Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC1
https://doi.org/10.5194/amt-2021-95-RC1, 2021

Comment on amt-2021-95

Anonymous Referee #1

Referee comment on "Interpreting estimated observation error statistics of weather radar measurements using the ICON-LAM-KENDA system" by Yuefei Zeng et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-95-RC1, 2021

Review of “Interpreting estimated Observation Error Statistics of Weather Radar Measurements using the ICON-LAM-KENDA System” by Zeng et al.

This work tries to estimate components of the observation error for radar reflectivity and radial wind measurements using two complementary techniques: the difference between high and low resolution forecasts as an indicator of representation error, and the well-known Desroziers approach that is intended to estimate the total observation error. There are substantial differences between the two estimates, which the paper attempts to explain. This is useful work on the tricky subject of assigning observation errors to precipitation-sensitive observations. However, there are major issues with the work as currently presented. The first is an important scientific issue: the results are given for a precipitation-affected sample based on a 5 dBZ reflectivity threshold; however little attention has been paid to how that threshold has been defined and how the definition might affect the sample of observations being examined. Given that the threshold definition must be different in the two different techniques, this might be a major cause of variation between the two. Second, there is a lack of high-level figures to help synthesise the results; some of the figures that are presented are given at a perhaps excessive level of detail. Hence, major revisions are recommended.

Major issues

1) Sample creation using the 5 dBZ threshold.

A first issue is that the precise application of the 5 dbZ threshold to identify precipitation is not described in detail. For the representation error calculation, the threshold described on line 128 may be applied to the low-resolution data, the high-resolution data, or both. Hence without further details, this is ambiguous. For the Desroziers calculation, the threshold may be applied to the observations or to the model simulations, or to both;

again this is not specified. It is also not fully clear if the threshold is applied per observation location, and how the threshold relates to the ensemble if one is used (is it the ensemble mean, control or ensemble members being used?)

In the literature on all-sky passive microwave assimilation, it is well-recognised that sampling issues need to be treated with care. Given that the location of precipitation in the forecast and the observation may be different, a "precipitation" sample based on the model precipitation alone will exclude many locations where the observations have precipitation but the model does not, and vice-versa. Depending on exactly how the thresholds are applied, very different bias and standard deviation characteristics may be observed. See for example:

"Observation errors in all-sky data assimilation", Geer and Bauer, 2011, https://doi.org/10.1002/qj.830

"Assessing the impact of pre-GPM microwave precipitation observations in the Goddard WRF ensemble data assimilation system", Chambon et al., 2014, https://doi.org/10.1002/qj.2215

In the current manuscript, a particularly striking difference is seen between the samples used in the representation error and Desroziers studies (Fig. 4c and 12c) at higher altitudes. Simulated reflectivity reaches 20 dBZ at 10km in the representation error sample but is just 12 dBZ in the observations in the Desroziers sample. This suggests that the sample at high altitudes in the representation error study is dominated by infrequent deep convection, since this is likely the only thing that can generate greater than 5 dBZ reflectivity at that altitude. The big difference to the observational sample could be explained by model error, but it could also just come from a major difference in the composition of the sample being analysed. These aspects need more attention.

2) Need for higher-level figures to summarise the results.

Figures 6, 8, 13 and 15 give a possibly excessive level of detail. The text has to do a lot of description and further analysis of these figures. It presents many lists of numbers derived from these plots, such as the correlation length scales (see e.g. lines 159 - 164). The results derived from these figures would be better presented and analysed on higher level figures, ideally comparing the representation error study with the Desroziers study on the same figures. This would reduce the need to present long lists of numbers in the text.

Similarly, the most interesting aspects of the study are the comparisons between figures 4 and 12, and Fig 7. versus Fig. 14. It is somewhat inconvenient to have to compare these figures manually. A summary figure combining some of the lines from both could be useful.

Minor issues

1) Line 22: the acronym ICON-D2 is not explained, nor its significance (presumably the first application of radar reflectivity assimilation in this framework?)

2) Line 65: The H() operators in equation 1 need some more explanation. Clearly they are not identical since the model inputs are on different grids. Any interpolation or coarse-graining within these operators needs to be explained. This is particularly important since the observation operator is non-linear. Hence, the application of H() to the mean of the model forecast may be strongly different to the mean of H() applied to individual locations in the forecast. This maps onto the well-known beam filling effect

3) Line 73: Although it's partly explained later on, it would be useful to have some words on how the model states x_a and x_b used in equation 2 relate to the model states defined in equation 1, given broadly as M(x^T). Even better, homogenising the notation between these equations 1 and 2 would help define the precise differences in methodology between the two halves of this work (such as highlighting the resolution and / or model differences involved, and/or differences in the observation operators being used).

4) Line 76-77: "R_est is optimal in case of …" - are there any further references to back this up or is it from Desroziers (2005)?

5) Figure 3, legend: "Scratch of…" is odd terminology - change the term or explain it.

6) All figures in the manuscript, but particularly Figs. 3 and 4: the point markers (such as a square or a circle) are so frequently sampled that they change the width of the lines, making them very thick in places and making it hard to do detailed comparisons between the lines. Consider showing all these graphs with only lines, not lines and symbols.

7) Figures 2 and 3 are not linked to the text where they appear. In any case there needs to be some early description of the processes of superobbing, and the details of the observation operator, in section 3. Instead these details appear partially, and too late, in section 4.2.1, for example.

8) Furthermore, there are some other introductory details missing on the model setups: for example for the models used in the representation error study, it is not clear whether data assimilation is applied to keep the forecast on track, whether (and where from) there are boundary conditions being applied to achieve the same result. An introduction to the

models being used in both halves of the study, their similarities and differences, would be very useful around section 3.

9) Line 107: “standard deviation and horizontal correlation” .. of what?

10) Line 113: “Around 10^3 … “ surely the authors mean “at 10^3 or below”?

11) Line 117: The training period concentrates on heavy thunderstorms. Somewhere, the authors should discuss the applicability of their results to other periods, such as wintertime cyclonic systems.

12) Line 131-133: “It is noticed that the variations of standard deviations are very comparable to the simulated reflectivities of the (high resolution) model run in Fig. 4c, indicating a systematic error that is proportional to the true value”. This does not make sense: Figure 4a and 4c do not look very similar, there is no strong similarity between the two. Also it is not supported why this should be explained by a systematic error.

13) Line 133-134: Comparison of Figure 4 to Figure 5 is not so helpful because the former is based on the <5 dBZ sample, and the latter is based on all data.

14) Line 208: “too big” - define?

15) Line 312 - 314: “the model produces too much ice” - this does not appear to have been much discussed or supported in the preceding text.

16) Line 314, 330: “this is mainly due to the application of scaling factor…”. Unless I missed it, this scaling factor has not already been discussed in the text, and its effect on the Desroziers-based observation error estimate has not been established.