To Reviewer 2

Zeng et al present an interesting paper about observation error statistics for radar reflectivity and Doppler radar wind measurements. Their study includes an estimation of the covariances arising from the error due to unresolved scales based on model data, and estimation of the full observation observation error covariances using the Desroziers et al (2005) assimilation diagnostics method.

Answer: Thank you very much for your kind acknowledgment of our work.

While the results for the Desroziers et al (2005) diagnosis of the Doppler radar winds are a little incremental (these have been published for a previous version of the DWD KENDA assimilation system by Waller et al, 2019), their comparison with the model-derived representation error statistics provided some fresh ideas. Furthermore, diagnosed estimates of the radar reflectivity error covariances have not been published in the mainstream literature before. I found the paper to be lacking a little background information, which might provide a deeper understanding of the results presented. I also had some minor questions about the experimental methods and results. I believe that these can be addressed very straightforwardly by the authors. My specific comments follow:

Answer: We added now following text in the introduction: In the present work, we use the Desroziers method to explore characteristics of the OE for radial wind and reflectivity in the operational ICON-KENDA system of the DWD. It is the first application of radar data assimilation using this framework (a similar study has been done by Waller et al. 2019 but for the COSMO-KENDA system and only for the radial wind). To authors' knowledge, it is also the first in-depth attempt to investigate the OE statistics (variances and correlations) of reflectivity data. However, the estimated OE statistics embraces contributions from the IE, FE and RE and it is not clear how much an individual error contributes. To approximate the RE, we assume that a high resolution model is the truth and we regard model equivalence of radar data calculated from the truth as observations (e.g., Waller et al 2014,Waller et al 2021) and evaluate the statistics from a set of samples of differences between observations and model equivalence of the low resolution model run, which can then be compared with the OE statistics estimated by the Desroziers method.

1. There was very little review provided of the expected sources of uncertainty for the observations. I believe that giving this background could provide more insight in the results. For instance:

(a) There is previous literature noting the dependence of the reflectivity error variability on the reflectivity value e.g.,

Doviak, R. J., and D. S. Zrnic, 1993: Doppler Radar and Weather Observations. 2nd ed. Academic Press, 562 pp.

Xue, M., Jung, Y., and Zhang, G. (2007). Error modeling of simulated reflectivity observations for ensemble Kalman filter assimilation of convective storms. Geophysical research letters, 34(10).

Answer: We added the references. The instrumental error of radar reflectivity observations is proportional to the measured values (Doviak and Zrnic, 1993 and Xue et al. 2007).

(b) Waller et al (2019) pointed out the contributions to the Doppler radar wind observation errors from the DWD superobbing scheme.

Answer: We mentioned "As noted by Waller et al. 2019, this superobbing technique may create error correlations since the same raw observations may be accounted for in neighboring superobservations". However, we do not fully agree with their interpretation of effects of superobbing on error correlations . The correlations arise since the wedges of the neighboring superobbing points overlap. In case of large size of overlap, strong correlations arise. To our understanding, there are several factors influencing the size of overlap for the the DWD superobbing scheme: First, the range or height to radar stations: on one hand, the closer the superobbing point is, the broader is the width of azimuth for supperobbing (see Fig. A7), which results in a larger overlap for neighboring superobbing points. On the another hand, two neighbouring superobbing points are further away when they are closer to radar stations, which results in a smaller overlap. Therefore, the size of overlap depends on these two competing factors. In addition, the size of the overlap also depends on the elevation. If the neighboring superobbing points have the same distance to the radar station, they are further away from each other if they are from higher elevations than from lower elevations, therefore, smaller overlap for higher elevations. The dependency of the size of overlap on the distance is not straightforward and for the moment we can not draw conclusion. However, we did another experiment with superobbing resolution of 10 km, which results in longer correlation length scales. This is probably due to larger overlapp of superobbing wedges. We mentioned this in the text.

(c) How might reflectivity attenuation (in a heavy storm) affect the results?

Answer: Reflectivity attenuation is an important error source and can have considerable impacts on the statistics of the OE, but this requires rigourous studies that can be done in the future.

2. A little more information about the form of the operator T is needed. The reader should not need to access Zeng et al (2019) in order to understand what this operator does.

Answer: T is the interpolation operator and the one used in this work is the iconremap utility from the DWD ICON Tools (Prill 2014). We added "using the the iconremap utility from the DWD ICON Tools (Prill 2014)," to Line 145.

Prill, F., 2014:DWD ICON Tools Documentation. Deutscher Wetterdienst (DWD),dwdicontools/doc/icontoolsdoc.pdf.

3. Localization: Waller et al (2017) pointed out an issue using the Desroziers et al (2005) method together with localization, and provided some criteria to es-

tablish which observation pairs can be used in the calculation. Was this method followed in this paper? What is the localization radius used in the experiments?

J.A. Waller, S. L. Dance, and N. K. Nichols, 2017: On diagnosing observation-error statistics with local ensemble data assimilation. Quart. J. Roy. Meteor. Soc., 143, 2677-2686, https://doi.org/10.1002/qj.3117.

Answer: The R-localization is applied in the LETKF, which is a type of domain localization as used in Waller et al. 2017. It is the same method used in Waller et al. 2019. The information on the localization radius is given in Line 236-237.

4. There is a further recent publication (Waller et al, 2021) providing model-based estimates of errors due to unresolved scales that are more appropriate for convection-permitting lengthscales than the earlier 2014 paper that is cited. The Waller et al (2021) statistics for zonal and meridional wind standard deviations decrease with height, as is also largely reflected by the black lines in the relevant panels in Figure 5 in this paper (above the first few km). However, the opposite holds for the radial winds in Figure 7 (i.e. the error standard deviations for the radial winds increase with height). I did not understand the explanation for this difference in the paper.

Waller, J.A., Dance, S.L. and Lean, H.W. (2021), Evaluating errors due to unresolved scales in convection permitting numerical weather prediction. Q J R Meteorol Soc. Accepted. doi:10.1002/qj.4043

Answer: There are several possible explanation for this difference. First, Waller et al, 2021 discussed the RE of horizontal wind and this article focuses on the radial wind. These are two different types of measurements. The latter one also contains information about the vertical wind. Therefore, the results of Waller et al, 2021 do not necessarily hold here. Second, intuitively it is not surprising that the standard deviations of errors increase with the height since the wind speed increase with the height. Third, the weather conditions can be also a factor. The study period in this article is with many thunderstorms and deep convective clouds. Fourth, the RE of lower elevation increase faster with the height in our study, indicating that the RE of horizontal wind is dominating with the height. This can be due to the fact that anvil regions are approached where divergent convective outflows occur and winds move in different directions and slight spacial shifts of cells in simulations can lead to large errors, therefore, the standard deviations increase.

5. L18 and L40 are a little out of date. The current operational system at the UK Met Office uses 4D-Var. Some more up to date references:

Milan, M, Macpherson, B, Tubbs, R, et al. Hourly 4D-Var in the Met Office UKV operational forecast model. Q J R Meteorol Soc. 2020; 146: 1281-1301. https://doi.org/10.1002/qj.3737

Hawkness-Smith, LD, Simonin, D. Radar reflectivity assimilation using hourly cycling 4D-Var in the Met Office Unified Model. Q J R Meteorol Soc. 2021;

1516-1538. https://doi.org/10.1002/qj.3977

Answer: Thank you. We added these references.

6. L44 The JMA have also used the Desroziers et al (2005) method with radar data and this should be noted. See for example,

Fujita, T., Seko, H., Kawabata, T., Ikuta, Y., Sawada, K., Hotta, D. and Kunii, M. (2020) Variational Data Assimilation with Spatial and Temporal Observation Error Correlations of Doppler Radar Radial Winds. Research activities in Earth system modelling. Working Group on Numerical Experimentation. Report No. 50.WCRP Report No.12/2020. WMO, Geneva.

Answer: We added the reference.

7. A short study estimating reflectivity variances using Desroziers et al (2005) was carried out by a Masters student using UK Met Office trial data. Some comparison could be made with these results.

Kouroupaki, V. (2019). Investigating radar reflectivity uncertainty in data assimilation for high impact weather prediction. MSc Thesis. University of Reading, UK.

Answer: We added the reference.

8. Fig 2. Is not referred to in the text until p14. It would be better to refer to this figure earlier, in section 3.

Answer: We switched the order of Figures 2 and 3.

9. Fig 3 caption - what is meant by "scratch"?

Answer: We changed "Scratch" to "Illustration".

10. Line 113 what do you mean by "statistically insignificant" here? How many samples are needed for reliable estimation?

Answer: We rephrased "As Waller et al. 2019, if the numbers of samples available for estimation are too small (e.g., < 1000), the estimated standard deviation and correlations might be considerably contaminated by the sampling error and therefore are not reliable".

11. Section 4.1 Do these RE estimation experiments include superobbing?

Answer: No superobbing has been applied in the RE estimation. We added this in the text. This is worth testing in the future

12. Section 4.1 Does the representation error exhibit a bias? At line 132 "systematic error" is mentioned, but the plots only show standard deviations,

so cannot give an indication of biases.

Answer: We calculated the means of the representation error for both reflectivity and radial wind, which are very close to zero, i.e., no bias. Therefore, we removed the term of "systematic error".

13. Fig 4a Why is there a sharp gradient at very low levels?

Answer: The sharp gradient at very low levels may correspond to increasing reflectivity values at these levels as shown in Fig 4c. Furthermore, it can be due to representation error of orography (Waller et al. 2021) or due to turbulence at the boundary layer, but it is not certain.

14. Reflectivity correlations: For the standard deviations a clear dependence on reflectivity value was shown. Might this also apply to the correlation lengths? Would it be appropriate to produce correlation plots separated by reflectivity value rather than beam elevation?

Answer: It is an interesting point but it is technically complicated to implement this. We may investigate this in the future.

15. Section 4.2 Please could you clarify what happens to the "dry" observations (zero/small reflectivity)? Are they assimilated? In the text there is some mention of "no reflectivity data" (line 193, 222) but it wasn't clear to me what this referred to.

Answer: Reflectivities with values smaller than 0 dBZ are set to 0 dBZ and treated as no reflectivity data and are assimilated. We discussed this in Line 94-103.

16. Section 4.2: Are the O-As and O-Bs used for calculating the Desroziers et al (2005) diagnostic unbiased? If not, is the bias subtracted before computing the covariances?

Answer: We now subtract the means from O-As and O-Bs.

17. Fig 11 (and later figures). The right hand panel (no of samples) displays a zig-zag pattern (most obvious for purple and blue lines). Why is this?

Answer: The reason for this is not clear (probably due to the superobbing methods) but it does not jeopardize the interpretation of results.

18. Line 255 you explain a difference in size of standard deviation compared with previous work due to a scaling factor in R. Can you explain this more clearly? Is this to do with the deficiencies of the diagnosis technique?

Answer: First of all, the scaling factor is introduced because the IE of radial wind measurements are usually large if onsite reflectivities are too small (Line 231-234), not due to the deficiencies of the Desroziers method. The scaling factor inflates R of radial wind where small reflectivities (i.e., between 0 and 10 dBZ) are observed, which makes the estimated standard deviation larger. However, it is recognized in practice that the Desroziers method tends to produce too small variances (Weston et al., 2014; Bormann et al., 2016), the introduction of the scaling factor also compensates the deficiency of the method (see Line 371-372).

19. The earlier paper on Doppler wind error estimation with the KENDA system (Waller et al, 2019) emphasizes the role of the superobbing procedure in generating error correlations. Is the superobbing procedure used here the same? Do some of the error correlations arising here stem from the overlapping superobbing wedges?

Answer: The same superobbing method is used here. We did another experiment with superobbing resolution of 10 km, which results in longer correlation length scales. This is probably due to larger overlapp of superobbing wedges. We mentioned this in the text.

20. Figure 9 is not referred to in the main text. It does not seem useful to include if it is not referred to.

Answer: We moved the Figure to Appendix.

21. Figure 10 could be cut as it does not tell us very much.

Answer: Fig. 10 is to demonstrate that superobbed observations are evenly distributed in horizontal. We put the figure in Appendix.

22. I feel that the paper would benefit from being a bit more selective about which figure panels to show to make the relevant points e.g., is it necessary to show correlations for every elevation? Or could the key points be made from one or two elevations, and the rest of the figures put in supplementary material.

Answer: We have selected five elevations from ten after rigorous consideration, with which dependency of important features of statistics can be well seen. Since we see the elevation 8° behaviors similarly as 5.5° and the numbers of samples available are not always sufficient, we removed the elevation 8°.

Typos and Small corrections

Line 46 "authors's"

Line 70 and throughout "setup" is rather informal.

Line 80 Parentheses needed around Waller et al (2016b)

Line 101 EMVORDAO

Line 137 grauple

Answer: Done.