Comment on acp-2021-1012
Andrew Sayer (Referee)

I am writing this review under my own name (Andrew Sayer) as I have collaborated with several on the author list and work at the same institution (NASA GSFC) as co-authors da Silva, Castellanos, and Choi. I disclosed this to the Associate Editor handling this manuscript on receipt of the invitation to review, and was advised that it is ok to proceed.

This paper presents results of airborne remote sensing (4STAR) and in situ (LARGE) observations of aerosols during the KORUS-AQ field campaign in and around Korea during summer 2016. They are analysed jointly with satellite retrievals (GOCI) and reanalysis (MERRA2). One key result is that Ångström exponent (AE) and aerosol fine mode fraction (FMF), proxies for columnar aerosol type, showed more rapid spatial variation than aerosol optical depth (AOD). This is in contrast to previous studies elsewhere which have generally observed that AOD varied on finer scales than composition. Potential reasons why are discussed.

My main expertise is in remote sensing and not meteorology or the in situ sampling. I recommend at least one reviewer with more of a focus on those areas, as they will be able to better judge some parts of the study than I can.
The topic is important and within scope for ACP. The quality of writing and presentation is high (though I have a few suggestions for changes to Figures). My overall recommendation is for minor revisions. I would be willing to review the revision if the Editor would like. My specific comments and suggestions for revision are as follows:

- The main metric used to quantify the spatial scale of variation is the distance at which the autocorrelation drops off to 85% of its value in the smallest distance bin. I was wondering why 85% was chosen? I would have thought it more common to state in terms of an e-folding distance – unless the autocorrelation profile doesn’t look like exponential decay (which some of them might not). Either way I’d appreciate some (brief – not repeating the whole analysis) discussion in the paper of why this particular threshold was chosen and if results qualitatively change if a different metric is used – for example the e-folding distance, or an autocorrelation drop to e.g. 70% of the max rather than 85% (given a correlation of 0.7 corresponds to about 50% of the variance in the field). Looking at curves my guess is in most cases the picture would be the same, but as thresholds are a bit arbitrary it is good to check sensitivity to them.

- Related to the above, it would be interesting to quantify at a couple of places what the typical variation in the field is for these autocorrelation drops (e.g. at the distance of 85% autocorrelation, what is the variance of the difference between AOD or FMF or AE at that point and at zero lag). This helps give an idea of how numerically important some of these variations are (with the understanding that these magnitudes might not be transferable to other regions or seasons). For example at the 22.7 km distance where AE autocorrelation has dropped to 85% is the AE difference about 0.1 or 0.3 or?

- KORUS-AQ also included a dense deployment of ground based AERONET sites (mostly around Seoul). I wonder if these could be used as an additional data source for the autocorrelation analysis to see if the overall picture of relative scales of variation holds as for the 20 DC-8 flights. While they would not be spatiotemporally collocated with the other data sources used, the data have low uncertainty and good temporal sampling. I am not sure if the inter-site spacings are sufficiently varied to fill out the autocorrelation distance profile, but it could be worth looking at the distance pairings to see if this could be a useful addition.
Line 355: the Abstract highlights average and variability of AOD/AE for flights below 500 m but the text here highlights those numbers for flights below 1000 m. Later in the paper there’s some discussion of profiles below/above 500 m but the main results here are all framed relative to 1000 m. I thought I’d mention as I’m not sure whether this difference in reporting altitude between the Abstract and main text was intentional.

Figures 3, 6: if I understand correctly the spectral plots are means and standard deviations. The data are shown on a log scale so the lower tails of the standard deviations often go down to the y axis. I think it could be more meaningful to plot geometric means (i.e. mean and standard deviation of log(AOD)) or else median and interquartile range (or central 68% of points). These, especially the latter, would be informative of the shape of the AOD distribution at each wavelength.

Figures B1, B2, B3 and lines 444-448: I am assuming that the regressions here are ordinary least squares (OLS) linear (unless I missed it, it’s not explicit). They should really be removed because this technique is inappropriate for these types of data. Some assumptions required for the validity of OLS linear regression include (a) an underlying linear relationship; (b) independent samples; (c) a single underlying (ideally Gaussian) distribution, (d) negligible uncertainty on the independent variable, and (e) equal variance of the dependent variable across the range of the independent variable. Looking at the clouds of points, the linearity assumption appears invalid for B1(b) and B2(b). The independence assumption is likely invalid throughout given the point of this paper shows high levels of correlation across the domain. The distribution shape assumption is likely invalid since AOD tends to be skewed and closer to lognormal, plus the different meteorological fields having different AOD distributions mean we don’t have draws from a single distribution but perhaps 4. The independent variable assumption is valid since, as noted, the 4STAR AOD uncertainty in the midvisible is about 0.03, which is not negligible relative to the low AODs commonly found for the bulk of the data. The AE is also uncertain. Note this assumption can be overcome by use of e.g. reduced major axis (RMA) regression accounting for the uncertainty in the independent variable, but this doesn’t help with the others. RMA might also be impractical in the present case because my guess is that a non-negligible fraction of the uncertainty in all the data sets here is systematic (e.g. radiometric calibration uncertainty through deployment) so would also be correlated. The equal variance assumption appears to be violated for panel B3(a) and possibly B1(a) (this can also be overcome using weighted regression if pointwise uncertainties are known beforehand). In short, all the data sets violate some of the assumptions, and the numbers and
uncertainties presented as regression results are not quantitatively correct. The OLS technique is often used in our field but this does not make it right. I recommend the authors remove the regressions from the plots. In any case I don’t think they are really needed to get to the main point about the level of comparability of the data. I think showing R2 is ok (as the collinearity of the data is of interest) but rather than regression equations perhaps some metrics like RMS difference, mean offset, mean absolute difference could be used instead. The discussion in lines 444-448 of the paper should be amended as a result. I don’t mean to harp on about this point but since inappropriate regressions are common in our field think it’s important to try and stop the practice when I get a chance in peer review.

- Lines 857-859: “Satellite algorithms that assume that aerosol size does not vary as much as aerosol optical depth should be reassessed.” I am not aware of data sets produced from algorithms that make assumptions like that, on the scales of tens of km being discussed here -- are there any? Most either operate on single pixels (i.e. no spatial constraints) or do multi-pixel processing at a much finer scale than the spatial scales reported on here (e.g. MISR at 4.4 km, GRASP applied to POLDER at 10 km). The VIIRS SOAR ocean algorithm assumes the same fine mode and coarse mode microphysics across 6 km grid cells but AOD and FMF are allowed to vary without spatial constraints for each 750 m pixel within that area. MAIAC used to have some constraints but now retrievals (at 1 km) are spatially independent. It sounds like the GOCI data set used here might (I’m not 100% certain by the way the model selection is described in the paper) but again that’s going from 0.5 to 6 km so a lot finer than the scales of variation here. It would be good to either give examples of algorithms here or else delete the comment if there are none using such constraints at the relevant scales.

Language comments:

- Title: I am not 100% on this, but I am a bit uneasy about “aerosol optical depth are”. I think it should either be “aerosol optical depths are” or “aerosol optical depth is”.
“Angstrom” should be typeset as “Ångström” throughout.

Line 428: “sporadic aerosols events” should be “sporadic aerosol events”.