MS No.: amt-2014-208

Referee comments for "SPARTAN: a global network to evaluate and enhance satellite-based estimates of ground-level particulate matter for global health applications"

We thank the reviewer for their helpful comments. We have revised the manuscript according to their suggestions with point-by-point responses below.

Answer to Reviewer #1 comments

Major comments

1. Statistics

The statistics presented in the paper is messy and to the best of my knowledge incorrect. What the authors want to show is a correlation between two parameters, PM2.5 (they use daily averages, which can be debated) and AOD. When they discuss the statistics on page 7582, they present r2 values for n versus 1/AOD and for n versus PM2.5. Since n is defined as PM2.5/AOD this seems to me ill-posed statistics since n is not an independent variable. Moreover, they write: "the variability in n is explained by neither AOD nor PM2.5 alone". I would argue that the variance in n (vn) is completely determined by the variance in AOD (vAOD) and PM2.5 (vPM2.5), like: vn/n^2 = vAOD/aod^2 + vPM2.5/pm2.5^2. So, it is fine to analyse the contributions of AOD and PM2.5 to the variance in n, but it is a statistical sin to correlate non independent variables. The same holds, more or less, to equation (4). Again, the variance in n can be decomposed in variance contributions of T1, T2, and T3 (please define these terms!). But the underlying questions is how well PM2.5, bsp24, and bsp,sat are correlated with AOD. So I strongly suggest to present results as correlations between independent variables. Now, figure 3 presents only one panel in which you see when the correlation between PM2.5 and AOD breaks down. It is virtually impossible to infer from the figure how the correlations between AOD and bsp24, and bs,sat would look like. It is clear from the figure, however, bsat/b24hour shows quite a pronounced variability (0.4-1.6), which again would imply that analysing PM2.5 at satellite overpass would clearly be preferable for SPARTAN. To move from PM2.5 to PM-24hours is a separate question. This clear message and associated analysis are not totally evident from the paper.

Reply: Thank you for your comments regarding our analysis of equation 3 and 4. As this network continues to grow, our perception is evolving of how to best analyze the data. As you mention, the correlation between η and AOD & $PM_{2.5}$ (equation 3) and the three terms in equation 4 are not orthogonal. The original statistical analysis in section 4.1 has been modified to express that the variables are intercorrelated and have been analyzed this way for their relative contribution to total variance. We continue to believe that more physical insight is established by focusing on the interpretation of eta than on other variables. The text in sections 4.1&4.2 uses the revised statistical methods as described on p7582, lines 10-15:

> "We calculated the contribution of AOD_{sat} and $PM_{2.5,24h}$ to the variation of the dependent variable η as the relative contribution to the coefficient of multiple determination (R^2), based on the product of the correlation coefficient ($r_{yx(j)}$) and standardized regression coefficients (a_j) for each variable j. In Beijing the contributions to η of $PM_{2.5,24h}$ and $1/AOD_{sat}$ are 0.07 and 0.51, respectively. The larger contribution from AOD_{sat} indicates the importance of accounting for aerosol aloft".

With regards to comparing 24-hour $PM_{2.5}$ to satellite overpass $PM_{2.5}$, we refer to Table A1, which uses three test cases for the relative error associated with both types of sampling. Relative sampling error decreases during midday-only sampling, as expected, since the nephelometer interpolates $PM_{2.5}$ for a much shorter period of time (4 hours versus 24 hours). Furthermore, we see from equation 4 that midday vs 24-hour scatter (term 2) does show a non-negligible contribution to total variance in η. However if SPARTAN were to focus on satellite overpass hours, we would inevitably introduce systematic biases in estimating long-term (annual) $PM_{2.5}$ averages. Since SPARTAN is ultimately aimed at global health applications, we hesitate to compromise this objective of the network.

2. Height profile

The authors should be more elaborate to explain the difference between surface PM2.5 and AOD. What are the sources of error? Although they recognize that the vertical pro- file is a main source of error, on page 7584 they discuss the low n in a sub-Saharan site in terms of coarse particles only, and not in terms of vertical profile (although it is said somewhere: "implying a pronounced aerosol layer above Dhaka"). Also the Bandung volcanic eruption case is interpreted as "sulphate particles grown in high relative humidity" and aerosols above the surface are not explicitly mentioned. I suggest a separate paragraph in which possible sources of error are discussed.

Reply: We agree that multiple factors influence the $PM_{2.5}$/AOD ratio. Based on your advice we have added a new paragraph in section 3.5 to discuss sources of uncertainty. We refer to topics affecting $PM_{2.5}$/AOD as factors, rather than errors, since understanding these factors is the objective of our network.

Minor comments

1. Maybe include some statement about clouds in the AOD measurements: are only cloud-free observations taken into account?
Reply: Good point, as all AOD measurements are from level 1.5 data or higher, hence cloud-screened. We amended this on page 7573 line 20-22:

> "AERONET provides temporally resolved cloud-free measurements during daylight hours at 0.01 to 0.02 mid-visible AOD accuracy and is extensively used for satellite validation (e.g. Remer et al., 2005).."

2. Page 7574, line 11: introduces→presents
Reply: Changed

3. Page 7575, line 2, sires→sites
Reply: Corrected

4. Page 7575, line 11: why not related to PM2.5 at overpass time?
Reply: Satellite overpass $PM_{2.5}$ is an interesting quantity, but deviates from our goal of measuring 24-hour PM, which is connected to long-term health.

5. Page 7577, line 16: nothing is being said about semi-volatile organic compounds. Maybe add a statement?
Reply: We consider semivolatiles together in this section, however we now point out this includes organics, p 7577 lines 13-15:

> "We choose to start sampling runs for each filter in the morning (9:00) when temperatures are lower, to increase retention of temperature-dependent semi-volatile inorganic and organic material that was collected overnight."

6. Page 7582, line 4: n has units of concentration, but should not be referred to as a concentration.
Reply: We have revised this section to longer make this implication. This sentence is now deleted.

7. Page 7582, line 5: select→ selectedPage 7583, line 24: again the discussion relates n to PM2.5: "a tendency for n(=PM2.5/AOD) to increase with PM2.5 . . .. . .sound awkward. What you really want to say is that PM2.5 and AOD are correlated. . . Also: PM2.5/AOD is better related to mean PM2.5 than AOD. This statement is unclear. Than AOD is related to PM2.5? Anyhow, reconsider this section in view of the statement above.
Reply: We have modified our discussion, now based on the statistics mentioned in the previous comment.

With respect to the PM2.5/AOD spatial and temporal ratios, on page 7582, line 12-17: and page 7585, line 1-3. We removed the subsequent statement of concern.

8. Page 7584, line 22: was the eruption modelled in GEOS-Chem?
Reply: We did not model the eruption in GEOS-Chem. As of now the eruption is only a potential explanation for the high AOD while ground-level did not increase dramatically.

9. Page 7584, line 26: than any → than at any
Reply: Corrected

10. Page 7587, lines 13-15: I think this statement leads to far. First, it will depend on the homogeneity of the scene. Second, of you would consider parings op 50 km apart in Taiwan, errors would hardly grow.
Reply: We have included 2003 work done by Anderson et al (Mesoscale Variations of Tropospheric Aerosols. *JAS*, *60*(1)) that shows similar spatial variability for city-scale aerosol distribution.

> Page 7588 Line 16-18, "Our analysis of spatial variability is consistent with the $R^2 > 0.8$ found by Anderson et al. (2003) for nephelometer scatter at distances less than 40 km".