Comment on gmd-2021-302
Anonymous Referee #1

Referee comment on "Implementation of an Ensemble Kalman Filter in the Community Multiscale Air Quality Model (CMAQ Model v5.1) for Data Assimilation of Ground-level PM2.5" by Soon-Young Park et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-302-RC1, 2021

This manuscript presented method comparison: EnKF and 3D-Var, for assimilating surface PM2.5 observations with two settings: IC and ICBC. It is a straightforward paper. One major issue is that the prediction model CMAQ has only 15 layers up to 20km, which is too coarse. How many layers below 1km? Could this coarse vertical resolution cause artificial dilution for near-surface pollutant concentration, and result in the systematic PM2.5 underestimation? Although this manuscript focuses on data assimilation (DA), the corresponding prediction model should be reasonable, too. Otherwise, the DA methods only show their effect on correcting the systematic underestimation.

Here are the detailed comments.

Section 2. PM2.5 is not a single species in CMAQ. How do you map the PM2.5 increment to individual CMAQ aerosol species?

Page 9, line 155. What's the vertical extent of the 50% perturbation being applied, to all layers? Considering that it is used to the assimilating surface observation, certain justification is needed.

Page 6, line 172. Do you think that the static horizontal width of 100km and vertical width of 2km fit for all scenarios, for day and night? Any discussion about it.

Line 164. Same as above. Is the 30% standard deviation of LBC perturbation applied to the all layers?

Section 2.2. The 3D-Var description in section 2.2 is too short, and needs to include more detail. What are the horizontal/vertical length scales, and model error covariance yielded by the NMC method? Could you show some plots about them?

Figure 4, it is better to include the corresponding 3D-Var increment for comparison.

Section 3.3. Does the evaluation use the same observation data as those used in DA?