The manuscript, titled “Improving estimation of a record breaking East Asian dust storm emission with lagged aerosol Ångström Exponent observations”, by Yueming Cheng et al. is a data assimilation (DA) case study for improving dust storm predictions in China using WRF-Chem. The authors chose a dust storm in Mar 2021 as a case study and successfully showed that using AERONET observations for DA could improve hindcasting ability of dust in WRF-Chem. Authors further showed that employing the Angstrom Exponent (AE) benefits more than employing the aerosol optical thickness (AOT) in improving DA results. The methodology is generally scientifically valid and clear in presentation, although there are occasional grammar issues or unclear descriptions that could use a little more editing. The main issue I thought was that there are insufficient science discussions to the results, such as why AE benefits more than AOT, or how could the WRF-Chem dust model be improved. The manuscript so far feels a little more like a technical report on improving hindcasts rather than a scientific development on our understanding of dust modelling. I have a few major comments below for science, and I have some other specific comments on technical questions or presentation. I suggest major revision for the current revision.

Major comments:

- I think there needs to be some science discussions on what caused the biases in WRF-Chem dust emissions in China. The paper currently leaves readers with puzzles regarding why WRF-Chem underestimates dust so much. Is it problems in simulated dust emissions, lifetime (dust deposition), or optical properties? If the bias comes from emission, is it a problem in wind speed, soil moisture, vegetation, or other met fields? If it's deposition or optics, is there anything to do with size distribution, dust particle shape, or dust refractive index? Does the dust underestimation also occur over the rest of the world in WRF-Chem? How does changing the *a priori* emissions (using other dust_opt or other emission schemes) alter the FR underestimations? Modellers would like to know how could our process-based dust understanding benefit from the insights from this DA study.

- It looks like authors attributed all the differences/biases between AERONET-measured and WRF-simulated AOT to dust. Could the biases be attributed to other natural and anthropogenic emissions? Although you only assimilated three days where dust was dominant, there must be some strong anthropogenic and natural emissions that got captured by AERONET, especially over Beijing, a heavily polluted metropolitan area. You only used AE values < 0.6 for evaluation to focus on dust, but it seems you didn’t do the same when doing the DA. From my point of view, it could be better to use the coarse-mode AERONET AOT from the spectral deconvolution algorithm (SDA) to do the DA since all fine-mode aerosols are truncated and only dust/seasalt remains.

- Even if all biases in AOT/AE were assumed to be due to dust, currently the authors attribute all differences/biases between observations and simulations to emissions and only correct emissions. This presumes there are no biases in dust settling/deposition and dust optical properties. But if so, assimilating AOTs should correct most of the error, and AE would not be needed. A science discussion is needed on why the AE information could further reduce bias. Authors could not just conclude that the more you use for DA the better the hindcast results. My thought is that if AE is additionally needed for DA, this either means

there are problems in AERONET AOT, or (more likely) WRF-Chem has problems not only on dust emissions but also on dust optics. Studies also have pointed out issues in both settling velocity (e.g., Huang et al., 2020) and optics (e.g., Di Biagio et al., 2019) in models. If so, it does not make so much sense to attribute all AOT/AE biases to emissions to compensate other errors in WRF-Chem. Maybe optics should also be inverted, not just emissions.

Other specific comments:

Line 78: It's a little difficult to grasp how many AERONET stations you used for assimilation. Please state in text. I suggest plotting out the locations of the AERONET stations, with values of AOT and AE, either in Fig. 1 or in SI.

Lines 87-88: Authors wrote two observation errors: "observation error is a sum of representation error and observation error", which is confusing. Maybe use another word like instrument error for the latter one?

Line 89: I glanced through Schmid 1999 but didn't see such characterization of representation error. Schmid 99 was not a WRF-Chem modelling study either. Why should representation error be 0.055τ? Should it be more related to WRF-Chem grid resolution? Also, please define τ.

Line 92: I think it is needed to state why authors chose to assimilate AERONET instead of SONET or CALIOP. It looks random to me.

Line 93: Please also plot out the locations of the SONET sites and their values of AOT and AE, in main text or supplement.

Line 117: Please define the acronym MOSAIC. I am not sure how important this modification is if you don't concern dust chemistry, since MOSAIC never appeared in the text again. How are the spatiotemporal distributions of metal ions changed through this modification?

Line 140: Again, it looks like the whole difference between AERONET AOTs and simulated AOTs are attributed toward dust emission biases. Can't there be biases from other natural and anthropogenic emissions?

Line 145: I am not sure if the error covariance includes forward model error (that is, error in the WRF-Chem H operator). Please clarify.

Line 155: A localization length of 600 km sounds a little too big to me for assimilating AERONET data. It is almost a meso-synoptic length scale and is much bigger than your WRF-Chem horizontal grid resolution.

Line 155: "Grid centroid" instead of "centre grid"?

Figure 1: It's interesting that using AE measurements in Beijing could lead to changes in the posterior AE (AOT+AE DA) over Taklimakan, or even in India, in comparison to the *a priori* AE

(FR). How does DA generate emission changes in Taklimakan and India, if you were using AERONET sites in Beijing?

Line 203-206: It is not clear what this means. There were no observations because AERONET sites were down, like because of the dust storm? Please rephrase. Does this mean if you use the observations on 14-15 March, dust emissions would be even higher? Please clarify.

Line 208: I suggest adding map plots on the prior error of the WRF FR emissions, as well as the posterior errors of inverted emissions for the AOT DA and the AOT+AE DA cases. It helps visualize how adding AOT and AE for DA reduces the posterior errors of DA emissions.

Figure 4: When you say “aggregate”, is this plot averaged across or summed across the domain? Please clarify in text.

Line 213: I suggest either saying posterior and prior dust emissions, or *a posteriori* and *a priori* dust emissions.

Lines 217-219: So, did you use more AERONET sites than listed here for DA above? Or are these all the sites used for DA? I am still confused, please clarify.

Line 236: Here authors should suggest scientific reasons for why using AE would benefit DA so much, while AOT DA would not.

Line 248: What's the reason of selecting these two SONET sites but not the other two? It looks a little random here. I suggest plotting the comparisons for Songshan and Jiaozuo in main text or supplement too.

Line 255: From here on, I start to find the message for the next few subsections a bit repetitive, stating that the AOT+AE DA run is better than the FR run and the AOT DA run. The manuscript could use a little rewriting to make the discussion and message more succinct.

Figure 6: It seems to readers there are insufficient SONET data points for the time series plot. I suggest authors also include SONET data points for AE > 0.6 since you used it for DA, in any color other than grey.

Figure 7-9: These are nice plots. Though, readers find the message across Figs. 7-9 a little repetitive. I would suggest showing the extinction coefficient (second rows) and skip the AOT plots (first rows), and maybe combine second rows of all three figures together. The first rows could be put in supplement.