Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC1
https://doi.org/10.5194/amt-2021-262-RC1, 2021

Comment on amt-2021-262

Anonymous Referee #1

Referee comment on "Deep-learning-based post-process correction of the aerosol parameters in the high-resolution Sentinel-3 Level-2 Synergy product" by Antti Lipponen et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-262-RC1, 2021

This paper provides a bias-correction to the Sentinel 3 synergy aerosol data. The method is based on a previously (and recently) published approach by the authors, Lipponen et al (2021), though I feel there is enough difference here to warrant a new publication. The previous application was to MODIS aerosol retrievals; the current work is to a higher resolution product (300 m) for which (I believe) the primary purpose is atmospheric correction for land cover retrievals. Atmospheric correction-based aerosol products are traditionally worse than aerosol-focused aerosol products as in surface-focused cases the atmospheric parameters are often used as an error sink. So doing a bias-correction of them is useful in that it provides a finer resolution aerosol data set than typically available from atmosphere-focused products (which are typically more spatially aggregated in the level 1 to level 2 stage). The bias correction is done using machine learning; a comparison is also made to a fully learned (i.e. level 1 to level 2) machine learning approach. Correcting for the "approximation error" (retrieval error) is expected to be better than a fully-learned approach as the former gains some benefit from retrieval skill (there is less to learn) and that is reasonably borne out by the results presented here. Having a finer spatial resolution is beneficial for eventual air quality applications.

The paper is well-written and in scope for the journal. I believe this, combined with Lipponen et al (2021), provide enough evidence that the technique is in principle generalizable. This is important as it implies a fast bias correction could be done for many data sets, which is better for most downstream applications. The authors mention air quality though this is also important for data assimilation which ideally needs unbiased inputs with understood uncertainty characteristics. I don't have any major concerns with what is presented here, and so recommend publication following minor revisions. I would be happy to review the revision if the Editor wishes. I applaud the authors for noting that the code will be available, as this can help speed uptake and transparency is in general a good thing. Hopefully it will be available by the time the final version of this paper is published such that it can be linked directly. I also downloaded the animation linked in the paper and confirm that works and is useful. I will note I am not a machine learning expert so have not commented on the details of that; I recommend at least one reviewer should have machine learning expertise in order to judge that aspect.

My specific comments are as follows:

- Throughout, the authors cite Lipponen et al (2020) for their prior work; this appears to be the preprint of the final Lipponen et al (2021) paper describing this technique applied to MODIS. I assume this is an oversight but it should be corrected.

- Section 3.1: The authors link to the ESA website to describe the Sentinel 3 source retrievals. Are there no publications or tech documents that can be cited here? The linked page is not informative (it's basically to a catalogue of products, no ATBD or validation report etc). I would like to know a bit about the general Sentinel 3 synergy algorithm approach, e.g. how the SLSTR and OLCI measurements (with different pixel sizes) are used and combined, what the main assumptions are (it looks from Figure 5 like a fixed value of AE is used, for example). The Conclusions notes that the standard synergy aerosol product is at 4.5 km but it is not clear to the reader why, especially if this is primarily an atmospheric correction algorithm which is normally done at fine resolution, and the SLSTR data are 1 km or finer – do they do the atmospheric correction at coarser scale than the surface retrieval? Or are the "land" and "aerosol" synergy products entirely separate? I know this is not the authors' algorithm but presumably the synergy product is not a mystery black box (someone somewhere knows what the algorithm is) and as an ESA product this information should be available to the public somewhere such that a summary can be given here. If not then please point this comment to the responsible ESA official because there really needs to be some documentation for a data product if it is put out to the public. It is all frustratingly opaque and, after clicking around the ESA site for some time, I was unable to satisfy my curiosity.

- Section 3.2: I believe the preferred citation for AERONET version 3 direct Sun is Giles et al (2019): https://amt.copernicus.org/articles/12/169/2019/ This should be given in addition to or instead of Holben et al (1998).

- Section 3.3: elsewhere in the paper the authors (rightly) note that some previous machine learning studies give an artificially high impression of performance by not having independent training and validation data sets. In this section the authors note that they split training/validation data by station, which is better than splitting individual observations within individual stations. I agree with this. However, it does seem a bit of a missed opportunity not to test the approach on something fully outside of the selected regions of interest, and more fully independent from the training set. Figure 2, for example, reveals many sites (individual or clustered) outside of these regions. I suggest the authors extend their validation to a few of these "untrained" sites or regions to see what the benefit of the networks is there – this will provide more evidence for how applicable the model is on a global scale with limited regional training. I know data volume is limited considering only 1 year of data but hopefully we can say something at least. I would suggest looking at sites in Amazonia (contrast between clear seasons and heavy biomass burning, in a somewhat cloudy environment), Korea/Japan (mixed aerosol types, good AERONET site density), and/or Australia (traditionally a difficult area for aerosol retrievals). The paper is not too long and I think adding this would add substantial further interest to the reader without making the length excessive. The Korea example dovetails well with my final point below.

- Figure 8: It is ok to have the scale different for each row because each region is quite different. But I think the scale for each panel in a given row should be made the same, for more direct comparability. I acknowledge that the scale is quite different because the uncorrected synergy product is a lot higher than the others, but if a logarithmic scale were used (as in figure 7) I feel the plots would be better without loss of contrast within and between them.

- Conclusions: the authors' fully-learned and bias-corrected approaches clearly work better than the standard synergy aerosol data at the AERONET sites. The regional maps also look more reasonable. But there is an uncertain middle ground on the scales of a few to tens of km. It's hard to know whether some of the fine structure in these maps is real variation, statistical noise, or surface-related artefacts. For example, returning again to Figure 8, there are AOD hotspots corresponding to the built-up locations. With only 1 or 2 AERONET sites in each area, how are we to know if this spatial structure is real? This is not a problem the authors can fully solve but it is something that should be acknowledged. I know there have been some regionally-dense AERONET deployments (dozens of sites in a comparable region); Korus-AQ (summer 2016) was early in the Sentinel 3 era with 20 AERONET sites (Choi et al 2021: https://doi.org/10.1016/j.atmosenv.2021.118301), maybe that could be looked at (here or elsewhere). There is also a network of shadowband radiometers providing aerosol properties distributed around the Southern Great Plains ARM site region in the USA which I believe were operational during 2019 (https://www.arm.gov/capabilities/instruments/mfrsr). For me this "variability on tens of km" scales is the key next step we need to solve as we move toward better fine-scale aerosol retrievals. In addition to expanding the text to draw more attention to this issue (which may attract further studies/funding on the problem) I encourage the authors to expand the paper by looking at one or both of these areas, if data are available, to take a first step.