Comments on Strawa et al. "Improving PM2.5 Retrievals in the San Joaquin Valley Using 3 A-Train Multi-Satellite Observations"

This reviewer is not a fan of Generalized Additive Models. Like multiple linear regressions or neural network techniques, they tell you very little about the physical mechanisms underlying the relationship between an odd set of variables, many of which are not physically independent. This paper chooses a set of GAM models that relate aerosol optical depth (AOD) from MODIS at a number of wavelengths and using varying algorithms (and from OMI), from NO_2 column from OMI, and adds the variables aerosol absorption optical depth (AAOD) from OMI and day of year. The result of this stew of measurements is that the authors report an improvement in PM correlations with these variables over a linear regression between PM and AOD in the San Joaquin Valley.

The paper is poorly written and appears to be cobbled together from a report or conference proceedings this were previously published and are not available openly. The paper needs a thorough reworking to be published in ACP and even after that, the results are questionable about whether they can be applied elsewhere. The GAM models chosen are "black boxes" with no reported parameters that allow the results to be applied elsewhere or reproduced by others. Unlike linear models which lay the slopes of the regressions out there for others to laugh at, the GAM model is a big matrix of coefficients, some of which are weighted, some which are not, some which may be linear and some not, and none are tabulated anywhere. Like neural net or "machine learning" techniques, the stew gets stirred by some witch's broom and the reader needs to accept that the result tastes better by adding some "eye of newt". This is not how science progresses.

Regarding the topics that ACPD asks to be addressed:
1) The paper does address questions which are appropriate to ACP.
2) It doesn't develop new techniques but rather applies Liu's prior GAM model to new data sets.
3) This reviewer doesn't find the conclusions applicable elsewhere.
4) There is a paucity of information on the technique itself and
5) there are results which are poorly described (especially Figure 5).
6) It would be practically impossible for others to reproduce these results since none of the parameters of the GAM are given.
7) The authors properly reference prior work.
8) The title of the paper is appropriate.
9) The abstract is satisfactory. The presentation needs work as noted in the previous paragraph.
10) The overall presentation is clear.
11) The language is fluent but there are typos.
12) Formulae are appropriate
13) Figures need much better captioning to be understood and there are legends in the tables which appear to be irrelevant
14) References are appropriate

15) There is no supplementary material.

The main criticism of the paper that I have is that two of the variables (day of year, and NO_2 column) are clearly not related to AOD in any physical way that can be mathematically represented with chemical or physical variables. They are surrogates for something. DOY could be a surrogate for emissions, for the inverse of the PBL height, for relative humidity in the column, for temperature (affecting nitrate, for example). NO_2 column could be an indicator for emissions, for the inverse of the PBL height, for temperature,.... wait, I just said that. And they could be highly correlated. If for example, one went to North Africa, would PM and AOD be related through NO_2? Hardly. This lies at this reviewer's concern with the uniqueness of the results.

Why not take a subset of the data, train it for coefficients and then run it on other years? It seems that the test of whether this is an improvement or not depends on whether it has any future utility.

Specific comments:

Pg 2, line 1: clarify the Pope conclusion that 10 $\mu g\ m^{-3}$ = 1 year. Is that 1 year per 10 $\mu g\ m^{-3}$, i.e. in an area with 100 $\mu g\ m^{-3}$ you lose 10 years on average? Do they know it is linear?

Pg 2, line 12-13: AOD is used to this but it more importantly is a measure of climate forcing potential from aerosols since it is related to the upwelling reflectance of the aerosol.

Pg 2, line 20: The references bounce back and forth between italics and normal font. Pick one.

Pg 2, line 31: r^2 may be near zero but 0.0? Have not seen that.

Pg 3, line 9: remove comma after Kacenelenbogen et al.

Pg 3, line 13: Engel-Cox is misspelled; also elsewhere in text... check

Pg 3, line 16: remove period from within parentheses

Pg 3, line 27: Kacenelenbogen misspelled

Pg 5: line 30: sites

Pg 6: line 19-20, I understand the reference for Levelt but this should be after OMI not MODIS. Reference for MODIS? Levy?

Pg 6, line 22: Aqua and Aura are not all capitalized.

Pg 7, line 2-5: why use a direct quote from Levy? Paraphrase.

Pg 7 line 23-24: good agreement is seen between OMI and ground-based measurements .... of aerosols? Did Celarier do this? How can it compare with ground based measurements when the AOD from OMI is in the UV and doesn't see the surface.

Pg 8 line 3: This reviewer does not argue with the choice of 50 km as a reasonable spatial scale for aerosols at the sub-daily averaging time scale, but it is still arbitrary. Use of a TOMS 50km/hr mid-tropospheric wind speed means nothing compared to the PBL comparison done here. Mean wind speeds in the PBL are probably closer to 5 km/hr. Matching the temporal scale of aerosols with the box size for AOD will depend on wind speed, degree of mixing, and local sources. I recommend leaving out the mid-tropospheric argument which is not relevant.

pg 8, lines 8-9: Ignore the MODIS quality flags at your peril. Many are there because of the position in the swath and the fact that the 10x10 MODIS AOD may only have a few non-cloudy pixels available to average. How much "better" is the correlation with the good-very good flag used compared to using all data? Why is more data better than good data?

pg 9, line 10: NO_2

pg 9, line 27: calling out Figure 4a before describing why it is used is awkward.

pg 10, line 21: speaking of Table 2, there are identical correlations between PM25_DAILY and OMI_AOD, is one daily or is one AAOD? What is OMI VAI? It is not described in the text to this point. Typo: DB_AOd_55. The correlation between AOD and Angstrom coefficient is not helpful at all. One is an extrinsic property of the aerosol and one is an intrinsic property. They would be decorrelated by the number density of the aerosol alone.

pg 11, line 3: Figure 2a, PM modeled from DB_AOD_47. Is this from a linear regression? Modeled how? What is the functional relationship between AOD and PM2.5?

pg 11, line 24-26: Saying that 12 ug m-3 is "the instrument sensitivity" is amazing. For which instrument, the BAM? Or the minimum sensitivity for an AOD-PM relationship? Clarify. As 12 ug m-3 would exceed a 24 hour average EPA level, this is not trivial. It would represent a 120 ug m^{-3}/AOD slope and this is clearly not physical given the index of refraction, size distribution and Mie extinction of most

$PM_{2.5}$ aerosols, unless there is a very compressed boundary layer. Chu et al. (2003) suggest a 60:1 slope for PM/AOD and Levy and Remer suggest that the AOD over land is ±0.1. That is 1/2 the result suggested above. Hoff and Christopher (2009) suggest that the intercept might not be zero because of the preponderance of data at elevated AOD values and a potential non-linearity in the PM/AOD relationship.

Looking at Figure 3a, the offset is not caused by low sensitivity to AOD. It is caused by high AOD and low PM. This is clearly due to the inclusion of aerosol aloft which has no correlation with surface PM. A child with a crayon would draw a different fit to that data than the computer.

pg 11, line 30-31: changing color in a table seems to be a complication which is unnecessary.... Shading the boxes but keeping everything in B&W?

Pg 12, line 6: sat >>> satellite

Pg 12, line 7: parameter set for Table 3.... what is this? Table 4? If so, these are out of order and Table 4 should be discussed before Table 3. In Table 3, the legend for the significance is not relevant to the table.

Pg 13, line 6: This step in the model needs justification. A model with all inputs is run and then a model with a limited number of inputs is run and only those points in the model with limited inputs are added (i.e. a subset of the data which does not satisfy the criterion for GAM1 is now added back into the mix). This step is necessary since you have added a GAM criterion requiring all $f_i(x_j)$ to exist.

Equation 4 could be reduced to a matrix multiplication of **f x** where f is a matrix of coefficients and x is a vector. **f** appears to a sparse matrix with many elements with either zero or NaN terms (you reduce your observations from >6000 to about 600). Your fourth step proceeds to fill those elements with data from another model (GAM3) say. This is not your equation (4). In fact, you now propagate other off diagonal elements which may actually violate GAM3. It would be the same as saying:

$$y = B_0 + \sum_{j=1}^{n-1} f_j^1\ (x_j) + \sum_{j=n-1}^{n} f_j^2\ (x_j) + \varepsilon$$

, for example. What logic is there that x_n subset of data has a different functional response than the x_{n-1} other points, just because it was not included in the data set because one variable only might be missing? Why should the relationship between PM and AOD47 change because you now have OMIAAOD? This "tuning" is very disconcerting.

Pg 13, line 12: Use of the p-statistic here is misleading, in my opinion. (1-p) would say that the relationship is "true" and yet, the correlations between variables that have $p=0.0000$ (four digits!) is poor. Explain that. There is literature on how the p-statistic is misused to infer statistical relevance (see: http://www.jstor.org/stable/2684655) and other techniques (Bayesian inference) may be more relevant in proving whether the missing data does or does not increase knowledge.

While we are on it, Table 4 has a number of errors (AOD for GAM1 is clearly wrong). There are no significance codes in Table 4... this seems to be a sloppy cut and paste from some other document (a report?). PM25? What is doy? Is it the previously defined θ?

Pg 14, in reference to Figures 3b and 3c, it looks to the naked eye that the relationship is decidedly non-linear. Did you just try to fit a parametric fit to that data? What if the PM(gam) goes like $PM_{2.5}^z$ where $z<1$? Doesn't this point to f(x) being non-linear?

Pg 14, line 20-21: Why if NO_2 (ground) does not correlate well with OMI NO_2 do you expect $PM_{2.5}$ (ground) to correlate with AOD?   AOD is far more complicated than column NO_2 and if they don't correlate, that indicates that vertical profile dominates any relationship in the Valley.

Pg 14, line 3: what is this "sensitivity"?   It needs definition and it is not possible to understand Figure 4 from the text. What are the "ticks" at the bottom of the figure?

Pg 15, line 32: "due to expedience"... unfortunately, this paper doesn't need to be published quickly, it just should be correct. It seems unphysical that the inclusion of RH will not help the regressions since it appears in equation 1. The fact that surface RH doesn't help is pretty obvious when one realizes that RH changes drastically in the PBL, often approaching 1 at the PBL top in convective situations where cumulus form. But the "attempt to use RH from assimilations [sic] models" is pretty weak justification. How "inaccurate"? $r^2= 0.7$ is inaccurate. Did it make the regressions worse? Include this data. Barnaba's paper is not available widely unless the conference abstracts are purchased.

Pg 17, line 3:   Hidy's paper (2009) explains why the PM/AOD method will NEVER replace PM measurements for regulatory purposes.   You state that the method cannot infer PM < 12 $\mu g\ m^{-3}$.   Regulatory agencies would seek better than a factor of 10 improvement in this and such an improvement is not going to happen.