	Reviewer 2	
	Comments	Response
1	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.	In Table 2, the equations used in the GAM models are tabulated and the coefficients are listed in the supplemental material. As stated in this and other papers, the science of the physical and chemical interactions of aerosol in the atmosphere are very complex and all methods of studying the problem need to be utilized. GAM models are useful because they allow the researcher to combine several different parameters that provide different information about the relationship between two variables. To my mind this is analogous to adjusting parameters in a model to tailor a calculation for a particular area or to match a particular observation. GAMs can also give insight into what parameters are important.
2	It doesn't develop new techniques but rather applies Liu's prior GAM model to new data sets.	The work of Liu, et al., (2009a,b), cited in the paper, uses a chemical transport mode and generalized linear models to give information about the vertical structure of aerosol. These CTMs use inputs from meteorological models that are also used in air quality models. We attempted to be more specific about our technique in sec. 1. In this paper we wanted to provide a retrieved PM that could be used to validate the air quality models. Therefore it was important that the retrieved PM be independent of the meteorological models. Our methods were inspired by those of Pelletier et al. (2007) who used GAMs to improve the PM retrieved from AERONET data. This study is different in that we are using multiple satellite observations. We attempted to make this clearer in the text.
3	This reviewer doesn't find the conclusions applicable elsewhere.	The meteorology, surface conditions, and pollution sources particular to the San Joaquin Valley can also be found in other areas of the planet. We have identified the Po Valley in Italy, China's Red Basin, and the Indo-Gangetic Plain as candidate area where the same

		mixture of parameters might work. In other areas, it is likely that the parameters would have to be changed to improve correlation of retrieved and measured PM, however, the methodology presented here should be applicable. This is discussed in greater detail in section 4.
4	There is a paucity of information on the technique itself	Sec 2.5 has been expanded providing more detail.
5	there are results which are poorly described (especially Figure 5)	The description and significance of Figure 5 are explained in more detail.
6	It would be practically impossible for others to reproduce these results since none of the parameters of the GAM are given.	The exact models used for the GAMs are now given in Table 2; R-language code and summary "objects" are now noted as available from the authors. (R is freely available and widely used, and preserves a history of all versions so that improvements to the codes and ample documentation do not lose precision of reference.) Tables of the spline fit parameters are now summarized in Supplementary Material.
7	Figures need much better captioning to be understood and there are legends in the tables which appear to be irrelevant	The captions have been improved.

8	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.	The reviewer has identified a key point of the paper which we are pleased to make explicit. It is not our contention that a statistical model can be blindly applied to every situation. In fact, we doubt that any type of model can ever do a decent job without some a-priori information about the particulars of a given region, or possibly similar regions. The following text has been added to Sec. 1. "The relationships between observed properties (spatial patterns of radiances at various wavelengths) and AOD and especially $PM_{2.5}$ are intrinsically regionally dependent and stated as statistical summaries of conditions typically prevalent in a given region in a given season. The optical properties of the aerosol components as a function of wavelength, are dependent on the character of surface bidirectional reflectance, the chemical composition, depth of lowest-layers mixing, and the presence, altitude, and transport times involved with elevated aerosol layers. All these factors make it unlikely that any simple equations can relate radiances to $PM_{2.5}$. When relationships are found, as we and others have found, they are statistical descriptions which may well have a more complex physical rationale." In the case of our study, day-of-year, is very likely a surrogate for the PBL height. Based on prior knowledge of the area, we know that it is dominated by highly seasonal meteorology with low PBL in winter and thick PBL in spring, summer, and fall, as discussed in Sec. 2.1. This matches the sensitivity shown in Fig. 4. Likewise, we know that the main source of pollution is vehicular traffic. Therefor it is not surprising that there should be a positive sensitivity between NO2 in our model. We provide a reasonable current understanding of a three-way correlation structure between OMI NO2, locally measured nitrogen oxides, and locally measured PM2.5, but admit that more detailed study is planned.

9	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.	One would not expect PM to be strongly related to NO2 in North Africa. The parameter set used in the model will be different based upon the area one is studying. This is discussed in Sec. 4. We have improved this discussion to be more clear.
10	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.	The mgcv routine we used makes this kind of evaluation of the splines; that is the GCV or generalized cross-validation portion of the name mgcv. This serves best for the spline terms. However, all terms are analyzed for idiosyncratic sub-populations using the standard method of plotting partial residuals. (Wood, 2008 and many other references). This is now described in the text.
11	Pg 2, line 1: clarify the Pope conclusion that 10 µg m-°©-3 = 1 year. Is that 1 year per 10 µg m-°©-3, i.e. in an area with 100 µg m-°©-3 you lose 10 years on average? Do they know it is linear?	The sentence is replaced by "The result of their population-based analysis shows that a decrease of 10 µg/m3 in the long-term exposure to fine-particulate matter concentration was associated with an estimated increase in life expectancy of approximately 0.61±0.20 year (Pope et al., 2009). "
12	Typos	Typos have been corrected
13	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. I recommend leaving out the mid-tropospheric argument which is not relevant.	Reviewer 1 took exception to our choice of 5x5 grid cells, so we conducted a study of the effect of grid size on the correlation coefficient. This is described in Sec. 3.1.
14	Pg 8, lines 8-°©-9: Ignore the MODIS quality flags at your peril.	Our decision to omit consideration of the quality flags was based on their effect on linear regressions. We have undertaken a study of the effect of MODIS quality flags using our GAMs. This is described in Sec. 3.1.

15	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.	I don t understand the comment referring to PM25_DAILY and OMI_AOD. Typo corrected. VAI and Angstrom references are deleted b/c they are not used in text.
16	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?	I believe the reviewer is referring to pg. 11, line 23. The ordinate is the PM retrieved from a model derived from a linear regression of DB-AOD and measured PM. We have elected to call the ordinate the 'modeled PM' plotted versus the 'measured PM' on the abscissa.
17	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.	Gupta and Christopher (2008) define the intercept in the regression equation as the minimum level of particle concentration for which satellite derived AOT is sensitive. Below the level of intercept, satellite signals are weak and detection of aerosols is difficult. They found a mean intercept value of 15.6 μgm−3 in a study of the correlation of seven years of MODIS data and surface PM measurements. We agree that is is confusing and inappropriate and the text has been eliminated.
18	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.	We have no basis for saying that aerosol aloft have caused the discrepancy for large values of PM. If we had, it would be a valid reason for eliminating those points. It seems to me that the slope of the regression line to driven by low AOD corresponding to high PM measurements.

19	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?	Good point. The color is black.
20	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.	Table 3 is not really valid b/c it compares different GAM models at the different locations. We have eliminated this figure.
21	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.	
22	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).	It is not true that "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" The f(x) indicates a function of x. The function could be a squaring operation or a cosine; the regression is still a linear regression even though it is not B x where B is a (constant) matrix.
23	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: see EQN.	We have elected to drop the combined model since it was difficult to explain and did not create very much benefit.

24	What logic is there that xn subset of data has a different functional response than the xn-°©-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.	There is a point that we do make clear. When linear or additive regression models are estimated with one term missing, the remaining parameters do very often change. We refer to classic as well as modern statistics texts on this point.
25	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**?	We thank the reviewer for pointing out some interesting limits in the interpretation of p values for anyone's results. We have added a note to our new table with the Schervim reference. (1) We also point out that the large sample allows the sampling statistics to describe confidence in the estimate even if the effect of the individual parameter is low. Consequently, we do not believe that there is anything to explain. (2) Bayesian confidence measures for p were indeed used. We include with our response a copy of the text that fully explains the usage; this description is easily available on the web. (3) We invite interested readers to contact the author for a machine-readable copy of our R-language gam "object," with which the reader may obtain further Bayesian information (See mgcv.pdf documentation.) (3) The basic point is made by the stars indicating p-values: these results certainly do not arise by chance using commonly applied measures, as is now noted.
26	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?	The p-statistic has been replaced with a significance code that is explained in the footnote to the figure. PM25 and doy are the computer names for PM2.5 and theta. These have been corrected.
27	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?	Sensitivity is described as the portion of the description, in PM 2.5 units, which is contained in the spline term.

28	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.	Poor choice of words. The text has been changed to explain the rationale for the seasonal parameter better.
29	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 μg m-°©-3. Regulatory agencies would seek better than a factor of 10 improvement in this and such an improvement is not going to happen.	We are not suggesting that the PM/AOD method will be used for regulatory purposes but for validation of air quality models. The GAM retrieval does considerable better in detecting exceedances.