Review of "A three-dimensional variational data assimilation system for aerosol optical properties based on WRF-Chem: design, development, and application of assimilating Himawari-8 aerosol observations" by Wang et al. (gmd-2021-215)

General comments

This study proposes a new development in assimilation of aerosol optical properties based on the foundation of three-dimensional data assimilation (3DVAR) scheme in the GSI and coupled with framework of MOSAIC in WRF-Chem. A heavily polluted event of one-week period occurred in northern China is selected for the demonstration. Referenced by several observational datasets, the results indicate that assimilation of Himawari-8 AOT retrieval can reduce the negative bias in modeling both AOT and PM2.5 over the region in comparison to the simulation without any aerosol DA.

In general, this manuscript is well-organized and quite readable. I enjoy reading the technical descriptions which provides sufficient information even for entry-level reader. Nevertheless, I realize the discussions in the sources of uncertainties in aerosol modeling, assimilation frequency of satellite data, and analyzed vertical distribution of aerosol properties are not included in current form and may need to be addressed to further enrich the contents. Thus, three major comments are made accordingly and provided below. Specific comments and suggestions are also given in the followed list.

1. The model resolution, meteorological conditions, and emission data could be other important sources of uncertainty in the air pollution modeling and in fact some of them can be identified in the diagram you show in Fig. 2. However, they are not discussed in the manuscript. Would you be able to quantify these uncertainties in relation to the impact of aerosol field initialization (DA) based on the design of model experiment? For instance, you may consider conducting additional experiment which assimilate meteorological states and aerosol to explore their relative impacts on the subsequent forecast.

2. The under-utilization of Himawari-8 AOT product (hourly data) in the context of assimilation frequency (24 h) seems to be obvious. I imagine a strategy with more frequent assimilations could be a unique point to make in this research as the geostationary satellite product used here has such a high temporal resolution. Nevertheless, the relevant discussion is not covered in the manuscript. I would suggest adding more content to address this comment.

3. Despite the vertical profiles of background error STDs and auto-correlations are given, the analyzed increments of each aerosol state variables are not seen anywhere in the document. Since the AOD is obtained through the integration of aerosol properties in the atmospheric column, it would be useful to show analyzed results in terms of their vertical

distributions and further discuss how would that contribute to the uncertainty of simulation.

Specific comments

L32: It is mentioned here that the developed DA system is able to assimilate lidar-based aerosol profiles. However, I did not find any relevant description with respect to the treatment in the followed sections. Would you clarify this?

L237-240: Have you conducted any experiment to test how sensitive this constant error is?

L260: Can you give an example of the minimization process, such as reduction of cost function in function of iteration numbers?

L288-289: Please include references to supplement statement here.

L291: Should be black car”b”on and organic car”b”on

L369: Would this introduce any inconsistency between nonlinear model and TL? Also, I am curious how did you deal with if statements in the code if there’s any.

L389: Since this manuscript documents the development of a DA package, it is of necessity to show the result of TL/AD test. For example, it is common to show the plot of gradient check with respect to various orders of perturbation.

L418: Please cite this reanalysis product and provide the link of the data source.

L422: The assimilation cycle time (24 hours) seems to be coarse in relation to data availability. Please discuss how it is designed and clarify if there’s any limitation on the data coverage or quality, etc.

L424-426: The statement here is contradictory to the design of assimilation cycles. Please explain.

L441: I am not sure this is the best treatment as it could further smooth out the observed data. Please address.

L443 and L463: Fig. 3b is mentioned earlier than Fig. 3a. I would suggest swapping them for the fluency of reading.

L492: It looks like the similar DA procedure is also carried out over the D01 but at least with different treatment in data thinning. Have you done any experiment without assimilation in D01? If true, what was the impact of additional DA in D01.

L532: Is it possible to estimate the correlation length with the observational data or alternatively the analysis after assimilation?

L577-579: Sentences such as these in the manuscript could be trimmed to shorten the length.

L587: Please elaborate more on this. Would the uncertainty mostly be on the magnitude or something else?

L606: The red triangles in Fig. 1b are hardly distinguished from one another as they are basically overlapped with each other. Please try to make them more visible. Add another zoomed-in map may help achieve that.

L610: What is the temporal resolution of AERONET observations? From the time series plot of Fig. 8, it looks like the data is mostly only available around 00 UTC of each day

L615-616: Any explanation why model has worse skill at XuZhou-CUMT? It seems the event on Nov. 25 is more severe than Nov. 26 at this site and not captured as well.

L617-618: Any guess on this? Have you looked at the meteorological conditions on these days? Could it be associated with the intensity of wind speed?

L622: It would be easier for reader to understand if the data distribution map of Nov. 26 is also provided. Along the same line, I would suggest adding information of available data amount in Fig. 8 to address this.

L643: You may remove "between analyses and the background field" since increment has been defined in the earlier paragraph.

L644-645: The of color bar scales in Fig. 3a and Fig. 9 are not consistent, which makes it hard to compare them visually. Please consider modify them.

L645: Need to mark where Tianjin is in the map, otherwise one may not know which location you talked about.

L644: Panels in Fig. 9 are not sufficient to conclude the underestimation in control experiment as no observation is provided.