Interactive comment on “A novel calibration approach of MODIS AOD data to predict PM$_{2.5}$ concentrations” by H. J. Lee et al.

Anonymous Referee #1

Received and published: 24 June 2011

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 PM$_{2.5}$ over the New England (US) region of about 200x200 km$^2$. 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’, $s_i$, takes into account the effect of specific site in the model.

Specific Comments and suggestions

1- Introduction
P9771; L20: Check if the Paciorek’s reference is the one related to the ‘2009’.
P9771; L25: It is better to substitute the sentence “...to calibrate...(MODIS) AOD data to accurately predict PM$_{2.5}$ ground concentrations.” With the following: “...to calibrate ...(MODIS) AOD data taking into account the daily variability in the prediction of PM$_{2.5}$ ground concentrations”.

2 – Methods
P9772; L6: Does “Collocated monitors” mean respect to the 10x10 km$^2$ of grid cell size? Introduce a brief explanation.
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?
P9772; L16: Being the data provision ‘under cloud-free condition’, substitute “..., provide aerosol data every...” with “..., provide aerosol data every...”
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.
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.
P9772; PL21: “...primarily employed...” Are there other work wavelength employed in aerosol retrieval over land? Introduce comment about this issue.
P9772; L24: see comment 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”.

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.

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.

Is the (Terra AOD / Aqua AOD) ratio monthly / seasonal depending?

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?

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.

An estimate of the error term seems not to be reported in the paper.

Is 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.

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?

3 – Result and discussion

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.

What is the ‘error’ on the estimated alpha and beta_1?

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.

put Fig. 2B, instead of Fig. 2b

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

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

It is necessary to add the reference of Wang and Christopher for the regression model.

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 9769, 2011.