****Reviewer's comments in black and italics   Author's response in red**

We are very grateful to the reviewer for well understanding our method and for providing many constructive comments, including the comparison to Anderson's work, and applications to special grids and location-varying observations. These comments are very insightful and helpful in improving our manuscript.

In this revised manuscript, we not only improved the clarification of the method and settings, and also incorporated the reviewer's comments in the discussion. The point-by-point responses to the questions are attached below.

Reviewer 2

The authors propose a localization scheme for prior correlations and compare it with the traditional localization scheme based on distance dependence. This localization scheme is of interest for the implementation of ensemble data assimilation methods. The manuscript is quite well written and meets the submission requirements of the journal NPG. Nevertheless, the manuscript has the following issues that need further clarification and improvement.

Specific comments:

1. *This new localization scheme for YK18 relies heavily on the statistical formulation of equation (5), so is there any similarity between this formulation and Anderson's work, and what are their similarities and differences? Please elaborate explicitly.*

Reply:

Thank you for the valuable comment. We have added Anderson's work in the Introduction section in this revised manuscript.

Both Anderson's work (hereafter AL13) and our method (YK18) are proposed to obtain a static localization function from posterior ensembles. Both methods deliver a flow-dependent localization and show comparable accuracy to the traditional localization method with the L96 model. However, there are several differences between our method and AL13.

AL13 tends to find the localization weight that performs a minimum analysis ensemble-mean RMSE. This work is achieved by minimizing the cost function for a group of ensembles (subset ensembles). The minimization is carried out with an OSSE with "truth" values. Moreover, iterations for solving the equation would be required. In addition, AL13 does not restrict its solution from 0 to 1, so the localization weight obtained from AL13 could be a negative value or exceed one.

In contrast, YK18 finds a localization by the prior error correlations from the ensembles generated by an offline run or from past data. It does not need a truth value nor runs iteratively like AL13, while it needs an additional cutoff function to filter out small perturbations in the prior error correlations. Besides, the localization weight obtained from YK18 is restricted to an interval between 0 and 1.

Does the statistical result of Equation (5) depend on the number of samples? If so, how much does this sample dependence affect the final results?

Reply:

Once the solution of Eq (5) is converged to the climatology, the final result would be almost the same. The required number of samples depends on the configurations (i.e., ensemble size) and model complexity.

In our revised manuscript, we have added a new Figure 2 (b) showing the convergence time for different models and ensemble sizes, and incorporated related discussion in Section 4.1.

Equation (5) counts the correlation coefficients between the model grid points and the observed points, but we know that the observed variables are hardly fixed in their positions at different moments. This situation is especially prominent when assimilating satellite data in NWP. Since the position of the observed data is difficult to be fixed, the observation operator H is actually difficult to be fixed as well. Then how should the correlation coefficients between the model grid points and the observed points, which are calculated by Eq. (5), be applied to other moments?

Reply:

Thank you for this great question! We have incorporated the related discussion in Section 5 as a future work.

For the observations that their position varies with time (i.e., satellite data), a possible solution is the application of machine learning. Yoshida (2019) showed that neural networks could be used to estimate the background error correlations and deliver the correlation function for YK18. This work can be further applied to varying observations for estimating their prior error correlations and will be our future works.

Similarly, the model in the validation experiment given so far is very simple, with only one variable. For a true NWP model, there are perhaps multiple model state variables such as U, V, P, T, Q, on the same model grid point. And due to the use of different grid schemes, these variables may not appear at the same location of the grid. So how to use Eq. (5) for statistics in this case and apply it to the real situation?

Reply:

Thank you for this valuable and important question. For a multivariate system, Eq. (5) can be directly calculated for different pairs of model variables. For example, Eq (5) can be represented as the correlation between U (x^u) and V (x^v) as:

$$corr_{i_j}(t) = \frac{\sum_{k=1}^{K}\left[x_{k\,i}^{u}(t) - \overline{x_i^u(t)}\right]\left[h_j\left(x_k^v(t)\right) - \overline{h_j\left(x_k^v(t)\right)}\right]}{\sqrt{\sum_{k=1}^{K}\left[x_{k\,i}^{u}(t) - \overline{x_i^u(t)}\right]^2}\sqrt{\sum_{k=1}^{K}\left[h_j\left(x_k^v(t)\right) - \overline{h_j\left(x_k^v(t)\right)}\right]^2}}$$

For special model grids such as the staggered grid in WRF or cell grids in MPAS, there are several strategies (i.e., convert to Cartesian coordinate, direct interpolation…etc.) proposed to deal with the grid structure problem in DA (Ha et al., 2017; Pattanayak and Mohanty,2018). Considering the fundamental statistics of YK18 (Yoshida and Kalnay, 2018), we would suggest calculating Eq (5) with the analysis grid and observation operator are consistent with the term of *HXXT* in the corresponding DA system.

How to properly apply localization, for either GDL or YK18, on special model grids is an important and challenging topic that needs more investigation. This advanced topic is beyond the scope of the current paper that we will include it in our future works.

The authors elaborate that one of the advantages of YK18 is that it is more computationally efficient. However, it can be seen from their analysis that in fact YK18 should essentially provide some new calculations of localization correlation matrices as well, so why does it make the improvement of computational efficiency?

Reply:

Thank you for this comment. Yes, both methods need to calculate the localization matrix; however, the pre-processing work for YK18 is more efficient than GDL in two ways:

1) Less trials-and-errors tunning for localization:

As we already mentioned in Section 4.3:

"Traditionally, the use of GDL requires multiple trial-and-errors to define the optimal localization length for the experiments of interest. In contrast, YK18 only needs one offline run to obtain the prior error correlations, whereas it provides a comparable analysis as GDL even with a faster spin-up."

2) Fewer computational loops between observations and model grid at each DA cycle

From the programming aspect, the most expensive part of GDL is the do-loop structure and data sorting. Looping over every model point and observation is essential for calculating the distance before each DA, and data sorting is commonly used for efficiently searching the neighbored observations for the analysis grid. When applying a large model and observations in NWP, the computational costs of looping and data sorting would be significant.

In contrast, the most expensive part for YK18 is developing the prior error correlation in the beginning. However, once the prior error correlation map is constructed, one can directly obtain the correlation value from the indices of i and j in the map (Eq (6)), then convert it to the corresponding localization weight. Therefore, the computationally expensive data sorting during DA can be avoided, making YK18 a more efficient method in the long term.

In the revised manuscript, we have improved the writing of this part in Section 4.3. We are grateful for this comment that helps us to improve the manuscript.

As for the "a faster spin-up" proposed in the manuscript, I do not quite understand it. The purpose of our data assimilation is to give a more accurate initial field and then drive the model to forecast. The spin-up seems to be more appropriate in the simulation of climate models.

Reply:

Thank you for pointing this out. The word “spin-up” in the manuscript is the “DA spin-up”, not the “model spin-up”. We have added sentences to better clarify the word “spin-up” in our revised manuscript [Line 234-236].

It seems that Section 2.3 of GDL appearing in Page7 should be Section 2.2.

Reply:

Thank you for pointing out this mistake. We have corrected it in this revised manuscript.

Reference

Ha, S., Snyder, C., Skamarock, W. C., Anderson, J., & Collins, N. (2017). Ensemble Kalman filter data assimilation for the Model for Prediction Across Scales (MPAS). Monthly Weather Review, 145(11), 4673-4692.

Pattanayak, S., & Mohanty, U. C. (2018). Development of extended WRF variational data assimilation system (WRFDA) for WRF non-hydrostatic mesoscale model. Journal of Earth System Science, 127(4), 1-24.