Comment: *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.*

Response: Thank you for the suggestion. We have added text to the early portions of the paper to reflect this.

Comment: *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). On 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).*

Response: Thank you for the suggestion. We followed your suggestion and compared maps of standard error of the trend (slope) to the trend maps (included as a supplement). We have added appropriate discussion in Sections 4.3 and 5.1, and have included the following figure, which compares the trends and trend standard errors, as a supplement to the paper.

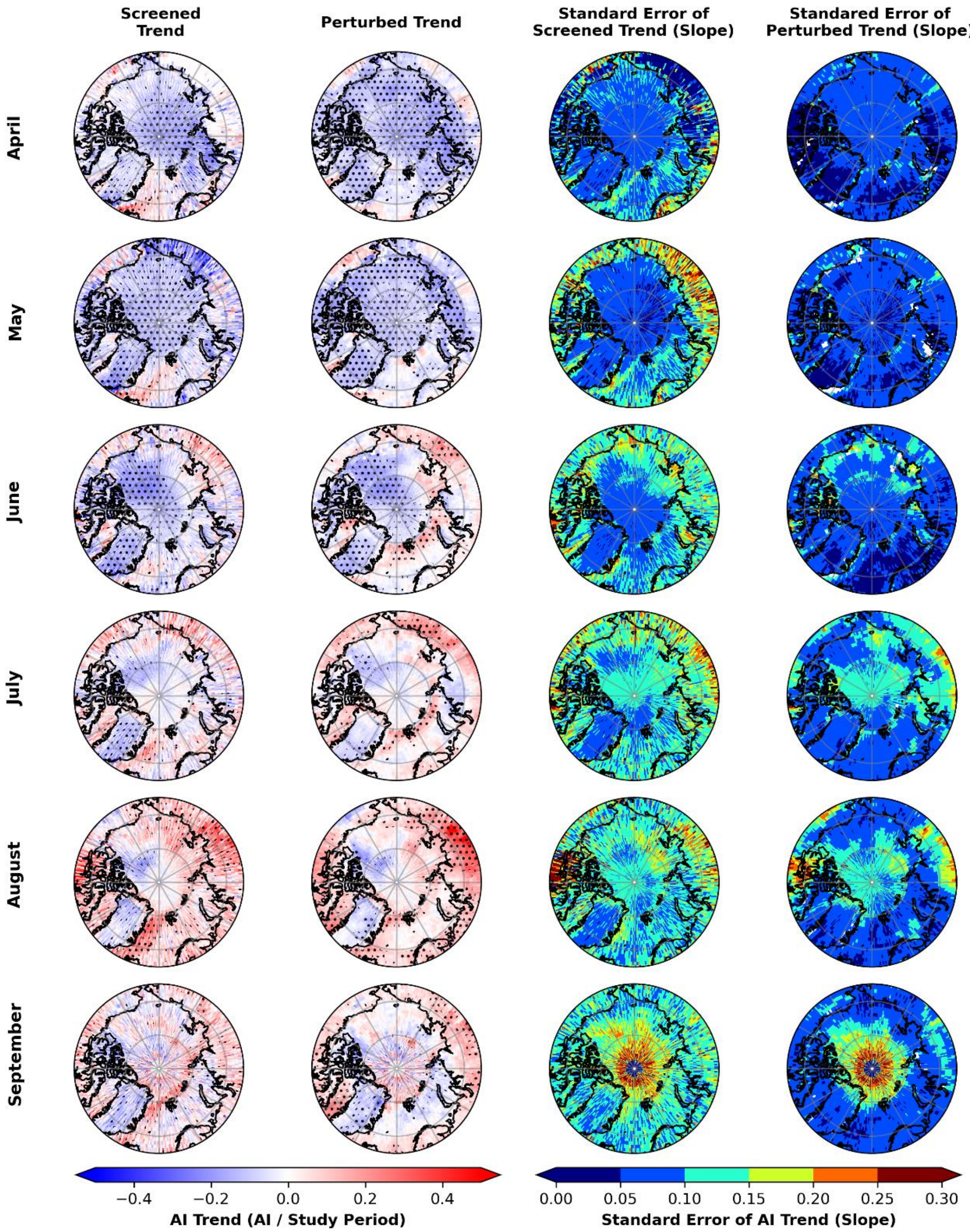

Figure S1: April (first and top row), May (second row), June (third row), July (fourth row), August (fifth row), and September (sixth and bottom row) monthly trend of screened OMI AI data (first column) and perturbed AI data (second column), as well as the standard error of the screened (third column) and perturbed (fourth column) AI trends. Trends are calculated between 2005 and 2020. The dotted regions in the left two columns denote trends that are statistically significant at the 95% confidence level.

Comment: *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.*

Response: Thank you for the suggestion. We added discussion emphasizing the importance of method choice for trend analysis to the text. We also mentioned that if trends are driven by a few extreme events, the trend may not make sense.

Comment: *The acronym QC should be defined at first use (I know what it means but some readers might not).*

Response: Thank you for the note. We have added a definition of the QC acronym.

Comment: *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.*

Response: Thank you for the suggestion. We have added discussion to the introduction about OMI's spatial resolution advantage over TOMS, and about TROPOMI's benefits over OMI for future Arctic aerosol analyses in the conclusion section. We have elected to leave the geo spectrometers out of the discussion, since, as you mentioned, they don't observe the Arctic.