Review of “10 yr spatial and temporal trends of PM2.5 concentrations in the southeastern US estimated using high-resolution satellite data” by X. Hu et al.
This paper shows the PM2.5 concentrations from 2001 to 2010 over an area in the SE U.S. and Atlanta metropolitan area using the MODIS MAIAC 1-km AOD data and a two-stage model that derives the surface PM2.5 concentrations from the AOD with a series of fitting parameters accounting for the meteorological fields, surface categories, and point emissions. The objectives are (1) to estimate PM2.5 concentrations in the study domain with MAIAC AOD as the primary predictor and other variables as the secondary predictors, (2) to generate maps of annual mean PM2.5 concentrations from 2001 to 2010, (3) to examine the 10-year temporal trends of PM2.5 in the study domain and Atlanta metro area, and (4) to investigate the potential impact of fires on PM2.5 levels. While it is valuable to use the high-resolution satellite AOD data for PM2.5 prediction and trend analysis, this paper has not shown the unique value of using such data and does not provide quantitative assessment on the connection between emission and PM2.5 level. My major comments are listed below. Major revision and a great deal of clarification are necessary before the paper can be considered for publication on ACP.

Response: We thank the reviewer for the valuable comments. All of them have been addressed in the revised manuscript. Please see our itemized responses below.

1. The use of 1-km resolution AOD data: I found that the paper has not demonstrated the merit of using high-resolution satellite data. Although it is stated in the “Introduction” that the standard MODIS and MISR products at 10-km and 17.6-km, respectively, are not sufficient for epidemiological studies and omit details of PM2.5 spatial variability, the results and analysis presented in this paper do not show any advantage of using the 1-km AOD data other than the visual structure in the maps. Considering that most analysis presented in the paper was done based on large area averages (either the study domain or Atlanta metro), what can the 1-km AOD data offer but the 10-km data cannot for the purpose of the present study? What is the spatial scale of AOD or PM2.5 variability that makes 10-km data insufficient?

Response: Our recently published paper (Hu et al., 2014) compared 1 km $PM_{2.5}$ concentrations estimated from MAIAC AOD with 12 km $PM_{2.5}$ concentrations estimated from MODIS AOD (in CMAQ grid). The results showed that 1 km $PM_{2.5}$ estimates can provide much more spatial details than 12 km $PM_{2.5}$ estimates. Within a single CMAQ grid cell (12 x 12 km^2), MODIS can only predict one $PM_{2.5}$ value and thus cannot reveal any spatial variability, while MAIAC can make ~144 predictions and thus can reveal a large amount of spatial variability of $PM_{2.5}$ exposure. For example, MAIAC predictions can distinctly show high $PM_{2.5}$ concentrations along the highway within the CAMQ grid cell while MODIS predictions cannot. Since we have already discussed the benefits of high resolution AOD data in $PM_{2.5}$ concentration estimation in (Hu et al., 2014), the primary objective of this study is to conduct a time-series analysis of $PM_{2.5}$ levels in Georgia and Atlanta metro area and facilitate epidemiological studies in this region. Since health data were geo-coded to small geographical units such as zip code and census block group that are in general smaller than the resolutions of MODIS and MISR, using 1 km AOD data to predict 1 km $PM_{2.5}$ concentrations should be more appropriate and beneficial for health effect analysis as 1 km data can reveal more spatial variability at the zip-code and census block group levels. In addition, in this study, we showed that large deceases of $PM_{2.5}$ levels between 2001 and 2010 occurred along major highways. We also related one pixel with large increase and another with large decrease between 2001 and 2010 in Atlanta metro area to a single emission source. Those cannot be achieved if using coarse resolution products.

We added the citation (Hu et al., 2014) in the manuscript.

We added the following sentence in the introduction "Hu et al. (2014) compared the performances of MAIAC and MODIS in $PM_{2.5}$ concentration prediction and found that MAIAC predictions can reveal much more spatial details than MODIS. In a single CMAQ grid cell, MODIS can only make one prediction, while MAIAC can make ~144 predictions. MAIAC predictions can distinctly show high concentrations along major highways, while MODIS predictions cannot."

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., Quattrochi, D. A., Puttaswamy, S. J., and Liu, Y.: Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model, Remote Sens. Environ., 140, 220-232, http://dx.doi.org/10.1016/j.rse.2013.08.032, 2014.

2. The two-stage model: This model depends on a large number of fitting parameters. It is not clear, however, from equations (1) and (2), how these parameters are obtained or derived. Did you use the observed PM2.5 and AOD plus other data to construct all the bi? What are the random and fixed intercept and slopes, by definition? Do different meteorological fields (winds, PBL, RH, etc.) associated with different b2s? How do road length associate with the site point location of individual site? Are all these coefficients "day-specific"? Among the secondary predictors, which ones matter the most? How are these secondary predictors chosen? The model and the methods are not clearly presented in the paper and clarification is necessary, especially there is no previous publication or documentation that might serve as a reference for the method.

Response: The method was adopted and can be found in detail in our recently published paper (Hu et al., 2014). Yes, we used observed $PM_{2.5}$, MAIAC AOD, meteorological parameters (e.g., boundary layer height, relative humidity and wind speed), land use parameters (e.g. point emissions, forest cover and road length) to build the model and established the relationship among them. Observed $PM_{2.5}$ was the dependent variable, while others were all predictor variables. Our first-stage model was a linear mixed effects model that includes both fixed-effects and random-effects terms. Fixed effects affect the population mean, while random effects are associated with a sampling procedure and contribute to the covariance structure of the data. In this study, fixed effects are used to examine the mean effect of predictor variables on the dependent variable for all days, while random effects investigate the day-to-day variability in the relationship between dependent and predictor variables for each individual day. Each meteorological field had one fixed slope (b_i) for all days and one random slope ($b_{i,t}$) for each individual day. Different meteorological fields were associated with different fixed and random slopes. To associate all the predictor variables (e.g., forest cover, road length, and point emissions) to the $PM_{2.5}$ monitoring sites, we create a 1 x 1 km^2 square buffer centered at each

$PM_{2.5}$ monitoring site, then calculated the total length of road segments within each square buffer in order to facilitate our $PM_{2.5}$ predictions over the 1 x 1 km^2 MAIAC grid cell (section 2.6 data integration). Only AOD and meteorological fields had "day-specific" random slopes to establish their daily relationships with observed $PM_{2.5}$, since they represent time-varying variables and we assumed that their relationships with $PM_{2.5}$ varied daily. In this study, prediction accuracy is critical and is the prerequisite for conducting a time series analysis. Therefore, we are more focused on the overall performance of the model rather than the performance of each individual variable. We calculated R^2, MPE, and RMSPE to evaluate the prediction accuracy. To assess which variable is the most critical is beyond the scope of this study. However, we only selected the secondary predictors that showed a statistical significance in predicting $PM_{2.5}$ concentrations. We also toggled on and off different predictors to find the model that have the best accuracy in $PM_{2.5}$ prediction by comparing the RMSPE, MPE, and R^2 values.

We changed "$PM_{2.5}$" to "observed $PM_{2.5}$" in section 2.7.
We added the following paragraph in section 2.7. "Fixed intercepts and slopes are the same for all days and generated via conventional linear regression, while random intercepts and slopes vary independently for each individual day by drawing samples with the same level of the grouping variable (day of year in this study) from the full set of observations. In this study, we generated fixed slopes for each predictor variable, while random slopes were only generated for AOD and meteorological fields, since they represent time-varying variables."
We added the following sentence in section 2.6 "road length and point emissions were summed over the 1 x 1 km^2 square buffer by calculating the total length of road segments and total point emissions within the buffer."
We added following sentence in section 2.7. "Only statistically significant variables were used."
We added a new table in the Supplemental Materials to show the final model structure for each year.

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., Quattrochi, D. A., Puttaswamy, S. J., and Liu, Y.: Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model, Remote Sens. Environ., 140, 220-232, http://dx.doi.org/10.1016/j.rse.2013.08.032, 2014.

3. Model fitting: It is said that "the model was fitted for each year individually" such that the predictors may vary for different years. I wonder why it was not fitted for each season, instead of for each year, since the seasonal variations of aerosol and meteorological variables are much stronger than the interannual variations, so doing seasonal fitting makes more sense.

Response: Our first-stage model was a mixed-effects model that incorporated both fixed-effects and random-effects terms for AOD and meteorological variables. In this study, the random effects were used to generate "day-specific" random slopes for AOD and meteorological variables. The daily slopes revealed daily relationships among observed $PM_{2.5}$, AOD, and meteorological variables. Since our model was capable of capture the daily variability, it also should be able to capture the seasonal variability. Thus, fitting for each season might not be necessary. In addition, we compared the models fitted for each season, each year, and all ten years and found that models fitted for each year generally yield better overall accuracy. Hence, we adopted the model fitted for each year in this study.

We added the following paragraph in section 2.7 “Although the $PM_{2.5}$-AOD relationship might vary by season, our first-stage linear mixed effects model was able to incorporate daily variability in the relationship by generating day-specific random slopes for AOD and meteorological fields and thus should be able to capture the seasonal variability. In addition, by comparing the performances of models fitted for each season, each year, and all ten years, we found that the models fitted for each year generally yielded better prediction accuracy. Hence, in this study, we fitted the model for each year individually.”

4. Error and uncertainty: There is no estimate of the range of error or uncertainty in this method, especially so many empirical fitting parameters have been used. It seems that aerosols above the PBL is not considered at all. Even though most time aerosols may be indeed concentrated in the PBL in the study region, such omission should be discussed.

Response: We adopted the method from our recently published paper (Hu et al., 2014). Our main objective in this paper is to make accurate $PM_{2.5}$ predictions, investigate the temporal and spatial trends of $PM_{2.5}$ in our study domain, and facilitate future epidemiological studies. Therefore, the prediction accuracy is critical. As a result, we more focused on the prediction accuracy of the model rather than examining the ranges of errors or uncertainties of the intercepts and slopes of fitting parameters. Instead, we calculated model fitting and cross validation R^2, MPE, RMSPE, and relative accuracy to indicate what levels of accuracy our predictions reached. We agree with the reviewer, aerosols above the PBL should be considered. However, to simplify the analysis, we assumed that the vertical distribution of particles above the boundary layer was relatively smooth.

We added the following sentence in the section 2.7. “For the model to be valid, we assumed that particles within the boundary layer were well mixed, and the vertical distribution of particles above the boundary layer was relatively smooth.”

5. PM2.5 trend and the cause of the decreasing trend: It is obvious from Fig. 3, 4, 5, and 7 that the PM2.5 started to drop in 2008. Before that there was just small fluctuations. There is no “generally decreasing trend” during the 10-year period; rather, it looks like a step function with a significant change occurring in 2008. What causes such change, however, is not adequately analyzed. It is mentioned a few times in the paper that the reduction of PM2.5 “might be due to recently enacted emission reduction program”, “is probably due to dramatically reduced number of emission sources”, etc., the quantitative relationship between emission and PM2.5 is not presented at all. I wonder why more quantitative analysis was not done, as the point emission is actually one of the variables used in the two-stage model on at least yearly basis, so the authors must have access to the emission data for all these years to see the year to year emission changes and link them to the PM2.5 changes.

Response: This has been corrected in the manuscript.

We changed the sentence in section 3.4 to “The $PM_{2.5}$ levels in the study region as well as the Atlanta metro area had a relatively small fluctuation from 2001 to 2007, while there was a significant drop in year 2008, which was in line with the trends of point emissions. The sharp decrease of $PM_{2.5}$ levels in 2008 was probably due to significant emissions reduction in 2008.”

We only obtained point emissions data from EPA National Emissions Inventory (NEI) facility emissions report for year 2002, 2005, and 2008, since EPA prepares NEI every three years. We added the emission trend to Figure 8. The emission trend was in line with the trend of $PM_{2.5}$ levels with a sharp decrease in 2008. We also made a new figure (Figure 9) to examine the relationship between $PM_{2.5}$ estimates and point emissions. The R^2 was 0.93, 0.69, and 0.99 for year 2002, 2005, and 2008, respectively, indicating a strong correlation between $PM_{2.5}$ and point emissions. Thus, we believe that the drop of $PM_{2.5}$ exposure in 2008 was probably due to significant emissions reduction. Emission reduction programs such as the Clean Air Interstate Rule (CAIR) issued by EPA in 2005 (http://www.epa.gov/cair/index.html) might played a significant role in the decrease of $PM_{2.5}$ levels in the region.

We added the following sentence in section 3.4 ” Figure 8 distinctly showed a sharp decrease of emissions from 2005 to 2008.”
We added following sentence in section 3.3 “recently enacted emission reduction program (EPA, 2011) such as the Clean Air Interstate Rule (CAIR) issued by EPA in 2005.”
We added a scatter plot (Figure 9) of $PM_{2.5}$ estimates vs. point emissions.
We added following sentence in section 3.4 “Figure 9 illustrated the relationship between $PM_{2.5}$ estimates and point emissions. The R^2 reached 0.93, 069, and 0.99 for year 2002, 2005, and 2008, respectively, indicating a strong correlation between $PM_{2.5}$ and point emissions.”

6. Impact of fire emission on PM2.5 level: This part of the study is particularly weak – basically there is no quantitative analysis of the fire impact. The only display that may suggest some fire impact is the difference of fire occurrence and PM2.5 levels between the two rural sites showing some coincidental peaks and valleys. Why is it necessary to show the difference between the two sites, instead of showing the variation of PM2.5 level and fire occurrence at the sites affected by the fire? Even if you choose to use the difference between the two sites, can you be more quantitative, e.g., make a scatter plot of the delta_PM2.5 vs. delta_fire? Is your study consistent or different from Zhang et al. 2010 that shows 13% PM2.5 in the SE U.S. is from fire?

Response: We found abnormal high concentrations in the southeast of our domain, which might be partially caused by fires. We conduct this analysis to examine the relationship between $PM_{2.5}$ estimates (not ground measurements) and fire counts, which has not been done before. Fire impacts were affected by fire size, meteorological conditions (wind speed and direction), the distance between fires and monitoring sites. Thus, to precisely determine which fire affects the sites is very difficult. Due to the lack of more comprehensive fire data such as fire intensity and fire size data, the quantitative analysis of the fire impact cannot be conducted at this point. Following the second reviewer’s suggestion, we decided to remove fire related results and discussion from the paper and focused on the trend analysis.

Other comments:

P 25618, line 8-9: “inherent disadvantage…”. But you really have not demonstrated such a disadvantage for your study. Also, what AOD products are considered as “current”? MODIS currently has 3-km product. MAIAC is also a current product.

Response: Chudnovsky et al. (2012) compared the 1km MAIAC AOD with the 10 km MODIS AOD and pointed out that the high resolution MAIAC data revealed a substantial spatial variability of AOD that cannot be captured by MODIS data. In addition, our recently published paper (Hu et al., 2014) compared 1 km $PM_{2.5}$ estimated from MAIAC AOD with 12 km $PM_{2.5}$ estimated from MODIS AOD (in CMAQ grid). The results showed that 1 km $PM_{2.5}$ estimates can provide much more spatial details than 12 km $PM_{2.5}$ estimates.

We changed the sentence in the abstract to “previous studies indicated that an inherent disadvantage of many AOD products is their coarse resolutions.”

Chudnovsky, A. A., Kostinski, A., Lyapustin, A., and Koutrakis, P.: Spatial scales of pollution from variable resolution satellite imaging, Environmental Pollution, 172, 131-138, 10.1016/j.envpol.2012.08.016, 2012.

P 25618, line 11-13, MAIAC: MAIAC is one of the MODIS products, which retrieves AOD from MODIS measurements using the MAIAC algorithm. This should be clarified to not mislead readers as MAIAC is an independent AOD product from a different sensor.

Response: It has been corrected in the manuscript.
We changed the sentence in the abstract from “a new AOD product with 1 km spatial resolution retrieved by the multiangle implementation of atmospheric correction algorithm was used.” to “a new AOD product with 1 km spatial resolution retrieved by the multiangle implementation of atmospheric correction algorithm based on MODIS measurements was used.”

P 25620, line 12-13, and line 25: Again, this is about the “coarse” resolution product: Can you elaborate why 10- or 17.6-km product cannot serve your purpose? What is the aerosol spatial variability that determines the adequacy of product resolution?

Response: Many urban areas can be covered by less than ten 10-km MODIS or 17.6-km MISR pixels. Each pixel only has one AOD value and lacks spatial variability of AOD. As a result, the resolutions of both MODIS and MISR are too coarse to characterize the spatial variability of AOD in urban areas. For example, studies showed that particle number concentrations decreased noticeably when moving away hundreds of meters from the freeway (Zhu et al., 2002). The distance is significantly smaller than the resolutions of MODIS or MISR and the variability cannot be captured by them. In addition, epidemiological studies typically use health data geo-coded to small geographical units such as zip code and census block group that are in general smaller than the resolutions of MODIS and MISR. Thus, MODIS or MISR are not suitable for examining the spatial gradients within a single metropolitan area when studying the relationship between AOD-based $PM_{2.5}$ estimates and health outcomes. The high resolution AOD product can provide more details about the spatial heterogeneity of $PM_{2.5}$ levels in the region, help to better understand the relationship between $PM_{2.5}$ exposure and each single emission source, and facilitate health effect analysis.

Our recently published paper (Hu et al, 2014) compared the MAIAC predictions with the MODIS predictions within a CMAQ grid cell. MAIAC predictions can reveal spatial variability of $PM_{2.5}$ within a 12 x 12 km^2 grid cell and distinctly show high concentrations along the highway, while MODIS can only predict one value in a single CMAQ pixel and thus cannot reveal any spatial variability. However, high resolution data also lead to larger data volume and longer processing time. For epidemiological studies, the resolution that is equal to or smaller than the geographical unit by which health data are collected should be adequate.

We added the following sentence in introduction "Hu et al. (2004) compared the performances of MAIAC and MODIS in $PM_{2.5}$ concentration prediction and found that MAIAC predictions can reveal much more spatial details than MODIS. In a single CMAQ grid cell, MODIS can only make one prediction, while MAIAC can make ~144 predictions. MAIAC predictions can distinctly show high concentrations along major highways, while MODIS predictions cannot."

Zhu, Y., Hinds, W. C., Kim, S., and Sioutas, C.: Concentration and Size Distribution of Ultrafine Particles Near a Major Highway, J. Air Waste Manage. Assoc., 52, 1032-1042, 10.1080/10473289.2002.10470842, 2002.

P 25621, line 12: It sounds like you have more than one objective. The paragraph should be re-phrased.

Response: It has been corrected in the manuscript.

We changed the sentence in the introduction from "The objective of this paper is" to "The objectives of this paper were".

P 25623, line 12-13: What differences it may introduce from using just Terra, just Aqua, or both Terra and Aqua?

Response: The differences among using only Terra, only Aqua, and both Terra and Aqua were rather small in terms of prediction accuracy. The differences of CV RMSPE between using only Terra and both Terra and Aqua ranged from 0 to 0.17 $\mu g/m^3$ for year 2001 though 2010, while the differences of CV RMSPE between using only Aqua and both Terra and Aqua ranged from 0 to 0.18 $\mu g/m^3$. We used both Terra and Aqua in this analysis in order to increase the spatial coverage. The increase in spatial coverage ranged from 30.2% to 72.4% for Aqua and from 17.2% to 26.3% for Terra from 2001 to 2010.

We added the following sentence in section 2.3 "In our study domain, the increase in spatial coverage ranged from 30.2% to 72.4% for Aqua and from 17.2% to 26.3% for Terra from 2001 to 2010."

P 25623, line 14-15: A combined use of AOD at 10:30 am and 1:30 pm can only produce the estimated PM2.5 averaged at these two particular time, not "between 10 am to 2 pm".

Response: Zhang et al. (2012) found that Terra and Aqua may provide a good estimate of the daily average of AOD; thereby the average of Aqua and Terra measurements should also be a good estimate of the daily average of AOD and thus should be able to be used to predict daily $PM_{2.5}$. In addition, Engel-Cox et al. (2006) built a linear regression between daily $PM_{2.5}$ concentrations and the average of Terra and Aqua AOD values, and Liu et al. (2012) successfully estimated $PM_{2.5}$ concentrations using the average of Terra and Aqua AOD. Furthermore, our analysis showed that using the average of Terra and Aqua AOD can achieve a slightly better accuracy in predicting daily $PM_{2.5}$ concentrations than using only Terra or Aqua.

To remove the ambiguity, we changed the sentence to "the average value, as pointed out by Lee et al. (2012), is likely to better reflect daily aerosol loading."

Zhang, Y., Yu, H., Eck, T. F., Smirnov, A., Chin, M., Remer, L. A., Bian, H., Tan, Q., Levy, R., Holben, B. N., and Piazzolla, S.: Aerosol daytime variations over North and South America derived from multiyear AERONET measurements, Journal of Geophysical Research: Atmospheres, 117, D05211, 10.1029/2011jd017242, 2012.

Engel-Cox, J. A., Hoff, R. M., Rogers, R., Dimmick, F., Rush, A. C., Szykman, J. J., Al-Saadi, J., Chu, D. A., and Zell, E. R.: Integrating lidar and satellite optical depth with ambient monitoring for 3-dimensional particulate characterization, Atmospheric Environment, 40, 8056-8067, http://dx.doi.org/10.1016/j.atmosenv.2006.02.039, 2006.

Liu, Y., He, K., Li, S., Wang, Z., Christiani, D. C., and Koutrakis, P.: A statistical model to evaluate the effectiveness of PM2.5 emissions control during the Beijing 2008 Olympic Games, Environment International, 44, 100-105, http://dx.doi.org/10.1016/j.envint.2012.02.003, 2012.

P. 25625, equation (1): As I mentioned earlier, this equation needs to be better explained.

Response: Equation (1) denotes our first-stage model, a linear mixed effects model. The linear mixed effects model not only examines the relationship between dependent and predictor variables for the entire observations (fixed effects), but also for groups of potentially correlated observations (random effects). In this study, the grouping variable was "day of year" to adjust for daily and seasonal trends. We defined the model with a fixed intercept and a fixed slope (e.g., b_0, b_1, …, b_6) for each predictor variable for the entire period, indicating the average relationship between the dependent variable (e.g., observed $PM_{2.5}$) and predictor variables across the entire period. In addition, we generated a "day-specific" random intercept and random slopes (e.g., $b_{0,t}$, $b_{1,t}$, $b_{2,t}$) for AOD and meteorological parameters (time-varying variables) for each day, denoting the relationship for each individual day. This model builds on early work detailed in our recently published paper (Hu et al., 2014).

We added the following paragraph in section 2.7 "Fixed intercepts and slopes are the same for all days and generated via conventional linear regression, while random intercepts and slopes vary independently for each individual day and are estimated by drawing samples with the same level of the grouping variable (day of year in this study) from the full set of observations. In this study, we generated fixed slopes for each predictor variable, but random slopes were only generated for

AOD and meteorological fields, since they represent time-varying variables. The fixed slopes (e.g., $b_1, \ldots, b_6$) denote the overall relationship for all days, and the random slopes (e.g., $b_{1,t}$, $b_{2,t}$) indicate the daily relationship among $PM_{2.5}$, AOD, and meteorological fields."

P. 25625, line 20: What are the definitions of "fixed and random intercept and slopes" and how are they obtained?

Response: Fixed and random intercept and slopes represent the fixed and random effects, respectively, in the mixed model of the relationship between dependent variable (e.g., observed $PM_{2.5}$) and our predictor variables. In this study, the fixed intercept and slopes were constant across days and generated via conventional linear regression, while random intercept and slopes vary independently for each individual day and are estimated by drawing samples of observations with the same level of the grouping variable (e.g, day of year) from the entire observations.

We added the following sentence in section 2.7. "Fixed intercepts and slopes are the same for all days and generated via conventional linear regression, while random intercepts and slopes vary independently for each individual day and are estimated by drawing samples with the same level of the grouping variable (day of year in this study) from the full set of observations."

P. 25626, line 1-13: "may include" – what are actually included? Do different met fields associated with different b0 and b2 values? It is hard to understand from eqn. (1). Maybe a detailed description (can be in a form of Appendix or Supplemental Material) is necessary if this method has not published in the literature. Are other land cover types considered other than forest cover?

Response: The actual model used for each year from 2001 through 2010 was included in the Supplemental Materials. Different meteorological fields have different fixed slopes (b_2), while the fixed intercept (b_0) was the same for each model. More details have been included in the manuscript and also can be found in our recently published paper (Hu et al., 2014). We also considered using impervious surface and cropland in the model in future studies.

We added a new table in the Supplemental Materials to show the final model structure for each year.

P. 25627, line 3-4: the sentence "That is,…" is confusing.

Response: a model is typically trained by maximizing its performance on training data. However, the efficacy of a model is not determined by its performance on training data but its ability of performing well on unobserved data. When a model has been overfit, it still can make perfect predictions using the training data simply by memorizing the training data in its entirety, but fail to make predictions using new or unobserved data because it has not learned to generalize. Consequently, an overfitted model generally has poor predictive performance.

We changed the sentence to "A model that has been over-fit could perform better on the data used to fit the model than unobserved data and thus generally has poor predictive performance."

P. 25627, line 15-16: As mentioned earlier, I don't understand why the fittings are done for each year individually, not for each season (or month).

Response: The dependent variable of our model was daily $PM_{2.5}$ concentrations. As a result, although the model was fitted for each year individually, the model generated daily $PM_{2.5}$ estimates first. Subsequently, the daily estimates were used to calculate the annual, seasonal, or monthly concentrations. Most importantly, our first-stage mixed effects model can effectively account for daily variability in the relationship between dependent and predictor variables by generating daily random slopes for AOD and meteorological fields, and our accuracy assessment showed that our model can predict daily $PM_{2.5}$ concentrations with satisfactory accuracy. Since our model can effectively capture the daily variability, it also can capture the seasonal or monthly variability. Fitting the model for each season or month was plausible. However, it inevitably reduced the sample size of the training data, which would increase model over-fitting and decrease the predictive performance. Comparison between models fitted for each season, each year, and all 10 years showed that models fitted for each year generally yield better prediction accuracy. As a result, we fitted the model for each year in this study.

We added the following paragraph in section 2.7 "Although the $PM_{2.5}$-AOD relationship might vary by season, our first-stage linear mixed effects model was able to incorporate daily variability in the relationship by generating day-specific random slopes for AOD and meteorological fields and thus should be able to capture the seasonal variability. In addition, by comparing the performances of models fitted for each season, each year, and all ten years, we found that the models fitted for each year generally yielded better prediction accuracy. Hence, in this study, we fitted the model for each year individually."

P. 25628, line 3-4: meteorological fields should have much stronger day-to-day, month-to-month, or season-to-season variations than year-to-year variations.

Response: We agree with the reviewer. We actually incorporated the day-to-day variations of meteorological fields in the calculation by including meteorological fields in the random effects terms of our first-stage model to calculate a daily random slope for each meteorological field for each individual day. The random effects of meteorological fields represent the daily variability in the relationship between $PM_{2.5}$ and meteorological parameters.

We added the following sentence in section 2.7 "The random slopes indicated the daily relationship among $PM_{2.5}$, AOD, and meteorological fields."

P. 25628, line 26: "…occur in the south of the study domain": From Figure 3 the high PM2.5 is the SE triangle area in the study domain, not "south".

Response: It has been corrected in the manuscript.

We changed "south" to "southeast".

P. 25869, line 1: What is the magnitude of the agriculture emission? Does it comparable with the urban industrial emission? Is the ag emission part of your predictors in equation (1)?

Response: Our recently published paper (Hu et al., 2014) examined the ground $PM_{2.5}$ measurements from five $PM_{2.5}$ monitoring sites located in the southeast of our domain and found high concentrations at those five monitoring sites. Hence, the high concentrations in this region could be real. However, this region does not contain large urban areas and major highways. Therefore, we suspect that high concentrations were due to agricultural emissions and biomass burning. Unfortunately, we do not have actual agricultural emission data at this point, and there are no ground monitors in those agricultural fields. It is hard to tell what the magnitude of the agriculture emission actually is and if it is comparable with the industrial emission. We will address this issue when the data is available. Due to the lack of data, we did not include agricultural emission in our model as a predictor, but we will include the data in the future studies when the data is available.

We added the following sentence in section 3.3 "However, actual agricultural emission data are needed for further validation."

P. 25869, line 5-6: Biomass burning emission is very seasonal. You should look the seasonal maps.

Response: We agree with the reviewer. We changed the sentence to "biomass burning also contributes to emissions of fine particles in the region with seasonal variations (Zhang et al., 2010)."
We have followed the second reviewer's suggestion and removed fire related results and discussion from the manuscript.

P. 25869, line 7: "corresponds well" – what is the criteria of "well"? Should have a quantitative measure instead of a subjective phrase.

Response: We added the following sentence in section 3.3 "the differences between estimated and observed $PM_{2.5}$ at on average 92% of the monitoring sites were within ± 3 $\mu g/m^3$ for the ten years, indicating a good agreement between them (Figure 5)." We also added a difference plot (Figure 5) to illustrate the difference between estimated and observed $PM_{2.5}$ at each monitoring site.

P. 25869, line 10, Fig. 5: What is the last panel in Fig. 5 that is never discussed?

Response: It has been corrected in the manuscript.
We added the following sentence in section 3.3 "Compared to the last panel of Figure 5 that illustrated the percentages of impervious surfaces and indicated the levels of urban development,".

P. 25629, line 16, "percent changes" - How do you obtain the % change? By linear fit of the time series, or by the difference between 2010 and 2001? It is said the change is "from 2001 to 2010" but on the next page it is said "between 2001 and 2010". Please clarify how the changes are calculated.

Response: The percent changes were the differences between 2001 and 2010 and were calculated as follows

$$(PM_{2.5,2010} - PM_{2.5,2001}) / PM_{2.5,2001} \times 100\%$$

We changed "from 2001 to 2010" to "between 2001 and 2010".

We added the formula in section 2.7.

P. 25630, line 1-15: The relationship between emission and PM2.5 should be better analyzed. As I mentioned at the beginning, if the two-stage model considers the emission as a predictor of PM2.5 concentration, why can't you pull out the emissions of each year to see if the increase or decrease is indeed due to the emission changes, and if they are of similar magnitudes?

Response: We obtained point emissions data from EPA National Emissions Inventory (NEI) facility emissions report for year 2002, 2005, and 2008, which were the only available emission data from 2001 to 2010 for us at this point. We added the emission data into Figure 8. The results showed that the fluctuations of $PM_{2.5}$ levels were in line with the increase or decrease of emissions in both the study area and the Atlanta metro area. The emissions were highest in 2005, and there were a significant drop of emissions in 2008. We also made a new figure (Figure 9) to examine the relationship between $PM_{2.5}$ estimates and point emissions. The R^2 was 0.93, 0.69, and 0.99 for year 2002, 2005, and 2008, respectively, indicating a strong correlation between $PM_{2.5}$ and point emissions. Thus, we believe that emissions might play a significant role in $PM_{2.5}$ levels in the region.

We added the emission data into Figure 8.
We changed the sentence in section 3.4 to "The $PM_{2.5}$ levels in the study region as well as the Atlanta metro area had a relatively small fluctuation from 2001 to 2007, while there was a significant drop in year 2008, which was in line with the trends of point emissions. The sharp decrease of $PM_{2.5}$ levels in 2008 was probably due to significant emissions reduction in 2008."
We added a scatter plot (Figure 9) of $PM_{2.5}$ estimates vs. point emissions.
We added following sentence in section 3.4 "Figure 9 illustrated the relationship between $PM_{2.5}$ estimates and point emissions. The R^2 reached 0.93, 069, and 0.99 for year 2002, 2005, and 2008, respectively, indicating a strong correlation between $PM_{2.5}$ and point emissions."

P. 25630-25631, section 3.4: Analysis in this section is too descriptive. More quantitative assessment is necessary. (a) Also there is no general declining trends – PM2.5 is significantly lower in the last three years, but there is no steady decline from 2001 to 2010. (b) For comparisons with the observation at the monitoring sites, you should compare your results with the obs at the same sites. Although you want to look at the trend at larger area, you could have shown the site comparisons on the same figure, maybe with a dotted line. (c) Again, do you have the emission to support your claim that the increase of sulfate due to the higher emissions from electric utilities and industrial boilers in 2005? (d) When was the emission reduction programs enacted? Which year is “recently”?

Response: We made a new figure (Figure 9) to examine the relationship between $PM_{2.5}$ estimates and point emissions. The R^2 was 0.93, 0.69, and 0.99 for year 2002, 2005, and 2008, respectively, indicating a strong correlation between $PM_{2.5}$ and point emissions. (a) It has been corrected in the manuscript. We changed the sentence to “The $PM_{2.5}$ levels in the study region as well as the Atlanta metro area had a relatively small fluctuation from 2001 to 2007, while there was a significant drop in year 2008, which was in line with the trends of point emissions.” (b) It has been added in Figure 8. We also added a dotted line for $PM_{2.5}$ estimates at monitoring sites to be compared with ground measurements. The results showed they are very similar. The average difference was 0.4 $\mu g/m^3$ for Atlanta metro area and 0.41 $\mu g/m^3$ for the entire study domain from 2001 to 2010. We added the following sentence in section 3.4 “our estimates over monitoring sites matched well with the ground measurements. The mean difference was 0.4 $\mu g/m^3$ for Atlanta metro area and 0.41 $\mu g/m^3$ for the study domain.” (c) Our point emissions data showed that emissions were the highest in 2005 (see Figure 8). The claim “the increase of sulfate due to the higher emissions from electric utilities and industrial boilers in 2005” was cited from the EPA report (EPA, 2008). At this point, we do not have detailed emissions data to support it. (d) We changed the sentence to “The decrease of $PM_{2.5}$ levels is due to emissions reduction program that have been enacted recently (EPA, 2007, 2011) such as the Clean Air Interstate Rule (CAIR) issued by EPA in 2005 (http://www.epa.gov/cair/index.html).”

EPA: National Air Quality - Status and Trends through 2007. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, RTP, NC 27711, 2008

P. 25631, section 3.5: This section does not tell us anything. Everyone expected that fire will have impact on PM2.5, so seeing some peaks and valleys of PM2.5 change with fire activity really is not any news. It would be more useful, given you have 10-year data, to estimate the quantitative magnitude of fire contribution to PM2.5 in fire-effected sub-domain from year to year, or even using only one-year data to show some quantified analysis.

Response: We intended to use MAIAC estimated $PM_{2.5}$ concentrations to examine the relationship between PM2.5 estimates and fire counts, which has not been done before. Due to the lack of more comprehensive data such as fire intensity and fire size data, we cannot conduct the quantitative analysis at this point. To follow the second reviewer’s suggestion, we decided to remove fire related results and discussion from this paper and focused on the trend analysis.

P. 25631 and Fig. 8: The fire activity certainly does not correspond to the PM2.5 changes anywhere within the study domain.

Response: Yes, fire activity does not correspond to the trend of overall $PM_{2.5}$ levels, which might be because fires were not the major contributor of fine particles in the region.

P. 25631, line 27: The estimated 13% contribution is not from the present work but from

Zhang et al., 2010. It should be clarified. This study did not show any quantitative number.

Response: It has been corrected in the manuscript. We changed the sentence from "although the contributions of fires to $PM_{2.5}$" to "although Zhang (2010) reported that the contributions of fires to $PM_{2.5}$".
We have removed this part from the manuscript.

P. 25632, line 8-9, the estimate at coarser scales "inevitably omit local spatial details" – but you did not use any local and spatial details in this study, so why the resolution matters?

Response: Our recently published paper (Hu et al., 2014) found that 1 km MAIAC predictions can distinctly reveal high concentrations along major highways, while 10 km MODIS predictions cannot. Likewise, in this study, we showed that large deceases of $PM_{2.5}$ levels between 2001 and 2010 occurred along major highways. In addition, we also related one pixel with large increase and another with large decrease between 2001 and 2010 in Atlanta Metro area to a single emission source. Those are all benefits of high resolution. However, this paper focused on the trend analysis, as the benefits of high resolution have been discussed in our recently published paper (Hu et al., 2014).

We changed the sentence to "inevitably omit local spatial details, as pointed out by Hu et al. (2014)."

P. 25632, last paragraph: Such statement can only be examined by comparing the PM2.5 changes over the entire domain as well as over the EPA monitoring sites.

Response: We compared the fluctuation of $PM_{2.5}$ from 2001 to 2010 over the entire domain with the fluctuations of both $PM_{2.5}$ ground measurements and our predictions over the EPA monitoring sites in Figure 8. The results showed that their trends were similar. However, $PM_{2.5}$ concentrations (both ground measurements and our predictions) over the EPA monitoring sites were generally higher than those over the entire domain. This might be because EPA monitoring sites were located in or near urban areas, while our estimates over the entire domain included both urban and rural concentrations. Hence, the $PM_{2.5}$ concentrations estimated from AOD over the entire domain might better and more thoroughly represent the true fluctuations of regional (including both urban and rural areas) $PM_{2.5}$ levels, since EPA measurements mostly represent urban situations.

We changed the sentence to "Most of the EPA FRM monitors are located in or near urban areas with generally higher $PM_{2.5}$ levels. Our results showed that both ground measurements and $PM_{2.5}$ estimates over the monitoring sites were generally higher than $PM_{2.5}$ estimates over the entire study domain. This is because observed and estimated $PM_{2.5}$ levels over the monitoring sites reflect mostly urban conditions, while $PM_{2.5}$ estimates over the entire study area account for both urban and rural areas, and therefore the temporal trends of $PM_{2.5}$ concentrations estimated from satellite AOD over the entire study domain might more thoroughly represent the true fluctuations of regional fine particle levels," in discussion.