Reply to reviewer #2

Specific comments

The article investigates the impacts of assimilating retrievals of polarimetric radar data on the assimilation cycles as well as on skills of short-term forecasts, using the ICON-D2 KENDA system of DWD. Authors' efforts are greatly appreciated and it is very encouraging to see positive results of this study. However, I have several specific comments for authors to consider:

1. Language: 1) The language is not concise. For instance, at places "Assimilating LWC estimates instead of Z data where possible (CNV+LWC/Z)", "assimilation of IWC instead of Z where possible (CNV+IWC/Z)" and etc. would be enough to use the acronyms of experiments. 2) There are too many acronyms that seriously disturb the readability of the article since it is often difficult to tell what they are associated with, e.g., LC, LH, LV, OE, LS, LL, MV and ect. If they are really necessary (some of them seem not), could you make them more self-explaining or make a list for acronyms. Besides, it is uncommon to call "Data assimilation parameter". "Settings of the data assimilation system" may be more appropriate.

   Dear reviewer, thank you very much for your time and helpful comments.

   Note that all new line numbers in the following refer to the manuscript file without tracked changes!

   1: Yes, we first tried to explain the assimilation configurations like "Assimilating LWC estimates instead of Z data where possible (CNV+LWC/Z)" repeatedly in the text to remind the reader, but we followed now your advice in the revised manuscript. (e.g. L396,397)

   2: Yes, we made the acronyms more self-explaining throughout the document: LC -> res_cartesian; LS -> winsize_avg; LH -> obsloc_hor; LV -> obsloc_ver; OE -> obserr_std; LL -> lower_lim; MV -> minnum_vals (see e.g. L292, 294, 297). Moreover, we use the wording "Data assimilation parameter" for winsize_avg, obsloc_hor etc. to clearly distinguish it from the data set configurations like CNV, CNV+LWC/Z etc., which also belong to the "settings of the data assimilation system". Therefore, we would like to keep the acronym "DAP" in our study.

2. Assimilating of Z at the melting layer: 1) It should be noted that attenuation is normally switched on in EMVORADO, but the observations are corrected for attenuation. Authors should check if they are consistent. 2) I assume that the Mie scattering scheme is used here. As shown in Zeng et al. 2022a,b, the default parametrizations for the melting layer in case of the Mie scheme result in unrealistically high reflectivities around the melting layer. Therefore, it is striking to assimilate Z only around the melting layer, which could be error-prone.

   1: Thank you for that important comment. Yes, it is true that attenuation of Z is turned on in ICON-EMVORADO by default. We talked to Ulrich Blahak from DWD about that topic. In his research group they are still elaborating on finding a way to compare observed attenuation-corrected Z with EMVORADO-simulated non-attenuated Z. However, at the moment Z assimilation works best when attenuation is switched on in EMVORADO while assimilating attenuation-corrected observed Z. This may be caused by the point that the ICON-EMVORADO tends to overestimate Z on average which may partly be compensated for by the applied attenuation. A brief view into the data showed a general consistency of the simulated and observed Z fields. Since this study assimilates Z as performed in the operational routine, attenuation-corrected observed Z and attenuated simulated Z are used. Besides the attenuation effect, inconsistencies between modelled and observed Z in *the ICON-EMVORADO have become evident e.g. in rain below the melting layer because of too large simulated drops, and around the melting layer because of excessive graupel production. These are important points to be addressed in the future but would by far go beyond the scope of this study.*

2: Yes, Mie scattering is used. In the operational ICON-D2 routine at DWD, Z is assimilated within the melting layer. Since the operational Z assimilation (without LWC or IWC) is used as reference in this paper, the configurations assimilating LWC/IWC also need to include the Z assimilation in the melting layer for comparability. However, results of assimilating Z data in the melting layer highly depend on the operator's melting scheme. Different flavours of the Maxwell-Garnett-, Bruggemann- and Wiener Effective Medium Approximations (EMA) can be chosen to explore the uncertainty in the melting layer. We are currently investigating deficiencies of the simulated melting layer signature, parts of which may be attributed to the fact that in the model microphysics parameterization, meltwater from snow, graupel, and ice is instantaneously shedded into the rain class, causing too small and too few remaining frozen particles in the melting layer. The current version of EMVORADO estimates a melted fraction as function of temperature and particle size (Blahak 2016) as part of the remaining frozen mass without "back-shuffling" some rain water to the particles. This leads to a systematic underestimation of the melting effect in all radar moments, despite the quite detailed consideration of various EMAs for the effective refractive index. We will explore better approaches, e.g. a wet snow class borrowing parts of the rain and snow mixing ratios and mix them (e.g. Jung et al. 2008a, 2008b, 2010), to reduce the bias to the observations. However, as soon as model microphysics with explicit mixed-phase snow, graupel, and hail (e.g, as in Frick et al. 2013 for snowflakes) becomes available in the future, their liquid fraction could be directly applied in the forward operator.

Blahak, U.: RADAR_MIE_LM and RADAR_MIELIB – Calculation of Radar Reflectivity from Model Output, COSMO Technical Report 28, Consortium for Small Scale Modeling (COSMO), available at: http://www.cosmo-model.org/content/model/documentation/techReports/cosmo/docs/techReport28.pdf (last access: 10 January 2022), 2016. a, b, c, d, e

Jung, Y., G. Zhang, and M. Xue, 2008a: Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part I: observation operators for reflectivity and polarimetric variables. Monthly Weather Review, 136 (6), 2228-2245, DOI: 10.1175/2007MWR2083.1.

Jung, Y., M. Xue, G. Zhang, and J. M. Straka, 2008b: Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part II: impact of polarimetric data on storm analysis. Monthly Weather Review, 136 (6), 2246-2260, DOI: 10.1175/2007MWR2288.1.

Jung, Y., M. Xue, and G. Zhang, 2010: Simulations of polarimetric radar signatures of a supercell storm using a two-moment bulk microphysics scheme. Journal of Applied Meteorology and Climatology, 49 (1), 146-163, DOI: 10.1175/2009JAMC2178.1.

Frick, C., Seifert, A., and Wernli, H.: A bulk parametrization of melting snowflakes with explicit liquid water fraction for the COSMO model, Geosci. Model Dev., 6, 1925–1939, https://doi.org/10.5194/gmd-6-1925-2013, 2013.

3. As shown in Zeng et al. 2016, Bick et al. 2016, the one produces smaller errors (likely due smaller localization radius, shorter update frequency and etc.) within assimilation cycles may not always win in forecasts, usually because of the imbalance issue that accelerates the error growth. Generally, it is inappropriate to say which one is the best only based on the performance in cycles but without looking at results of forecasts. Therefore,

authors should avoid saying, e.g., ”most successful assimilation settings”, ”finding the best DAP sets” and similar phrases. Maximally, it can be said that they result in smallest errors in terms of ...

You are absolutely right. An “optimal” or “best” DA setting should refer to the quality of the forecast and not on the quality of first-guesses in assimilation cycles. Therefore, we changed the formulation at all positions in the document. (e.g. L405-407)

4. Redundancy in appendix: It is unusual to describe the LETKF, FBI, FSS and BSS in full length since they are well-known. It would be enough to describe them shortly but provide their references.

We appreciate your suggestion and removed the appendix. In the document text the references to LETKF, FBI, FSS, and BSS are cited and should be sufficient for understanding.

Minor comments

1. Line 54: the error covariance matrix ==> the background error covariance matrix

   Thanks! (L55)

2. Line 69: Bonavita et al. 2010 ==> Gastaldo et al. 2021

   Thanks! (L71)

3. Line 115: EMVORADO may not stand for ”Efficient Modular Volume Radar Forward Operator”. Please ask Dr. Blahak Ulrich for this.

   Yes, it should be “Efficient Modular VOlume scan RADar Operator”. Thank you! (L116,117)

4. Line 160: reference for one-moment scheme

   Citation now included: Doms et al., 2011 (L161)

5. Line 175: Add the inflation method, RTPP and additive noise are applied for inflation.

   We included one sentence. (L175-177)

6. Line 199: Please check if the used radar observations in 2021 are really of radial resolution of 0.25 km.

   Yes, they have a resolution of 0.25 km while the 2017 cases have a 1 km resolution.

7. Line 236; The American climatological value 0.02 can be used for German radars?

   As there is no comparable value for central Europe it may be used approximatively.

8. Line 290: The superobbing described in Bick et al. 2016 is a bit outdated, the updated one is given in Zeng et al. 2021.

   We included the newer reference. (L296)

9. Line 304: better results than what?

   We clarified the point. (L309-311)

10. Line 332: Please provide the frequency of boundary data

    We included the frequency of one hour. (L338)

11. Line 350: Is the JQS a new metric introduced in the work? If yes, please explain more why this hybrid metric JQS is more interesting or important than FSS and BSS, otherwise, provide the reference.

    Yes, this metric is new. We modified the respective sentence: “The results of using the DAP configurations/values are compared with each other in terms of both first-guess deterministic and ensemble QPF quality via a single univariate measure newly introduced here – the joint quality score (JQS)..” (L355-360); “While changes in deterministic and ensemble QPF quality with respect to the CNV+Z configuration are not always consistent, the JQS provides a useful measure for the overall intercomparison of DA settings..” (L361-363)

12. Line 361: How are weights determined?

    “weights are determined by the fractions of threshold exceedances for a given time and threshold of the total number of exceedances at all thresholds (0.5, 1.0, 2.0, and 4.0 mm h-1) and events (C2017, S2017, and S2021) in the RADOLAN data (see Fig. 3)” (L371-374)

13. Line 387: Could authors be more specific about ”which may be due to discrepancies between true and model microphysics”?

    We removed the sentence about the discrepancies. (L400)

14. Line 485: Remove colors

    We removed the colors. (L498)

15. Line 534: Strange phrase ”first-guess precipitation forecasts” ==> firstguess of precipitation

    We modified it. (L523)