Reviewer #1

Review of "Reduced non-Gaussianity by 30-second rapid update in convective-scale numerical weather prediction " by J. Ruiz et al..

General comments: The manuscript investigated the degree of non-Gaussianity of forecast error distributions and how it is affected by the DA update frequency and observation number. This article used 1000 ensembles, but the generation of ensembles is not very clear. The introduction of the KLD method is not clear. Some words are not rigorous. This paper theoretically provides evidence that increasing the update frequency and the observations can improve the accuracy of assimilation in convective-scale. I think this manuscript can be considered for publication if these concerns could be addressed:

We would like to thank the reviewer for the comments which help to improve the clarity of the presentation and also bring additional discussions to the paper. Based on the reviewer's comments we expand the description of the methodology and add additional explanations and discussions.

Specific comments:

1. The generation of ensembles need to be described in detail, because it is very important for this article and the ensemble DA method.

We agree that this is an important aspect of the experimental design so following the reviewer's comment we expand the description on how the initial perturbations for the first assimilation cycle as well as the boundary conditions are generated in our experiments.

The initial ensemble at the first assimilation cycle and the ensemble boundary condition ensemble are created by adding random perturbations which preserve the hydrostatic and nearly geostrophic equilibrium (Necker et al. 2020a, Maldonado et al. 2021). These perturbations are generated from a sample of continuous 6-hourly analysis states provided by the Climate Forecast System Reanalysis (CFSR, Saha et al. 20202), [$X_{CFSR}(t_1)$, $X_{CFSR}(t_1)$,...,$X_{CFSR}(t_N)$], where N=5840 (4 years). The horizontal grid spacing of the CFSR data is 0.5º. At the beginning of the assimilation cycle ($t=t_s$), the initial condition perturbation of the *i-th* member $X'^{(i)}$ is computed as:

$$X'^{(i)}(t_s) = \alpha\left[X_{CFSR}(t^i_{n_1}) - X_{CFSR}(t^i_{n_2})\right]$$

where α is a multiplicative factor equal to 0.1 so that the amplitude of the perturbations is roughly equivalent to 10% of the climatological variability. The two CFSR analysis states are chosen by randomly selecting two numbers $n_1^{(i)}$ and $n_2^{(i)}$ from the N elements satisfying the condition that $t_{n1(i)}$ and $t_{n2(i)}$ correspond to the same time of the year and time of the day. In the following assimilation cycles at time $t>t_s$, we obtain the boundary perturbations as:

$$X'^{i}(t) = \alpha\Big[(1-\beta)\Big(X(t_{l_1^{(i)}}) - X(t_{l_2^{(i)}})\Big) + \beta\Big(X(t_{u_1^{(i)}}) - X(t_{u_1^{(i)}})\Big)\Big]$$

where $l^{(i)}_{1,2} = n_{1,2}^{(i)} + m$ and $u^{(i)}_{1,2} = n_{1,2}^{(i)} + m + 1$, with m = floor[(t - t_s)/6 h] and β = [(t-t_s)/6 h] - m being a temporal linear interpolation factor to compute perturbations at arbitrary times (not necessarily a multiple of 6 h). In this way we obtain perturbations that are smoothly varying in time and consistently with the large scale dynamics of the atmosphere. This procedure is applied to all atmospheric and soil state variables.

2. The introduction of KLD should contain how to operate in this article? What should be noted? So that readers can repeat your experiment

We agree with the reviewer and we expanded the description on how KLD is computed in this work. In particular, we added more details on how KLD is computed from the ensemble-based sample, and an equation expressing how KLD is computed in this paper.

To measure the degree of non-Gaussianity of the error distributions we compute the Kullback-Leibler divergence (KLD, Kullback 1951) which is defined as follows:

$$KLD(P\|Q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx,$$

where p(x) and q(x) are the probability density functions (PDFs) of P and Q respectively. The KLD is 0 if P and Q are the same and takes positive values if P and Q differ. In our case p(x) is either the first guess or analysis error distribution for the state variable x (e.g. temperature, wind components, etc), and q(x) is a Gaussian distribution whose mean and standard deviation are equal to the ones of p(x). Therefore, a low KLD value corresponds to the first guess or analysis error distribution close to a Gaussian.
In the EnKF we do not have access to the continuous PDF p(x) but to a finite sample of it. For each state variable x and at each model grid point, we approximate p(x) with the sample histogram populated from the 1000-member ensemble using 32 equally-sized bins covering the range where p(x) is greater than 0. This range is defined by the minimum and maximum values of x at each model grid point and time. Then we can approximate the KLD as follows:

$$KLD(P\|Q) \approx \sum_{j=1}^{j=32} p_j \ln \frac{p_j}{q_j},$$

where p_j is the relative frequency of x at the *j-th* histogram bin. q_i is the integral over the *j-th* histogram bin of a Gaussian PDF whose mean and standard deviation are given by the ensemble-based sample estimates. After implementing this we end up with an estimation of the KLD of the analysis and first guess error distributions with respect to the Gaussian for each grid point location, vertical level, and time.

3. Is the solution of KLD grid point by grid ? Using the assimilated ensembles to statistics?

Following this comment and the previous one we extend the description on how KLD is computed so that the experiments can be reproduced. The specifications on how KLD is computed from the ensemble members is included (see also the answer corresponding to the previous comment).

4. Please give the formula of relative KLD difference.

We include an equation for the relative KLD difference.

The impact of DA frequency on non-Gaussianity is investigated by means of the relative KLD difference between the 5MIN and all the other experiments, computed as:

$$\overline{KLD_{diff}} = \frac{\overline{KLD_E} - \overline{KLD_{5MIN}}}{\overline{KLD_{5MIN}}}$$

where $\overline{KLD_{diff}}$ is the relative difference between the averaged KLD in the 5MIN experiment ($\overline{KLD_{5MIN}}$) and on each of the other experiments ($\overline{KLD_E}$), where E can be either 5MIN-4D, 2MIN, 1MIN, 1MIN-4D or 30SEC).

5. Please explain "closest" quantitatively.

Following the reviewer's comment we expand the discussion to better explain possible time differences between the observations and the analysis time.
Here, only a single volume scan closest to the analysis time is used per analysis. Namely, more frequent updates assimilate more data. In all cases the time difference between observation time (center time of the radar volume scan) and the analysis time do not differ by more than 15 seconds.

6. The higher the update frequency, will it break the balance between physical variables? How to understand and explain.

We agree with the reviewer that this is an important aspect that has been overlooked in the previous version of the manuscript. In the revised version we include a discussion about the effect of imbalance.

As has been shown in the previous studies, more frequent assimilation can produce a larger degree of imbalance in the initial conditions which can degrade the quality of the forecasts (e.g. Lange and Craig 2014, Bick et al. 2016). Therefore, despite the potential benefits of a more Gaussian model error distribution on the analysis accuracy, other factors may degrade the forecasts initialized from more frequent data assimilation cycles. Imbalance and non-Gaussianity can also possibly be related. Gaussian error distributions can lead to more physically meaningful assimilation updates in the context of an EnKF and thus more balanced initial conditions. However, a larger imbalance in the initial conditions can contribute to faster error growth and increased departure from the Gaussian in the forecast distribution. Possible interactions of these mechanisms in a data assimilation cycle have not been investigated, and are a subject for future research. Our results suggest that despite the effect of a larger imbalance, the increase of DA frequency reduces non-Gaussianity in the sample distributions obtained with the EnKF. This is even true for variables like the vertical velocity within convective clouds which are frequently used to measure the effect of imbalance in the initial conditions.

Added references:

Lange, H. and Craig, G. C.: The impact of data assimilation length scales on analysis and prediction of convective storms, Mon. Wea. Rev.,
142, 3781–3808, 2014.

Bick, T., Simmer, C., Trömel, S., Wapler, K., Stephan, K., Blahak, U., Zeng, Y., and Potthast, R.: Assimilation of 3D-Radar Reflectivities with an Ensemble Kalman Filter on the Convective Scale, Quart. J. Roy. Meteor. Soc., 142, 1490–1504, 2016.

7. This paper used the super-observations. What is the relationship between observation scale and grid scale matching and more observations?

We agree with the reviewer in that this point deserves further clarification. We expand the discussion in the paper. Reflectivity and Doppler velocity observations are superobbed to horizontal resolution of 1 km and vertical resolution of 500 m to approximately match the model resolution. This is done to reduce the amplitude of spatial scales, in the observed data, which are not well resolved by the model as well as that of small scale noise. This procedure can also reduce the impact of possible spatial correlations in the observation errors.

8. This article needs polishing.

Based on the reviewer's comment and on the comment of the other two reviewers we perform an extended revision and polishing of the text and improvement of the figures. We hope that these changes improved the overall presentation quality of the manuscript.