Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-100-RC1, 2020

Earth System Science Data Discussions

Interactive comment on "A Six-year long (2013–2018) High-resolution Air Quality Reanalysis Dataset over China base on the assimilation of surface observations from CNEMC" *by* Lei Kong et al.

Anonymous Referee #1

Received and published: 24 July 2020

This study presents high-resolution air quality reanalysis products over China for 2013-2018. The air quality reanalysis assimilated the country-wide surface observations using the regional EnKF data assimilation. The assimilated results were evaluated against the assimilated and independent measurements. The topic of this study is very interesting, and the produced data sets can be useful for various applications. The paper is generally well written. However, because this is the first paper describing the system and data, more careful description of the system and its performance would be useful for readers and future developments. Please see my recommendations below.

1. The representativeness error estimation is not clear. How did you estimate L_repr for each station and Epsilonˆabs for each species? Urban and rural observations could be (or should be) used in a different way, but this is not mentioned. Were any temporal averages applied to the observations? Temporal variability information could be used a part of representativeness errors. Further explanation is needed.

2. The assimilated results are compared with the independent observations for PM but with the assimilated observations only for other species (they only demonstrate self-consistency. CAMS is not observation). This provides limited information on the performance of the developed system. The Chi-square diagnostic can be used to see whether the Kalman filtering worked properly. OmF & OmA statistics can also be demonstrated. Given limited validation data, more efforts are required to demonstrate the performance.

3. Inter-species correlation was totally neglected in background error covariance. This setting is extremely conservative and does not fully utilize the advantages of EnKF data assimilation that produces comprehensive background error patterns. I'm wondering if the authors have tried to implement inter-species correlations. Further discussion is needed (e.g., why it is so conservative, what is the disadvantage of the current setting).

4. Please clarify whether there are any variations in inflation factor and how it was optimized for different species. In most regional ensemble data assimilation systems, fixed lateral boundary condition tends to limit the effectiveness of data assimilation near their boundaries (and also inside when horizontal advection is strong) because of reduced spreads. Did you find any problem with it?

5. Using automatic outlier detection method, how much observations were rejected? What was the impact in data assimilation?

6. Because of the fine-scale variability and large degree of freedoms, the high-region data assimilation would require larger ensembles. I'm wondering if 50 members are sufficient. Further discussion is needed to demonstrate whether the background error

is produced properly to propagate observational information in space.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-100, 2020.