

Atmos. Chem. Phys. Discuss., referee comment RC2
<https://doi.org/10.5194/acp-2022-743-RC2>, 2023
© Author(s) 2023. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on acp-2022-743

Andrew Sayer (Referee)

Referee comment on "Ozone Monitoring Instrument (OMI) UV aerosol index data analysis over the Arctic region for future data assimilation and climate forcing applications" by Blake T. Sorenson et al., Atmos. Chem. Phys. Discuss.,
<https://doi.org/10.5194/acp-2022-743-RC2>, 2023

Summary and recommendation

I am writing this review under my own name (Andrew Sayer) as I know the authors from previous collaborations. We don't have any funding, common projects, or recent papers in common, and I don't believe there is any conflict of interest here. Just being transparent.

The authors use the OMI UV absorbing aerosol index (AI) data product to look at trends in Arctic aerosols during daylight months over the period 2005-2020. The motivation for the choice of OMI is that the AI provides semi-quantitative estimates of aerosol burden over this surface while standard imager-based retrievals do not (due to high cloudiness and/or surface snow cover) and the CALIOP lidar has difficulties quantifying the weak background aerosol loading commonly found.

One challenge working with OMI is the row anomaly, whereby certain pixel rows are not quantitatively useful due to a sensor issue that, in general, gets worse through time. There are row anomaly flags in the standard product but they have a few limitations. The authors deal with this through additional manual filtering of the OMI time series based on across-track AI variation. A second challenge is that AI is not the same as e.g. aerosol optical depth (AOD): it is a semi-quantitative parameter dependent on aerosol amount, absorption, altitude, solar/view geometry, presence of clouds, absorbing gases, surface type, surface pressure, etc. The Arctic is a particularly challenging environment as extrema of the above can be encountered. These features lead to artefacts from the point of view of doing a climatology/time series analysis because they can be systematic (though they are arguably not really artefacts – AI is not a measure of aerosol loading – they are a feature a way of the definition of the AI forward calculation, which is explicitly designed not to account for such things, and why there are e.g. separate OMI UV AOD, etc products). Therefore as a second step the authors bin the data from applicable OMI rows as a function of various geometric and surface parameters to generate a time series of AI perturbations. This perturbation time series is used to look at trends in aerosols over the Arctic, and additionally compared to the source (unscreened and screened) OMI data. This 'cleaned' data set shows generally smaller AI and trends from the raw OMI AI, which are analysed and discussed.

The topic is important and relevant to the journal. The quality of writing and presentation is high. The authors provide their processed OMI data in a supplement in NetCDF format,

which is appreciated (I checked and the files seem as described; I did not attempt to reproduce the detailed calculations). References are appropriate, though I feel more should have been added to note that the AI is not really a retrieval of aerosol properties but a forward calculation of an aerosol effect that is nonlinearly related to a large number of aerosol and surface parameters. This is more an issue with the framing of the analysis than with the analysis itself.

As a result, my recommendation is for minor revisions. I would be happy to review the revision if the Editor wishes. My comments focus on the data preparation and analysis as this is my main interest (so I recommend involving at least one other reviewer who is familiar with Arctic aerosols and their trends more generally).

Specific comments

- As noted above, I think the framing of the discussion of the AI product itself is a bit wrong. Is it not the same sort of thing as an AOD retrieval and doesn't pretend to be. It's a semi-quantitative measure of the perturbation to UV reflectance coming (mostly) from absorbing aerosols. I think the analysis the authors have done here to transform it into something that can be looked at for trends is a good one. But I think the initial discussions of the AI product might make an unfamiliar reader feel like OMI AI is a bad data set that's full of artefacts. That's not the case: it's just if you want to use it in a meaningful way for quantitative climatology and trend analysis, you have to take all these extra steps, to account for these geometric/surface, etc dependencies baked in. I think this could be better articulated in the early part of the paper. See for example Torres et al (1998), Hsu et al (1999) which discussed these issues (talking about both aerosol index and AOD). The authors write "semi-quantitative" in the Abstract, which is at least something, but I think this needs to be given more space in the paper itself.
- The trend analysis (section 4.3) is done in a common way: do ordinary least squares linear regression on the time series of perturbations, and do a T-test to identify grid points where the p-value is below 0.05. The results are framed in terms of this linear AI perturbation gradient and the locations of low p-value (which correspond to points for which, if there is truly no trend in the time series, the chances of observing an apparent linear trend at least this large are lower than 5% - at least I believe this is the correct interpretation). This is a common way of doing things but has a few issues which should be acknowledged. One is that my doing these tests pointwise on a map we are not doing single hypothesis testing but rather multiple; further, since the source data are highly spatially correlated, 'noise' in the fit can be correlated as well, leading to blobs of apparent significant trend which may or may not be real but look realistic because they are spatially coherent. Wilks (2016) has an important discussion of this and some suggestions (references therein) to use a dynamic p-value to control on the false discovery rate instead. Another approach (which I personally prefer) is not to focus on significance but rather look at estimated trends and uncertainties on those estimates (which should be provided by whatever linear regression routine is used). One reason is because 'insignificant' is not one thing: if you have an 'insignificant' trend with a low uncertainty on the trend estimate you can fairly confidently rule out there being a large trend; if you have an 'insignificant' trend with a high uncertainty on the trend estimate then any true trend might be large or small (and we might not know the sign). It is not clear from the analysis done how much of the 'insignificant' trend areas might fall into each sub-category. I suggest the authors try looking at maps of AI trend and AI trend uncertainty and see if they can make some assessment of this (it doesn't need to be shown in the paper, just some statements of what is the typical level of

precision on the trend estimates in various cases and therefore where we can/can't rule out some missed important trend).

- A further issue that I think should be mentioned is that trends on a time series of monthly mean perturbations might not make sense if trends are driven by changes in the number of extrema rather than the baseline AI (since we know aerosol distributions tend to be skewed with a long tail). Plus, it's not clear that a linear model is appropriate for the same reason. This ties into the above as significance testing and uncertainties are predicated on the assumed model. I do appreciate that the analysis was done separately for each month (since trends can differ between months). Plus, the authors do not infer too much from the quantitative AI trends – more when and where they are happening – which alleviates those quantitative concerns a bit. But, the fact that choice of model for trend construction is important should be acknowledged.
- The acronym QC should be defined at first use (I know what it means but some readers might not).
- I was a bit surprised there was no mention of e.g. TropOMI here. Not expecting it to be included in the analysis given it was launched in 2017 but it could be useful to point to its advantages for this type of work over OMI (e.g. spatial resolution, no row anomaly) for the future. Likewise OMI's advantages over TOMS, etc (again spatial resolution) could be mentioned. I don't know that much discussion is needed but a mention wouldn't be amiss. Not sure the geo spectrometers need mentioning, though, since they won't observe the Arctic.

References

- Hsu, N. C., Herman, J. R., Torres, O., Holben, B. N., Tanre, D., Eck, T. F., Smirnov, A., Chatenet, B., and Lavenu, F. (1999), Comparisons of the TOMS aerosol index with Sun-photometer aerosol optical thickness: Results and applications, *Geophys. Res.*, 104(D6), 6269– 6279, doi:10.1029/1998JD200086.
- Torres, O., Bhartia, P. K., Herman, J. R., Ahmad, Z., and Gleason, J. (1998), Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation: Theoretical basis, *Geophys. Res.*, 103(D14), 17099– 17110, doi:10.1029/98JD00900.
- Wilks, D. S. (2016). "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It, *Bulletin of the American Meteorological Society*, 97(12), 2263-2273. Retrieved Jan 24, 2023, from <https://journals.ametsoc.org/view/journals/bams/97/12/bams-d-15-00267.1.xml>