Atmos. Chem. Phys. Discuss., 14, C8274–C8281, 2014
www.atmos-chem-phys-discuss.net/14/C8274/2014/

Atmospheric Chemistry and Physics Discussions
Open Access

ACPD
14, C8274–C8281, 2014

Interactive Comment

Interactive comment on "Study of satellite retrieved aerosol optical depth spatial resolution effect on particulate matter concentration prediction" *by* J. Strandgren et al.

Anonymous Referee #1

Received and published: 20 October 2014

This paper is a correlation analysis between aerosol optical depth (AOD) retrieved using the MAIAC algorithm applied to MODIS measurements and ground-based $PM_{2.5}$ data, over the contiguous USA. After reading through several times, I am sorry to have to recommend rejection of this paper.

From a statistical point of view the analysis is unsound. From a scientific point of view I feel that there is nothing new here, and no clear scientific result of use to the broader community. AOD/PM correlation analyses have been published several times in the past (such as the two Chudnovsky et al. papers cited, which cover much of the same

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

ground), and it is well-established that the two are related but have a lot of scatter for a variety of reasons. We are already at the point where it has been established that more advanced methods (accounting for e.g. humidity and aerosol vertical profiles) can be used to give better predictive power for AOD/PM relationships. Other groups have looked into this for several years, see e.g. work by van Donkelaar (which the authors do not cite) as one prominent example. By comparison the current research is very simplistic: a correlation and (inappropriate) least-squares linear regression, with some plots of meteorological parameters added and discussed in a qualitative manner. I see no real novelty or potential application for the results in this paper, and the authors do not really highlight any themselves (beyond saying it is interesting to see how correlations change with averaging size). The results are also likely to be algorithm-specific, and may change as the algorithm evolves (since it is an analysis on an in-development dataset). Additionally, the quality of writing is poor, some statements are not substantiated with evidence, and it is unclear in some cases what exactly was done.

Below are some technical comments in support of this recommendation. After thorough consideration I do not feel that revisions can bring this paper to a standard worthy of publication, because even if it were rewritten to fix methodological issues, the lack of scientific novelty/utility would remain.

P25871, line 26 and onwards: Saying both 'size distribution' and 'effective radius' here is redundant because the effective radius is a weighted average of the size distribution. An obvious omission from this list is particle density (to relate volume to mass).

P25872, line 5: With this phrasing, are the authors really suggesting that nothing new has been done in this subject area since 2009? A Google or Web of Science search will show this is false (see e.g. above comments that we are able to go beyond the

ACPD
14, C8274–C8281, 2014

point of simple correlation analysis).

P25872, lines 12-19: The list of references here seems somewhat self-serving. It acts as a vehicle to boost the citation count for the authors' own published algorithms, which are (to my knowledge) not in routine processing or available to the broader scientific community, while ignoring some publicly-available well-used datasets (e.g. NASA MISR, OMI, SeaWiFS to name but three). Further, the references are largely unnecessary since the specifics of the datasets mentioned are not discussed at all in the paper. If the authors want to make a tangential comment about the fact there are many ways to retrieve AOD from space then citing a relevant review paper or two would suffice.

P25873, line 6: I take issue with this statement. It is true that higher AOD spatial resolution does not necessarily mean a better result. But I do not believe that higher AOD resolution means a worse result, which is what the authors' statement here implies ('cannot be expected to be as good'). Clearly the optimal resolution is dependent on context and application. This statement is unsubstantiated and may mislead a nonexpert reader.

P25874, lines 23-25: The MODIS FMF was found to have very little quantitative skill over land some years ago, and should be treated only as a diagnostic parameter about the retrieval solution, i.e. should not be used for studies like this. This is discussed in the Levy et al. (2013) reference the authors cite here (perhaps they missed that), as well as Levy et al. (ACP, 2010). The authors attempt to interpret it quantitatively (e.g. P25884, 'the amount of coarse particles is considerably higher'), which is ill-advised. It is also statistically dubious to average a fraction in this way, especially over multiple seasons, where aerosol regimes may change. This reinforces my impression of the authors not having a good understanding of this dataset.

ACPD
14, C8274–C8281, 2014

P25875, line 8: Why 8×8 degrees for meso-scale? Does this correspond to some expected spatial scale of variability, or the size used for some forecast/assimilation model, or not? The authors may be interested in T. L. Anderson et al. ('Mesoscale Variations of Tropospheric Aerosols', JAS, 2003) as a useful reference about aerosol spatiotemporal variability. Also, what is the point of doing a correlation against the whole contiguous US? Do the authors really believe this has any practical value? I don't think that the questions which the authors are attempting to answer in this study are ones that the PM community is asking.

P25875, section 2.2: I am not certain that this outlier removal is justified. Clearly there is still a huge amount of scatter even after outliers are removed. The justification for outlier removal seems to be taking away cases where the AOD might not good a give representation. But isn't one of the points of the analysis to see how representative the AOD is for these cities? Also, by using PM to identify and remove AOD outliers, the analysis becomes unrepresentative of any potential predictive application, because in a predictive sense the PM data would not be available to do this screening.

P25876, section 2.3: This section has methodological problems. Linear least-squares regression is not appropriate for analyses of these type (just because people do it sometimes, does not mean it is right). See for example the Wikipedia page on the topic: http://en.wikipedia.org/wiki/Ordinary_squares.There are several reasons for this:

1. The errors on the AOD data are not Gaussian. This is because AOD cannot be negative, so in low-AOD areas the low tail on the AOD error distribution is truncated. In high-AOD cases we don't know if they are Gaussian or not for this algorithm.

ACPD
14, C8274–C8281, 2014

2. The points may not be independent in all cases. This in part depends on how the analysis was done (which is not clear). For the 'urban' scale were AOD points averaged all across the city and PM between sites, or was each retrieval matched to a PM site? Looking at Figure 2, for some cities it looks like there are PM monitors within 10 km of each other. This means that, at least for some spatial scales, depending on how the analysis was done, some AOD pixels may have counted against multiple PM points. Again, it's not clear because it is not clear how the analysis was done.

3. The effect of error in the datasets is not accounted for by this regression technique. It is known that generally AOD retrieval uncertainty is AOD-dependent (the specifics of MAIAC AOD retrieval error are not discussed quantitatively in this study so it is hard to say how true it is for this case). The regression technique the authors uses treats each point equally. Additionally, there is no mention I could find in the manuscript about the uncertainty on the PM datasets. Is this negligible or not, both in terms of measurement error and sampling error (i.e. what is the variability around the daily mean)? Again, the technique the authors use assumes zero error on the PM data.

4. Because AOD is distributed roughly lognormally (see work by e.g. O'Neill and others), much of the data is in the low-AOD regime and the number of high-AOD points (which will strongly affect the slope) is limited. It is therefore likely that sampling effects from a small number of extreme effects is driving these. For example the high-AOD events may come from e.g. smoke plumes which have a different aerosol composition and vertical profile to the background places.

5. Following from the above, these extreme events are not really part of the same underlying population as the bulk of the data. So the calculated correlation and slope are really not representative of the ability to track typical variations, more about how well the unusual high values are spotted. Related to this the grouping

of seasons together for parts of the analysis may be making correlations different than they would be for any real day-to-day predictive use of this type of relationship, because seasonal variability in aerosol and meteorological conditions becomes conflated.

The lack of quantitative discussion of MAIAC AOD uncertainty in the later analysis of the results is also problematic, because it limits the interpretation for some of the scatter in the comparisons beyond hand-waving justifications. All the discussion is qualitative. Presumably if the authors have four months of MAIAC data then they could easily do a validation against e.g. AERONET. It may be that the strongest driver for regional variability in correlation is in fact related to regional errors in MAIAC rather than meteorological factors: the authors present no evidence either way, and without a quantitative evaluation of the AOD, we cannot tell.

This section also takes some space to give the definition of correlation coefficient, which is basic stuff (is it really necessary?) and I am left wondering if the authors fully understand it or have just written out a definition. For example, the authors talk about the 'null hypothesis' but don't actually state what this is (which I assume is that there is no correlation between AOD and PM), or what their 'alternative hypothesis' is. The p value measures confidence in decisions about this hypothesis, nothing about the strength of it (statistical significance does not necessarily equate to scientific importance because p alone does not tell us about the magnitude of an effect). In any case it looks like the p value is used only to judge whether these correlations are significantly different from zero, which they are, and it isn't really needed at all in this analysis because the sample sizes are quite large and the positive correlation between AOD and PM is well-established. Again, this points to a poor understanding of the technique the authors are using.

ACPD
14, C8274–C8281, 2014

I also wonder if the focus on correlation as the analysis metric is appropriate. This is something which is not even discussed in the manuscript. It depends on what one wants to do with this type of dataset. For some applications (e.g. hazard warning) I would imagine detection of extreme events is the goal, but for other applications (e.g. determining compliance within some PM regulation threshold) I would imagine that the bias and RMS error would be more useful. The analysis is shallow, as well as outdated compared to more recent work which e.g. attempts to improve the predictive power of relationships by correcting for humidity and vertical profile.

P25878, lines 9-11: 'poor' and 'good' are weasel words. They are subjective, and need to be defined relative to some requirement.

P25878, lines 14-15: The authors claim that correlation changes are 'not significant', but they do not provide any statistical tests to determine whether, in fact, the correlations are statistically different from one another. There are methods to estimate this. I suspect want they want to say is that the increases are 'not large' but again going to the previous two points this is subjective and application-dependent.

P25878, lines 26-28: Averaging a MAIAC retrieval to 10 km is not the equivalent of looking at a 10 km MODIS standard Dark Target retrieval, so this claim about contamination at 10 km is not justified. The Dark Target dataset filters data at the radiance level, then averages and retrieves, to remove clouds (and other unsuitable pixels). Averaging retrievals to coarser resolution, as is done here, is the other way around. They might be equivalent in some circumstances, but not necessarily. This is another example of an unsubstantiated statement made by the authors.

In the rest of the analysis section the discussion is again often shallow and just refers

ACPD
14, C8274–C8281, 2014

back to other similar studies to note the similarity of results, again highlighting that nothing really new is done in this study.

P25880, line 26 and onwards: This is an example of where the authors do use quantitative correlation changes and perhaps overinterpret them, given there is no statistical test on whether these values are statistically distinguishable from each other given this sample size, and again there's little discussion of any real-world relevance of this result.

Interactive comment on Atmos. Chem. Phys. Discuss., 14, 25869, 2014.