**Referee #1**

General comment

In this paper the authors propose a statistical mixed effects model to calibrate MODIS/Terra aerosol optical depth (AOD) useful to predict daily PM2.5 over the New England (US) region of about 200x200 km2. The adopted model takes into account the day-to-day variability of the 'calibration coefficients'. The main hypothesis underlying the study is that these coefficients present a minimal spatial variability on a given day over the considered spatial scale. The paper is interesting because it derives the day-specific regression parameters (slope and offset) as a combination of (i) a fixed effect explaining the average effect for all study days overall sites and (ii) a random effect explaining the daily variability, the same for each site. The introduction of a 'site term', si, takes into account the effect of specific site in the model.

**The authors thank Referee 1 for the valuable comments. We hope that we have well addressed all of Referee 1's comments and improved our manuscript in response to the comments.**

Specific Comments and suggestions

1- Introduction

P9771; L20: Check if the Paciorek 's reference is the one related to the '2009'.

**We have added reference to a review paper, Hoff and Christopher (2009), and removed Paciorek et al. (2008). The review paper seems to be more appropriate for the discussion of previous studies.**

P9771; L25: It is better to substitute the sentence ": : :to calibrate: : :..(MODIS) AOD data to accurately predict PM2.5 ground concentrations." With the following: ": : :to calibrate : : : (MODIS) AOD data taking into account the daily variability in the prediction of PM2.5 ground concentrations".

**We have revised the sentence as recommended.**

2 – Methods

P9772; L6: Does "Collocated monitors" mean respect to the 10x10 km2 of grid cell size? Introduce a brief explanation.

**We have revised "monitors" to "samplers" and added a brief explanation in the manuscript to make the statement clear. Some monitoring sites have two $PM_{2.5}$ samplers, which are set up next to each other, to determine precision by comparing measurements**

**from those two samplers. Thus, we did not mean "collocated monitors" with respect to the 10x10 $km^2$ of grid cell size, but two identical $PM_{2.5}$ samplers at a single monitoring site.**

P9772;L9-10: ": : : including collecting samples every day, every third day, and every sixth day." It is not clear the reason of these different sampling frequencies: Is it a feature of a typical site?

**Yes, it is a feature of a typical site, and we utilized all the data available from the monitoring sites. The U.S. EPA decides a sampling schedule every year, and each site has one of the sampling frequency, every day, every third day, or every sixth day. Daily sampling may generate a rich dataset for $PM_{2.5}$, but it will be very labor- and cost-intensive. Thus, many EPA monitoring sites collect $PM_{2.5}$ samples with 1-in-3- or 1-in-6-day sampling schedules unless the increased amount of data are needed at a specific site.**

P9772; L16: Being the data provision 'under cloud-free condition', substitute ": : :, provide data every: : :" with ": : :, provide aerosol data every: : :"

**We have revised the sentence as recommended.**

P9772; L17: The "under cloud free" condition is not the only one condition for which aerosol data are provided. Depending on the region/season, maybe this condition is likely the most frequent. Suggestion: introduce briefly the other conditions.

**We have added the other conditions of data retrieval in the manuscript. In addition to cloudy conditions, aerosol data are missing in the conditions of high surface reflectance (e.g., snow- and ice-cover) and retrieval errors.**

P9772; PL19: The satellite features 10:30 a.m. and 1:30 p.m. are parameters setting the satellites orbit and not necessary the satellite overpass time over the study region.

**We agree with the referee, and have revised the sentence to the following not to confuse readers: "The Terra and Aqua satellites cross the equator at about 10:30 a.m. (descending orbit) and 1:30 p.m. (ascending orbit) local sun times, respectively, with a scanning swath of 2,330 km (cross-track) by 10 km (along-track at nadir)."**

P9772; PL21: ": : :primarily employed: : :" Are there other work wavelength employed in aerosol retrieval over land? Introduce comment about this issue.

**We have added the following discussion in the manuscript. The MODIS has 36 different wavelength channels, and seven of them (between 0.47 and 2.12 μm) are used for the aerosol retrieval. Among the 7 channels, three wavelength channels (0.47, 0.66, and 2.12 μm) are used to finally report AOD values at 0.55 μm, and many other channels are used for screening procedures (e.g., cloud and snow- and ice-cover).**

P9772; L24: see comment in the General Comment section.

**We have read the referee 1's comments in the General Comment section.**

P9773; L1: substitute “More details about MODIS satellite data: : :” with “More details about the retrieval of MODIS satellite aerosol data”.

**We have revised the sentence as recommended.**

P9773; L3: Please, it is proper to explain briefly the choice of 387 grid cell: for instance by using the criteria of the maximum distance from a monitoring site and considering the opportunity to exclude sea pixel.

**We have added discussion about the choice of the 387 grid cells in the manuscript. The predicted $PM_{2.5}$ concentrations from this study will be used for health effect studies, and we first selected all the grid cells where study subjects resided. Then, we excluded any grid cells along the coast that had sea surface of at least 2/3 of the cell area.**

P9773; L5-L15: The authors privilege Terra AOD data and use AOD Aqua to estimate the missing value of the same day of Terra considering the (Terra AOD / Aqua AOD) ratio. It is not clear the reason for which the mean AOD of the two satellite is not considered for each selected day. The same estimated ratio can be used also to estimate both Terra or Aqua AOD missing value. If the days where only one of the two retrievals is available are not so many, also the accuracy of the estimate of the (Terra AOD / Aqua AOD) ratio could be not so good.

**It is likely to be the most reliable to use the averages of the two satellites if both AOD values are daily retrieved. However, approximately 27.5% of the total Terra and Aqua AOD observations had both Terra and Aqua AOD values. Since there are many days where only one of the two satellites is available, this may cause bias in daily AOD values depending on the availability of two satellites on a given day. It is also possible to use Aqua AOD as a primary satellite and estimate missing Aqua AOD from Terra AOD using the (Aqua AOD / Terra AOD) ratio. Nevertheless, we decided to use the Terra as a primary satellite because there is a period of time that only Terra AOD is available (Year 2000, 2001, and a part of 2002). If our approach is to be applied to any of those years by us or others, it will be a better choice to use the Terra AOD as a primary satellite to be consistent with those potential studies.**

P9973; L14: Is the (Terra AOD / Aqua AOD) ratio monthly / seasonal depending?

**We have added the following discussion in the manuscript, “Due to the limited satellite data in winter (January, February, and December), the data were grouped into two seasons (warm and cold). The estimated (Terra AOD/Aqua AOD) ratios for the warm and cold seasons were similar, 1.15 and 1.17, respectively.”**

P9774; Eq(1): maybe the subscript ‘1’ in the fixed slope(beta_1 is not needed. While, it is necessary to better explain the symbol Sigma_beta variance-covariance matrix. In particular, is it necessary to introduce the beta subscript in this last symbol?

**We have removed the subscript '1' in the fixed slope and the beta subscript in the matrix. The variance-covariance matrix is a 2 × 2 matrix with diagonal terms representing the variances of the day-specific random intercepts and slopes and the off-diagonal term representing the covariance between these two terms.**

P9774; L8: It is better to use the italic 'N' instead of 'N'. Author should also consider that 'N' indicates the number of samplings in the Tables 1,3,4.

**We have changed 'N' to the italic '*N*' and added notes to the Tables: "N indicates the number of $PM_{2.5}$ samples" in Table 1, and "N indicates the number of days with both measured and predicted $PM_{2.5}$ concentrations" in Tables 3 and 4.**

P9774; L8: An estimate of the error term seems not to be reported in the paper.

**We have added an estimate ($\sigma^2$=3.70) of the error term ($\varepsilon_{ij} \sim N(0, \sigma^2)$) in the manuscript (P9774; L8 and P9777; L20).**

P9774; L25: Explain the reason of the choice of the limit of 5 microg m-3.

**We have added the following text "RMSE > 5 μg/m$^3$, which was approximately the 95$^{th}$ percentile of the RMSE values" in the manuscript. The 95$^{th}$ percentile of RMSE was 4.39 μg/m$^3$, and we were more conservative to choose the limit of 5 μg/m$^3$. However, the choice of either one did not make a large difference, since there was only one value of 4.82 μg/m$^3$ between the 95$^{th}$ percentile of 4.39 μg/m$^3$ and 5 μg/m$^3$.**

P9776; L1-5: It is not clear if the comparison between measured and predicted annual mean PM2.5 uses for each site the measured values of 99 days or of the whole data set of 365 days of 2003 years. It can be useful to report the difference in the two cases.

**The term 'annual mean' seems to be confusing, and we have changed to 'site mean' in the manuscript. As shown in Table 4 and Fig. 4, the comparison was based on those days with both the measured and predicted $PM_{2.5}$ concentrations at each site in 2003, and the number of the measured and predicted values varied by site. Consequently, we may not be able to report the difference in the two cases since the comparison was not performed based on either 99 days or 365 days.**

P9776; Section 2.5: In my opinion it is better to re-write this section. It must be clear that for each analyzed day (99) a map of 387 grid cell is obtained, using specific-day calibration coefficient and specific site term (daily independent). It is not so clear the choice of the 'regional' mean of predicted PM2.5. Does Fig.5 represent a 'mean difference' over the whole set of daily map presenting more than 50 grid cell predictions?

**We have revised Sections 2.5 and 3.3 to reflect Referee 1's recommendation to make the sections clear to readers. When there were 50 or more grid cell predictions on a given day, the average of the grid cell predictions was used as the regional mean $PM_{2.5}$ for that day. All the other days with less than 50 grid cell predictions were not considered. For the**

**selected days (50 or more grid cell predictions), we estimated a mean grid-specific difference (grid minus regional mean $PM_{2.5}$ concentrations) shown in Fig. 5.**

3 – Result and discussion

P9777; Section 3.1 It could be useful to introduce in this section some information about the distribution of the PM samplings throughout the year (at least in warm season and cold season) and also some information about the sampling technique and some related references.

**We have added the distribution of the $PM_{2.5}$ samplings by season (warm and cold seasons) in Section 3.1 and the sampling technique and the related reference (U.S. EPA, 2011) in Section 2.1. The average number of $PM_{2.5}$ samples across the monitoring sites was 136 (SD=78) with 73 (SD=41) in warm season and 63 (SD=39) in cold season. 24-hr $PM_{2.5}$ samples were collected on Teflon filters and were analyzed using gravimetric analysis.**

P9777; L18: What is the 'error' on the estimated alpha and beta_1?

**We have added the standard error estimates for alpha (SE=0.93) and beta (SE=1.53) in the manuscript.**

P9778; L5. It is necessary to remember that Table 3 and Fig. 2A (not 2a) report the comparison between the studied model and the fixed model of Wang and Christopher, 2003. Moreover, it is not easy to understand the reason of 576 measured and predicted daily PM2.5 in the caption of Fig.2.

**Table 3 and Fig. 2A were more intended to show the performance of the mixed effects model, and we reported the comparison between the mixed effects model and the fixed model of Wang and Christopher (2003) in Table 4 and Fig. 3 and 4. We have added the reference (Wang and Christopher 2003) in P9779; L20 to remind readers of the fixed model. We have also added the following text in the caption of Fig. 2, "All pairs of the measured and predicted $PM_{2.5}$ concentrations at the 26 monitoring sites (576 pairs in total) are plotted."**

P9778; L12: put Fig. 2B, instead of Fig. 2b

**We have changed Fig. 2b to Fig. 2B as recommended.**

Table 1: At least in this table add the geographical references of the monitoring site.

**We have added the site coordinates (latitude and longitude) in Table 1.**

Figures 1 and 5: Please add a geographical grid reference and at least the indication of the interstates I-91 and I-95.

**We have added a geographical grid reference and the indication of the interstates I-91 and I-95 in Figures 1 and 5.**

Figure 3 and Table 4: It is necessary to add the reference of Wang and Christopher for the regression model.

**We have added the reference in Figure 3 and Table 4.**