Responses to the 2nd review for the paper entitled

North Atlantic Subtropical Mode Water properties:
Intrinsic and atmospherically-forced interannual variability

that we submitted to Ocean Science for potential publication.

Preliminary comment :
we have corrected a few typos in the text. As requested, we also have removed 2 out of the 3 footnotes present in the former manuscript: we have not found any way to re-introduce adequately into the main text the remaining footnote (section 2.1.2), since [i] it concerns a point that branches out of the main discussion, but [ii] its content might be useful for some readers.

Main comments/questions:
1. The response to this comment #1-3: “In contrast, it is not obvious to us why the ensemble with EA forcing “might overestimate the strength of intrinsic variability”
My point is that there is no damping (or very weak damping) in the “ensemble averaged” forcing, and then the strength of the intrinsic variability might be overestimated. As I mentioned in my previous comment “observational data show enhanced upward surface heat flux over warm meso-scale eddies (e.g., Tomita et al. 2019, doi:10.1007/s10872-018-0493-x)”. Although it is not in situ observation, this suggests a damping of intrinsic variability.
I agree with “its upper ocean enhancement obtained with the EA method goes in the right direction”, and just wonder about the possibility of overestimation.

Our ensemble averaged forcing strategy indeed exerts no damping on intrinsic variability, but this option is preferable to the member specific strategy which damps it much too strongly: here is a rough estimate of the damping timescale of warm eddies through air-sea fluxes in the real ocean (and presumably in coupled simulations resolving mesoscale). Tomita et al’s Fig 11 suggests that on average over their typical 3°x3° extent, warm eddies in the Kuroshio have temperature anomalies of about ~0.7K over which anomalous upward air-sea heat fluxes are about 50W/m2. Using Barnier et al (1995)’s equations 2, the damping of warm eddies (taking ~400m for their thickness) may have a timescale of about ~220/250 days; the damping timescale of cold eddies is more difficult to evaluate, but it is presumably much longer since their heating via anomalous air-sea fluxes is very likely to remain superficial and barely able to erode their subsurface cold core. In both cases anyhow, these estimates of damping timescales are much longer than the ~40-day timescale associated with member-specific fluxes, which is thus too short.

Note that we focus on spatiotemporal scales that clearly extend those of mesoscale eddies, and that the timescale of a possible damping of interannual intrinsic thermal anomalies is unknown in the real ocean. In practise, we have summarized these considerations by replacing the last sentence of section 2.1.2 by the following:

Intrinsic thermal anomalies are not damped with the ensemble averaged forcing approach; such anomalies in the real ocean may be slightly damped by air-sea interactions, but much less strongly than in the member-specific approach. We thus hypothesize that the amplitude of upper-ocean temperature interannual CIV in nature sits between those simulated with both forcing strategies, and argue that the ensemble-averaged forcing method lets it evolve in a more physically-consistent and realistic way.

2. Line 136-137: “the heat capacity … in nature”

I agree with this. It should also be noted that, given the wind speed, the air affected by the sea surface heat flux would move away quickly by advection.

Indeed, our ocean simulation is forced by the atmosphere and thus does not represent the feedback of the ocean on atmospheric fields, such as that mentioned by the reviewer. It seems to us that we clearly mention this specificity in section 2.1.2, which also presents our dedicated approach to improve our forcing method in this context.

3. Line 326-327: “the simulated partition … in ARMOR3D.”
Figure 6 becomes very easy to see, and it is now very clear that the frequency of overlaps are very different between the thermohaline properties and the others. From Fig. 6, I think it would be better to summarize the results slightly more carefully.

Here is our proposed rewriting of the last paragraph of section 3.1.3:

Figure \ref{fig:Taylor_mems_as_ref} exhibits an overlap between the distributions of member-ARMOR3D correlations (blue) and member-member correlations (grey) for most interannual STMW properties. Member-ARMOR3D and member-member correlations overlap over the range 0.5--0.75 for STMW volume for instance, and over much wider ranges for STMW thermohaline properties. In particular, member-member correlations do not largely fall below member-ARMOR3D correlations, suggesting that the ensemble is not clearly over-dispersive. The opposite is however found for STMW depth, for which the ensemble seems to be under-dispersive. Besides this main exception though, we conclude that it is unlikely that a signal-to-noise paradox contaminates the statistics of STMW properties in our simulation. In other words, the simulated partition between forced and intrinsic interannual variabilities of STMW properties are consistent with their counterparts in ARMOR3D.

Specific comments/questions:

Line 15: “STMW” should be “EDW.”

Yes indeed. Corrected.

Line 141: “to 1/12 degree simulation” The meaning of the comparison with the higher resolution simulation is not very clear to me. Is the underestimation in the lower resolution simulation not just due to the limited representation of eddies? It would be better to add a brief explanation.

Indeed, the fact that intrinsic variability is stronger at 1/12° than at 1/4° with the same forcing is very likely due to a better representation of eddy processes. The fact that switching to ensemble averaged fluxes enhances surface intrinsic variability at 1/4° resolution demonstrates the damping exerted by the member specific forcing. Both biases are thus at work at 1/4° with the member specitifc forcing, and the ensemble averaged forcing removes one of them. To keep this paragraph concise and clear, and to avoid any risk of confusion, we propose to drop the reference to 1/12° simulations in this particular sentence, and focus it on the forcing-related argument as follows:

Indeed, previous 1/4°-resolution NEMO simulations driven by classical (member-specific) forcing have been shown to underestimate surface intrinsic variability at all scales compared to observations \citep[see e.g.][]{penduff2010}. The use of ensemble averaged fluxes enhances surface intrinsic variability and compensates for this bias.