egusphere-2025-2543

Impact of reduced non-Gaussianity on analysis and forecast accuracy by assimilating every-30-second radar observation with ensemble Kalman filter: idealized experiments of deep convection

Arata Amemiya and Takemasa Miyoshi

General comments:

This study conducted observation system simulation experiments (OSSEs) idealized for high-frequency radar reflectivity data assimilation (DA) with local ensemble transform Kalman filter (LETKF) and compared 3 experiments: (i) 3D-LETKF every 5 min, (ii) 4D-LETKF every 5 min with observations every 30 sec, and (iii) 3D-LETKF every 30 sec. As a result, (iii) had the smallest non-Gaussianity of first guess ensemble and the best analysis accuracy. Since it is interesting to clarify the advantages of high-frequency DA with the idealized OSSEs, this study is valuable to be published. However, the causal relationship between the non-Gaussianity of first guess ensemble and the analysis accuracy is not clear even in the OSSEs conducted in this study. Therefore, it should be clarified that this study does not investigate pure impact of non-Gaussian distribution but imply the advantages of high-frequency DA partially in the viewpoint of non-Gaussianity. To prevent misunderstanding about it, I recommend major revision. The order of the following comments is not related to importance.

Specific comments:

1. L4-5, L50-51, L75-77, and L286-287: The OSSEs conducted in this study do not completely exclude the impact other than non-Gaussianity because the DA method, the number of assimilated observations, the ensemble spread, etc. are different between the 3 experiments compared in this study. This study does not clarify the pure impact of non-Gaussian distribution but imply the advantages of high-frequency DA partially in the viewpoint of non-Gaussianity. The sentences should be revised not to be misunderstood.
2. L20-23: This description is not necessarily correct in 4D-EnKF. Since the experiments with 4D-EnKF are conducted in this study as well as 3D-EnKF, the development of 4D-EnKF and its advantage also should be explained.
3. L53-54: Why is this study useful for the non-Gaussian DA? Could you cite any previous studies?
4. L73-77 and L288-290: This experimental design does not completely exclude modification to the ensemble perturbations because spatial localization is applied to reduce the effect of sampling error. This limitation should be stated.
5. Figure 1: To add axis of height (km) is helpful to understand the height of convections shown in other figures.
6. L90-91: The forcing by the warm bubble should be stated more concretely and quantitatively here.
7. L137-138: Does it mean that 5-dBZ reflectivity is assimilated even where the first guess < 5 dBZ? In this case, the precipitation becomes stronger in the analysis. Is it no problem?
8. L143-146: This description is redundant and difficult to be understood. Does it mean that reflectivity is assimilated only where at least one ensemble member > 10 dBZ in the first guess?
9. L160-161: Why was the potential temperature perturbation over the entire domain in addition to the perturbation in the warm bubble?

10. L161-162: Were the Gaussian perturbations added in the warm bubble and the whole domain at the first assimilation cycle? If so, why is the first guess ensemble expected to be non-Gaussian? How to add the perturbations should be explained more clearly.
11. L183: What determines this kernel bandwidth? Could you cite any previous studies?
12. L191-192: The ensemble spread in the 5MIN-4D case should also be stated. In addition, it is better to show the time series of the ensemble spread until 00:50:00 in all cases to confirm that the filter divergence has not occurred.
13. Figure 5: To make the discussion in Section 3.2 deeper, the ensemble spread should also be shown in addition to ensemble mean and the difference from the nature run. I think main difference between the 3 experiments is the ensemble spread.
14. L213-214, L276-277, L279-280, L283-284, and L309-314: If the ensemble spread or the KLD is largely different between the 3 experiments, these descriptions are not precise for ensemble forecasts. The difference should be shown also for the ensemble forecasts.
15. Figure 6: The time of the 100 samples should be shown as well as the position of the grid point.
16. L221-223 and L304-306: The nonlinear cross-variable relationship between graupel mixing ration and vertical velocity looks caused by difference of non-Gaussianity of these variables (vertical velocity is closer to Gaussian). If this interpretation is correct, it should be stated.
17. L227-228: Could you show the mathematical definition of “the mutual information between the ensemble members of graupel and vertical velocity at every grid point, after removing the linear dependency”?
18. Figure 7: The caption should include “in the 5MIN-3D case”.
19. L230-231 and L306: The nonlinear relationship between the two variables may be one cause of the error of the analysis. However, it may simply be caused by smaller impact of assimilation in the part of smaller ensemble spread. Or, the smaller number of observations assimilated with low-frequency may also make the large error of the analysis. It is better to show various possible causes.
20. L237: “pertrbed” -> “perturbed”
21. L253-255, L266-269, L323-324, and L325-326: These disadvantage of 5MIN-4D and 30SEC are general and should be found also in the experiments in Section 3. Why are they found only in the additional experiments in Section 4?