

Nonlin. Processes Geophys. Discuss., referee comment RC1
<https://doi.org/10.5194/npg-2022-5-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on npg-2022-5

Zheqi Shen (Referee)

Referee comment on "Applying prior correlations for ensemble-based spatial localization" by Chu-Chun Chang and Eugenia Kalnay, Nonlin. Processes Geophys. Discuss., <https://doi.org/10.5194/npg-2022-5-RC1>, 2022

This manuscript applies a newly developed spatial localization method (YK18) for the ensemble-based data assimilation using Lorenz (1996) model. This correlation cutoff method is developed in their previous work as a variable localization strategy for coupled systems. They declared that it can be further utilized as a spatial localization method. They performed twin experiments with Lorenz 1996 model, and compared the YK18 method with the conventional spatial localization method. Overall, the work is useful and the manuscript is well written. However, some details of the method should be better explained. And I still have questions about the foundation of the method. The authors should answer my questions and make some revisions before it could be accepted for publication. Please see my questions and comments below.

1. In section 2.3, the authors imply that "The prior square error correlations are collected from a preceding offline run." According to section 3.2 (line 185), I realize the offline run is a DA experiment using a relatively large ensemble (50 in this case) without localization for a long period. Of course it works with the toy models such as the Lorenz model in this work and that in YK18. However, for more complicated models (such as GCMs), it is meaningless to perform DA experiment without localization, and it is impossible to use an ensemble large enough to get rid of localization. That would weaken the argument about the usefulness of the method.

2. Equation (6) implies that the temporal mean of the squared correlation over all analysis steps is computed to serve as "prior error correlation" to estimate the localization function. I have some questions about that:

- 2.1 Why do you compute the temporal mean over all analysis steps, instead of all steps including forecast and analysis? Considering the analyses are from the data assimilation without localization, does that indicate the ensemble size is large enough such that the true correlation can be recovered without localization? This is still impossible for large models.
- 2.2 You use the temporal mean of the squared correlation. So does the period of the offline run have an impact on the correlations?

- 2.3 Is the assimilation process necessary in the offline run? I wonder, is it possible to compute the correlation using an EnOI-like idea? i.e., running a single model and computing the correlation using members at different time steps. This seems much more practical for real applications. You have already use the temporal mean in the current method anyway.

3. About Figure 2a. I cannot see any connection between error correlation and observation size from equations (5) and (6). But the connection between error correlation and ensemble size is very clear from equation (5). Do you use this figure to explain the parameters in the offline run?

4. The comparison results in figure 5 and figure 6 are not impressive. Though the authors declaim that YK18 can accelerate the spin-up. However, the parameters in GDL and YK18 may be not optimal, so the conclusions are not very persuasive.

Detailed comments

- Line 175 "GDL: Distance-dependent localization introduced in Section 2.3." I think it is section 2.2.
- For equations (1) - (3), there are linear observation operator H , for equation (5), it is a potentially nonlinear operator $h(x)$
- Line 99 "Equation (4) is a smooth and static Gaussian-like function that offers the same localization effect as the GC99 when applied to LETKF." It is inaccurate, because GC99 uses a compact-support function, and it cutoff at some distance, but Eq. (4) does not cutoff.
- Line 91 and line 102, whether the R localization multiplies the elements of R inverse or R itself? Please clarify that.