Identifying Optimal Temporal Scale for the Correlation of AOD and Ground Measurements of PM$_{2.5}$ to Improve the Model Performance in a Real-time Air Quality Estimation System

Hui Li[a], Fazlay Faruque[a], Worth Williams[a], Mohammad Al-Hamdan[b], Jeffrey Luvall[b], William Crosson[b], Douglas Rickman[b], Ashutosh Limaye[b]

[a] University of Mississippi Medical Center, Jackson, Mississippi 39216
[b] NASA Marshall Space Flight Center, Huntsville, Alabama 35812

Abstract

Aerosol optical depth (AOD), an indirect estimate of particle matter using satellite observations, has shown great promise in improving estimates of PM$_{2.5}$ air quality surface. Currently, few studies have been conducted to explore the optimal way to apply AOD data to improve the model accuracy of PM$_{2.5}$ surface estimation in a real-time air quality system. We believe that two major aspects may be worthy of consideration in that area: 1) the approach to integrate satellite measurements with ground measurements in the pollution estimation, and 2) identification of an optimal temporal scale to calculate the correlation of AOD and ground measurements. This paper is focused on the second aspect on the identifying the optimal temporal scale to correlate AOD with PM$_{2.5}$. Five following different temporal scales were chosen to evaluate their impact on the model performance: 1) within the last 3 days, 2) within the last 10 days, 3) within the last 30 days, 4) within the last 90 days, and 5) the time period with the highest correlation in a year. The model performance is evaluated for its accuracy, bias, and errors based on the following selected statistics: the Mean Bias, the Normalized Mean Bias, the Root Mean Square Error, Normalized Mean Error, and the Index of Agreement. This research shows that the model with the temporal scale of within the last 30 days displays the best model performance in this study area using 2004 and 2005 data sets.

1. Introduction

Aerosol optical depth (AOD), derived from satellite measurements using Moderate Resolution Imaging Spectrometer (MODIS), is a measure of atmospheric extinction of radiance, due to the presence of aerosols, through a vertical column in the atmosphere. It offers indirect estimates of particle matter. Previous research showed a significant

1

positive correlation between satellite-based measurements of AOD and ground-based measurements of particulate matter with aerodynamic diameter less than or equal to 2.5 micrometers ($PM_{2.5}$) and aerodynamic diameter less than or equal to 10 micrometers (PM_{10}) (Chu, 2006; Gupta et al., 2006; Li et al., 2003; Engel-Cox et al., 2004). In addition, satellite observations have shown great promise in enhancing air quality event monitoring (Engel-Cox et al., 2004, Hutchison et al., 2003), and improving estimates of $PM_{2.5}$ air quality surface (Gupta et al, 2006; Kumar et al., 2007), and improving national air quality forecasts (Al-Saadi et al., 2005). Research shows that correlations between AOD and ground $PM_{2.5}$ are affected by a combination of many factors, such as inherent characteristics of satellite observations, aerosol optical depth algorithms, errors of estimate of regression models, terrain, cloud cover, height of the mixing layer, relatively humidity, wind velocity, temperature, and sea-level atmospheric pressure conditions (Kumar et al., 2007; Gupta et al., 2006), thus may vary widely in different regions, different seasons, and even on different days in one location. For example, Engel-Cox's study (2004) finds that the correlations are stronger in the eastern half of the United States, while the correlations are weak in the western United States. They believed that some of the general variation between AOD and PM measurements is caused by the artifact of linear analysis, different terrain conditions, and inherent differences in the datasets.

The temporal scale of the correlation between AOD and $PM_{2.5}$ in this study is defined as the number of immediate latest days to cumulate AOD data for a given day's run (see Figure 1). Analysis of correlating AOD with ground measured $PM_{2.5}$ on a day-to-day basis in this research suggests the temporal scale in determining their correlations needs to be considered to improve air quality surface estimates, especially when satellite observations are used in a real-time pollution system. In addition, correlation coefficients between AOD and ground $PM_{2.5}$ cannot be predetermined in real-time daily air quality estimation and need to be calculated for each day's run in a real-time system, because the coefficients can vary in different seasons and even different days. Few studies have been conducted to explore the optimal way to apply AOD data to improve model accuracies of $PM_{2.5}$ surface estimation in a real-time air quality system. We believe that two major aspects may be worth considering when applying satellite data to

2

improve the performance of pollution surface models: 1) the approach to integrate satellite measurements with ground measurements for the pollution estimation, and 2) identification of an optimal temporal scale in calculating the correlation of AOD and ground measurements. This paper will focus on the second aspect and discuss the best temporal scale to calculate the correlation of AOD and ground particle matter data to improve the results of pollution models in a real-time system.

Figure 1. Schematic diagram of temporal scales in calculating the correlation of AOD and ground measurements in this study

2. Real-time PM$_{2.5}$ estimation system

The near real-time PM$_{2.5}$ estimation system used in this research is built from a PM$_{2.5}$ surface model, originally developed by NASA Marshall Space Flight Center (MSFC). This surface model was improved to integrate with a real-time geo-spatial health surveillance system developed at the University of Mississippi Medical Center. The system estimates daily average PM$_{2.5}$ concentration for Mississippi and its neighboring states of Arkansas, Tennessee, Alabama, Florida, and Texas, using NASA MODIS AOD data on board Terra and Aqua and EPA air quality ground measurements from the AirNow gateway system. The system calculates daily average ground-level PM$_{2.5}$ in a batch mode on a daily basis with 2-days delay due to the delay of satellite data received in the system. The model adopts the same spatial resolution as that of satellite data as its grid surface outputs (about 10*10 km). Ground measurements of air quality from EPA monitoring stations and satellite-derived AOD from MODIS instruments on aboard NASA's Earth Observing System Terra and Aqua satellites are two major data sources to estimate daily ground-level pollutant surface in the system. The ground measurements of daily average PM$_{2.5}$ data are obtained automatically daily from the AirNow gateway system using the File Transfer Protocol (FTP). MODIS-derived AOD data from Terra and Aqua, stored in a Hierarchical Data Format (HDF), are obtained automatically from the NASA-Goddard Earth Sciences (GES) FTP site using the FTP. Satellite-derived AOD is a Level 2 atmospheric product from MODIS instruments on board Terra and Aqua platforms at a spatial resolution of about 10*10 km nearly covering the entire Earth surface everyday. AOD data include day-time and night-time

hour products. In the system, only day-time hour products are used. Because MODIS AOD data from Terra are found to have better relationship with ground measurements of $PM_{2.5}$ than MODIS AOD from Aqua, AOD from Terra are used in the model by default. However, whenever the relationship is found not statistically significant, the system will automatically switch to use AOD data generated onboard Aqua. If neither relationship is statistically significant, then only ground measurements are used in the model.

The system includes the following three main components: 1) AOD-$PM_{2.5}$ linear regression models for AOD-derived $PM_{2.5}$; 2) a surface model to interpolate AOD-derived $PM_{2.5}$ and ground measurements of $PM_{2.5}$ to a continuous grid surface respectively using B-spline algorithms; and 3) an approach to integrate the two interpolated surfaces above into a final surface output if a significant relationship is found between them on each calculated day; otherwise, only ground measurements are used for the model output. The model domain is shown in Figure 2, which also shows the distribution of the monitoring stations used in the air quality models.

Figure 2. Model domain and monitoring stations for the air quality system

3. Methodology

To identify the optimal temporal scale for the AOD-$PM_{2.5}$ correlations, we chose the following five different temporal scales to evaluate their impact on the performance of the daily-basis pollution surface models in both 2004 and 2005: 1) within the last 3 days, 2) within the last 10 days, 3) within the last 30 days, 4) within the last 90 days, and 5) the time period with the highest correlation in a year (August-October in 2004 and June-September in 2005). For the first four temporal scales, the regression analysis was conducted on the fly to determine the significant relationship between AOD and $PM_{2.5}$, based on the p-value at a significant level of 5% on each model-running day. First, each EPA monitoring station is identified in the study area by its longitude and latitude. Second, all corresponding pixels of satellite observations within the 0.1 degree distance range of each station were identified in the AOD data set on each accumulated day, which was inside the evaluated temporal scale range. Third, for each involved AOD daily data, only the fist three identified pixels, closest to their paired station, are kept for

further process. Fourth, the pairing AOD value of each station is estimated by averaging the AOD values from all identified pixels in the above process on each accumulated day. Once the satellite measurements are paired with all stations on a modeled day, a linear regression model is fitted to the identified paired data on a day-by-day basis. When their relationship is considered statistically significant (p-value less or equal to 5%), AOD data are determined to be used in the model. As to the last temporal scale, a predetermined regression model is used for the model estimation in the defined time period in each evaluating year (August to October, 2004 and June to September, 2005)).

To make the accuracy assessment subjective, a station site with the ID 280810005 (seen in Figure1) was left out in the air quality estimation and was only used for the performance evaluation. The model performance is evaluated for its accuracy, bias, and errors based on the following selected statistics: the Mean Bias (MB), the Normalized Mean Bias (NMB), the Root Mean Square Error (RMSE), Normalized Mean Error (MNE), and the index of agreement (IOA). They are defined below:

$$MB = \frac{1}{N} \sum_{1}^{N} (C_{m-} C_o) \tag{1}$$

$$NMB = \sum_{1}^{N} (C_m - C_o) / \sum_{1}^{N} C_o 100\% \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{1}^{N} (C_m - C_o)^2} \tag{3}$$

$$MNE = \sum_{1}^{N} | C_m - C_o | / \sum_{1}^{N} C_o 100\% \tag{4}$$

$$IOA = 1 - \sum_{1}^{N} (C_m - C_o)^2 / \sum_{1}^{N} (| C_o - \overline{C_o} | + | C_m - \overline{C_o} |)^2 \tag{5}$$

Where C_m and C_o are modeled and observed values, respectively. $\overline{C_o}$ is average observed value with the sample size N.

4. Results

4.1 The model performance

The results of the model performance for each evaluating temporal scale are displayed in Table 1 and Figure 3. Since the first and second temporal scales (within the last 3 days and within the last 10 days) are the two closest temporal scales in reference to the modeled day, they are expected to show a better model performance. Surprisingly, the models with these two temporal scales showed the highest biases (MB and NMB), consistent in both 2004 and 2005. The model with the temporal scale of last 3 days also had the highest errors (RMSE and MNE) in both 2004 and 2005, and thus was believed to have the worst model performance. Its IOA value, the lowest among the five chosen temporal scales, also supports this conclusion. The model with the fifth temporal scale (the predefined seasons with the highest correlations) had higher biases (MB and NMB) in both 2004 and 2005. This result is expected, because it only used satellite observations in the predefined time period and failed to use those observations having a significant correlation with ground measurements outside the predefined time period; thus, it is a poor strategy to utilize satellite data for building a model.

The temporal scale of last 30 days generated higher model biases than did the temporal scale of last 90 days, whereas it caused lower model errors in 2004. The IOA index suggests that the model with the temporal scale of last 30 days might have better performance in 2004. However, these two models showed different performance in 2005. The temporal scale of last 30 days caused the same biases and IOA as did the temporal scale of last 90 days, but the first temporal scale caused higher model errors. Thus, it is difficult to determine the performance of these two models by just looking at those chosen statistics.

Table 1. Accuracy assessment of the air quality models using different temporal scales for AOD-PM$_{2.5}$ correlations in 2004 and 2005

4.2 Distribution of R-Squared values across different temporal scales

A key factor possibly impacting the performance of these models is the correlation coefficients of AOD and ground PM$_{2.5}$ calculated for a model day's run. Better correlation coefficients will certainly improve the model performance, whereas poorer correlation coefficients will degrade the model performance. To analyze their correlation coefficients, a histogram of R-Squared values of AOD and ground measurements of

PM$_{2.5}$ for each evaluated model (except the fifth temporal scale) in 2004 and 2005 is displayed in Figure 4. It clearly shows that the first and second temporal scales have the least days, with a significant correlation between satellite observations and ground data in each year. Moreover, their R-Squared values are also generally lower in 2004 and 2005 compared to other models with different temporal scales. This fact indicates that a short temporal scale is not a good choice to determine the correlation of satellite and ground observations. As mentioned before, their correlation is affected by many factors. One possible reason is that the correlations in short temporal scales contain more noises because of the impacts of other factors such as inherent characteristics of satellite observations, weather conditions, errors of the regression models, etc. When a longer temporal scale is used for the correlations, those noises might be smoothed by the time factor and reduced by larger sample data (see further discussion in Section 5.2), and thus the correlation may have better quality. This explains why a short temporal scale is not a good choice in the model construction. However, if a temporal scale is too long, the correlation might be over smoothed by the time factor, and thus it will not reflect their real relationship in a specific short time period. This explanation can be confirmed by looking at the following two distribution patterns of the R-Squared values in different temporal scales: 1) the longer a temporal scale, the greater number of days showed significant association between AOD and ground data in both 2004 and 2005; and 2) the highest correlation (R-Squared larger than 0.6) only appeared in a middle-range temporal scale in 2004 and 2005. None of the days showed the R-Squared values larger than 0.6 when the temporal scale of last 90 days was used in both 2004 and 2005, which possibly contributed to over smoothing by the long time period. That might explain why the model with the temporal scale of last 30 days tended to have a higher IOA value than did the model with the temporal scale of last 90 days in 2004. By considering the model performance as well as the distribution patterns of the R-Squared values, it is also believed that the model with the temporal scale of last 30 days is the best model in utilizing satellite data in 2005, because it has the highest frequency (23 days), with R-Squared values larger than 0.6 in this model (the highest R-Squared level), compared to none in the model with the temporal scale of last 90 days in 2005.

Figure 3. Accuracy assessment of the air quality models using different temporal scales for AOD-PM$_{2.5}$ correlations in 2004 and 2005

Figure 4. Histograms of statistically significant R-Squared values between AOD and ground measurements of PM$_{2.5}$ at a significant level of 5% with five different temporal scales in 2004 and 2005.

4.3 Examination of the model outputs

Figure 5 shows the modeled results against the ground data from July to August, 2005 using the temporal scale of last 30 days and the temporal scale of last 90 days. Generally, MB and RMSE values ranged from 1.0 to 12.0 and 2.0 to -13.0 respectively in both temporal scales used in the models. MB and RMSE were low in most of the days. However, on some specific days, the errors and biases of the model were increased, which showed a repeated pattern. Many potential factors may be responsible for this phenomenon, such as inherent characteristics of satellite observations, errors of estimate of regression models, and weather conditions, etc. More research is needed to determine the real reasons for the observed patterns of the model performance.

Figure 5. Comparison of daily time series of the modeled and observed PM2.5 concentrations (Grand data from an EPA monitoring station site: 280810005)

5. Discussion
5.1 Impact of data fusion on the model performance

The five selected statistics of the model performance show only slight differences among the five evaluated temporal scales for the correlation of AOD and ground data, especially RMSE and NME. The reason is not because the temporal scales of the correlation do not have much impact on the model performance, but because the weight of satellite observations was only given 10% compared with the weight of ground data given 90% when integrating two interpolated surfaces of satellite observations and ground data into the model output. Therefore, the major contribution of the model

outcome comes from the ground data. Consequently, it is reasonable to believe that the slight differences of the selected statistics still truly represent the impact of the temporal scales on the model performance, therefore the conclusion is reliable. Although this paper does not cover the topic of the integration approach of these two data sets (satellite and ground data), it might be worth pointing out that their weight should be dependent on their correlation instead of a prefixed value. More research is recommended to focus on developing an optimal solution to integrate satellite-derived AOD and ground measurements to improve the model performance in air quality surface estimation.

5.2 Optimal Temporal scale for the correlation of AOD and ground data

This research shows the optimal temporal scale for the correlation of AOD and ground data of $PM_{2.5}$ is the latest 30 days among the five chosen temporal scales in the study area. However, many factors mentioned above may have potential impact on this conclusion. One of those factors is the errors of regression models. To analyze the regression models and their potential impact on the model performance at different temporal scales, fitted regression models and their corresponding paired data (AOD from Terra) at the three temporal scales the latest 3 days, latest 10 days, and latest 30 days were displayed respectively in a selected time period (August, 2005) in Figure 6 and Figure 7. The relationship between AOD and ground PM2.5 was not statistically significant on many days when the small temporal scale (latest 3 days) was used in the model, which might be influenced by the noises contained in the data contributed by other confound factors mentioned in Section 4.2. In contrast, their relationship was statistically significant in the examined period when the temporal scale of the latest 30 days was used in the model [Figure 7 (4)]. According to Figure 6 and 7, it was also found that the model outcomes tended to be affected by the outlier points, which might be linked with clouds and other atmospheric conditions, at the short temporal scales [Figures 6(1), 6(2), 6(4), 6(6), 6(9), 7(2), and 7(3)] because of the small sample data set. Therefore, it is confirmed that the regression models tended to have more errors when small temporal scales were used in the model. This further explains why the model had higher errors and biases when the two small sample scales, the latest 3 days and 10 days, were used.

Insert Figure 6.

Insert Figure 7.

Although the conclusion was based on the linear regression model for the AOD-$PM_{2.5}$ correlation, a non-linear model through power transformations of predictors and dependent variables, such as logarithmic, square root, or cube root, had not shown any obvious improvement of the model performance. Therefore, it is believed that it is a good strategy to use linear regression models, determined on a monthly basis, for estimating particulate matter in the models. However, the finding in this study area might not apply to other areas considering the multiple factors that influence the correlation of AOD and ground measurements of $PM_{2.5}$ as well as their variation over space and time. Similar research in other areas will be valuable to conduct in the future.

5.3 Areas to improve

Previous research shows that the effect of weather conditions, such as wind velocity, relative humidity, temperature, and atmospheric pressure, can confound the AOD-$PM_{2.5}$ association (Kumar et al., 2007). However, the identified optimal temporal scale in this study did not consider the potential impact of these confound factors; thus, it is not clear what impact the weather factors might have on our conclusion. Future study to incorporate these factors to determine the optimal temporal scale is likely to answer this important question and may improve the model performance through a better strategy on using satellite observations.

Second, the MODIS AOD data were acquired at a specific time once on a day, whereas the ground measurements of $PM_{2.5}$ are daily average values over 24 hours. Therefore, the time frame for these two data was not matched. The uncertainty in AOD-$PM_{2.5}$ association is likely increased as the time of $PM_{2.5}$ observation deviates from the overpass time of satellites (Kumar et al., 2007). Since the AOD data used in this research were acquired in the daytime, it is reasonable to expect that the ground measurements of daily average $PM_{2.5}$ over 8 hours or ground $PM_{2.5}$ recorded matched the satellite overpass time will improve their relationships and thus improve the model

performance. Therefore, it is also needed to evaluate whether these two new measurements of $PM_{2.5}$ will impact our conclusion on the optimal temporal scale in this research.

6. Conclusion

This research shows that the model with the temporal scale of the last 30 days displays the best model performance in estimating surfaces of $PM_{2.5}$; thus, the temporal scale of the last 30 days is believed to be the best strategy to utilize satellite observations to improve estimation of particle matter in the study area. It is necessary to point out that this conclusion is not considering the confounding impact of weather conditions on their association. It will be a valuable study that incorporates these weather conditions in determining the optimal temporal scale in future research.

Reference

Al-Saadi Jassim, J. Szykman, R. B. Pierce, C. Kittaka, D. Neil, D.A. Chu, et al., 2005. Improving national air quality forecasts with satellite aerosol observation. AMS Journals 86 (9): 1249-1261.

Chu, D.A., 2006. Analysis of the relationship between MODIS aerosol optical depth and $PM_{2.5}$ in the summer-time US. *Proceedings of the SPIE* 6299: 629903.

Engel-Cox, A. Jill, C.H. Holloman, B.W. Coutant, R.M. Hoff, 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment* 38: 2495-2509.

Gupta P., S. A. Christopher, J. Wang, R. Gehrig, Y. Lee, N. Kumar, 2006. Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment* 40: 5880-5892.

Hutchison, K.D., 2003. Applications of MODIS satellite data and products for monitoring air quality in the state of Texas. *Atmospheric Environment* 37: 2403-2412.

Kumar, N., A. Chu, A. Foster. 2007. An empirical relationship between $PM_{2.5}$ and aerosol optical depth in Delihi Metropolitan. *Atmospheric Environment*. 41: 4492-4503.

Li, Chengcai, J. Mao, K.H. Lau, 2003. Research on the air pollutants in Beijing and its surroundings with MODIS aerosol products. Proceeding of SPIE. Vol 4891: 419-430.
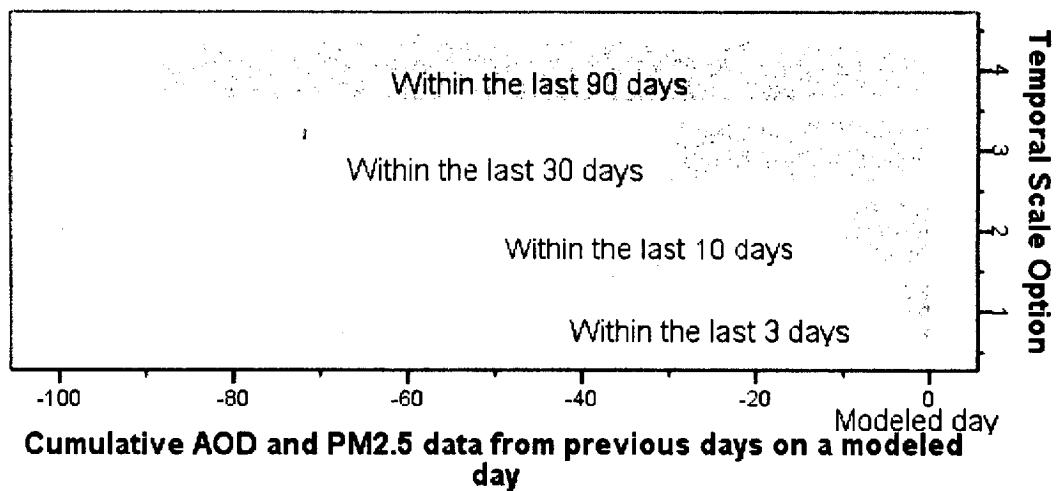
Figure 1. Schematic diagram of the first four temporal scales in calculating the correlation of AOD and ground measurements in this study
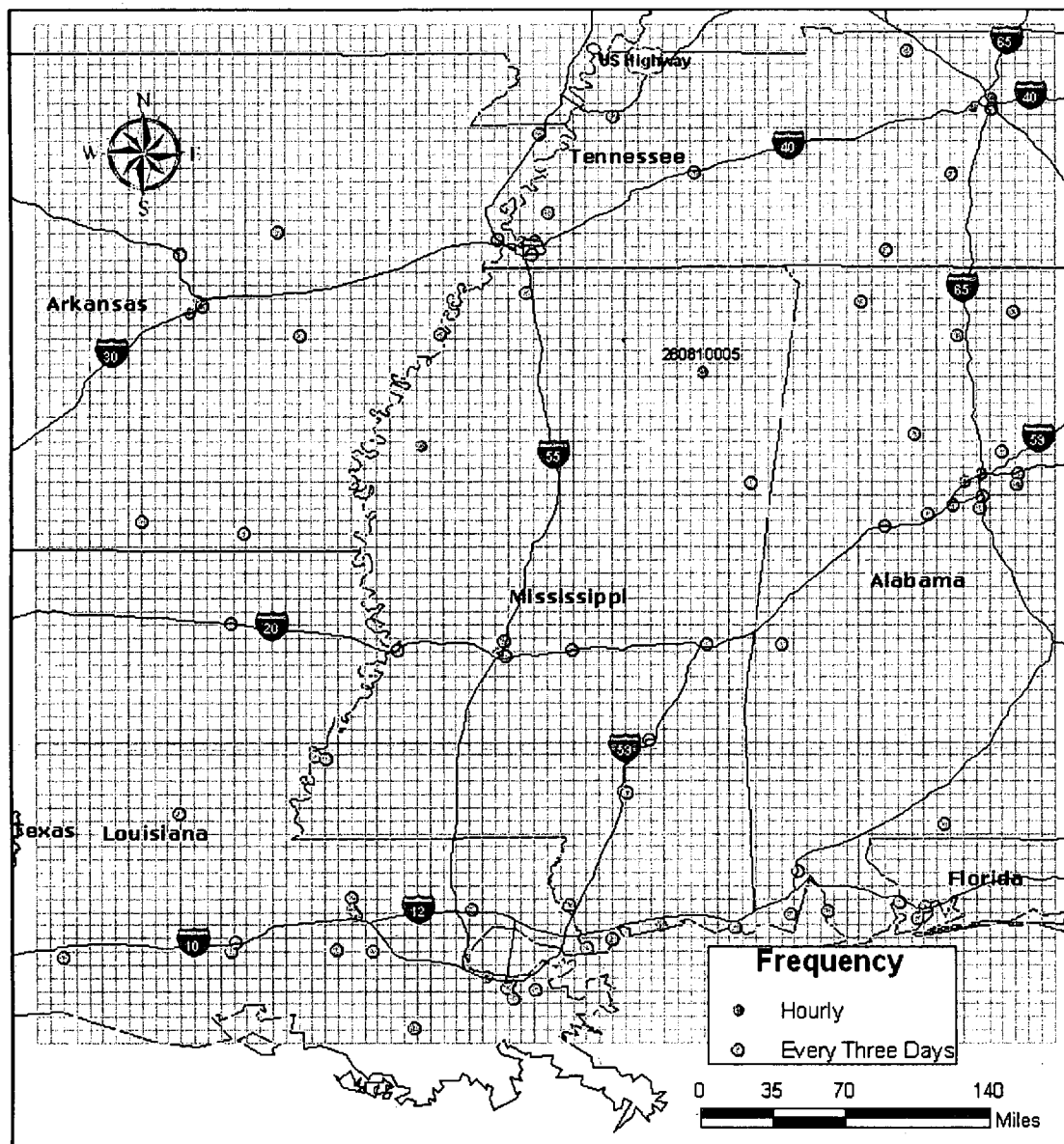
Figure 2. Model domain and monitoring stations for the air quality system

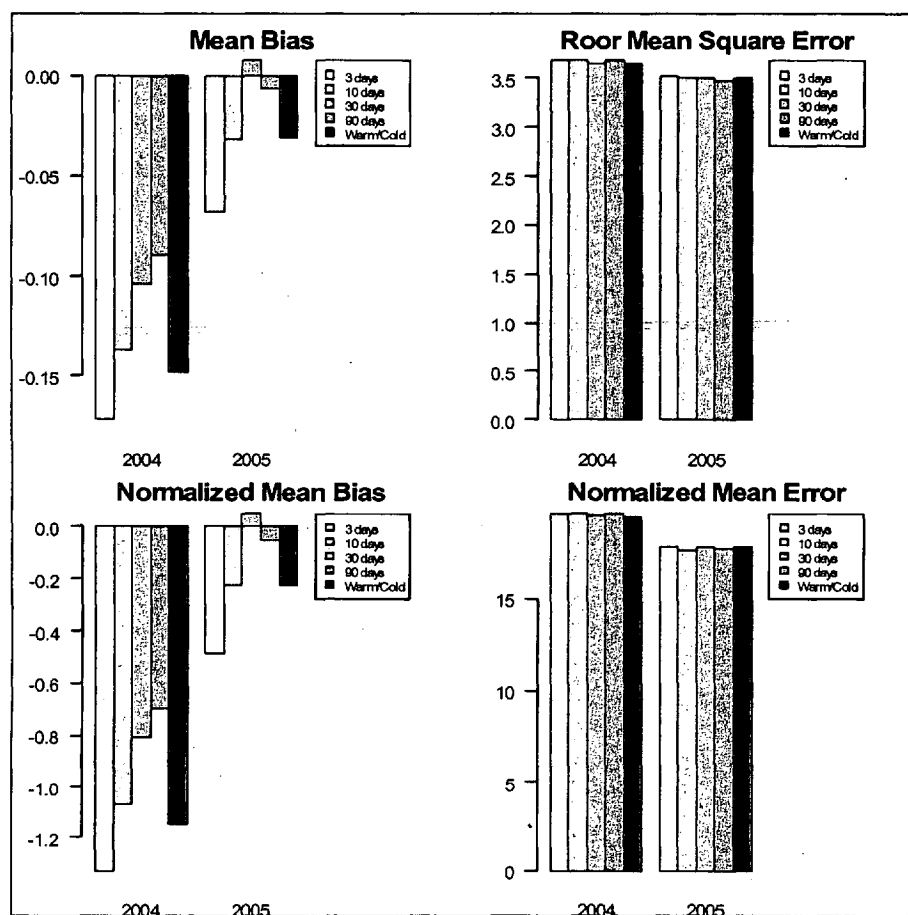Table 1. Accuracy assessment of the air quality models using different temporal scales for AOD-PM$_{2.5}$ correlations in 2004 and 2005

| Year	Temporal scales (Cumulative previous days)	MB	NMB	RMSE	MNE	IOA
	3	-0.172	-1.33	3.68	19.70	0.906
2004	10	-0.137	-1.07	3.68	19.70	0.906
	30	-0.104	-0.81	3.65	19.60	0.908

	90	-0.090	-0.70	3.68	19.70	0.907
	Season with highest correlation	-0.148	-1.15	3.65	19.50	0.908
2005	3	-0.068	-0.49	3.52	17.90	0.943
	10	-0.032	-0.23	3.50	17.70	0.944
	30	0.007	0.05	3.50	17.90	0.944
	90	-0.007	-0.05	3.47	17.80	0.944
	Season with highest correlation	-0.031	-0.22	3.51	17.90	0.943

Figure 3. Accuracy assessment of the air quality models using different temporal scales for AOD-PM$_{2.5}$ correlations in 2004 and 2005
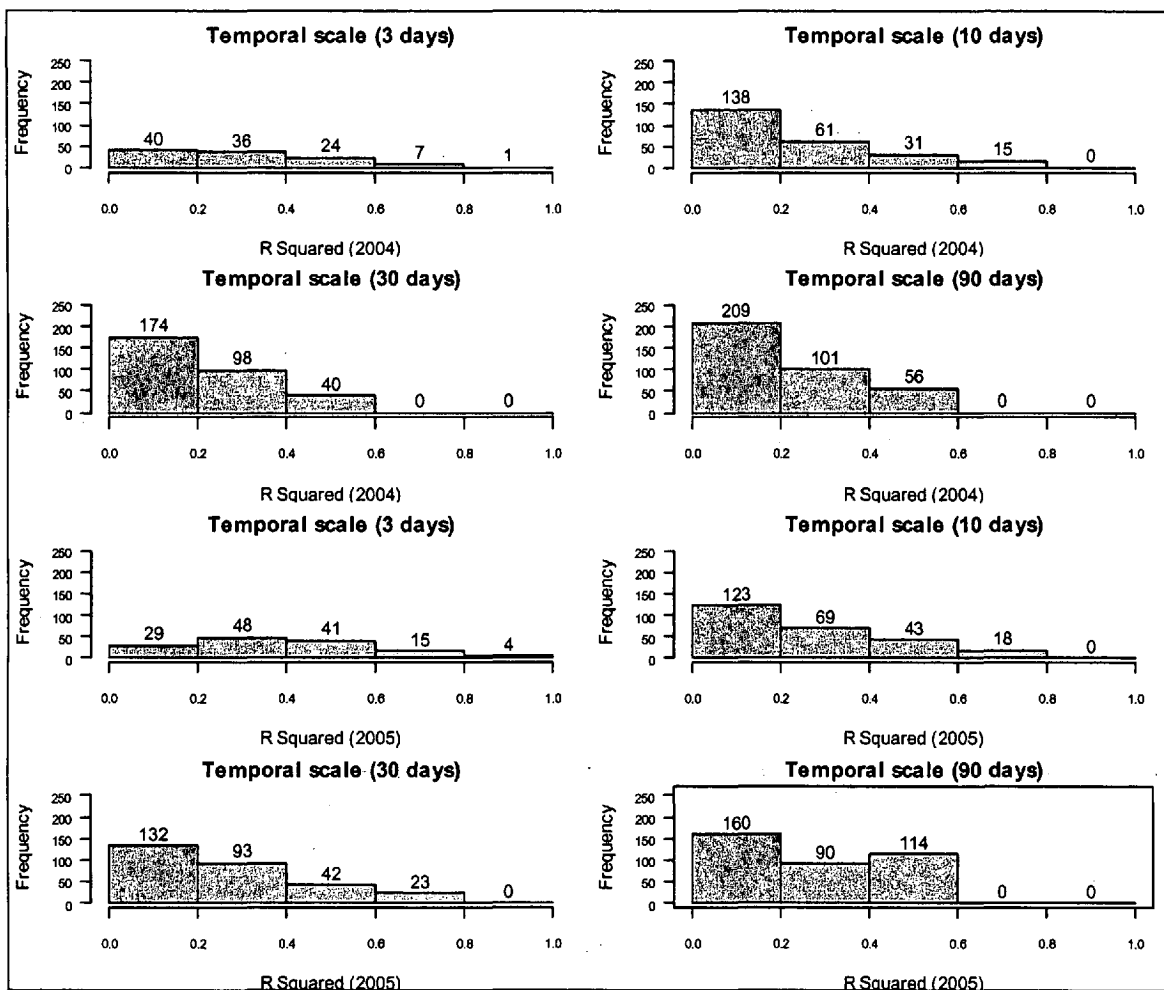
Figure 4. The histogram of R-Squared values of AOD and ground measurements of $PM_{2.5}$ with five different temporal scales in 2004 and 2005.
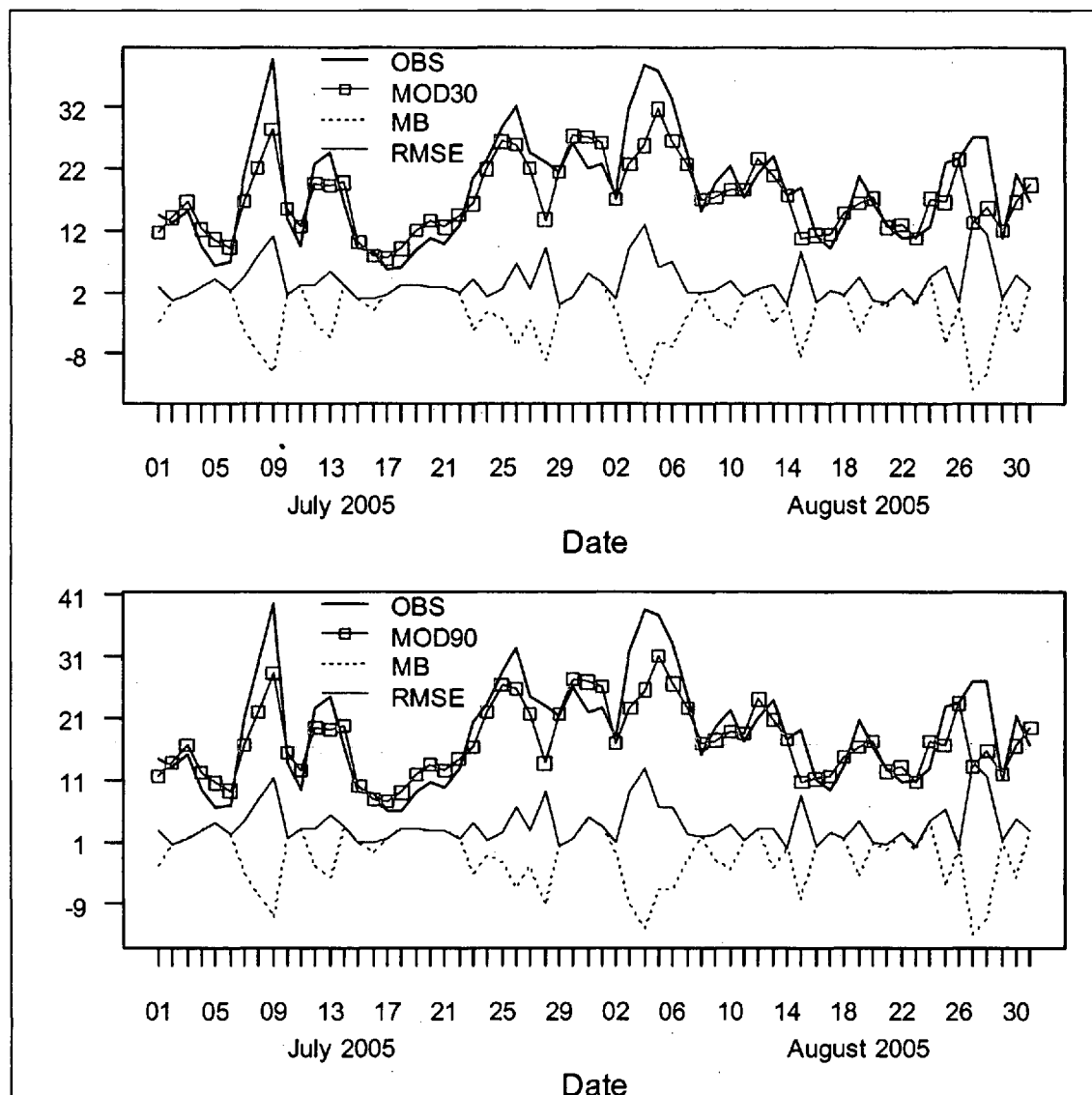
Figure 5. Comparison of daily time series of the modeled and observed PM2.5 concentrations (Grand data from an EPA monitoring station site: 280810005)
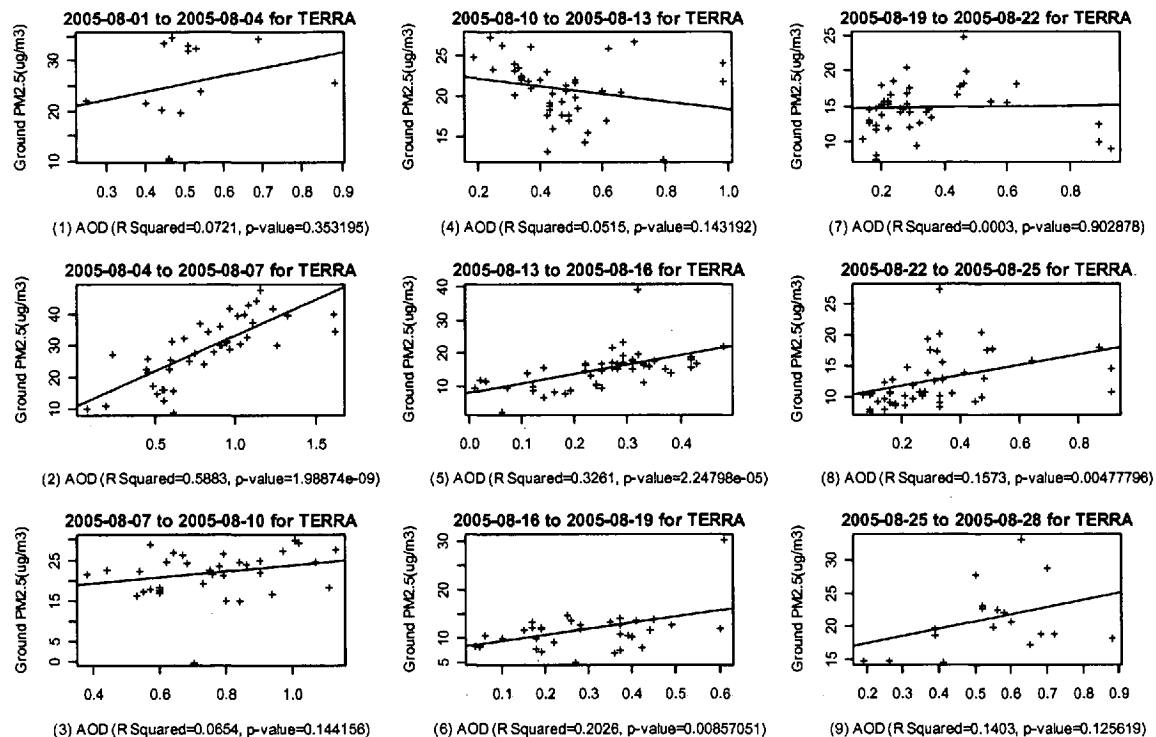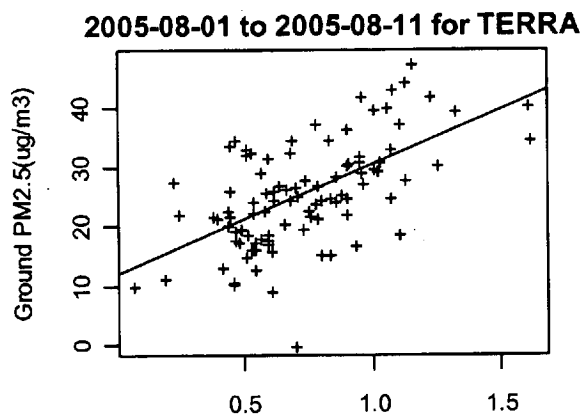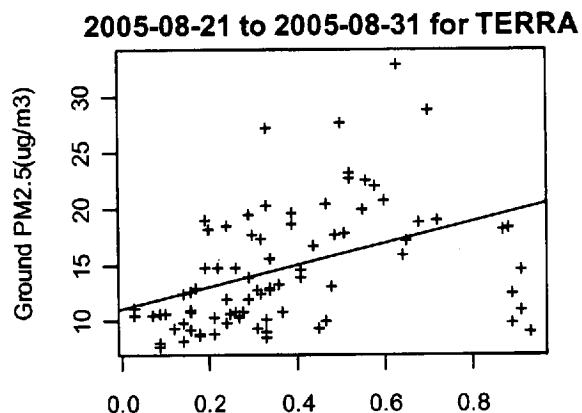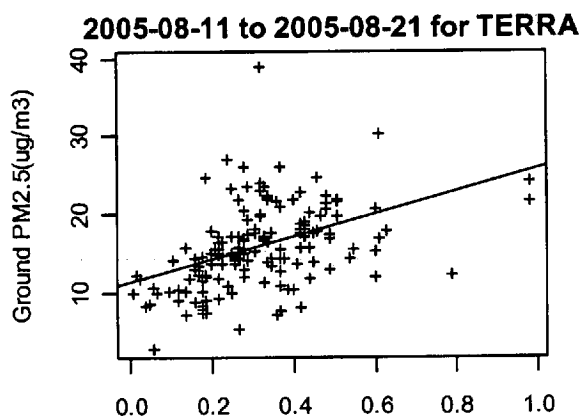
Figure 6. Outcomes of fitted regression models and their corresponding sampled data (MODIS AOD from Terra satellite) in August, 2005 using the temporal scale of the last 3 days.
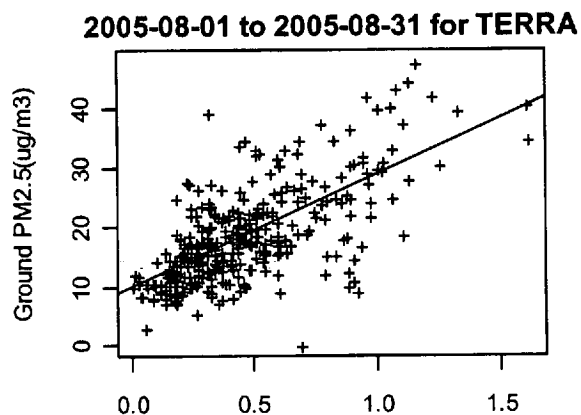
Figure 7. Outcomes of fitted regression models and their corresponding sampled data (MODIS AOD from Terra satellite) in August, 2005 using the temporal scales of the last10 days and 30 days.