Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-335-RC1, 2018

Atmospheric Measurement Techniques Discussions

Interactive comment on "A bulk-mass-modeling-based method for retrieving Particulate Matter Pollution using CALIOP observations" *by* Travis D. Toth et al.

Anonymous Referee #1

Received and published: 24 October 2018

This study represents a credible attempt at a new way to infer surface PM2.5 levels from CALIOP data, on a regional, two-year average basis. An advantage of CALIOP over passive sensors for this sort of analysis is the fact that it measures vertical profiles of backscatter and depolarisation, so bypasses a limitation inherent with imager data in partitioning between total column and near-surface aerosol loadings. In contrast, an acknowledged limitation is the curtain sampling of CALIOP vs. the broad-swath sampling of MODIS, etc. The authors introduce their technique and explain the relevant assumptions, and show results over the USA, evaluated with EPA monitors. This is a sensible, strong first step in this direction. The topic is important and relevant to AMT.

I have a number of comments (below) but on the whole recommend that the paper can be accepted after minor revisions. Hopefully this will be a springboard for further studies refining the technique and expanding to other regions and time periods.

As a general comment, much of the quantitative evaluation is presented as scatter plots with linear regression fits, and the discussion is often framed in terms of r2 and slope. I'm not sure that this is the right thing to do here. One reason is that my understanding is that there can be non-negligible uncertainties on the PM data. Indeed, Ayers (2001, https://www.sciencedirect.com/science/article/pii/S1352231000005276) recommends using reduced major axis (RMA) regression instead of ordinary least squares when comparing PM monitors, for that reason. But also, the analysis in section 3 indicates that the CALIOP-derived estimates seem to have PM-dependence on their uncertainties too, so standard RMA may not be right either (as that assumes independent identically-distributed errors). For this reason I'd recommend Deming regression as a reasonable alternative (https://en.wikipedia.org/wiki/Deming_regression) when trying to compute the best-fit line. This should be more appropriate for this case, has packages in standard programming languages (and is not hard to code anyway), and is not hard to interpret. So this should be a pretty straightforward change to make which would improve the rigor of the manuscript. I recommend this is done throughout. Or, alternatively, don't fit a line but report something like mean ratio and RMS across certain ranges by binning the data.

I think it is important that appropriate statistical methods be used; continued publication using techniques we know to be deficient for our analyses just normalizes and encourages bad practice in the future. There isn't really a good justification for not fixing this.

My remaining comments are given as PXX, LYYY referring to page and line numbers respectively.

P1L21: I suggest replacing "sizes" with "diameters", as that is my understanding of the definition, but the remote sensing community often refers to radius instead when discussing size.

P4L95: I am curious as to why, with over 10 years of data, the two-year period 2008-2009 is used here? If sampling is a limiting factor in some areas, surely adding a few more years would help with this? Is there something special about these two years, or some a priori reason why two years provides sufficient sampling? I realize that running the whole mission is probably not feasible at this stage. But I would imaging that in the time between this comment being posted and the close of Open Discussion, there wouid be sufficient time to download and analyze an additional few years of data. This should mostly be a matter of storage and CPU time, since the code is already written (and since the first author is at Langley where CALIPSO is based, I doubt computational concerns would be significant here).

P6L123: somewhere in this initial paragraph, I'd ideally like some more discussion of the EPA data. For example, what are the uncertainties, is there any significant difference in these between the TEOM and BAM methods, and is there a difference in the siting of these two instrument types? If they're super-accurate and precise and equivalent, that's important to know. But if one is better than the other, and there's some spatial/temporal clustering in when TEOM vs. BAM is employed, that is also important to know. Recently, Kiss et al (2017, https://www.atmos-meas-tech.net/10/2477/2017/) published an analysis showing biases in hourly PM10 measurements. Is that relevant here? It might be, especially since that some daily averages in the EPA data correspond to a single sample. These are examples of things I'd like to see covered in the opening part of this section.

P8L189-190: This assumption (negligible mass above 10 micron size) is probably reasonable. But it would be fairly easy to try and quantify with AERONET. Take the inversion product from a half-dozen AERONET sites and count the fraction of the volume size distribution above 10 microns (and note here that the AERONET retrievals report size in terms of radius, while PM definitions are in diameter). You have to make some

assumption about the density of particles being the same across the size range, but otherwise that gives a first order estimate at how big the effect might be, which could be compared to the other parts of the uncertainty analysis in section 3.2. I think AERONET dust radius peaks somewhere like 2.5 microns so in the western US, it might be that there's some dust contribution from the tail of the distribution which is being systematically missed here and would lead to an overestimate in the CALIOP-derived PM levels. Maybe it is negligible, but it would be fairly easy to show that it is negligible, and the authors have not.

P13L299: An alternative to this (whether for the sensitivity analysis or the analysis as a whole) might be to look at the whole boundary layer (determining on a case by case basis) rather than testing different height ranges. Assuming that boundary layer depth is included as part of the MERRA2 meteorology being used here? This would go from assuming "the surface level of PM is represented well by the atmospheric layer from 0.1-1 km" to assuming "the boundary layer is well-mixed so represents the surface PM well", which is subtly different and might work better. I do agree that it seems reasonable to exclude the lowest 100 m, though.

P14L323: This section made me wonder why the authors do not estimate PM10 from CALIOP, and evaluate that, in addition to PM2.5? This would remove the need for an assumption of the ratio (taken as 0.6 here), and line 326 notes that there are 409 EPA stations providing both data on a daily basis. Given that this ratio seems to be one of the more uncertain parts of the error budget, it might be that there is more skill in predicting PM10 from CALIOP. Or it might go the other way. That would also be a worthwhile result, since right now we don't know.

P17L382: No particular comment here other than to say I am glad that the authors included this specific analysis. It's a point well-made that CALIOP uncertainties propagate downwards so, while CALIOP can see through thin clouds, that does not mean that the data quality is the same as for cloud-free columns.

P19L424: This isn't really an uncertainty analysis, so I suggest promoting it from a section 3.2.9 to a section 3.3 by itself. I also have a few suggestions for expansion of this section. It's good to know the correlation lengths across the western vs. eastern USA, but there's a lot of scatter in the plots. Some of this is probably due to limited sampling but some is probably also due to real changes in correlation length. So I wonder if the authors can pull out data from one or two large cities, and one or two remote areas, and highlight the correlation lengths for these (as well as the more general case of east vs. west). This would provide a bit more context about typical correlation lengths in these conditions, which would be helpful for future research built around this analysis.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-335, 2018.