

Thanks for all the valuable comments, the replies are listed below

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.

Response: The main purpose with this paper is to analyze the spatial resolution effect on the relationship between satellite-derived AOD with different spatial resolutions and ground-based PM_{2.5} measurements for different scales of coverage (urban scale, meso-scale and continental scale), in order to provide a better understanding to the atmospheric science community. The paper does not aim to present a method for the retrieval of PM_{2.5} or an improved correlation between AOD and PM_{2.5}, but to answer the two questions listed below: (1) For certain regions (not only a limited region over Boston like presented by Chudnovsky et al. (2013)), how the spatial resolution of satellite-derived AOD can affect the relationship between AOD and PM_{2.5}. (2) How the spatial scale of the study region can affect the relationship between AOD and PM_{2.5}, in another words, how representative is an empirical relationship between AOD and PM_{2.5} derived at urban scale for meso- and continental scale and vice versa? The reason why we introduce other factors like BLH, RH and FMF is mainly to obtain the characteristics of the study regions. For a limited region like presented by Chudnovsky et al. (2013), those factors may be stable, which is not the case if the study region increases from urban scale to meso-scale and continental scale. Some more qualitative analysis related to the RH and BLH

corrections can be presented as follows using the AERONET (AOD), EPA (PM_{2.5} and RH) observation datasets together with BLH data from ECMWF and the MAIAC 1km AOD data. According to Fig.1 and Fig.2, the correlation between AOD and PM_{2.5} increases when the two types of RH and BLH corrections are applied, both when the contribution from wet particles are removed from the AOD and added to the PM_{2.5} concentration respectively (Fig. 1), as well as when the AOD is scaled to a surface AOD value and when the PM_{2.5} concentration is scaled by the BLH to represent the PM_{2.5} concentration within the boundary layer (Fig. 2). We can also see that AERONET provides better correlation with PM_{2.5} compared with MAIAC, which further proves the importance of understanding the spatial resolution/scale problem related to the coupling between AOD and PM_{2.5}. This case study is done for three nearby ground stations in Maryland, which provide AOD (AERONET), PM_{2.5} (EPA) and RH (EPA) respectively, together with the MAIAC AOD data at 1 km and ECMWF BLH data at 0.25 degrees.

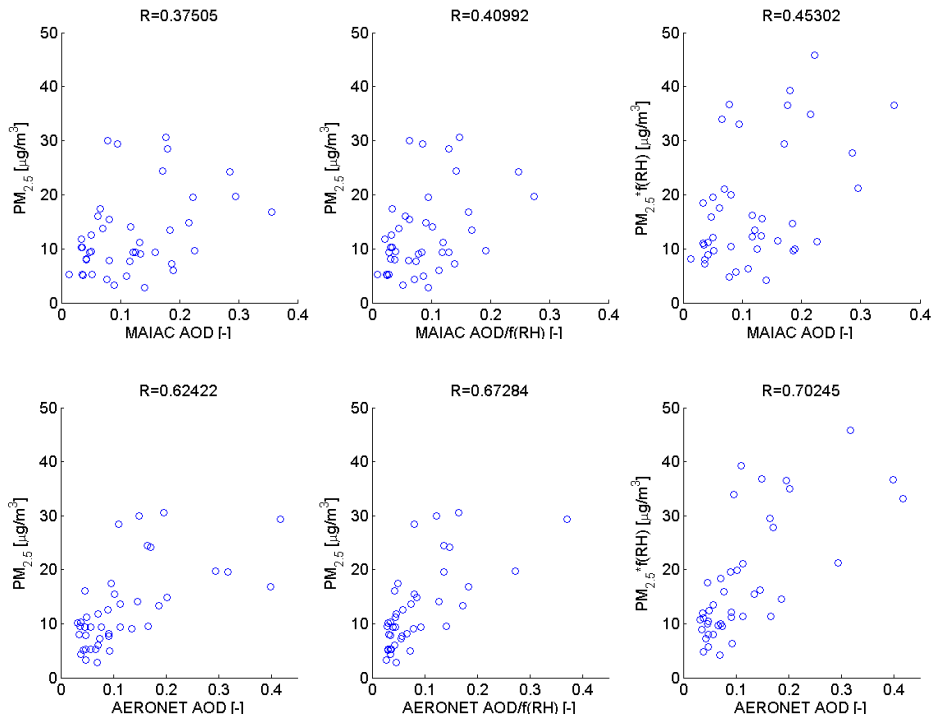


Fig.1 The correlation between MAIAC/AERONET AOD and $PM_{2.5}$ with/without the RH corrections.

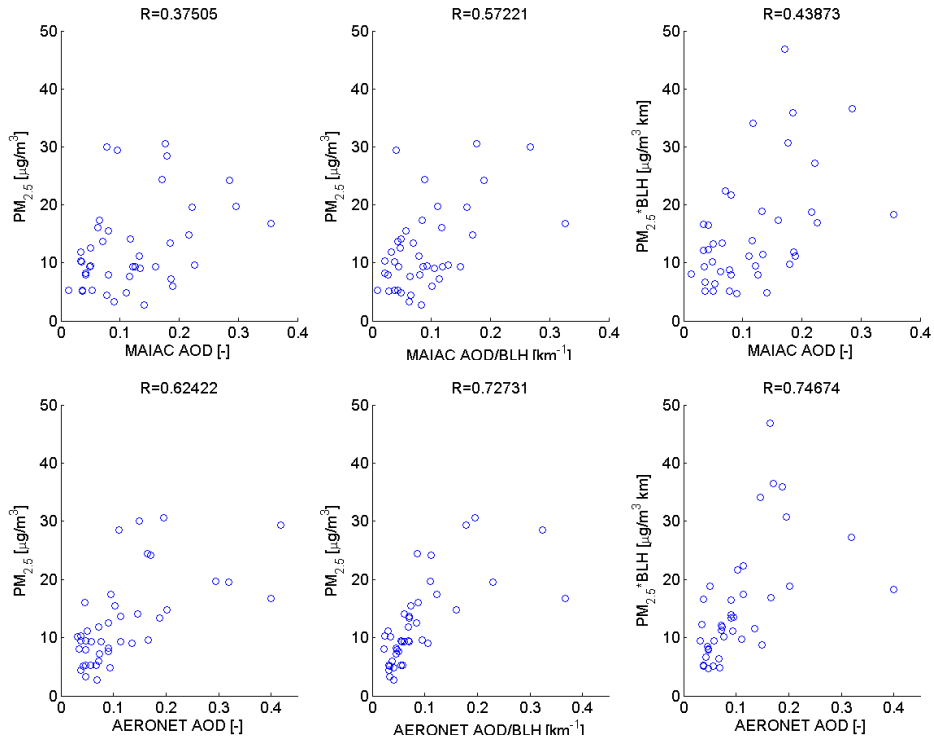


Fig.2 The correlation between MAIAC/AERONET AOD and $PM_{2.5}$ with/without the BLH corrections.

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 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.

Response: This work is an extension of the work by Chudnovsky et al. (2013) and actually the authors have discussed with Dr. Chudnovsky during the writing of this paper. The main difference between this paper and the work from Chudnovsky et al. (2013) is that

this paper tries to extend the understanding of the AOD spatial resolution effect on the relationship between AOD and $PM_{2.5}$ from limited regions (like for instance Boston) to different regions inside the US with both different surface characteristics and atmospheric conditions and different scales of coverage (urban scale, meso-scale and continental scale).

The work presented by van Donkelaar et al. (2006, 2011, 2013 and 2014) together with Liu (2005, 2013) presented the two possibilities of a large coverage retrieval based on mode or statistical methods, all the work related to AOD retrieval and $PM_{2.5}$ retrieval will be updated in the revised version, however, as we mentioned before, this paper does not aim to present anything related to $PM_{2.5}$ retrieval or how any kind of correction can improve the relationship between AOD and $PM_{2.5}$, but how the AOD spatial resolution affects the relationship between these two parameters over different regions and scales of coverage.

Thanks for the comments about the aerosol vertical profile effect to the prediction of $PM_{2.5}$, this problem has been discussed in detail by Liu et al. (2011) and the coarse horizontal resolution GEOS-Chem simulated aerosol profile has been proved to be very useful and van Donkelaar et al. (2013) have tried to adjust the shape of the GEOS-Chem aerosol vertical profile using the CALIOP product for a better $PM_{2.5}$ estimation. Lary et al. (2014) proposed a machine-learning method for the prediction of $PM_{2.5}$ with the prior information of 8329 *in-situ* measurements. In another words, the retrieval accuracy is definitely related to the prior information, van Donkelaar et al. (2013) especially relies on the prior

information of aerosol vertical profile while Lary et al. (2014) emphasizes the “direct link” between AOD and $PM_{2.5}$ within the “black box” (machine training). All these effects have been added in the revised version using the AERONET and EPA *in-situ* measurements, before we test the AOD spatial resolution effect on $PM_{2.5}$ prediction as well as the representativeness of the linear empirical relationship between AOD and $PM_{2.5}$ obtained for one specific spatial scale (urban scale, meso-scale or continental scale).

Thanks for the suggestion of a validation of the MAIAC AOD product. The quality of MAIAC data has been evaluated by Lyapustin (2012) and the MAIAC dataset will be released soon, so we can trust the dataset. As suggested by the reviewer, we validated the MAIAC product for the time period (January, April, July and October 2008) that we used for the analysis in the paper, and the result is presented as in Fig.3. According to Fig. 3, the monthly mean difference between MAIAC and AERONET for the selected time is in nearly all cases inside ± 0.04 , except for the mid-western part of the US where the surface is bright and the problem with satellite AOD retrieval well known. This is accurate enough for our further analysis. More validation for the MAIAC AOD product can be found in Lyapustin et al. (2011), Arvani et al. (2013) and Lyapustin et al. (2013).

Some conclusions in the paper have been clarified according to the comments from the reviewer.

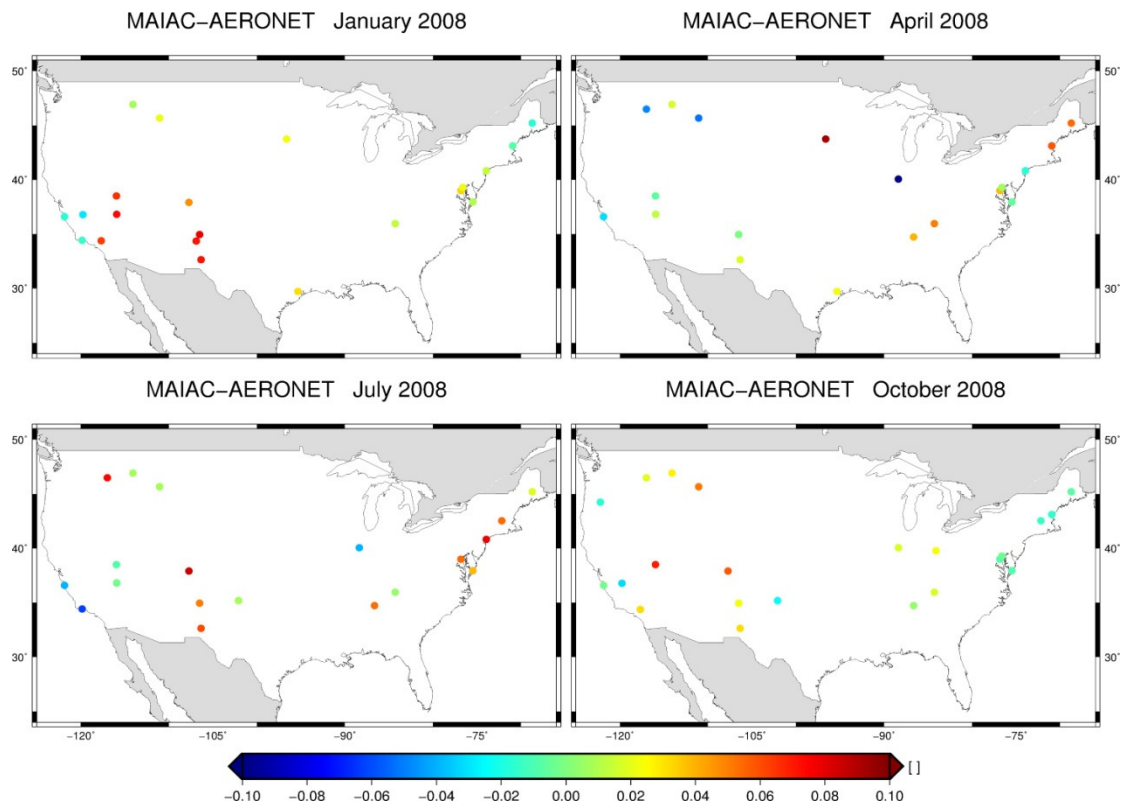


Fig. 3 The comparison between MAIAC and AERONET AOD for January, April, July and October 2008. Each point refers to the monthly mean difference between MAIAC and AERONET AOD.

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).

Response: Thanks for the suggestion; “particle density” is used in the revised version.

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 point of simple correlation analysis).

Response: Thanks for the suggestion. The authors have included a comprehensive review of both PM_{2.5} retrieval and AOD retrieval in the revised version. AOD retrieval algorithms can be divided into (1) Multi-spectrum method like MODIS Dark-Target (Kaufman et al., 1997; Remer et al., 2002; Levy et al., 2007; Levy et al., 2013), DeepBlue (Hsu et al., 2004, 2006, 2013; Sayer et al., 2014), MAIAC (Lypustin et al., 2012); (2) Dual/Multi viewing method like AATSR (Holzer-Popp et al., 2013; de Leeuw et al., 2013) and MISR (Kahn et al., 2005); (3) High-temporal retrieval method for geostationary satellites like MSG/SEVIRI (Govaerts et al., 2010; Wagner et al., 2010); (4) Polarization method for e.g. Parasol (Dubovik et al., 2011).

As to PM_{2.5} retrieval, the algorithms can be divided into at least four groups (1) Two variable (AOD and PM_{2.5}) empirical algorithms (Wang et al., 2010; Wang et al., 2013); (2) Multiple variable algorithms (besides AOD and PM_{2.5}, other parameters like meteorological parameters, local emission parameters and so on) (Liu et al., 2005, 2007, 2013); (3) Machine learning retrieval algorithms (Guo et al., 2009; Lary et al., 2014); (4) Mode-based algorithms like van Donkelaar et al. (2006, 2011, 2013 and 2014). And the

method proposed by van Donkelaar et al. (2014) is now mature enough for trend analysis (Boys et al., 2014).

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.

Response: The introduction about AOD retrieval is updated in the new version as mentioned above.

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.

Response: Thanks for the suggestion related to the AOD spatial resolution effect, which is the essential problem that we would like to investigate. We definitely agree that the spatial

resolution is related to the application. For the effect on $PM_{2.5}$ prediction, there is no doubt that a higher spatial resolution is urgently needed, especially for the health issues caused by $PM_{2.5}$, which was firstly presented in IPCC5. One of the main objectives of MODIS 3km AOD product is to provide the atmospheric science community a chance of obtaining high-spatial resolution $PM_{2.5}$. Since of course, for $PM_{2.5}$ prediction, we need higher spatial resolution. From the remote sensing point of view, 1km is preferred. However, due to the ill-posed problem, aerosol algorithms always decrease the spatial resolution in order to provide higher Signal-Noise-Ratio (SNR). MAIAC is the only operational algorithm providing 1km currently (DeepBlue 1km product is under development). There is no doubt that 10km AOD product will provide better SNR compared to 1km. However, 1km AOD product provides a better chance of distinguishing the aerosol variability inside a super-pixel like 10km, but at the same time, introducing larger uncertainties. This is the key issue to discuss in this paper. So actually, the key issue is which AOD spatial resolution that is the optimal choice in order to find a good balance between high resolution (that it urgently needed for better $PM_{2.5}$ prediction) and low uncertainty of the AOD, in order to predict $PM_{2.5}$ with the highest accuracy.

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.

Response: Thanks for the suggestion. We have checked all SDS from C5, C5.1 and C6 and for sure some SDSs are less accurate (like Fine-Mode-Fraction (FMF), Angstrom coefficient, Aerosol types) compared with others SDS (like AOD), even for AOD, the accuracy is different from pixel to pixel, which is related to both surface and atmospheric conditions. The point is that FMF is the only available dataset which can be used to distinguish the fine mode aerosol proportion for a relative large area. The required accuracy of the dataset is of course related to the application and since we do not intend to give any quantitative conclusion like for instance that the correlation between AOD and $PM_{2.5}$ can increase by 10% if FMF increase by 10%, but instead use the FMF to qualitatively distinguish the coarse and fine particles, the accuracy of the MODIS FMF product is good enough. But some more clear explanation related to the FMF will be included in the revised version.

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.

Response: Thanks for the suggestions. Several AOD-PM_{2.5} studies over the US have been presented for the whole US, the eastern and western parts and for urban scales. The meso-scale of 8x8 degrees was chosen as an intermediate size of the study region, for which the understanding of the correlation between AOD and PM_{2.5} is less understood. Of course, different authors focus on different size for the meso-scale, like for instance Wang et al.(2006), Christopher et al. (2009), Chen et al. (2014) for different applications. In Christopher et al. (2009), they evaluate PM air quality quantitatively from satellites and ground-based monitors, near and far away from fire source regions, using a meso-scale study region of approximately 8x8 degrees.

PM_{2.5} is a parameter related to daily life, thus plenty of previous publications like Hu et al. (2014), Martinez et al. (2012), Davidson et al. (2007), Vinikoor-Imler et al. (2011), Lee (2014), Li (2014) show the PM retrieval from satellite and corresponding applications for different scales. And even for the PM_{2.5} concentration trend analysis, plenty of publications/reports including reports from the EPA, report the PM_{2.5}/PM₁₀ trend for different scales, like eastern us, western US and the whole US.

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 give a good 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.

Response: Thanks for the suggestion. Actually quite few points (less than 1%) have been removed, which has almost no effect to the analysis and conclusions presented in the paper. However, those outliers should be removed to make the further analysis more physically reasonable and robust. Actually, there is a physical assumption behind the removal of the outliers; we assume that in the linear correlation analysis like presented in the paper, the uncertainty/error from the satellite-derived AOD product can also be linearly propagated into the $PM_{2.5}$ concentration. The uncertainty of MODIS standard AOD product Collection 6.0 (Levy et al., 2013) can be expressed as $\pm(0.05+AOD \times 15\%)$, so the two boundaries for the acceptable AOD product related to the accuracy is $-0.05+0.85\tau \leq \tau \leq 0.05+1.15\tau$. Hence the uncertainty of $PM_{2.5}$ concentration (η) can be expressed as $A(-0.05+0.85\tau)+B \leq \eta \leq A(0.05+1.15\tau)+B$, so the final form of the uncertainty of $PM_{2.5}$ is $A'\tau+B' \leq \eta \leq A''\tau+B''$, which is finally related to the regression coefficient between AOD and $PM_{2.5}$ concentration. After testing plenty of daily-based atmospheric conditions (clear, moderate pollution, strong pollution), the outlier removal criteria given in the discussion version are proposed.

P25876, section 2.3: This section has methodological problems. Linear least-squares

regression is not appropriate for analyses of these types (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.

Response: Firstly, AOD cannot be negative in reality, but it can be negative in the retrieval (See detailed explanation in MODIS ATBD). We thank the reviewer for pointing out the problem of the statistical analysis, which is also highlighted by the other reviewer. So we would like to update the simple RMSE linear fitting by comparing with the York fitting (an overview of the usage of different regression methods can be found in Cantrell, 2008). Fig. 4 presents an example of 1km MAIAC AOD product with $PM_{2.5}$ concentration using the two fitting methods mentioned above. According to Fig. 4, we found that the slope and intercept for these two methods shows great difference due to the consideration of uncertainties from the input data. However, since the input data are the same, there is no change for the correlation for the input data. Some more detailed analysis will be contained in the updated version.

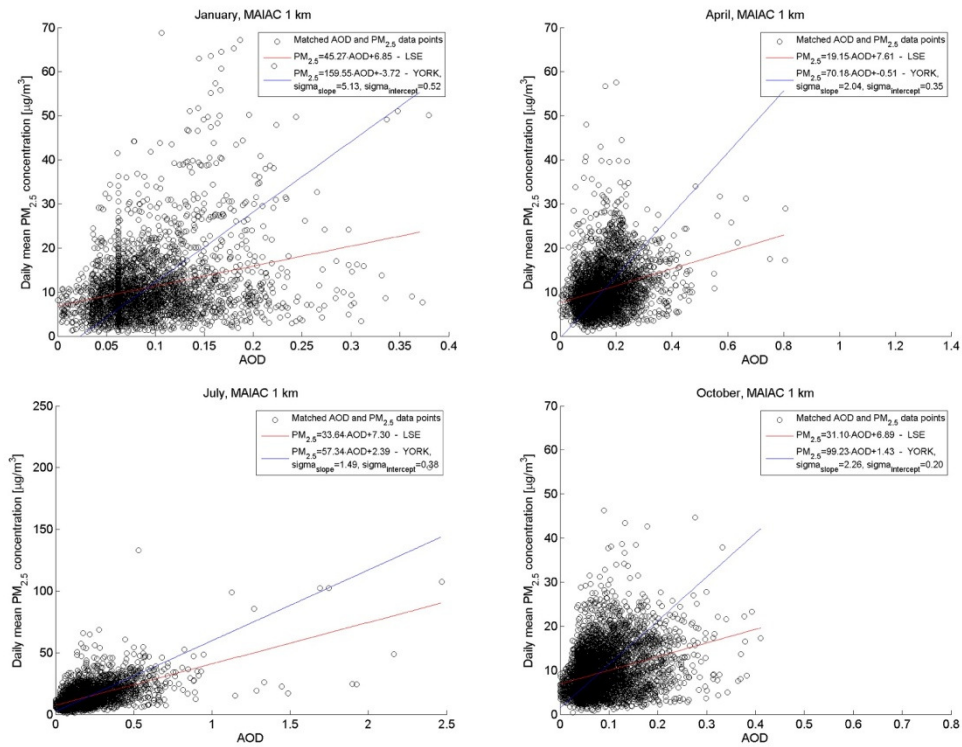


Fig. 4 The AOD and PM_{2.5} linear correlation using two fitting methods, one is the standard RMSE fitting (red), the other one is the York fitting (blue).

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.

Response: We do not average AOD for the match-ups. The match-ups are obtained as

follows: The location (longitude, latitude) of EPA PM_{2.5} monitor is used as the reference location, and then the AOD for the gridded satellite pixel covering the EPA PM_{2.5} monitor is retrieved. The same way for obtaining match-ups between satellite-derived AOD and EPA PM_{2.5} data has been used in previous papers like Chudnovsky et al., (2013). Except for the gridding of the AOD data, averaging is only performed when there are several PM_{2.5} monitors providing PM_{2.5} concentrations at the same location (same latitude and longitude).

Thanks for the nice point related to the match-up problem. The situation that is mentioned by the reviewer occurs due to the scale problem of “area”-represented satellite-derived AOD and “point”-represented *in-situ* PM_{2.5} measurements. There may be several PM_{2.5} monitors located in one satellite pixel if the spatial resolution is not high enough. That means that one AOD value from a given satellite pixel can be matched with several PM_{2.5} records from different PM_{2.5} monitors located within the 10x10km pixel. This also proves the importance of the objective of this paper. We would also like to mention that for bright surfaces the MAIAC algorithm provides a climatology AOD value of ca. 0.06 that is used for the atmospheric correction part of MAIAC. This also results in vertical lines in the scatter plots. Fig. 5 clarifies these two features.

Some more figures related to the situation above together with more detailed analysis will be included in the revised version.

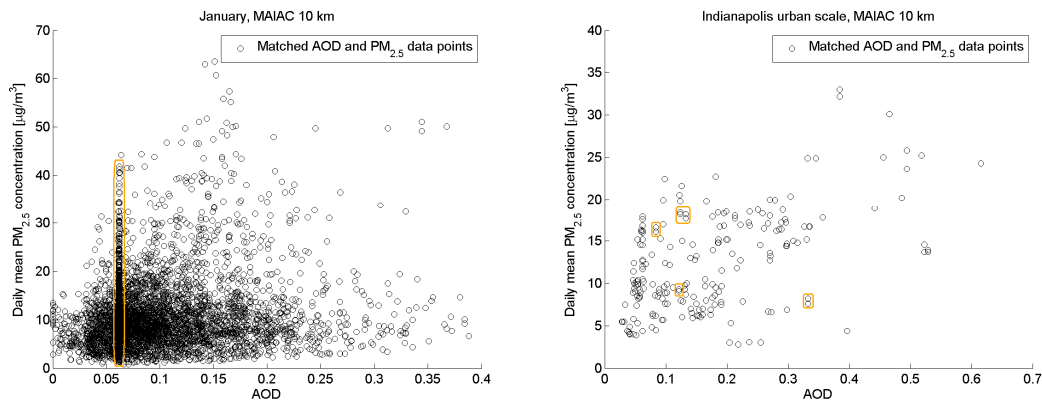


Fig. 5 The left plot shows the case where the MAIAC climatology AOD value is matched to several PM_{2.5} observations and the right plot shows some example cases where multiple PM monitors are covered by the same MAIAC AOD pixel.

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 use 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.

Response: All AOD retrieval, no matter if satellite or AERONET, is AOD-dependent. And this is the physical theory behind the removal of the outlines as presented above. The errors from the MAIAC dataset can produce different slopes and intercepts using a different regression method like the York regression as mentioned also by the other

reviewer. But this will not change the correlation between these two dataset. The comparison using simple linear regression and York regression is shown in the revised version.

Since in-situ measurements are the only reference for the satellite-derived product, it is reasonable to assume that PM monitors provide an un-biased PM dataset, which is enough for our analysis. Especially compared to a satellite-derived product, the error of ground-based measurements can be assumed to be negligible.

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.

Response: From a statistical point of view, the correlation between AOD and $PM_{2.5}$ should not be affected so much if the relationship is well-defined. As to the slope, it can be easily affected by some extreme cases. In the revised version, the York regression, which considers the uncertainty of MAIAC AOD, is shown to compare with the simple linear regression. And some quantitative analysis related to the slope change is presented.

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.

Response: We agree with the comments here. The statistical analysis cannot show very much related to extreme cases. As we mentioned in the introduction part, those four month were chosen in order to represent the seasonal variability and we caught the seasonal pattern in the paper. And the main analysis is to analyze the same month with different AOD spatial resolutions, and seasonal variability does not play any role here.

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.

Response: Thanks for the suggestion. The comparison between MAIAC AOD and

AERONET is done as shown before.

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.

Response: Thanks for the suggestion. We rephrase the presentation here.

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.

Response: We agree with the reviewer. The analysis is always related to the application. As we emphasize several times. The objective of this paper is neither to retrieve $PM_{2.5}$ nor to provide a method for improving the correlation between AOD and $PM_{2.5}$. And for this purpose we mean that the linear correlation coefficient as the main analysis tool is sufficient. Additional parameters like the RMS as suggested by the reviewer is also included in the revised version.

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

Response: More detailed and quantitative analysis related to the correlation has been included in the revised version, especially linked to the AOD spatial resolution effect.

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 what 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.

Response: Thanks for the suggestion. The problem highlighted by the reviewer can be further improved from two aspects: (1) More quantitative analysis related to the change of the correlation between AOD and $PM_{2.5}$ will be contained in the revised version; (2) The significance of the statistical test like the p-values are also included. According to the p-values, all the statistical tests in the paper are significant.

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.

Response: We never tried to judge MAIAC and DT or other mature AOD retrieval algorithms. Different algorithms definitely provide different products with different accuracies for different conditions like surface/aerosol conditions; this is why C6 merged DB and DT. This is also the case for MAIAC. Here 10km refers to the aggregated 1km MAIAC AOD product and the analysis regarding the cloud contamination is updated in the revised version.

In the rest of the analysis section the discussion is again often shallow and just refers back to other similar studies to note the similarity of results, again highlighting that nothing

really new is done in this study.

Response: We would like to highlight again that very few studies focus on the AOD spatial resolution effect on the relationship between AOD and PM_{2.5} (previous knowledge is mainly from Chudnovsky et al. 2013). But to our best knowledge, there is no publication that tries to investigate the AOD spatial resolution effect for different regions and scales of coverage simultaneously. Some more detailed and quantitative analysis has been added.

P25880, line 26 and onwards: This is an example of where the authors do use quantitative correlation changes and perhaps over interpret 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.

Response: Thanks for the suggestion. However, it seems like the reviewer emphasizes a point which is slightly different from some of the comments above. In the suggestion above, the reviewer said that we cannot analyze based on real-world relevance application, but that we have to move to quantitative correlation analysis, but here the reviewer changed. We thank for the suggestions, and some more analysis related to the real-world relevance are added.