

Answers to reviewer #1

We thank the reviewer for his careful reading of the manuscript and for his appreciation of the work done in this study. Please find below a detailed answer to all comments.

- Yes, it is true that the ensemble was not produced specifically for this study. Important sources of uncertainties (e.g. in the atmospheric forcing and in the boundary conditions) were not taken into account; this is why we could not rely on the prior variance (basically we only use the local correlation structure), and why an adjustment of the ensemble statistics is even more important. In any case, it is difficult to produce model ensembles with an appropriate spread, which is why an adjustment procedure is useful. This difficulty was already emphasized in the last paragraph of the conclusions, but it is true that this should be done before that.

We have thus added the following sentences in section 2.2 (line 119) to warn the reader about this from the very beginning: "This ensemble was indeed not specifically produced for this paper, but was dedicated to the specific study of intrinsic variability by Héron et al. (2026). Important sources of uncertainties (e.g. in the atmospheric forcing and in the boundary conditions) were not taken into account; this is why we could not rely on the prior variance, and why an adjustment of the ensemble statistics is even more important."

- In principle, yes, any parameter could be included in the estimation vector. But, with our method, this is not true for the parameters controlling the covariance localization (l_3 , l_{loc} , τ_{loc}), because the ensemble of vectors with correlation given by Eq. (13), which depends on these parameters, and which are used to compute perturbations with the right localized covariance, must be computed once and for all before the MCMC iterations.
- Yes, we have made the notation uniform, i.e. "log" to mean the natural logarithm.
- Yes, the effective number of degrees of freedom (d) in our system is much smaller than N , and it indeed depends on the ratio between the size of the domain (in space and time) and the localization scales. Moreover, only a tiny fraction of the N possibilities for the perturbations are effectively used in the iterations. However, if we want to produce a quasi-random sample in a space of dimension d , it is always good if the number of directions of perturbation N (which is here discrete because of the finite sample) can be very large.

On the other hand, yes, it is also true that the size of the ensemble of vectors that are used for the localization of the prior ensemble [with correlation given by Eq. (13)] does not need to be the same as the size of the prior ensemble itself. It would be quite easy to use a larger sample (with just more memory), and this could make N even larger. We have used the same size here because it is technically more simple if the two ensembles have the same size, and to limit the memory requirements. It can also be noted that this "localizing ensemble" was originally intended to be a smoother version of the original ensemble and had thus the same size by construction (see Brankart, 2019).

This has been clarified by adding the following sentences in section 3.4 (line 273): "Actually, the effective number of degrees of freedom (d) depends on the ratio between the size of the domain (in space and time) and the localization scales, and it is very tiny as compared to N . The idea behind this large value of N is to produce a quasi-random sample of this target space of dimension d , and thus ideally among the largest possible number of directions of perturbation (which has here to be discrete because of the finite samples). See Brankart (2019) for more details. It must also be mentioned that the size of the localizing ensemble could also be larger than the size of the original ensemble, which would also increase N . We used here the same size for simplicity and to limit the memory requirements."

Answers to reviewer #2

We thank the reviewer for her careful reading of the manuscript and for her appreciation of the work done in this study. Please find below a detailed answer to all comments.

- The typo has been fixed.

- Yes, it is indeed much better to make the dependence in Δt explicit in the LHS. This has been corrected.
- Yes, indeed. This can be a source of problem and can affect the convergence of the Markov chains. In principle, the two factors can be estimated simultaneously if prior uncertainties and observation uncertainties have distinct correlation structures and if there are enough observations (as in adaptive Kalman filters). But the non-convergence of alpha in the results means that information is missing. Maybe adding an additional constraint on alpha would be necessary to control correctly the two parameters.

It is indeed useful to add a caution about this already in section 3.1. We have added the following text: "It must be noted here that it may be challenging to estimate these two factors simultaneously, since it is always the sum of the prior and observation uncertainties that is observed. The joint estimate is still possible in principle (as in adaptive Kalman filters) if they have distinct correlation structures, but it is always possible that information is missing to obtain accurate estimates."

- Yes, indeed, there is a constant missing in Eq. (11). It has been added.
- This paper is only focused on dynamic height. There are also biases in the model temperature and salinity. Part of them are related to dynamic height. Others are not. We made preliminary attempts (not included in the paper) to work on the multivariate problem, but biases are more difficult to identify in this case, because of the ever changing coverage of temperature and salinity observations.
- Yes, this is related to a question raised by Reviewer 1. This is because the ensemble was not produced specifically for this study, but for the paper by Héron et al. (2026), which was dedicated to the study of the intrinsic variability whose emergence is triggered by initial condition perturbations.

The following sentences has been added in section 2.2 (line 119) to clarify this point: "This ensemble was indeed not specifically produced for this paper, but was dedicated to the specific study of intrinsic variability by Héron et al. (2026). Important sources of uncertainties (e.g. in the atmospheric forcing and in the boundary conditions) were not taken into account; this is why we could not rely on the prior variance, and why an adjustment of the ensemble statistics is even more important."

- Regarding the top panels Fig. 5, yes, it would have been an option to keep the same scale for the Y-axis. But what matters here is more to compare the shape of the histograms, which would be less obvious is we adjust the scales. The total number of observations is more like a secondary information here, which is even often dropped by normalizing the histograms.

Regarding the bottom panels of Fig. 5, yes, indeed, the total amount of verifying variables must be the same in the two panels. The integral of the two histograms is the same. In this case, it is indeed much better if we use the same scale for the Y-axis. The two panels have been redrawn to correct this. The close similarity between the two histograms is now more clearly visible. Only the external maxima are notably different.