The paper entitled '10 yr spatial and temporal trends of PM2.5 concentrations in the southeastern US estimated using high-resolution satellite data' by Hu et al., demonstrate the capabilities of current earth observations from satellites. Authors have selected very important and critical issue of spatial and temporal trends in surface PM2.5 mass concentration. Although, it is well written paper but authors fails to justify their PM2.5 estimation method and analysis. The methodology is poorly discussed in the paper. Authors do not provide proper reference or analysis on the claims about accuracy of satellite AOD product over urban regions. The paper in the current form is not recommended for the publication and I strongly suggest to resubmit. Followings are major concerns:

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. Page 4, Line 25: Chudnovsky et al., 2012 shows a correlation of 0.62 and 0.65 for MYD and MAIAC AODs with ground PM2.5. First, it is equivalent to existing 10km product and I would not call 0.65 strong correlations. The authors should reexamine their claims and avoid such statements. I would strongly suggest that author provides inter-comparisons of MAIAC and MYD AODs with AERONET measurements in their study region.

Response: While the number R~0.65 by itself is not impressive, it is quite a strong correlation considering the mismatch between above-ground $PM_{2.5}$ and column AOD. The regression analysis against AERONET data over the available AERONET sites on the East Coast USA shows that MAIAC does at least as good as MOD04, and usually a little better. A new analysis involving DRAGON-USA 2011 campaign in Baltimore-Washington area confirmed a very close performance of the two algorithms over heavily vegetated surfaces but a significant improvement of MAIAC over residential-urban areas (publication in preparation). We cannot provide the same comparison for the specific area of this study as it has no AERONET stations, however the general similarity of the aerosol type and of the study area in terms of surface "greenness" and urban cover (metropolitan Atlanta) to those of the Baltimore-Washington region allows us to conclude that MAIAC performance should be very similar to that of the Dark Target algorithm.

To account for the reviewer's comment, we changed the sentence from "MAIAC AOD has been demonstrated to be strongly associated with monitored $PM_{2.5}$ levels in the New England region " to " MAIAC AOD has been demonstrated to be correlated with monitored $PM_{2.5}$ levels in the New England region ".

We added the following sentence in section 2.3 "Due to the lack of long enough data records from AERONET sites in this region, a comparison between MAIAC AOD and AERONET measurements in the study area is not possible."

2. In the model (eq 1) there are several input parameters and each have different impacts on PM2.5 fields. I have failed to obtain any information on which of this input parameter is most critical in PM2.5 estimation? After including all non-satellite parameters as input in the model, how much AOD improves the PM2.5 estimation? Can this

model works in the areas where satellite AOD retrieval is not possible due to cloud cover.

Response: The objective of this study is to conduct a time series analysis of $PM_{2.5}$ in the study region, and thus accurate predictions are critical. Previous studies have shown that AOD, meteorological fields, and land use variables are all significant predictors of $PM_{2.5}$. In this study, we only selected statistically significant predictors that have significant contributions to the model. We assessed our prediction accuracy using model fitting and cross validation R^2, MPE, and RMSPE, and the results showed a good agreement between estimated concentrations and observations. We more focused on the overall performance of the model rather than the performance of each individual variable. To assess which parameter is the most critical is beyond the scope of this study. We conducted a test in our recently published paper to evaluate the impact of AOD on $PM_{2.5}$ estimation. The results showed that AOD was a significant predictor of $PM_{2.5}$. Without AOD, the prediction accuracy of the model dropped. Please see (Hu et al., 2014). We did not incorporate an AOD-filling method in this study, thereby this model did not work in areas where there were no AOD retrievals. We will address this issue in future studies.

We added the following paragraph in the introduction "Hu et al. (2014) introduced a geographically weighted regression (GWR) model as the second stage to account for the spatial variability in the $PM_{2.5}$-AOD relationship. The model used the MAIAC AOD as the primary predictor and meteorological fields and land use variables as the secondary predictors. Hu et al. (2014) futher pointed out that AOD is essential in the two-stage model framework in terms of predcition accuracy. The model can predict $PM_{2.5}$ with high accuracy and thus was adopted in this study."

3. Separate model for each year and then using the estimated data for trend analysis is hard to absorb. In this case it is very important to make sure there is no year to year bias in the estimations. It is not very clear why there is need to develop separate model for each year? Is there any specific relationship, which is changing from year to year? I believe AOD-PM2.5 relationship shows variability with seasons but why it should behave different in 2003 than 2005? Or there are other input parameters, which behaves differently in different years? Figure 7 clearly demonstrate that there is more variability in PM2.5 with different season than different year. I would develop models for seasons rather than years.

Response: The model could be built for each season, each year or all ten years. However, the predictions were first made for each day. Thus, how to improve the accuracy of daily predictions is critical. We calculated R^2, MPE, and RMSPE to evaluate the prediction accuracy, and the results showed a good agreement between $PM_{2.5}$ estimates and observations. In addition, we compared the model built for each season, each year, and all ten years and found that the model built for each year can generate better accuracy. As a result, we built the model for each individual year in this study. We agree with the reviewer that $PM_{2.5}$-AOD relationship showed variability by season. In fact, we assumed that $PM_{2.5}$-AOD relationship varied by day. Therefore, our first-stage linear mixed effects model was capable of capturing the daily variability in the $PM_{2.5}$-AOD relationship by calculating the daily random slopes for AOD.
Fixed slopes indicated the overall $PM_{2.5}$-AOD relationship, while daily random slopes revealed the daily variability in the $PM_{2.5}$-AOD relationship. Since our model can reveal the daily variability, it also should be able to capture the seasonal variability. Thus, it might not be

necessary to build models for each season, and moreover we found the models built for each year can reach better prediction accuracy. Hence we developed models 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. It is not very clear how the Figure 6 has been generated? Authors report the annual PM2.5 numbers for each year from 2001 to 2010 on Page 13, line 5. It is very clear from these numbers that there was hardly any changes in PM2.5 values from 2001 to 2007, in 2008 there was sudden decrease in PM2.5 values (also visible in Fig 7). Therefore, mapping 2001 to 2010 is miss leading as decreasing trend is not linear. There is hardly any discussion on 'why year 2008 has sharp decrease? Probably, difference between mean of 2001-2007 and 2008-2010 will be more meaningful.

Response: Figure 6 the percent change was generated as follows

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

We added 2002, 2005, and 2008 emissions data to Figure 8, which showed that there was a sharp decrease of point emissions from 2005 to 2008. EPA prepared NEI emission report every three years, and thus those are the only emission data we have at this point. We added the following sentence in section 3.4 "The sharp decrease of $PM_{2.5}$ levels in 2008 was probably due to significant emissions reduction in 2008."

We added the formula in the section 2.7.
We followed the reviewer's suggestions and added the differences between annual mean of 2001-2007, 2007-2008, and 2008-2010 in Figure 7.

We added the following paragraph in section 3.3 "We also illustrated the percent changes between 2001 and 2007, between 2007 and 2008, and between 2008 and 2010 in Figure 7 (c)-(h), since the decreasing trend between 2001 and 2010 was nonlinear with small decreases between 2001 and 2007 (on average ~5% for the entire domain and ~6% for the Atlanta metro area) and between 2008 and 2010 (~1% both for the entire domain and for the Atlanta metro area) and a sharp decrease between 2007 and 2008 (~14% for the entire domain and ~17% for the Atlanta metro area). Figure 7 (c) and (d) illustrated the percent changes between 2001 and 2007 in the entire domain and the Atlanta metro area, respectively. It showed that large decreases (> 10%) mainly occurred in the northern part of the domain and the mountainous region as well as urban built-up areas and along major highways in the Atlanta metro area, while increases (> 5%) appeared in the southern and southeastern parts of our domain as well as some residential and suburban regions in the Atlanta metro area. By comparing the 2002 with 2008 point emissions data, this might be due to the addition of extra emission sources in the region despite the fact that total emissions dropped significantly during this period. Figure 7 (e) and (f) illustrated the percent changes between 2007 and 2008. It showed that large decreases (> 10%) occurred in most of our domain. We could not confirm whether this was related to emissions reduction in the absence of 2007 emissions data. Figure 7 (g) and (h) illustrated the percent changes between 2008 and 2010. It revealed < 5% increases in many areas in the eastern part of the domain with increases in some areas exceeding 5%. On the other hand, many areas in the western part of the domain had < 5% decreases with decreases in some parts of the mountainous region exceeding 15%, In the Atlanta metro area, our results showed decreases (< 10%) in urban built-up areas and along major highways with some residential and suburban areas showing < 5% increases. Similarly, in the absence of 2010 emissions data, we could not examine whether these changes were associated with changes of emissions."

5. I would strongly suggest that temporal trends should be analyzed for each season separately. Also, impact of fires on PM2.5 does not fit here and can be left out for separate analysis/paper.

Response: We added the seasonal trend of $PM_{2.5}$ in Figure 8, and we added the following sentence in section 3.4 "The results also revealed seasonal variations of $PM_{2.5}$ levels with the highest concentrations occurring in summer and the lowest appearing in winter."
We followed the reviewer's suggestion and removed the fire related results and discussion from the paper and focused on the trend analysis.