General comments

Dear editor and authors,

The manuscript presents an interesting application of aerosol data assimilation for a high pollution event in the eastern China for the period 23 to 29 November 2018. Aerosol Optical Depth (AOD) from the Himarawari-8 satellite is assimilated in an effort to improve AOD and surface particulate matter smaller than 2.5μm (PM2.5). Assimilation experiment is evaluated with independent observations (MODIS, AERONET and China's nationwide monitoring network system) and shows improvement in comparison to Control experiment, either for AOD or PM2.5.

Overall the manuscript is well written, describes most of the aspects of the system in detail and the result section is clear and well presented. However the authors are not discussing adequately the role of observation representation error, the possible misrepresentation of aerosol state when assimilating just AOD while adjusting 20 aerosol state vectors, as well as the impact of the temporal assimilation cycle (1 day). I briefly discuss this three points below. Further I have included some specific comments that can improve quality and readability of the manuscript as well as some technical corrections.

- How can AOD distinguish and constrain 20 different aerosol state variables? What is the impact of using only AOD? There is no mention of other studies that assimilate more information than just AOD (e.g. AOD in other wavelengths or Angstrom Exponent, Absorption Aerosol Optical Depth or Single Scattering Albedo as well as direct radiances assimilation). Although the authors acknowledge the need for combine assimilation of various optical properties in their closing statement in conclusions (L746-751), many recent studies that are related to that are not mentioned. To name a few ones: (Chen et al., 2019; Escribano et al., 2017; Tsikerdekis et al., 2021)

- The spatial aggregation of observations that the authors describe (aggregating observations in the spatial resolution of the model) is indeed often used in data assimilation studies. Although was there any consideration regarding the representation error of this aggregated observations? For example, was the observational error inflated by X amount because you were not using the original resolution of Himawari-8? (Lines 437-442)

- As a geostationary satellite, Himawari-8 is known for its high temporal frequency. Since the data assimilation cycle is in daily frequency (updating analysis once a day), are you fully exploiting this satellite capabilities or rather its strong point? I realize that the daily assimilation step was chosen for practical reasons (computational speed), nevertheless I would expect some discussion about it. Further related to this topic, I did not find any discussion related to temporal collocation of observation in the data assimilation system.

Specific Comments

L60: Missing references.

L65-67: Refence, name and accessibility (or the lack of) for this dataset should be provided.

L73: Probably mean "remote sensing optical properties can cover a much larger domain". Because just optical properties can be retrieved also from AERONET stations.

L189-192: In principle PM2.5 can be estimate from the modes that the MADE scheme uses, assuming you know the median and the standard deviation of the distribution for each mode. In that case MADE would be superior to MOSAIC since it will also include mixing of different species within each mode. So I would suggest to emphasize only the numerical efficiency of MOSAIC against MADE. Further, indicating how much faster it is could really promote that argument and it could be easily estimated with two forward simulations, one with MADE one with MOSAIC (no DA required).

L211-213: Authors could mention here that the vertical axis is on hybrid sigma-pressure levels, if that is the case.

L237-238: It would be really helpful to briefly mention here how Yumimoto et al. (2016) estimated this error for Himwari-8 AOD and what this error actually describes (e.g. instrument error, retrieval error, representation error) ?

L491-493: It would be interesting to compare the D02 and D01 estimated background error standard deviation. It would show how important is the model horizontal resolution for this metric. If possible an additional plot for the D01 over the domain of D02.

L562: I would strongly recommend to replace "improvements" with "changes" in that sentence or rephrase. Figure 6 shows the differences of the Analysis – Control. It is not an evaluation with observations (assimilated or independent) where we can truly determine if there was an improvement by the data assimilation.

L585-587: It would be beneficial to provide how much this difference in AOD wavelength (500nm and 550nm) is affecting your evaluation. Maybe you can use Angstrom Exponent from AERONET to determine that and provide a number? Usually AOD at higher wavelength (550nm) is smaller than AOD at lower wavelength (500nm). Which means that the bias would be even more negative if you were comparing MODIS and Model at the same wavelength at Figure 7b. I think it is worth discussing in the manuscript (L595+) although it may enhance the negative bias you get for both Control and Analysis.

L604-606: AERONET sites at Figure 1b are hardly visible (probably because 4 of them are in the Beijing area). It would be visually better to enlarge them a bit.

L664-669: Good point, spatial availability of AOD in contrast to PM2.5 can play a role. I would also add that AOD is an atmospheric column measurement while PM2.5 is a surface measurement. Therefore, if you have an aerosol plume which is not close to the surface AOD

can be increased by increasing the aerosol concentration of that plume while PM2.5 can remain almost unaffected by that change.

Technical Corrections

L140: "3DAVR" to "3DVAR"
L173: "back carbon" to "black carbon"
L203: "/MADE/" is some kind of typo?
L291: "black carton, organic carton" to "black carbon, organic carbon"
L609: Something is missing in the sentence. Probably "used to" to "used them to"
L1185: Figure 11: Do you mean "average over 7 analysis steps" instead of "average over 7 single experiments"?

References

Chen, C., Dubovik, O., Henze, D. K., Chin, M., Lapyonok, T., Schuster, G. L., Ducos, F., Fuertes, D., Litvinov, P., Li, L., Lopatin, A., Hu, Q. and Torres, B.: Constraining global aerosol emissions using POLDER/PARASOL satellite remote sensing observations, Atmos. Chem. Phys., 19(23), 14585–14606, doi:10.5194/acp-19-14585-2019, 2019.

Escribano, J., Boucher, O., Chevallier, F. and Huneeus, N.: Impact of the choice of the satellite aerosol optical depth product in a sub-regional dust emission inversion, Atmos. Chem. Phys., 17(11), 7111–7126, doi:10.5194/acp-17-7111-2017, 2017.

Tsikerdekis, A., Schutgens, N. A. J. and Hasekamp, O. P.: Assimilating aerosol optical properties related to size and absorption from POLDER/PARASOL with an ensemble data assimilation system, Atmos. Chem. Phys., 21(4), 2637–2674, doi:10.5194/acp-21-2637-2021, 2021.