Air Qual Atmos Health (2015) 8:55–65
DOI 10.1007/s11869-014-0271-x

Satellite-based estimates of outdoor particulate pollution (PM_{10}) for Agra City in northern India

Shikha Chitranshi · Satya Prakash Sharma · Sagnik Dey

Received: 23 February 2014 / Accepted: 5 June 2014 / Published online: 14 June 2014

Abstract Air quality of north Indian cities has worsened over the last few decades which has been posing a great risk to consequential health-related issues. Ground-based monitoring of particulate matter smaller than 10 μm (PM_{10}) in Indian cities has been limited to few selective sites at local hot spots, and thus, related health studies at regional scale were constrained. To overcome this issue, we utilized the aerosol optical depth (AOD) from Moderate Resolution Imaging Spectroradiometer (MODIS) onboard EOS Terra and Aqua satellites to estimate the regional PM_{10} concentration in Agra City located in the northern part of India. The approach envisaged the developments of linear, log-linear, and multi-linear regression models to estimate PM_{10} using AOD_{MODIS} and in situ measured meteorological parameters by utilizing the data of years 2010 and 2011. The results indicated that both hourly and 24-h average PM_{10} had a weak correlation with AOD_{MODIS} when chosen as the only regressor. However, hourly PM_{10} showed better correlation with AOD_{MODIS} (R ~0.45) than 24-h average PM_{10} (R ~0.24). The log-linear estimation of PM_{10} utilizing AOD_{MODIS}, relative humidity, wind speed, and atmospheric temperature as regressors had the highest correlation coefficient (R=0.81) and a minimum relative standard error as 8.93 %, and thus, it was able to provide the best estimates of PM_{10} among all the models considered in this study. However, the model adequacy checks suggested the further scope of strengthening of these linear and log-linear models by adopting some suitable transformations in them.

Keywords Hourly and 24-h average PM_{10} · AOD_{MODIS} · Linear and log-linear regression modelling · Meteorological parameters

S. Chitranshi (✉)
Institute of Engineering and Technology, Uttar Pradesh Technical University, Lucknow, India
e-mail: shikhachitranshi@gmail.com

S. Chitranshi
e-mail: Shikha_c24@rediffmai.com

S. P. Sharma
Civil Engineering Department, Institute of Engineering and Technology, Uttar Pradesh Technical University, Lucknow, India

S. Dey
Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India

Introduction

Outdoor air pollution has been widely recognized as one of the leading factors to the global burden of diseases (Cohen et al. 2005). Particulate matter smaller than 10 μm (PM_{10}) is considered to be one of the major criteria pollutants to indicate the air quality (World Health Organization 2006). Recognizing the importance of this issue, the Central Pollution Control Board (CPCB) of India under the National Air Quality Monitoring Program (NAQMP) has been monitoring PM_{10} across the entire country through a network of 342 ground-based monitoring stations. These 24-h ground-based measurements are being carried out mostly by high-volume air samplers with a frequency of twice a week. Most of the locations of ground-based monitoring samplers are limited to urban areas adjacent to some local hot spots. The data are often missing at many sites due to operational problems. Considering the high spatial and temporal variability in aerosol concentration in India (Dey and Girolamo 2010), the existing network is not sufficient to cover the entire country seamlessly, thereby raising concerns about the applicability of such data in regional-scale studies. Satellite remote sensing techniques for estimating the ground level PM_{10} are now becoming more and more popular where surface PM_{10}

monitors are not available and also because of the advantage of their wide spatial and temporal coverage.

The Moderate Resolution Imaging Spectroradiometer (MODIS) onboard EOS Terra and Aqua satellites has been retrieving aerosol properties routinely over land and ocean including aerosol optical depth (AOD) at 0.55 μm (Remer et al. 2005). AOD_{MODIS} retrievals (level 2 data of collection 5) are at 10 km by 10 km spatial scales (Levy et al. 2007a, 2007b), whereas ground measurements of PM_{10} are at point locations with high temporal resolution. As AOD reflects total columnar aerosol optical properties, it has been used as an input parameter in correlative models estimating the PM_{10} concentration measured at surface level. The correlation between them is severely influenced by the vertical distribution of aerosols and the meteorology that impact AOD. Thus, to estimate PM concentrations simply from AOD would have large uncertainties. To reduce these uncertainties, the atmospheric boundary layer height and ambient relative humidity should be introduced into the correlative models. Furthermore, several meteorological factors are recommended to be assimilated to exactly depict the ambient impact and the linear regressive models are, therefore, to be updated with nonlinear/poly-parameter models to improve the correlation between AOD and PM concentrations. Thus, the widely used approach of estimating PM concentration from AOD is simply an empirical analysis where in situ PM measurements are regressed suitably with the corresponding AOD and the other in situ parameters including the meteorological parameters. Some recent studies supporting the above developments are briefly mentioned in the proceeding paragraphs.

Numerous recent studies have suggested the use of regression models to estimate the ambient particulate matter of different sizes using satellite-based AOD (Wang and Christopher 2003; Chu et al. 2003; Lee et al. 2011; Yap and Hashim 2013). In order to improve the predictive power of regression models, related parameters such as local meteorology and land use information were also used as additional inputs (Li et al. 2005; Gupta et al. 2006; Gupta and Christopher 2008, 2009a; Kumar et al. 2007, 2008; Hoff and Christopher 2009; Liu et al. 2009; Van Donkelaar et al. 2010). When both AOD_{MODIS} and PM_{10} experienced a highly skewed distribution, these were transformed into log scales for regression analysis (Kumar et al. 2008, 2011). In a study for Jefferson County, Alabama, by Wang and Christopher (2003), the AOD results from Terra and Aqua satellites were correlated with both hourly and 24-h averaged $PM_{2.5}$ (particulates smaller than 2.5 μm aerodynamic diameter) from seven locations within 100 km. They found a coefficient of correlation (R) of 0.70 when hourly $PM_{2.5}$ was linearly related with AOD_{MODIS} for the aggregated data of all the sites, and when the data were aggregated for daily means of $PM_{2.5}$, R increased to 0.98. On the other hand, Justice et al. (2009) examined the impact of diurnal variation in $PM_{2.5}$ on R^2. It was found that R^2 values from hourly $PM_{2.5}$ were higher than those from the daily averages.

Recently, Dey et al. (2012) have utilized the Multi-angle Imaging SpectroRadiometer (MISR)-retrieved AOD data to indirectly estimate the annual mean of $PM_{2.5}$ concentration over India for the last decade so as to assess the potential health implications. They have identified the Agra City as a hot spot for $PM_{2.5}$ with a mean annual concentration of 88 μg m^{-3}. With regard to the profiles of particulate matter in and around Agra City, Parmar et al. (2001) and Satsangi et al. (2007) have reported the mean annual PM_{10} concentration as 131 μg/m^3 at semi-urban area of Dayalbagh and at St. Jones College, a heavy traffic site. In the same city, the annual mean of total suspended particulates was reported as 368.5 μg/m^3 near the Taj Mahal (Kulshrestha et al. 1995) in the range of 173–973 μg/m^3 at the industrial area (Khare et al. 1996) and 441.1 μg/m^3 at Dayalbagh (Kumar et al. 2007). These results clearly indicated that PM_{10} in Agra City has been well above the Indian standard for 24-h average as 100 μg/m^3. Singh and Sharma (2012) and Pachauri et al. (2013) have reported the seasonal variations of PM_{10} concentration in Agra City with a high value in winter (December–February) followed by summer and the lowest in monsoon (June to September). Presently, the air quality of the city is being monitored hourly by Uttar Pradesh Pollution Control Board (UPPCB), a state regulatory board of India, with the help of a single automatic air quality monitoring station (AAQMS). However, this single site may not be enough to provide the insight of the space–time variation of hourly PM_{10} concentration with a reasonable resolution in and around this historical city. Therefore, the present study aims to develop a suitable regression model which could provide reliable estimates of regional PM_{10} concentration with the help of the AOD_{MODIS} and coincident in situ measurements of meteorological parameters. One of the main implications of such models could be their applicability in studying the long-term trends of PM_{10} on regional basis and the associated health impacts.

Study area, data, and approach

Study area

The present study was focused on Agra City (latitude 27° 12′ 12.26″, longitude 78° 00′ 21.03″, and elevation 122.26 m), where the historic monument “Taj Mahal” is located. The monitoring station AAQMS has been placed by UPPCB on the Agra Nagar Nigam Building near Sanjay Palace to monitor the air pollution in the city on long-term basis. This monitoring station is surrounded by many office buildings (not high rise) and by moderately green cover. It is about 200 m away from the main road which carries relatively high-density traffic. The pollution generated from this main road traffic is expected to get attenuated before it reaches the site of the monitoring station. Thus, the particulate level at this site is mainly influenced by traffic activities on the main road and diesel generators (DGs) installed in the premises of the office buildings. There was no other remarkable anthropogenic activity in that area responsible for the particulate pollution.

Data availability

The data availability is discussed here keeping in mind the approach of estimating PM_{10} with the help of various regression models utilizing the AOD_{MODIS}, PM_{10}, and meteorological parameters such as relative humidity (RH), wind speed (WS), and atmospheric temperature (AT). All these data were obtained from the UPPCB monitoring station coincidental to the measurements of AOD_{MODIS}.

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) (Remer et al. 2005). In the collection 5 aerosol retrieval algorithm, three different channels of 0.47, 0.66, and 2.1 μm are primarily employed for over land aerosol retrievals while other channels are used for screening procedures (e.g., detection of cloud and snow and ice cover, etc.). AOD_{MODIS} is reported at the wavelength of 0.55 μm at 10 km×10 km spatial resolution as level 2 data (Levy et al. 2007a, 2007b) The global validation of AOD_{MODIS} product has already been described very well by Levy et al. (2010). In brief, more than 66 % of AOD_{MODIS} was found to be comparable to AErosol RObotic NETwork (AERONET)-retrieved AOD within an expected error envelop, $\Delta AOD=\pm(0.05+0.15\ AOD)$ with a high correlation (R=0.9). Both Terra- and Aqua-retrieved AOD were similarly comparable to AERONET retrievals; however, Terra's global AOD bias changed with time, overestimating by ~0.005 before 2004 and underestimating by a similar magnitude after due to calibration uncertainty (Levy et al. 2010). Many studies have evaluated the quality of AOD_{MODIS} over the Indian subcontinent. Jethva et al. (2007) have found nearly 70 % of the retrievals falling within the aforesaid error envelope with a high correlation (R=0.91) based on the comparison with Kanpur AERONET site which is the only AERONET site within 300 km radius from Agra (India). However, Tripathi et al. (2005) found the AOD_{MODIS} bias for the same location as low as 0.12±0.11 during the nondust loading season (R ~0.84) and much higher (0.4±0.2) particularly during the dust-dominated seasons (R ~0.85). The present study employed the AOD values retrieved from MODIS sensors for the level 2 pixels overlying the ground-based monitoring station. These AOD values were retrieved with quality control flags (Quality Assurance Land (QAL)) 2 and 3 where flag 2 indicates as good confidence and 3 as very good confidence.

The ground-based monitoring site employed Met One Instrument model BAM-1020, which automatically measured and recorded the PM_{10} concentration and the meteorological parameters every hour for the entire day (i.e., 24 h). Met One Instrument model BAM-1020 records airborne particulate concentration levels using the principle of beta ray attenuation. A measured amount of dust-laden air is pulled through a filter tape, and then, dust-loaded filter is automatically placed between the source of high-energy electron known as beta particles and the detector there by causing an attenuation of the beta particle signal. The degree of attenuation of beta particle signal is used to determine the mass concentration of particulate matter (Met One Inc. 2008). The working principle of Met One Instrument model BAM-1020 is different than that of another automated method for PM monitoring known as tapered element oscillating microbalance (TEOM) which is widely used in the USA, Australia, Hong Kong, and Europe. In this method, the measuring signals are clearly related to the inertia of deposited mass (Hauck et al. 2004).

During the study period, the monitoring station could record PM_{10} only for 15 days a month and, for the remaining period of 15 days, it was recording $PM_{2.5}$. These data were analyzed within ±15-min time window of Terra and Aqua overpasses. The AOD_{MODIS} and PM_{10} data were generally scantly available for our study area during the monsoon period.

Approach

The approach in brief envisaged the application of various linear, multi-linear, and log-linear (logistic) regression models to the given data sets and then finding out the best of these models after analyzing their results to have the reliable estimates of PM_{10}. While formulating the multi-linear regression models, it was also envisaged to include the in situ measured meteorological parameters as regressors in addition to the AOD_{MODIS}. Some recent studies have supported the use of the meteorological parameters to improve the correlative estimation of PM with the help of AOD. The meteorological and other ancillary data sets were used by Gupta et al. (2006) to assess the effect of the wind speed, cloud cover, and height of mixing layer (MH) on PM air quality. Their study demonstrated that $PM_{2.5}$–AOD relationship was strongly dependent on aerosol concentration, ambient RH, fractional cloud cover, and MH. The study finally concluded that these data were necessary to further apply satellite data for air quality research (Gupta et al. 2006). The weather conditions can greatly influence aerosol loading. Thus, the effect of weather conditions such as wind velocity, relative humidity, temperature, and atmospheric pressure can confound the AOD–PM association (Kumar et al. 2007; Grguric et al. 2013; Gao and Zha 2010). It has been further emphasized to control for meteorological conditions to estimate PM mass by using AOD_{MODIS} as these conditions can influence AOD in a number of ways. RH and dew point have a direct impact on particle size, whereas the wind speed effects the mixing of aerosols (Kumar et al. 2011). Based on all these findings, the present study used only three meteorological parameters, namely, RH, WS, and AT which were most commonly used in the above studies. The MH was not considered in this study since, directly or indirectly, it is related to these main meteorological parameters.

The data of years 2010 and 2011 were used for the development of regression models, while the data of year 2012 were used exclusively for the purpose of model testing and/or validation studies. The study first proposed to evaluate the linear regression models as presented below in set I.

Set I: linear regression models

Initially, the correlation of 24-h average (avg) PM_{10} with AOD_{MODIS} alone was checked with the help of the simplest form of regression model (a) below. This model is more usable practically as the PM air quality standards in India are defined in terms of 24-h avg PM_{10}.

$$24-\text{h avg PM}_i = \alpha + \beta_1 \text{AOD}_i + e_i \tag{a}$$

where suffix i denotes the ith observation, PM is PM_{10} measured by AAQMS along with the meteorological parameters, AOD is AOD_{MODIS}, α and β are the estimators, and e_i is the residual error in the estimation.

The hourly PM_{10} data coincidental to AOD measurements are likely to give better correlation with AOD alone than 24-h avg PM_{10} data. This is to be evaluated with the help of the model (b) below

$$\text{Hourly PM}_i = \alpha + \beta_1 \text{AOD}_i + e_i \tag{b}$$

The better of the above two associations is proposed to be studied further to assess the impact of meteorological parameters on the estimation of PM_{10}. This impact was evaluated step by step for each of the meteorological parameters. First, the impact of RH was studied through the set I (c) model and then meteorological parameters WS and AT were added in subsequent steps through the set I (d) and (e) models given below. These models were expected to show some incremental improvement in the correlation as a result of the addition of specific meteorological parameter.

$$^{*}\text{PM}_i = \alpha + \beta_1 \text{AOD}_i + \beta_2 \text{RH}_i + e_i \tag{c}$$

$$^{*}\text{PM}_i = \alpha + \beta_1 \text{AOD}_i + \beta_2 \text{RH}_i + \beta_3 \text{WS}_i + e_i \tag{d}$$

$$^{*}\text{PM}_i = \alpha + \beta_1 \text{AOD}_i + \beta_2 \text{RH}_i + \beta_3 \text{WS}_i + \beta_4 \text{AT}_i + e_i \tag{e}$$

where asterisk "*" denotes the 24-h avg PM_{10} value or hourly PM_{10} value whichever had the better association with AOD_{MODIS} through the linear regression set I models (a) and (b). All other terms have their usual meanings defined earlier.

The observed values of PM, AOD, and meteorological parameters were found to have some substantial skewness (more than 1) in their variations about their mean values. This finding suggested that the logistic regression models on the lines of the linear regression models of set I may lead to some improvement in the correlation over the linear regression models. Some researchers have tried the similar logistic regression models to have the reliable estimates of PM_{10} with the help of AOD_{MODIS} and other factors including the meteorological parameters (Kumar et al. 2008, 2011). Therefore, this study also proposed below the application of logistic models (set II) on the lines of the set I models. Further, the reasoning behind adopting the following five logistic models is same as that of the set I models.

Set II: logistic regression models

$$\log(24-\text{h avg PM}_i) = \alpha + \beta_1 \log(\text{AOD}_i) + e_i \tag{a}$$

$$\log(\text{hourly PM}_i) = \alpha + \beta_1 \log(\text{AOD}_i) + e_i \tag{b}$$

$$\log\left({}^{**}\text{PM}_i\right) = \alpha + \beta_1 \log(\text{AOD}_i) + \beta_2 \log(\text{RH}_i) + e_i \tag{c}$$

$$\log\left({}^{**}\text{PM}_i\right) = \alpha + \beta_1 \log(\text{AOD}_i) + \beta_2 \log(\text{RH}_i) + \beta_3 \log(\text{WS}_i) + e_i \tag{d}$$

$$\log\left({}^{**}\text{PM}_i\right) = \alpha + \beta_1 \log(\text{AOD}_i) + \beta_2 \log(\text{RH}_i) + \beta_3 \log(\text{WS}_i) + \beta_4 \log(\text{AT}_i) + e_i \tag{e}$$

where double asterisk "**" denotes the 24-h avg PM_{10} or hourly PM_{10} whichever had the better association with AOD_{MODIS} through the logistic models (a) and (b) of the above set II. All other terms have their usual meanings.

Set I and set II models are also proposed to go through the adequacy checks and model validation studies as briefed below. Finally, the statistical performance of adequate and validated models shall be compared to finally select a suitable model for estimating the better and reliable estimates of PM_{10}.

Model adequacy checks

As per the standard practice, the straight line plot between the estimated responses vs observed responses is sometimes employed to assess the suitability of the model application. After analyzing in detail the statistics of this kind of plots, it has been found that the information and inferences provided by such plots can be derived theoretically also by the regression results itself. It can easily be proved that the slope of such plots is nothing but the coefficient of determination (R^2) of the developed linear regression model, and its intercept value is always $1-R^2$ times the mean of the observed responses. Further, the correlation between predicted and observed responses is also the same as that of the developed regression model. The exercise of the said straight line plotting thus becomes redundant, and hence, it has not been used in the present study. For checking the model adequacy, a better approach by Johnson (2005) was adopted. As per this approach, the assumptions involved in the least square-based classical linear regression essentially ensure the residual errors e_i to be independent of the estimated responses. The adequacy of the regression model can thus be checked by plotting the residuals against the estimated responses. For models to be adequate, residual plots should be confined within a horizontal band around the zero residual line. Any other defined distribution patterns of residuals may lead to the suggestion of transforming the proposed linear form of the regression model and/or addition of the square terms of the independent variables in it. Any undefined pattern may lead to the decision to declare the model as inadequate.

Model validation

The validation studies of the models under consideration were carried out using the data for the year 2012 (validation year) to examine whether the developed models were capable of estimating PM_{10} appropriately or not. In the model validation studies, first, the PM_{10} (estimated) was computed using the observed AOD_{MODIS} and the meteorological parameter(s) of the year 2012 using the estimators of the corresponding regression model from the data of years 2010 and 2011. Then, the coefficient of correlation (R) and standard error of estimates (SE) were computed using the estimated and observed PM_{10} values for the year 2012. These two parameters (R and SE) were then compared with those computed during the regression stage as a part of the validation process. It is noteworthy that the magnitude of SE is very much influenced by the mean of the observed responses. Hence, it will not be appropriate to compare SE as such for the validation of regression stages. In this study, an improvised term—"relative standard error" (RSE)—has been defined as "SE as a percentage of the mean of the observed PM_{10} values," i.e., (SE/mean of observed PM_{10})×100, which is a more rational statistical parameter for comparing the validation results with those of the regression. Now, these parameters (R and RSE) cannot only assess the models' accuracy of estimation but also be used to compare the performance of different types of models when applied to data sets of different time periods.

Results

Descriptive statistics

Table 1 depicts the descriptive statistics in brief of all the relevant parameters involved in the regression models as mentioned in the approach. Mean, standard error, standard deviation, range of values, and total count of the coincident data sets for Agra City have been presented in this table for the

Table 1 Descriptive statistics of the input data for Agra City

Year	Statistical parameter	24-h avg PM_{10}, μg/m^3	Hourly PM_{10}, μg/m^3	AOD	RH (%)	WS (m/s)	AT (°C)
2010	Mean (μ)	70.48	72.98	0.62	38.41	3.86	26.71
	Standard error	2.86	7.00	0.04	1.25	0.19	0.84
	Standard deviation (σ)	32.00	39.70	0.20	9.90	1.57	6.67
	Range	18.42–196.83	21.5–271	0.25–1.97	19.35–66.9	1.4–7.94	19.3–45
	Count	125	63	63	63	63	63
2011	Mean (μ)	76.78	80.12	0.68	40.86	3.83	29.07
	Standard error	3.25	6.67	0.02	0.94	0.19	0.81
	Standard deviation (σ)	40.14	47.77	0.25	9.48	1.80	8.18
	Range	28.83–179.45	16–342	0.20–1.41	22.3–67.55	1.52–9.68	8.45–45.65
	Count	152	102	102	102	102	102
2012	Mean (μ)	90.83	141.75	0.89	38.63	4.48	27.78
	Standard error	5.17	11.18	0.03	0.94	0.19	0.77
	Standard deviation (σ)	71.63	85.24	0.19	9.34	1.91	7.63
	Range	14.5–325	25–337	0.46–2.00	22.4–70.4	0.58–8.76	11.6–41.5
	Count	192	98	98	98	98	98

whole study period. The descriptive statistics of 24-h avg PM_{10} values have also been presented in this table. The Agra City had 165 coincident observations, i.e., data sets of hourly PM_{10}, AOD_{MODIS}, RH, WS, and AT during the period, i.e., 2010–2011, for the purpose of regression studies and 98 such data sets during the year 2012 for the validation studies. The annual mean (±1 standard deviation, σ) of hourly PM_{10} ($\mu g/m^3$) ranged from 72.98 (±39.70) to 80.12 (±47.77) in the regression period, while it suddenly jumped to 141.75 (±85.24) in the validation period. The annual mean of the 24-h avg PM_{10} ($\mu g/m^3$) ranged from 70.48 (±32.00) to 76.78 (±40.14) in the regression study period and then rose to 90.83.75 (±71.63) in the validation period 2012. Both these annual means were found to be increasing year by year but with a sudden rise in the year 2012. The annual mean of hourly PM_{10} during the validation period was found almost 1.5 times the mean of 24-h avg PM_{10} raising the possibilities of increased daytime activities of pollution generation and/or transport in the year 2012. The annual mean of AOD_{MODIS} (dimensionless) also varied almost in a similar fashion, from 0.62 (±0.20) in the year 2010 to 0.89 (±0.19) in the year 2012, thereby indicating a year by year increase in Agra City. The annual mean of the meteorological parameters, i.e., RH, WS, and AT, did not have much yearly variation, and their values centered around 40 %, 4.0 m/s, and 28 °C, respectively. These meteorological parameters have substantial variation both month wise and season wise in the study area.

Figure 1 shows the variation in the mean monthly values of PM_{10} and AOD_{MODIS}. The mean values were computed for hourly PM_{10} data for those months which had at least 12 data sets. AOD data are not available generally for the monsoon season and in month of January. Mean monthly PM_{10} ($\mu g/m^3$) was observed to be lowest as 45 (±10.97) in the month of February/March and highest as 159.33 (±25.93) in the month of November/December. The months of June and October also witnessed relatively higher concentration of PM_{10}. The

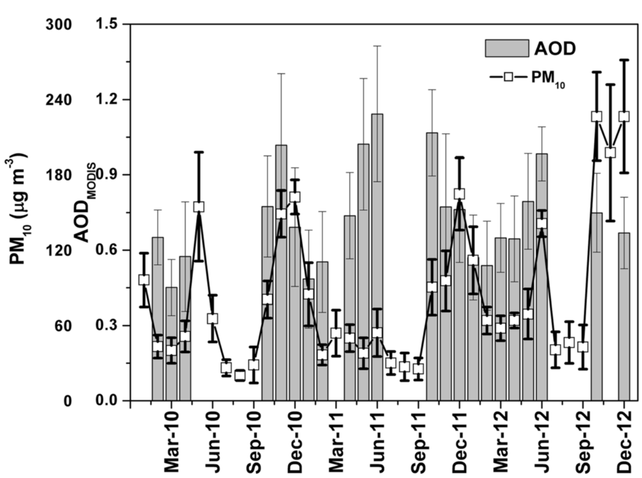

Fig. 1 Monthly variations of PM_{10} and AOD_{MODIS} during the study period 2010–2012. Bars indicate ±1 standard deviation

similar findings for PM_{10} variations have also been reported by Pachauri et al. (2013) and Singh and Sharma (2012) for the city of Agra. Mean monthly AOD_{MODIS} varied from 0.45 (±0.12) in the month of February/March to 1.05±0.26 in the month of June. These findings were supported by the findings by Ahmad et al. (2012) who studied the AOD_{MODIS} variation for Aligarh, a nearby city of Agra wherein they attributed the decrease in AOD_{MODIS} during February/March to the changed weather conditions and attributed the high AOD_{MODIS} in May/June to the dust storms coming along with hot winds from western Rajasthan.

Estimation of PM_{10} using linear regression models

Five types of linear regression models (set I) were studied, and the results are presented in Table 2. From the close perusal of these results, it was evident that both hourly PM_{10} and 24-h avg PM_{10} had a weak correlation with AOD_{MODIS} alone. However, the correlation of hourly PM_{10} (R=0.41, $p \leq 0.05$) was found to be almost twice as compared to that of 24-h avg PM_{10} (R=0.22, $p \leq 0.05$) with AOD_{MODIS} alone. Wang and Christopher (2003) correlated the AOD results from the Terra and Aqua satellites with both hourly and 24-h averaged $PM_{2.5}$. They found a coefficient of correlation (R) of 0.70 when hourly $PM_{2.5}$ was linearly related with AOD_{MODIS} for the aggregated data of all the sites, and when the data were aggregated for daily means of $PM_{2.5}$, R increased to 0.98. On the other hand, Justice et al. (2009) examined the impact of diurnal variation in $PM_{2.5}$ on R^2. It was found that R^2 values from hourly $PM_{2.5}$ were higher than those from the daily averages. For this reason, the linear regression of hourly PM_{10} with AOD_{MODIS} was further extended to include the meteorological parameters as additional regressors. This study showed a substantial improvement in the linear correlation when the meteorological parameters such as RH, WS, and AT were added as successive regressors in addition to AOD_{MODIS} for the estimation of hourly PM_{10}. It has been found that the RH had the maximum incremental impact of 0.19 on the coefficient of correlation (R), whereas WS has the least impact of only 0.06. AT has a moderate impact of 0.09 on R. Thus, a moderate value of R (~0.60) or above could be achieved only by the models (c), (d), and (e) of set I as 0.60, 0.66, and 0.75, respectively. Nevertheless, all the linear regression models as studied above were found to be significant ($p \leq 0.02$). The estimators (α and β) of all these five models were also found to be significant ($p<0.05$). The RH can affect the AOD_{MODIS}–PM_{10} association through changing the optical properties of the aerosols. The higher the RH, the larger the proportion of light is scattered and, hence, the larger AOD. Therefore, the slope should be smaller with larger RH (Zhang et al. 2009). Surface level wind speed has been found to be a highly significant regressor in the regression model types (d) and (e), and negative sign of its estimator showed that

Table 2 Statistical results of linear regression models

Model reference	Statistical parameters						
	R (sig.)	RSE (%)	α (sig.)	β_1 (sig.)	β_2 (sig.)	β_3 (sig.)	β_4 (sig.)
Set I (a)	0.22 (0.02)	52.9	65.41 (0.00)	36.73 (0.01)	–	–	–
Set I (b)	0.41 (0.00)	74.3	15.08 (0.09)	94.71 (0.00)	–	–	–
Set I (c)	0.60 (0.00)	65.6	−120.91 (0.00)	123.44 (0.02)	2.93 (0.00)	–	–
Set I (d)	0.66 (0.00)	61.9	−41.58 (0.02)	106.13 (0.03)	2.30 (0.00)	−11.04 (0.05)	–
Set I (e)	0.75 (0.00)	54.5	131.60 (0.04)	130.42 (0.00)	0.74 (0.03)	−10.80 (0.01)	−4.01 (0.00)

AOD_{MODIS} estimated lower PM_{10} concentration at higher wind speed. Ambient temperature was also found to be a significant predictor, although its impact on the estimation of PM_{10} concentration is relatively small. The negative sign of its estimator also showed that AOD_{MODIS} estimated lower PM_{10} concentration at higher temperature which may become very unstable at smaller sample size (Liu et al. 2006). The successive addition of the meteorological regressors did not result in any specific trend in the variation of the estimators. The model adequacy checks and its validation studies were carried out only for the models having at least a moderate *R* value.

Model adequacy checks were applied only to the models (c), (d), and (e) of set I which had a moderate *R* value. The residual errors were plotted against the estimated PM_{10} for these models (Fig. 2). All the three plots in this figure clearly showed an increasing trend of residual error (positive/

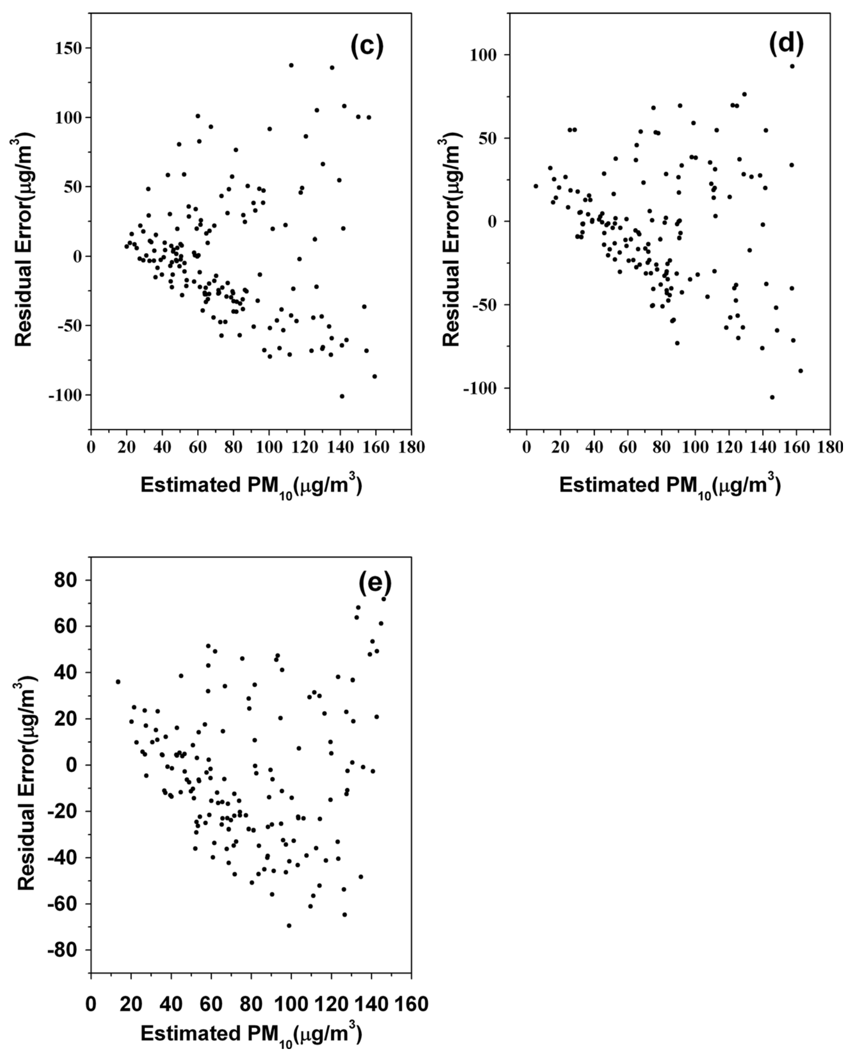

Fig. 2 Residual plots for the linear regression models of set 1

negative) with increasing PM_{10} values. These increasing trends indicated the need of model transformation for improved estimation of PM_{10} concentration.

The validation of the models (c), (d), and (e) of set I was carried out with the data of 2012. The results of this study (Table 4) showed that the coefficients of correlation for the validation period were close to that of the regression stage for all the three models. Similarly, the RSE values of the validation stage were also found to be close to that of the regression stage. The model set I (e) had the maximum R (0.75) and the least RSE, and therefore, it could be claimed to provide a better estimate of PM_{10} with the help of AOD_{MODIS}, RH, WS, and AT.

Estimation of PM_{10} using logistic regression models

The results of the regression studies for the log-linear models of set II are presented in Table 3 which showed the similar findings as that for the set I models. The log transformation reduced the skewness in the data distribution, consequently increasing the accuracy of the estimated regression coefficients (estimators) and their standard error (Liu et al. 2006). AOD_{MODIS} alone has been found to have a very weak log-linear correlation with 24-h avg PM_{10} (R=0.26, p=0.05) but a better log-linear correlation with hourly PM_{10} (R=0.48, p<0.05). The impact of the meteorological parameters was, therefore, studied only on the log-linear correlation of hourly PM_{10} with AOD_{MODIS}. Like linear regression results, this study also showed significant improvements in the log-linear correlation when RH, WS, and AT were added successively to AOD_{MODIS} as additional regressors for the estimation of hourly PM_{10}. RH has been found to have the maximum incremental impact of 0.19 on R followed by the impact of WS as 0.09 and then AT having the least impact of only 0.05. The level of significance for all the regression models in this case and also for their estimators was found to be less than or equal to 0.05. In this case, the models which could achieve a moderate value of R were (c), (d), and (e) of set II with R values as 0.67, 0.76, and 0.81, respectively. Here, also the impact of meteorological parameters and of their estimators is experienced and analyzed in a similar manner as has been described for the multi-linear regression models above.

The residuals vs estimated PM_{10} scatter plots are shown in Fig. 3 (a), (b), (c) for the models (c), (d), and (e) of set II respectively. None of these residual plots showed an ideal scatter to declare the models as perfectly adequate. All the three scatter plots are similar in pattern. These plots are a fusion of a constant band and a divergent trend of residuals, meaning thereby a further transformation in these models may facilitate a better estimation of PM_{10} concentration. Nevertheless, these plots suggested the better adequacy of the log-linear regression models over the linear regression models.

The results of the validation studies of the models (c), (d), and (e) of set II are presented in Table 4. The validation R for all the three models was found to be constant about 0.76, but not much different from the regression R. The RSE was almost the same for both stages and regression as well as validation. The model (e) of set II had the maximum R and the least RSE, and thus, it was able to provide the best estimates of PM_{10} among the models presented in this study. However, the validation results were found to be satisfactory here also for all the three models (c), (d), and (e) of set II.

Discussion and conclusions

From the analysis of the results presented above, it is evident that the PM_{10} has a weak correlation with AOD_{MODIS} when chosen as the only regressor in the linear and log-linear form. However, AOD_{MODIS} has shown a better correlation with hourly PM_{10} (R ~0.45) than with 24-h avg PM_{10} (R ~0.24). Log-linear regression models (set II) are able to provide slightly better estimates of PM_{10} than the linear regression models of set I with the help of AOD_{MODIS} and in situ measured meteorological parameters. The models are able to attain a moderate R value only when RH is added to the AOD_{MODIS} as the second regressor. Further improvements in R with the addition of WS and AT as the third and fourth regressors are not substantial. The increments are only 0.06 and 0.09 in case of linear models and 0.09 and 0.05 in case of

Table 3 Statistical results of logistic regression models

Model reference	Statistical parameters						
	R (Sig.)	RSE (%)	α (sig.)	β_1 (sig.)	β_2 (sig.)	β_3 (sig.)	β_4 (sig.)
Set II (a)	0.26 (0.05)	11.64	1.96 (0.00)	0.33 (0.01)	–	–	–
Set II (b)	0.48 (0.00)	13.4	1.96 (0.00)	0.77 (0.00)	–	–	–
Set II (c)	0.67 (0.00)	11.73	0.03 (0.04)	1.03 (0.01)	1.25 (0.00)	–	–
Set II (d)	0.76 (0.00)	10.05	0.97 (0.00)	0.83 (0.04)	0.83 (0.00)	−0.59 (0.00)	–
Set II (e)	0.81 (0.00)	8.93	0.31 (0.05)	0.90 (0.00)	0.46 (0.03)	−0.61 (0.00)	−0.84 (0.00)

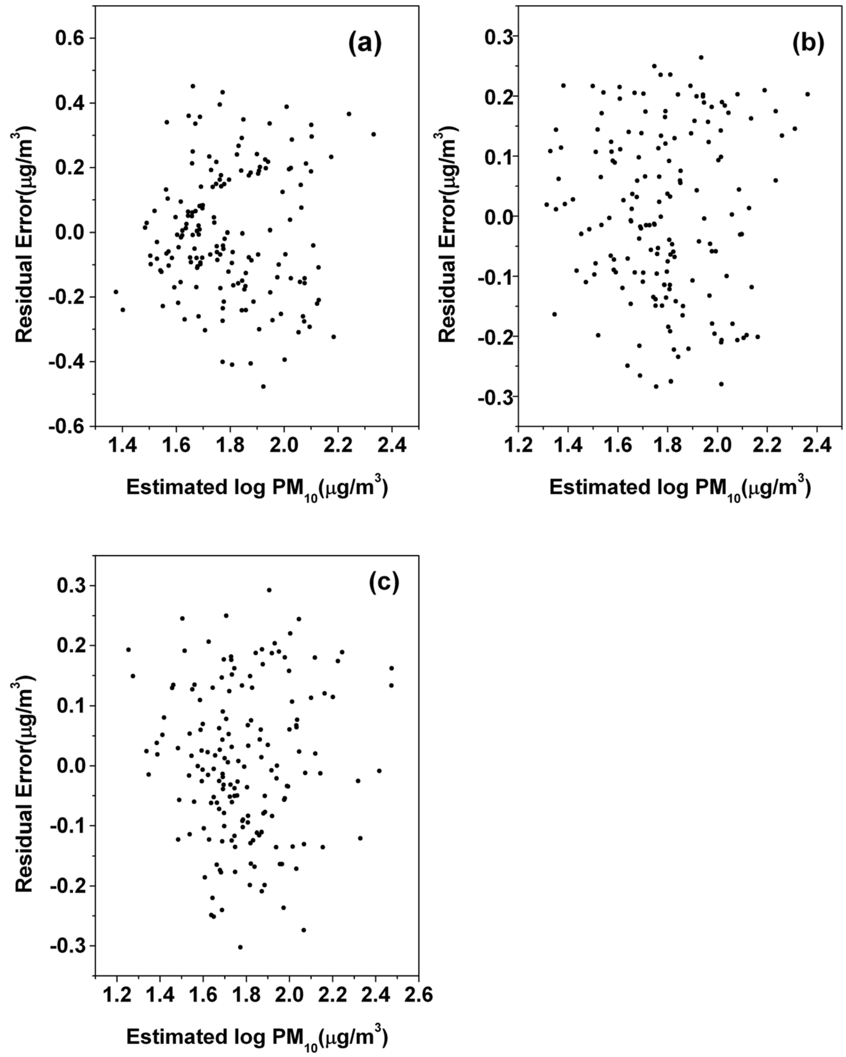

Fig. 3 Residual plots for the logistic regression models of set II

log-linear models, respectively. The model (e) of set II, i.e., log-linear model with AOD_{MODIS}, RH, WS, and AT as regressors, had the highest R as 0.81 and minimum RSE as 8.93 % and thus was declared as the best model to estimate the hourly PM_{10}. However, the model adequacy checks suggested some scope of improvements in these linear and log-linear models.

Table 4 Brief comparison of regression and validation results

Model equation	Coefficient of correlation, R		RSE (in %)	
	Regression	Validation	Regression	Validation
Set I (c)	0.60	0.68	65.6	42.90
Set I (d)	0.66	0.71	61.9	37.71
Set I (e)	0.75	0.72	54.5	36.9
Set II (c)	0.67	0.75	11.73	11.46
Set II (d)	0.76	0.76	10.05	9.44
Set II (e)	0.81	0.76	8.93	8.86

PM_{10} was being monitored only by one observation site and was being correlated to the average AOD_{MODIS} value of the pixel (10 km×10 km) covering the ground-based monitoring station. This situation is not very favorable for a very robust correlation until we have at least four to ten ground-based monitoring stations for each pixel. Further, there were limitations in the availability of the PM_{10} data as it was monitored only for 15 days a month as per the prevailing practice of the monitoring by the AAQMS. For all these reasons, the above PM_{10}–AOD_{MODIS} association is not as strong as has been reported for the other parts of the world. Therefore, there is a strong possibility to have more accurate and robust models once the data in desired volume are made available for the city each year. Nevertheless, this study is unique in a sense that it has used hourly values of PM_{10} which have been found to have a better association with AOD_{MODIS} than 24-h avg PM_{10} values.

The surface measurements of air quality cannot identify the real problem of regional transport of pollutants and hence cannot identify the source of pollutant as well. Since satellite measurements are routinely available on a global basis, the transportation of pollution can easily be examined by the proposed methodology. Further, the climatology of PM_{10} and its inter-annual variation in the city can be examined using the above estimation models with the help of AOD_{MODIS} and the meteorological parameters for the last decade. All these would facilitate a better air quality management in the vicinity of the historic monument Taj Mahal which is very important from the tourism point of view. The proposed methodology could also be useful in the air quality management of big cities in India in a cost-effective manner where ground-based monitoring data are scanty. However, direct extrapolation of the results from this study to other regions may not be done without further analysis. Thus, the proposed methodology can be helpful in producing PM_{10} concentration maps for the cities of Uttar Pradesh for the better satellite-based health management strategies.

Acknowledgments The data of particulate matter and meteorology were provided by Uttar Pradesh Pollution Control Board, Lucknow, India, and the data of aerosol optical depth from Moderate Resolution Imaging Spectroradiometer (AOD_{MODIS}) were provided by Goddard Space Flight Center, NASA (USA). The authors highly feel grateful to these organizations for supporting the present research. Special thanks are extended to Mr. Sadbodh Sharma, the Project Associate, Department of Computer Science and Engineering at Indian Institute of Technology, Kanpur, India, for developing a computer code in MATLAB to retrieve the AOD_{MODIS} values. The authors are also thankful to Dr. Vijaya Lakshmi Sharma for her valuable suggestions which improved the quality of this manuscript. One of the authors (Sagnik Dey) also acknowledge the support by research grant from the Department of Science and Technology, Government of India, under contract SR/FTP/ES-191/2010 (Fast Track Scheme) through a research project operational at IIT Delhi (IITD/IRD/RP02509).

References

Ahmad S, Khan MA, Husain A, Matloob H, Warsi TR (2012) Monitoring of the air quality in Aligarh City, India using spot: VGT data. Journal of Environmental Research and Development 6(4):1000–1005

Chu DA, Kaufman YJ, Zibordi G, Chern JD, Mao JM, Li C, Holben HB (2003) Global monitoring of air pollution over land from EOS Terra MODIS. J Geophys Res 108:4661

Cohen AJ, Anderson HR, Ostra B, Pandey KD, Krzyzanowski M, Künzli N (2005) The global burden of disease due to outdoor air pollution. J Toxicol Environ Health 68:1–7, Part A

Dey S, Girolamo LD (2010) A climatology of aerosol optical and microphysical properties over the Indian Subcontinent from 9 years (2000–2008) of Multiangle Imaging SpectroRadiometer (MISR) data. J Geophys Res 115, D15204. doi:10.1029/2009JD013395

Dey S, Girolamo LD, Donkelaar AV, Tripathi SN, Gupta T, Mohan M (2012) Decadal exposure to fine particulate matters ($PM_{2.5}$) in the Indian subcontinent using remote sensing data. Remote Sens Environ 127:153–161

Gao J, Zha Y (2010) Meteorological influence on predicting air pollution from MODIS-derived aerosol optical thickness: a case study in Nanjing, China. Remote Sens 2:2136–2147

Grguric S, Krizan J, Gasparae G, Antonic O, Spiric Z, Mamouri ER, Christodoulou A, Nisantzi A, Agapiou A, Themistocleous K, Fedra K, Panaylotou C, Hadjilmitsis D (2013) Relationship between MODIS based aerosol optical depth and PM_{10} over Croatia. Central European Journal of Geosciences. doi: 10.2478/s13533-012-0135-6

Gupta P, Christopher SA (2008) Seven year particulate matter air quality assessment from surface and satellite measurements. Atmos Chem Phys 8:3311–3324

Gupta P, Christopher SA (2009a) Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. Journal of Geophysical Research-Atmosphere 114:D14205. ISI: 00271318200002.

Gupta P, Christopher SA, Wang J, Gehrig R, Lee Y, Kumar N (2006) Satellite remote sensing of particulate matter and air quality assessment over global cities. Atmos Environ 40:5880–5892

Hauck H, Berner A, Gomiscek B, Stopper S, Puxbauma H, Kundib M, Preining O (2004) On the equivalence of gravimetric PM data with TEOM and beta-attenuation measurements. Aerosol Science 35: 1135–1149

Hoff RM, Christopher SA (2009) Remote sensing of particulate pollution from space: have we reached the promised land. J Air Waste Manage Assoc 59:645–675

Inc MO (2008) BAM-1020 continuous particulate monitor: operation manual. Grant Pass, Oregon

Jethva H, Satheesh SK, Srinivasan J (2007) Assessment of second generation MODIS aerosol retrieval (collection 5) at Kanpur, India. Geophys Res Lett 34, L19802. doi:10.1029/2007GL029647

Johnson RA (2005) Miller & Freund's probability and statistics for engineers, 7th edn. PHI Learning Private Limited, pp 367–369

Justice E, Huston L, Krauth D, Mack J, Oza S, Strawa AW, Skiles JW, Legg M, Schmidt C (2009) Investigating correlations between satellite-derived aerosol optical depth and ground $PM_{2.5}$ measurement in California's San Joaquin Valley with MODIS deep blue. ASPRS Annual Conference, Baltimore.

Khare P, Kulshrestha UC, Saxena A, Kumar N, Kumari KM, Srivastava SS (1996) The source apportionment of particulate matter using enrichment factor and principal component analysis. Indian Journal of Environmental Health 38(2):86–94

Kulshrestha UC, Kumar N, Saxena A, Kumari KM, Srivastava SS (1995) Identification of the nature and source of atmospheric aerosols near the Taj Mahal (India). Environ Monit Assess 34(1):1–11

Kumar N, Chu A, Foster A (2007) An empirical relationship between $PM_{2.5}$ and aerosol optical depth in Delhi Metropolitan. Atmos Environ 41:4492–4503

Kumar N, Chu A, Foster A (2008) Remote sensing of ambient particles in Delhi and its environs: estimation and validation. Int J Remote Sens 29:3383–3405

Kumar N, Chu A, Foster A, Peters T, Willis R (2011) Satellite remote sensing for developing time and space resolved estimates of ambient particulate in Cleveland, OH. Aerosol Sci Technol 45:1090–1108. doi:10.1080/02786826.2011.581256

Lee HJ, Liu Y, Coull BA, Schwartz J, Koutrakis P (2011) A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentration. Atmos Chem Phys 11:7991–8002. doi:10.5194/acp-11-7991

Levy RC, Remer L, Dubovik O (2007a) Global aerosol optical properties and application to moderate resolution imaging spectroradiometer aerosol retrieval over land. Journal of Geophysical Research-Atmosphere 112(D13), D13210. doi:10.1029/2006JD0078151

Levy RC, Remer L, Mattoo S, Vermote E, Kaufman YJ (2007b) Second-generation operational algorithm: retrieval of aerosol properties over land from inversion of moderate resolution imaging spectroradiometer spectral reflectance. Journal of Geophysical Research-Atmosphere 112(D13), D13211. doi:10.1029/2006JD007811

Levy RC, Remer LA, Kleidman RG, Mattoo S, Ichoku C, Kahn R, Eck TF (2010) Global evaluation of the collection 5 MODIS dark- target aerosol products over land. Atmospheric Chemistry & Physics. doi: 10, 5194/acp-10-10399.

Li C, Lau AKH, Mao JT, Chu DA (2005) Retrieval, validation and application of 1-km resolution aerosol optical depth from MODIS data over Hong Kong. Trans. Geoscience Remote Sensing 43:2650–2658

Liu Y, Franklin M, Kahn R, Koutrakis P (2006) Using aerosol optical thickness to predict ground level $PM_{2.5}$ concentrations in the St. Louis area: a comparison between MISR and MODIS. Remote Sens Environ 107:33–44

Liu Y, Paciorek CJ, Koutrakis P (2009) Estimating regional spatial and temporal variability of $PM_{2.5}$ concentrations using satellite data, meteorology, and land use information. Environmental Health Prospect 117:886–892

Pachauri T, Singla V, Satsangi A, Lakhani A, Kumari KM (2013) SEM-EDX characterization of individual coarse particle in Agra, India. Aerosol and Air Quality Research 13:523–536

Parmar RS, Satsangi GS, Kumari M, Lakhani A, Srivastava SS, Satya P (2001) Study of size distribution of atmospheric aerosol at Agra. Atmos Environ 35:693–702

Remer LA, Kaufman YJ, Tanre D, Mattoo S, Chu DA, Martins JV, Li RR, Ichoku C, Levy RC, Kleidman RG, Eck TF, Vermote E, Holben BN (2005) The MODIS aerosol algorithm, products, and validation. Journal of atmospheric Science 62:947–973

Satsangi PG, Kulshrestha A, Taneja A, Rao PS P (2007) Chemical characterization of PM_{10} in Agra, a semi-arid region. A seminar on aerosol climate connections: our present understanding and future directions. IASTA, National Physical Laboratory.

Singh R, Sharma BS (2012) Composition, seasonal variation, and sources of PM_{10} from world heritage site Taj Mahal, Agra. Environmental Monitoring Assessment 184:5945–5956

Tripathi SN, Dey S, Chandel A, Srivastava S, Singh RP, Holben BN (2005) Comparison of MODIS and AERONET derived aerosol optical depth over the Ganga Basin, India. Ann Geophys 23:1093–1101

Van Donkelaar A, Martin RV, Brauer M, Kahn R, Levy R, Verduzco C et al (2010) Global estimates of ambient fine particulate matter concentrations from satellite based aerosol optical depth: development and application. Environ Health Perspect 118(6):847–855

Wang J, Christopher SA (2003) Intercomparison between satellite derived aerosol optical thickness and $PM_{2.5}$ mass: implications for air quality studies. Geophysical Research Letter 30: 2095.

World Health Organization (2006) Health risks of particulate matter from long-range trans boundary air pollution. European Centre for Environment and Health, Copenhagen, p 113

Yap XQ, Hashim M (2013) A robust calibration approach for PM_{10} prediction from MODIS aerosol optical depth. Atmos Chem Phys 13:3517–3526

Zhang H, Hoff RM, Cox JAE (2009) The relation between Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth and $PM_{2.5}$ over the United States: a geographical comparison by U.S. Environmental Protection Agency regions. Journal of the Air and Waste Management Association. doi: 10.3155/1047-3289.59.11.1358.