climate models (Hirschi et al., 2020; Jackson et al., 2020, 2023; Gou et al., 2024). As a result, many simulations produce an overturning circulation that is too shallow and a western boundary current that is overly strong, while underestimating AMOC variability on interannual to decadal timescales (Jackson et al., 2023; Gou et al., 2024). Substantial model uncertainty also persists in estimates of the AMOC strength at 26.5° N, and models differ markedly in their depiction of overturning within the western subpolar gyre, as well as the water exchanges with the Indo-Pacific Ocean (Weijer et al., 2020; Jackson et al., 2023; Baker et al., 2023). Beyond mean-state biases, differences in properties such as Labrador Sea salinity—sensitive to model and resolution—further influence AMOC by modulating the intensity and variability of dense water formation and, consequently, the strength and variability of the overturning itself (Jackson et al., 2020, 2023). It is therefore important to keep in mind the biases of these models when interpreting the results.*" (l. 87-98)

"*The second increase, starting at the very end of the 20th century, seems due to the important AMOC decline and to the differences of decline magnitude among models. For example, it has been shown that the rate of AMOC weakening is linked to surface salinity and dense water formation in the Labrador sea and further influenced by model horizontal resolution (Jackson et al., 2020, 2023). Historical AMOC strengths and pathways observed in models also appear to correlate with their rate of weakening (Baker et al., 2023).*"(l. 322-326)

"*This illustrates that a fraction of internal variability is present in the model dimension and that averaging across models removes a part of internal variability—consistent with the logic previously discussed for the scenario dimension (see Section 2.2.2). This behavior reflects the fact that inter-model differences also manifest in terms of internal variability, which recent studies have linked to salinity biases and differences in dense water formation in the Labrador Sea, as well as to model horizontal resolution (Jackson et al., 2020, 2023).*" (l. 501-505)

We also agree that the Result section was lacking some reference to the literature where the common decline among scenario at the beginning of the 21st century were observed, this has been updated:

"*The absence of direct scenario main effect variability is also evident by S, that remains at zero before 2040-2050 (Fig. 5c and d). This particular behavior was highlighted by Weijer et al. (2020) and acknowledged in the latest Intergovernmental Panel on Climate Change (IPCC) report (AR6, Lee et al., 2021, p. 576).*" (p. 431-433)

2. Seminal papers such as Hawkins et al 2009 are also not discussed neither in the introduction not in the discussion. It should be introduced right at the beginning as the most classical method used to separate various contributions to spread, and then in the conclusion to compare results and advances.

This has been added to the manuscript.

Introduction:

"*The analysis benefits from a large body of work, over the last decades, dedicated to partitioning the variability (or the uncertainty) in climate simulations (Hawkins and Sutton, 2009, 2011; Yip et al., 2011; Sévellec and Sinha, 2018; Lehner et al., 2020; Zhang et al., 2023).*" (l. 52-54)

"*Traditionally, internal variability was assessed by computing the variance of the residuals of a climate variable (after removing trends) over a fixed period, and was therefore often assumed to be stationary (e.g., Hawkins and Sutton, 2009). However, this hypothesis had already been highlighted in the seminal article (Hawkins and Sutton, 2009), and after being challenged in numerous articles, the latter demonstrated that the climate phase space evolves over time and that internal variability is anything but stationary when a forcing is applied (e.g., Cheng et al., 2016; MacMartin et al., 2016; Coquereau et al., 2025).*" (l. 59-63)

"*The ANOVA approach also enables us to move beyond another key assumption of the widely-accepted methodology (Hawkins and Sutton, 2009, 2011; Lehner et al., 2020)—namely, the additivity of variability/uncertainty components—by explicitly exploring their interactions.*" (l. 73-76)

Conclusion:

"*The ANOVA method employed in this study enables a detailed analysis of interaction terms between sources of variance, which is not possible with the widely used approach of Hawkins and Sutton (2009). This approach allows us to go beyond the assumption of additivity of variance associated with each individual dimension and directly examine cross-terms and interdependencies between different dimensions. With the ANOVA, by combining the main effect of each dimension with their interactions, we achieve a full reconstruction of the total*

variance. This contrasts with the previous method, where total variance was simply the sum of explicitly computed components." (l. 590-595)

3. Adding time variability as an additional dimension is interesting but adds a real complexity as to what is called variability. I strongly advise the authors to check the vocabulary more precisely and possibly use the term "spread" when relating to the models or members dimension in particular.

In response to the reviewer's comment, we have updated our terminology. The phrase "model variability" has been replaced with "inter-model variability," which is more commonly used in the literature. Similarly, "ensemble variability" has either been explicitly defined as the inter-member spread or replaced with "ensemble spread," depending on the context.

4. Description of the figures and the Conclusions are sometimes very wavy and unclear.
See for example comments on Fig 2 below. I have had difficulties to follow?
Other example: L295 I am not convinced with the argument that the different behaviour at the end of the period is due do different sensitivity to external forcings? . Why would the large ensembles exhibit similar sensitivity and not the small ones? Averaging or not internal variability can also lead to very different behavior (see for example Bonnet et al 2021 https://www.nature.com/articles/s41467-021-26370-0 discussing this point). This also somewhat contradicts something that is said later in the text (l448-450) that there is not intrinsic differences between models that have performed large and small ensembles. One way to back up your assertion would be to test more carefully the effect of the ensemble size, by checking the results of large ensemble taking only a subset of members. This is partly done in B5 but I suggest a bootstrapping on more selections of members for more robust results. Finally, l296: I don't understand the argument of the reference period, please clarify.

We agree that internal variability is another plausible explanation for the differing AMOC trends among the small ensembles, as suggested by Bonnet et al. (2021). The paragraph has been updated accordingly.

"*After 2040, the trajectory of the time series becomes increasingly uncertain, as evidenced by the growing divergence between small and large ensemble models (grey dashed line). This uncertainty partly arises from differences in AMOC sensitivity to external forcing between the two groups. Specifically, the three large ensemble models show similar AMOC sensitivity, while the seven small ensemble models exhibit greater diversity in their responses to external forcing (e.g., Fig. 2 in Weijer et al., 2020). This study, which focuses on AMOC anomalies relative to the historical state (defined as the 1850-1900 mean), emphasizes the AMOC's sensitivity to external forcing, i.e. its tendency to shift states as forcing deviates from the historical baseline. Consequently, the small ensemble models, representing a broader range of sensitivities, exhibit an increasingly larger inter-model spread over the second half of the 21st century. In addition to the distinct characteristics of the individual models within each group and the greater number of small ensemble models analyzed, this divergence is also potentially associated to the poorer*

separation of internal variability in small ensembles which may contribute to differences between models (Bonnet et al., 2021), although selecting only three members from each large ensemble model does not increase model variability (Fig. B5)." (l. 327-338)

5. Scenarios and time variability: some elements of the external forcing are known to influence time variability: volcanoes, modulations of the atmospheric aerosols loadings etc. A discussion of how these components are treated in the scenarios and implications for the results is needed.

Volcanic eruptions are indeed not included in the future scenarios, thereby removing one source of time variability. However, the absence of this contribution appears to be relatively minor compared to other sources of variability, since the main time effect (T), where the influence of volcanic eruptions would be expected, remains stronger at the end of the period than at the beginning. The sources included in the scenarios are detailed in O'Neill et al. (2016), that have been added to the manuscript. We have also updated the manuscript to acknowledge the absence of volcanic eruptions in future scenarios.

"*Accounting for time variability is particularly critical in the historical period where there are no forcing scenarios. In such cases, neglecting the time dimension prevents the detection of externally driven changes in the climate state, for example those associated with volcanic eruptions.*" (l. 84-86)

"*For the 21st century, three major SSP are studied to estimate the forcing-scenario variability: SSP1-2.6 ("Sustainability"), SSP2-4.5 ("Middle of the road") and SSP5-8.5 ("Fossil-fueled development"). In the future scenarios, volcanic eruptions, which constitute short-term external forcings common to all members and can induce substantial climate variations, are not included, thereby removing a source of time variability. For more information on sources of climate variability in the SSP, see O'Neill et al. (2016). Overall, four dimensions are investigated: model, scenarios, ensemble, and time."* (l. 123-128)

6. The English level is in general quite poor and sometimes makes the understanding pretty hard. I recommend a thorough reading by an English-native. I corrected a few obvious grammar mistakes below, but the improvement required is more general.

<u>Specific comments:</u>

7. L6 «the numerous terms » : which ?

The expression "numerous terms" has been replaced by "variance components".

8. L8-10: what is the difference between what you call "internal variability" and "time variability" (same question in other instances in the text)

Time variability refers to changes in the system through time, whereas internal variability refers to variations among the possible system states, as represented by the ensemble spread. This distinction has now been explicitly clarified in the manuscript:

"*The first regime, spanning most of the historical period, is characterized by a relatively stable AMOC dominated by internal variability (i.e., ensemble spread).*" (l. 8-9)

9. L10 forcing differences: do you mean scenarios?

The expression "forcing differences" has been replaced by "forcing scenario differences".

10. L12: model variability:the models don't vary. Do you mean "spread"?

We agree with this comment, "model variability" has been replaced by "inter-model variability" throughout the text.

11. L40: the link between AMO and subpolar gyre is not really expalained

A detailed description of all AMOC mechanisms is beyond the scope of this statistical study. However, we agree that AMO is a good example to cite in terms of significant climate mode. It has been added to the introduction.

12. L40: change-> changes

This has been updated.

13. L68: I have had a hard time understanding the concept of dimensions here during my first reading. I don't think it was introduced before. Lease clarify.

We agree and have therefore detailed the dimensions concerned in order to facilitate understanding:

"*The ANOVA approach also enables us to move beyond another key assumption of the widely-accepted methodology (Hawkins and Sutton, 2009, 2011; Lehner et al., 2020)—namely, the additivity of variability/uncertainty components—by explicitly exploring their interactions. Zhang et al. (2023) computed an ANOVA decomposition involving three dimensions (3-way; i.e., models, realizations, and scenarios) and focusing on ensembles.*" (l. 73-77)

14. L70: some comments on the result of Zhang et al 223, not just their methods, is needed.

Thank you for this interesting comment, which we have included as:

"*They applied this decomposition to temperature and precipitation, with a separation between ensemble members, scenarios, and models and showed that interactions account for almost half of the variance in surface temperatures."* (l. 77-79)

15. l. 75-77 I don't understand

We clarified this sentence by adding an example:

"*Accounting for time variability is particularly critical in the historical period where there are no forcing scenarios. In such cases, neglecting the time dimension prevents the detection of externally driven changes in the climate state, for example those associated with volcanic eruptions.*" (l. 84-86)

16. l93-95: why is this a problem? Model ensembles averaging 1 member / model exist and are found to be robust for certain features. In this case, one does not care which member model is averaged with which. The numbering contains nothing physical

This is not a problem. We simply tried to provide a rationale on how inter-model and inter-scenario averaging should be performed when applied to ensemble simulations. Our rationale builds on the initialization procedure, and the fact that '*one does not care*' is due to this procedure not providing any clock to the system. Hints of this initialization procedure sensitivity can be found in Hawkins et al. (2016).

"*This initialization strategy does not impose a specific clock to individual ensemble members such that it is not possible to statistically distinguish one ensemble member from another (although they might have dynamical peculiarities with important implications, e.g. Hawkins et al., 2016).*" (l. 112-114)

Hawkins, E., Smith, R. S., Gregory, J. M., & Stainforth, D. A. (2016). Irreducible uncertainty in near-term climate projections. Climate Dynamics, 46(11), 3807-3819.

17. l109-110: I still don't understand the dimensions at this stage

This has been clarified in the updated paragraph and in response to comment #13.

"*It enables us to investigate the explanatory power of several qualitative factors (or dimensions, e.g. the various time steps, ensemble members, models, scenarios) on a quantitative variable (here, the AMOC intensity)*" (l. 132-133)

18. l139: vary -> varies

This has been fixed.

19. l144: why main? The direct effect?

The "main effect" is the standard term used in ANOVA, and we now introduce it at its first occurrence:

"*This method allows us to separate the total variability into the direct effect of each dimension (referred to as the "main effect" in the ANOVA framework) and the interactions among them.*" (l. 72-73)

20. L147: Nr is not introduced I think

We have added:

"with N_a the number of samples in dimension a, e.g. N_r is the number of ensemble members." (l. 174)

21. L160: (7) is just another formulation of (6) right? I don't understand the phrasing 'similarly…'

We agree, "Similarly" has been removed.

22. L161: "described" -> "named after"

The change has been made.

23. L170: the discussion of Hawkins et al and Lehner et al should be moved to the introduction

We have now introduced these seminal papers in the introduction.

24. L187: want -> wants

The change has been made.

25. L195 and around: I don't understand the articulation between eq (10) and (11). I am expecting that there is one mathematical computation for the variance, so that either (10) or (11) is correct. Here, you seem to say that this decomposition is partly empirical. Please clarify. If (10) overestimates the total variance, how can it be written as an equality?

Equation (10) corresponds to the sum of all ANOVA components that involve a given dimension. The example is provided for the ensemble dimension, where all components involving "R" (for realizations) are summed. We explain that this sum is exactly equivalent to computing the ensemble variance, then averaging over the other dimensions—a classical approach for defining internal variability when ensemble simulations are available.

We then discuss this grouping of components by comparing it with the method proposed in Zhang et al. (2023), highlighting the advantages and limitations of both approaches. Equation (10) enables the reconstruction, from the ANOVA, of quantities commonly used in the literature and that have clearer physical interpretations—for example, the ensemble variance. However, it cannot be applied systematically to all dimensions and then summed to obtain the total variability, as this would overestimate the total variance (though not the mean ensemble variance). In contrast, Equation (11) separates each component, which prevents overestimation

of the total variance when all dimensions are treated and then combined. The drawback, however, is that its physical interpretation is less intuitive, since variability is partitioned among dimensions without explicitly accounting for the underlying physical mechanisms.

This distinction is reflected in our use of two different terms: “Mean Ensemble Variance,” and “ENSEMBLE,” the latter corresponding to the reconstruction method of Zhang et al. (2023).

We have added the following equation between the two equations mentioned above in order to clarify this aspect:

“Total Variance < Mean Ensemble Variance + Mean Time Variance + Mean Model Variance + Mean Scenario Variance.” (11)

26. L208-210: related to that, do you mean here that (11) is not physical? Why? Furthermore, why does (11) prevent from comparison with traditional variance computation? What are the traditional variance computation? Do you mean just R?

This has been clarified in the paragraph:

“*In the next section, we will employ this separation method to provide an overview of the variance distribution. However, this method does not allow for a direct comparison with traditional variance computation (e.g., Mean Ensemble Variance). Moreover, the statistical separation relies on a dimensional-wise approach rather than physical considerations. As a result, a single physical mechanism driving changes in variance may influence several variance components simultaneously. For instance, the ST components, which might highlight differences in trends between scenarios and therefore appear naturally linked to the scenario dimension, would nevertheless be partitioned equally into the time dimension. Finally, compensations can occur between ANOVA components and cause spurious results with the statistical separation method.*” (l. 234-240)

27. L222 more than what?

The word “more” has been removed.

28. L223 “phase shift”: what are you referring to here? A phase shift between time series? Which ones? Where do you see that?

This phase shift reflects the fact that internal variability spreads the individual simulations, each exhibiting a different temporal phasing of its internally generated variability relative to the others. We have now explicitly stated that this phase shift arises from internal variability.

“*When the trajectories separate under different forcing scenarios, a slight phase-shift of internal variability appears.* ” (l. 252-253)

29. L236 is B2 computed also for the synthetic model

B2 is calculated for CMIP6 data. This has been clarified in the paragraph:

"*During the CMIP6 historical period, R represents the part of the internal variability that is not captured in RT , i.e., the internal variability with a period larger than the rolling time window of 30 yr. The sensitivity test on time windows depicts an anti-correlation of R with RT , and with the length of the time windows (Fig. B2).*" (l.266-268)

30. L247 is "mixture" the correct term? It sounds a bit like a recipe no?

We have updated this paragraph:

"*To obtain robust results and assess the model uncertainty it is important to mix the different models. As the ensemble models have different sizes, we used a Bootstrap methodology to aggregate them. Only models of comparable size are combined.*" (l. 276-278)

31. L258 I am not sure the "first regime" was introduced at this stage

We have updated this paragraph:

"*Specifically, during the historical period, we observe a transfer from R and RT to T , and a transfer from RT and SRT to ST in the 21st century.*" (l. 288-289)

32. L269: "slow": do you mean "weak"?

This has been updated.

33. L269: associated with increased aerosol concentration: this is an hypothesis proposed by some studies that you do not demonstrate here. Please temperate the affirmation

We have updated this paragraph:

"*Some models then present a weak increase often attributed in the literature to increasing aerosol concentration (Menary et al., 2013, 2020; Robson et al., 2022).*" (l. 299-300)

34. L276: studies by Weijet et al, Jackson et al etc should be cited here

The references has been included in the paragraph:

"*This brief overview reveals substantial AMOC variability across time, scenarios, and models—consistent with previous studies (e.g., Weijer et al., 2020; Jackson et al., 2023)—from the last decades of the 20th century. This behavior is expected to persist and even intensify throughout the 21st century.*" (l. 306-308)

35. L284-285: the first increase that I see occurs in 1930-1050. Is this what you call the 2nd half of the century? And this is not a typical aerosol period. PLease clarify

We agree that the original sentence was unclear, and it has now been clarified as follows:

"*Two consecutive periods of increase detach from this time series. The first, which reached its peak in the second half of the 20th century, is likely associated with the differences of AMOC response to aerosol forcing among models (Menary et al., 2013, 2020; Robson et al., 2022).*" (l. 319-321)

36. L290: why do you attribute this later increase in small ensembles to the same causes? How do you know?

We do not exactly "attribute" the changes; we only refer to the literature that emphasizes the dominant influence of aerosol forcing during the second half of the 20th century. The later peak observed in the small ensembles may arise because its timing is partially obscured by the limited separation of internal variability and by the larger inter-model differences in the initial AMOC trends.

37. L299" model variability". This term is very confusing. "spread" rather?

This sentence has been updated to clarify:

"*Consequently, the small ensemble models, representing a broader range of sensitivities, exhibit an increasingly larger inter-model spread over the second half of the 21st century.* " (l. 333-334)

38. L319: "analysis of the model-associated interactions": where is this done?

The section 3.3 has been referenced:

"*As it shows the robustness of the study, model-associated variability and interactions are discussed later in subsection 3.3. In that subsection, the analysis of the model-associated interactions shows that they are mainly governed by physical factors. This makes it more natural to analyze the physical factors first, even if their associated variance is smaller.*" (l. 356-359)

39. L326: "significance of the results": where is this at all done?

We agree that the term "significance" can imply "statistical significance." We have replaced it with "consistency."

40. L326 "analysis of the model-associated variability": you have just said that this is not done here for these small ensembles?

We have updated this sentence to clarify:

"*However, the small ensemble models remain important both for assessing the consistency of the results and for the analysis of model-associated variability (since there are more models with a small ensemble) performed in section 3.3.*" (l. 363-365)

41. L349 "In this context, internal variability dominates: Over what? Why? I don't understand

We have updated this sentence to clarify:

"*In this context, internal variability is the dominant component among the "physical factors" of variability.*" (l. 377-378)

42. L355 Please specify which panel you are looking at here. Why not commenting the large ensembles as well?

We agree that both ensembles should have been discussed and the panel specified; this has now been updated.

"*Going into detail, the ensemble-to-time variance ratio appears slightly smaller than one for small ensemble models, whereas it is slightly greater than one for the large ensembles (Fig. 4a).*" (l. 393-394)

43. Fig 5 is largely redundant with Fig 3. Isn"t there a way to merge the two?

We agree that Fig. 5 is somewhat redundant with Fig. 3. However, it greatly simplifies the analysis of interactions and main effects between two important pairs of dimensions: time and realizations, and time and scenarios. We believe this figure is particularly useful for clarifying the message and understanding the role of interactions and their evolution over time, which is not straightforward when all interactions and main effects are combined.

44. L371 "This period" -> "the first century of the historical period"

This has been updated.

45. L379 please check the grammar of this sentence

This has been done:

"*Around 2000, the increase in the time main effect caused it to exceed the ensemble main effect in the large ensemble models (Fig. 5a).*" (l. 418-419)

46. L387: "slight differences": quantify and be more precise. What type of perturbations are these? In general, discussion of how forcings that can induce variability (volcanoes, aerosols etc) are treated in the scenarios is welcome

These slight differences correspond to small differences between scenarios (for instance in terms of CO2 emissions, e.g. O'Neill et al.; 2016) that are amplified by the chaotic nature of climate models. This has been clarified in the manuscript:

"*This increase of scenario-associated variability before stabilizing, is therefore not the direct effect of the forcing but, rather, a sort of chaotic internal variability triggered by slight differences among scenarios allowing the simulations to spread. Indeed, during the historical period, the three time series corresponding to scenarios for each member of models are perfect replicas of the historical simulation. However, when the scenarios formally begin in 2015, small differences or perturbations in forcing emerge among them due to the progressive separation of scenario trajectories—for instance in terms of CO2 emissions (O'Neill et al., 2016). These small differences are amplified by the chaotic nature of the system, causing the three scenario time series to spread similarly to pseudo-ensemble members. This internal and chaotic nature of SRT variability is underlined by the magnitude of this variability, which appears to be three times as large in large ensemble models as in small ones, consistently with RT that is greater in large ensembles.*" (l. 422-431)

47. L388 "pseudo ensemble members": I don't understand (or I am not convinced) why the effect is different from large ensembles?

The fact that SRT (associated with internal variability) is stronger in large ensembles is consistent with the observation that RT is also stronger in large ensembles. This is because when many realizations are averaged, internal variability is better isolated. This clarification has been added to the manuscript, as indicated in our response to comment #46.

48. L418: this sentence is very confusing in my view. Please clarifiy what is declining and increases: AMOC strength, variance ; the effect of time variability on the total variance, …

We have updated this sentence to clarify:

"*Beyond the AMOC decline associated with the transient increase in time variability, the T main effect initially decreases and then stabilizes together with SRT around 2050.*" (l. 460-461)

49. L430: "be thus directly" -> thus be direcly

This has been updated.

50. L436 and below: "mode associatedl variability" is a misleading phrasing. "variance associated with model differences", of "inter-model spread"?

We have updated the text to use "inter-model variability" instead of "model variability," and the sentence has been revised as follows:

"*After focusing on the physical factors of variability, we can investigate variability associated with inter-model differences and the drivers of these differences throughout the multisecular simulations.*" (l. 478-479)

51. L440 "This follows" -> "It follows"

This has been updated.

52. L448-450: see general comment above + are you sure it is due to inherent characteristics of small ensembles? Larger modelling groups generally are the ones having enough resources to run large ensembles… Or run large ensembles on their model version that is more robust, more carefully adjusted etc. R. Knutti et al have probably discussed this idea of model democracy

Thank you for this comment. To the best of our knowledge, such a bias in small ensembles has not been reported for AMOC. We have nonetheless rephrased as:

"*Taking aside potential inherent characteristics of small ensembles (which would break model democracy ; Knutti, 2010) and the impact of internal variability on the model trends, we interpret this as being also associated with the particular combination of models in the "small ensemble" and "large ensemble" category, which happen to exhibit different AMOC sensitivities to forcing.* (l. 490-493)

"*When performing inter-model statistics, we adopt a model democracy approach (Knutti, 2010), assigning equal weights to each model.* " (l. 121-122)

Knutti, R. (2010). The end of model democracy? An editorial comment. Climatic change, 102(3), 395-404.

53. L450: please recall what "reference" you are refering to here

The reference has been specified:

"*These results are also sensitive to the reference chosen for computing AMOC anomalies, as discussed in the method section.*" (l. 493-494)

54. L451 please check the grammar (and the position of the ",")

This sentence has been updated.

"*Indeed, when considering the absolute AMOC intensity, we observe a convergence in AMOC intensity among the small ensemble models and a divergence among the large ensemble models after 2040 (Fig. B3).*" (l. 494-496)

55. L452: I don"t understand this point and I don't understand how B3 shows this

The convergence or divergence among the ensemble models refers specifically to the AMOC anomalies. This means that the small ensembles tend to converge toward similar absolute AMOC intensities, whereas the large ensembles exhibit increasingly divergent values. This clarification has been incorporated into the sentence mentioned in the previous comment.

56. L465 what are the first two ways"?

This sentence has been rephrased for clarity as follows:

"*To summarize the previous findings on the ANOVA components, we propose a third approach—complementary to statistical separation and variance reconstruction—that organizes the variance components following a physical attribution.*" (l. 511-512)

57. L468-469: a verb is missing in this sentence + L471-472: here as well

The initial phrasing was unclear, so we have updated the sentences accordingly.

58. L497: I think you mean scenario spread rather than variability. See general comment above.

Yes, this has been updated with "scenario separation".

59. L 499: please be more temperate: "we argue that tis AMOC decline is therefore a lagged response to…

The sentence has been updated according to the main comment of reviewer #1.

60. L 501: do you think there is also a change in the frequency of internal variability?

This is an interesting comment, and a change in frequency of internal variability is indeed possible, as shown for example in Coquereau et al. (2025) for other climate quantities. However, we believe that addressing this question would require a dedicated study focusing specifically on internal variability — its various characteristics, frequency, and spatial patterns — which lies beyond the scope of the present work.

Coquereau, A., Sévellec, F., Huck, T., & Fedorov, A. V. (2025). Increase in ENSO frequency and intensity under 20th and 21st century warming: Insights from CMIP6 large ensembles. Geophysical Research Letters, 52, e2025GL116541. DOI:10.1029/2025GL116541

61. L525-527: isn't that point obvious, given that eache simulations are just one realization of the time variability ?

While this may be apparent to some readers, we consider it important to emphasize that inter-model variability includes a component of internal variability, which underscores the need to account for interactions.