Reply to Referee #1 (Prof. Lili Lei)

Atsushi Okazaki, Diego S. Carrio, Quentin Dalaiden, Jarrah Harrison-Lofthouse, Shunji Kotsuki, Kei Yoshimura

We sincerely thank the reviewer for the positive assessment and the helpful comments and excellent suggestions. Below, we provide a point-to-point responses to all the reviewers' comments. The reviewers' comments are in *blue and italic*, and the replies are in black.

Lines 195-200, it is hard to follow from (18) to (19). How the innovation statistics link to the covariance inflation? Using (17), is the numerator the same as the denominator, which give delta = 1? Moreover, in the following discussions, the role of the inflation, especially the relation with the observation error variance, is not clearly discussed.

Thank you for the comment. The numerator in Eq. 19 represents the estimated size of trace($\mathbf{HBH}^{\mathrm{T}}$). In other words, it is the expected size of trace($\mathbf{HBH}^{\mathrm{T}}$). On the other hand, the denominator is the one represented by the ensemble. We can estimate how much the **B** (represented by the ensemble) should be inflated by calculating their ratio.

As for the second point, to accurately estimate the observation error, **B** should be estimated as well (Li et al., 2009). Depending on the magitude of the error variance **B**, the size of the estimated **R** differs. For instance, when **B** is underdispersive, the estimated **R** can be too small as well. To deal with this issue, we used the covariance inflation. We will update the sentences to make these points clearer in the revised manuscript.

Lines 218-220, do you mean 136 annual mean simulations are used as ensemble priors? Are the simulations or anomalies used?

Yes, 136 annual means (i.e., model states) are used as ensemble priors. While we use the raw simulated value for the state vector (**x**), the assimilation is performed using anomaly fields for the comparison between model states (**Hx**) and observation. We will modify the sentences for clarity in the revised manuscript.

Lines 246-247, this is unclear. Do you mean the climatological mean is computed as a smoothing averaging with adjacent years? If yes, how many years are used to compute the climatological mean?

Yes, exactly. We set a criteria of 30 year to calculate the climatological mean. If the overlapping period shorter than 30 years, the observation will not be assimilated. We will modify the sentences for clarity in the revised manuscript.

Lines 269-275, till now, it is unclear why 'BIAS' is designed?

Thank you for the comment. In general, the structure of the background error is considered to be different from the nature. We conducted the experiment 'BIAS' to investigate the impact of a misrepresented background error covariance.

Lines 355-360, with too large (small) R, small (large) inflation values are expected. It would be nice to show the estimated inflation given different R.

Thank you for the comment. The figure below shows the estimated inflation factors for (a) underestimated **R** and (b) overestimated **R**. As you expected, they are small with too large **R** and vice versa.

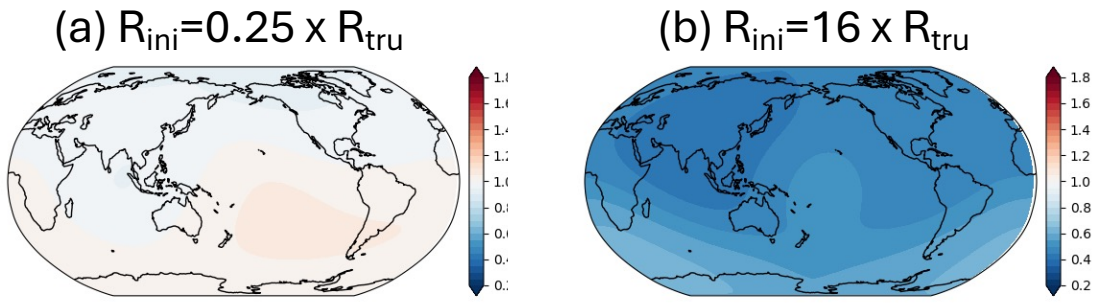

Figure 9, please give some potential explanations for the regions with negative skills.

We appreciate your valuable suggestion. If the estimation works well, the detrimental impact should be attributed to the bias in the prior covariance. We will add the discussion in the revised manuscript.

Lines 414, 'remarkably' -> 'remarkably worse'?

Thank you for finding the typo. We will correct it in the revised manuscript.