Journal of Environment Sciences (2023), **Volume IX**, 1-12
https://doi.org/10.3126/jes.v9i1.56417

Relation between Modis-based Aerosol Optical Depth and Particulate Matter in Kathmandu using Regression Model

Saurav Timilsina, Pawan Gautam* & Kundan Lal Shrestha
Department of Environmental Science and Engineering, Kathmandu University, Dhulikhel, Kavre, Nepal
*Email: gautampawan240@gmail.com

Abstract

Ambient fine Particulate Matters have been linked to various adverse health outcomes. Exposure to the high level of such particles would increase the risk of premature death, especially for people with weak immune systems, such as children and elder people. This research derives the relation between particulate matter and AOD from the Regression model on the seasonal (Pre-monsoon season (March 2020) and winter season (December 2019) basis of Kathmandu. Here two models have been developed one linear single-variable regression model and the other multivariable regression model. For the multivariable regression model, meteorological factors like Wind speed, Temperature, and Relative Humidity were adopted from the wunderground and the Planetary boundary layer height was simulated from WRF. Particulate matter ($PM_{2.5}$) was adopted from the US Embassy air quality station and MODIS Level 2 AOD having 10 km resolution was analyzed for regression modeling. The linear single variable and linear multivariable regression model were developed seasonally one from December 1st to December 31st, 2019 (winter season) and the other from March 1st to March 31st, 2020 (Pre-monsoon season) using Python. The seasonal correlation coefficient of these two models was obtained. In both seasons, the multivariable linear regression model showed a good correlation between AOD and Particulate Matter R^2 (Pre-monsoon) = 0.72657, R^2 (winter) = 0.4687) compared to the single variable regression model having R^2 (Pre-monsoon) = 0.45, R^2 (winter) = 0.133). In both these regression models using the evaluated regression coefficients, two seasonal equations were derived from which Particulate Matter can be estimated.

Keywords: *Multivariate Linear Regression, PM-AOD Relationship, Regression Model*

Introduction

Airborne particulate matter is a crucial pollutant affecting the environment, human health, and the climate (Ferrero et al., 2019). Particulate matter (PM), also defined as an atmospheric aerosol, is the general term used to define a complex mixture of solid and liquid particles. These particles vary in size and composition and remain suspended in the air for a long period (Arvani et al., 2015). Aerosol Optical Depth (AOD) is the extinction of radiation in the atmospheric column at a certain wavelength while atmospheric aerosols are a complex and multiphase system formed by gases, liquid, and solid particles suspended in the atmosphere, at a scale ranging from 10^{-3} to 10^2 microns (Chen et al., 2014; Stirnberg et al., 2018). PM including fine particles; $PM_{2.5}$ and coarse particles; PM_{10} (Particulate Matter with aerodynamic diameters less than 2.5 μm and 10 μm, respectively) have proven to have strong associations with adverse health effects (Chen et al., 2014; Ghotbi et al., 2016). Short and long-term exposure to PM causes an increase in mortality rates and morbidities such as a variety of cardiovascular diseases (Ghotbi et al., 2016).

PM_{10} is a major component of aerosol and is suspended in the air under a dispersed phase. Some particles are emitted directly from both human activities and natural events, while others are formed in the atmosphere through secondary chemical transformation (Tsai et al., 2011; Zhao et al., 2018). Inhaling fine particulate matter with an aerodynamic diameter of less than 2.5 μm is a serious health hazard. Health studies demonstrate that $PM_{2.5}$ has substantially greater toxicity than larger particles (Goldberg et al., 2019). Many recent epidemiological studies have shown that fine particles in populated regions are emitted primarily from anthropogenic and biogenic sources, and are associated with various health outcomes, including increased risk of cardiovascular and respiratory diseases, myocardial infarction, and significantly reduced heart rate variability. Aerosols influence the radiation balance of the earth-atmosphere system through direct and indirect radiation effects, which is one of the most important factors in weather and climate change (Chen et al., 2014). The increasing level of air pollutants has become a complex issue affecting public health and the environment in various cities in developing countries in recent years. When sunlight passes through the atmosphere, aerosol particles can scatter and absorb sunlight, reducing atmospheric visibility and solar radiation, thereby changing the temperature of the environment and affecting the growth rate of plants (Chen et al., 2014). Particulate Matter (PM) is nowadays one of the major air quality issues in South Asia. In many developed and developing nations, air pollution has caused an estimated side effect of approximately two million premature deaths worldwide per year. The unfriendly health hazards of Particulate Matter (PM) on the human respiratory and cardiovascular system are notable and incorporate asthma, emphysema, and lung cancer (Tian & Chen, 2010). Aerosols, both natural and anthropogenic, play an important role in air quality and the climate. Their presence leads to pollution events, and they have a direct and indirect role in modifying the Earth's radiation budget and cloud/precipitation properties, respectively and dominate the health effects of air pollution, as well as affecting the energy balance of the Earth-atmosphere system (Lennartson et al., 2018). Respiratory and cardiovascular diseases are provoked by particulate matter pollution.

Ground-level measurements just provide PM values within a small area that may not be a good representative for areas far from monitoring stations. The sparse spatial distribution of monitoring stations and lack of dense monitoring network due to economic and feasibility considerations could cause bias in epidemiological research. To overcome these issues, researchers have tried to find new approaches to attain accurate predictions in addition to ground PM measurements. In the recent decade, satellite remote sensing has been used as a powerful and cost-effective tool to estimate PM concentrations. AOD data, representing PM loading in the air, are non-dimensional parameters calculated by integrating the light extinction of aerosols from ground level up to the top of the atmosphere (Ghotbi et al., 2016). As the world continues to industrialize and increase in population (especially in developing countries), it is imperative to understand and mitigate the effects pollutants have on air quality, climate, and human health, on various spatial and temporal scales (Lennartson et al., 2018). It is necessary to monitor particulate matter pollution. Satellite remote sensing can step in to monitor regional air quality where ground monitors are not available or sparsely distributed and satellite-derived aerosol optical depth (AOD) is related to ground-level PM concentration and can be empirically converted into PM mass (You et al., 2015). In urbanized and populated Kathmandu city there are only five air quality measuring stations. Hence using satellite data estimation of the PM of every corner could be possible. Satellite data have the potential to complement air quality stations.

Materials and Methods

Study Area

The main study area of our research is the Kathmandu District. We have chosen the US

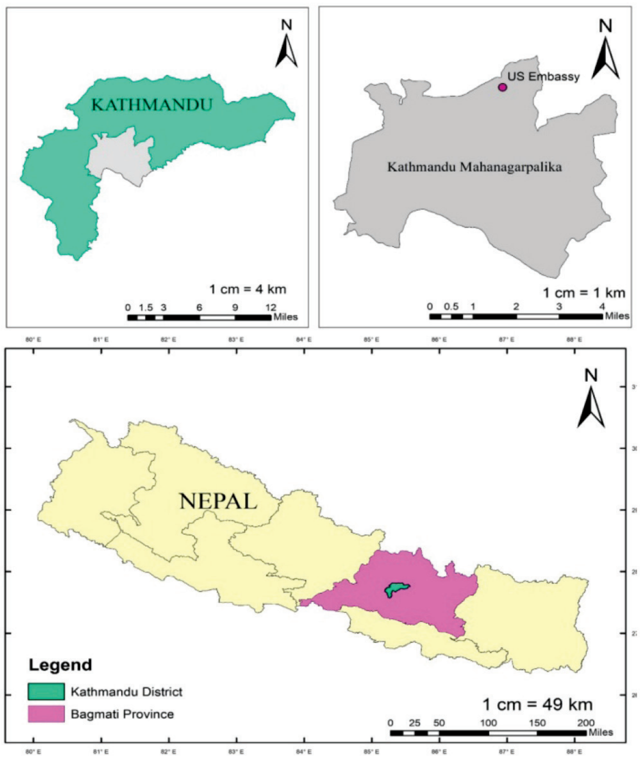

Figure 1: Study Area of the research

Embassy (latitude of 27.75805°N and longitude of 85.336383°E) situated in Maharajgunj, Kathmandu, Bagmati Province, Nepal as our primary station for the collection of essential data. It is around 1300 m above sea level spreading with an area of 433.6 km².

The study area of the project is shown in Figure 1 above. For our study period, we have chosen one month of the winter season i.e., December 2019 and March 2020 as the pre-monsoon season. To observe the variations in PM and AOD in two different seasons, we have adopted these months.

PM Data Extraction

The daily PM concentration data for December 2019 and March 2020 were collected from the air pollution monitoring network operated by the U.S. Department of State (DOS) and the U.S. Environmental Protection Agency (EPA). There are two air quality monitors (AQMs) installed by the US Embassy in Kathmandu (Embassy, 2020). We choose the US embassy at Maharajgunj station and the open data was downloaded from their website. The $PM_{2.5}$ data were downloaded from https://www.airnow.gov/international/us-embassies-and-consulates/. Initially, the raw data given to us were continuous (hourly) $PM_{2.5}$ mass concentration measurements and for the daily analysis, a 24-hour average of each mass concentration data was done through Python.

MODIS AOD Extraction

With a 2330 km wide swath, the MODIS sensor onboard Terra and Aqua satellites provide near-global coverage each day with AOD retrieval limited to cloud/snow/ice-free regions. MODIS sensor provides AOD, which is a unitless quantity and represents the integrated extinction of light by aerosols in the entire atmospheric column (Duncan et al., 2014). Here AOD 550 Dark Target Deep blue combined algorithm is used to extract the AOD from the HDF file using Python.

MODIS Level 2 aerosol data (MOD04, Collection 6) were obtained from the Atmosphere Archive and Distribution System (LAADS) at NASA's Goddard Space Flight Center (GSFC) (Gupta & Christopher, 2009). To fill out the missing data Level 2 collection 6 AOD at a spatial resolution 10 km from both Terra and Aqua satellite MOD04_10k and MYD04_10k of the winter season (December 2019) and pre-monsoon season (March 2020) respectively. HDF files were downloaded from https://ladsweb.modaps.eosdis.nasa.gov/. During extraction pyhdf and numpy libraries were used in Python.

Meteorological Data Extraction

Fine particular matter in the atmosphere is produced by gas-to-particle conversion mechanism as well as through various sources due to anthropogenic and natural activities. The meteorological conditions that strongly influence the concentration of PM particles include Temperature, Relative Humidity (RH), Wind speed, and Planetary Boundary Layer Height (PBLH). The variability in these meteorological conditions is primarily governed by large-scale high and low-pressure systems, diurnal heating and cooling, and topography (Gupta & Christopher, 2009). So, the open data of wind speed (*m/s*), temperature (°*C*), and relative humidity (%) were adopted from https://www.wunderground.com/ in December 2019 and March 2020.

But in the case of PBLH open-source data was not found so WRF simulated PBLH date of December 2019 and March 2020 was adopted. Mues et al., 2017 found that the mixing layer height changes on a diurnal basis (increasing during daytime and decreasing during nighttime and morning) at Kathmandu.

WRF Model

The Weather Research and Forecasting (WRF) Model is an atmospheric model designed, as its name indicates, for both research and numerical weather prediction (Powers et al., 2017). WRF produces atmospheric simulations. The process has two phases, with the first to configure the model domain(s), ingest the input data, and prepare the initial conditions, and the second to run the forecast model. The forecast model components operate within WRF's software framework, which handles I/O and parallel-computing communications. WRF is written primarily in Fortran, can be built with several compilers, and runs predominately on platforms with UNIX-like operating systems, from laptops to supercomputers (Powers et al., 2017).

Our WRF model used a Mercator projection with a grid resolution of 3 km. Here 66×58 grid was used with a vertical layer of 38 levels as shown in Table 1. We have done WRF only to simulate the value of PBLH in December 2019 and March 2020.

Table 1: WRF model setup

WRF Parameters	Domain
Grid resolution	3 km
Projection	Mercator
Grid	66×58
Vertical layers	38 levels

Regression Model

As the indicators of the changes in particle composition and vertical profile, the sensitive impact factors (e.g., relative humidity and temperature) can influence the association between satellite-retrieved AOD and ground-measured $PM_{2.5}$ significantly. To describe the numerical or quantitative relationship between these predictors and $PM_{2.5}$ effectively at the regional scale, two statistical models were developed: a general linear regression model and a multivariate regression model (Song et al., 2014).

Estimation of particulate matter is done by using two different models:

Single Linear Regression Modeling

We applied the daily-calibration model approach, allowing the AOD–$PM_{2.5}$ relationship to vary daily assuming that on any given day the relationship does not vary spatially within each of the study domains (Sorek-Hamer et al., 2015). First, we developed a simple linear equation where MODIS AOD is used to estimate surface-level $PM_{2.5}$ mass concentration.

$$[PM_x] = \alpha_0 + \alpha_{AOD}[AOD] \quad \text{............................} \ (1)$$

Where x represents $PM_{2.5}$ and PM_{10}, $\alpha_{0,}$ and α_{AOD} are the intercept and slope of single-variable linear models respectively.

Regression coefficients are calculated with the help of Python script using Python libraries like Pandas, NumPy, sklearn, linear_model, and then using the calculated regression coefficients from and AOD, particulate matters are estimated.

Multivariable Regression Modeling

Mirzaei et al., 2020 have shown that the spatial relationship between AOD and $PM_{2.5}$ varies daily and that is due to time-varying variables such as temperature, humidity, or $PM_{2.5}$ optical properties. Then, meteorological parameters are added to the analysis to form multiple linear regression equations to estimate $PM_{2.5}$ and PM_{10} mass concentration. Regression coefficients were calculated from Python code for equations (2) and (3) and then these equations are used to calculate $PM_{2.5}$ and PM_{10} mass concentration using input parameters from satellite and meteorological fields. We have used daily measurements of $PM_{2.5}$, matched with the MODIS Terra AOD closest to the satellite overpass time (Christopher & Gupta, 2010). The meteorological parameters are also obtained for each AOD-$PM_{2.5}$ data point by using the same spatial and temporal matching approach. Similar equations of Ghotbi et al., 2016 are adopted as multiple linear regression modelling.

$$[PM_{10}] = \alpha_0 + \alpha_t(T) + \alpha_w(W) + \alpha_{Dir}(Dir) + \alpha_{RH}(RH) + \alpha_{AOD}(AOD) + \alpha_{PBLH}(PBLH) \ (2)$$

$$[PM_{2.5}] = \alpha_{21} + \beta_{21}(AOD) + \beta_{22}(T) + \beta_{23}(RH) + B_{24}(W) + \beta_{25}(PBLH) \ (3)$$

where, *T, W, Dir, RH, AOD, PBLH* are the temperature, wind speed, wind direction, relative humidity, aerosol optical depth, and planetary boundary layer height parameters, respectively. α_0 is the intercept of the general equation and α_{is} are the regression coefficients of the independent variables and $[PM_{10}]$ is the ground concentration measured. And, $\alpha_{21, \ldots mn,}$ and $\beta_{21, \ldots, mn}$ represent the regression coefficients associated with corresponding variables. *AOD* is Aerosol optical depth (unitless), $PM_{2.5}$ is particulate matter concentration (μg/m^3), *T* is the temperature (°C), *W* is wind-speed (m/s), *PBLH* is plate boundary layer height (m) and *RH* is relative humidity (%).

Results and Discussion

Particulate matters were estimated through the Regression model. Both these models are validated with the nearest air quality monitoring station. The time series plot of $PM_{2.5}$ of the US Embassy Monitoring Station and MODIS AOD of the winter season and monsoon season of our study period is shown in Figures 2 and Figure 3. In these two figures, we can see the AOD missing which was due to the presence of the cloud. To reduce the missing data of AOD, we took an average of both aqua and terra satellite AOD. Even after doing so, we were not able to get the AOD for the whole month but we were able to fulfill the AOD for a few missing days.

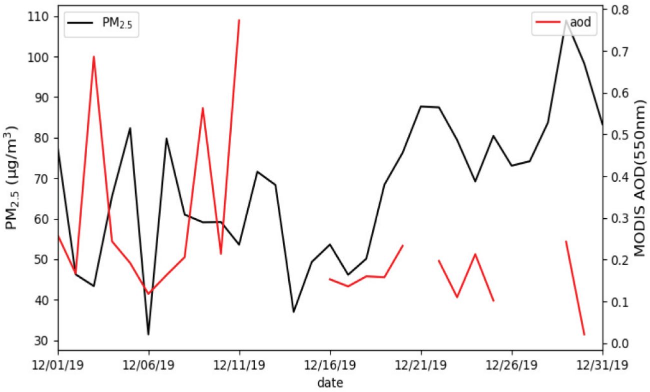

Figure 2: Observed Winter AOD-$PM_{2.5}$ Time series

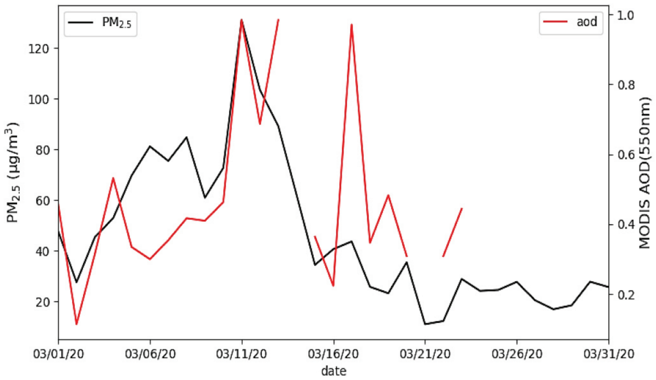

Figure 3: Observed Pre-monsoon AOD-$PM_{2.5}$ Time Series

Linear Single Variable Regression Model

Simple Linear regression analysis was performed between AOD and $PM_{2.5}$ in both seasons and the coefficient of determination (R^2) and the correlation coefficient was determined. Figure 4 and Figure 5 show the regression model graph of our specified winter and pre-monsoon seasons month

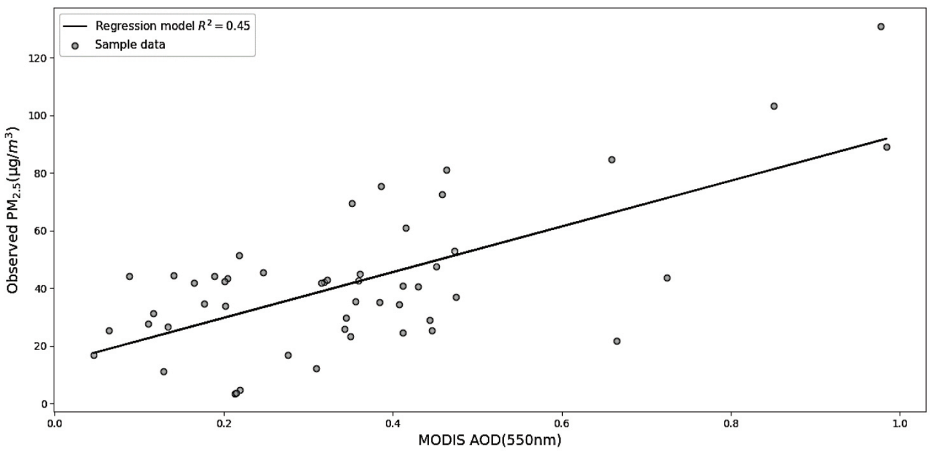

Figure 4: Pre-Monsoon Linear Regression Model

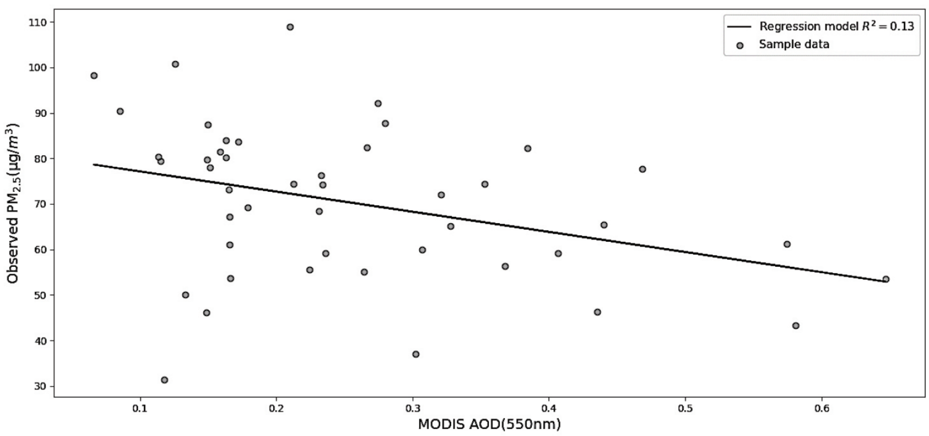

Figure 5: Winter Linear Regression Model

respectively. Here in this figure, US Embassy $PM_{2.5}$ was taken as the primary data.

The result of a single linear regression model is shown in Table 2. Pre-monsoon season (March and April) has good relation coefficients compared to that of the winter season (December and January) due to more available data on AOD in the pre-monsoon season. 27 data were used in evaluating the premonsoon season and there were more MODIS AOD pixels of the premonsoon season during MODIS AOD extraction. While 24 data sets were used in determining winter season R^2 and during the extraction of MODIS AOD in the winter season, it had fewer pixels in comparison with the premonsoon season. So, due to this factor premonsoon season had a good correlation coefficient compared with the winter season

Table 2: Results of Single Linear Regression Model

Parameters	R^2	α_0	α_{AOD}
Pre-monsoon Season (March 2020)	0.45	19.7976	73.5447
Winter Season (December 2019)	0.13	74.5447	-30.121

Now the equation to estimate the $PM_{2.5}$ directly from AOD for the pre-monsoon season is shown below in equation (4)

$$[PM_{2.5}] = 19.7976 + 73.5447[AOD] \quad (4)$$

Similarly, the equation to estimate the $PM_{2.5}$ directly from AOD for the winter season is shown below in equation (5)

$$[PM_{2.5}] = 74.5447 - 30.121[AOD] \quad (5)$$

From Table 2 we can see the results of a single-variable regression model. Here regression coefficients of AOD are positive in the premonsoon season while negative in the winter season. This is due to the presence of clear pixels of aqua and terra satellites in the premonsoon season. During the extraction of AOD in the premonsoon season there were more pixels with less standard deviation and the winter season had fewer pixels centered at our desired location having more standard deviation. So, this could be one cause for less correlation coefficient in the winter season compared to the premonsoon season.

Linear Multi-Variable Regression Model

In this model, other four variables like temperature, wind speed, relative humidity, and plate boundary layer height along with the AOD are taken from open source and $PM_{2.5}$ is referenced from US Embassy Maharajgunj Station. Using equation (2) in Python using the least square method, regression coefficients and correlation coefficients were obtained. The values of the regression coefficients of respective variables and the correlation coefficient are shown in Table 3. The Planetary Boundary layer height of March and December was only modelled so these two months were selected.

Now the equation to estimate $PM_{2.5}$ for the Pre-monsoon season will be shown in equation (6):

$$[PM_{2.5}] = 79.4116 + 83.0779(AOD) - 0.1932(T) - 1.6006(RH) - 1.6670(W) + 0.0907(PBLH) \quad (6)$$

And the equation to estimate $PM_{2.5}$ for the winter season is shown in equation (7) below:

$$[PM_{2.5}] = 267.674 + 6.6281(AOD) - 4.9866(T) - 1.3103(RH) - 12.0314(W) - 0.1442(PBLH) \quad (7)$$

From Table 3, looking at the values of correlation coefficients for March and December, it is seen that the correlation coefficient of March 2020 is greater than that of December 2019. we can see the positive regression coefficient in AOD in both seasons which means that AOD and PM are proportional to each other (increase in PM concentration, increase in AOD). Comparing this coefficient of AOD pre-monsoon season is dominant compared to the winter season. It is due to the good and predictable value of

Table 3: $PM_{2.5}$ Regression Coefficients for the Linear Multivariable Regression Model

Month	α_0	α_{Temp}	α_{RH}	α_{Wind}	α_{AOD}	α_{PBLH}	R^2	Adj R^2
March	79.4116	-0.1932	-1.6006	-1.6670	83.0779	0.0907	0.727	0.646
December	267.674	-4.9866	-1.3103	-12.0314	6.6281	-0.1442	0.469	0.358

AOD in the pre-monsoon season due to clear days compared to that of the winter season (more cloudy days). Wind can play some role in the dilution of PM concentration; it prevents the stable condition of PM concentration area and could flush away and dilute the particulate matter concentration in the wider region with height and area. In some cases, wind can carry suspended mineral and dust particles to the measuring station which in turn could increase the Particulate matter concentration in that station. In our research, the first case was the dominant regression coefficient had negative values (increase in wind value, decrease in PM concentration) in both months.

The generation of secondary particles through photo-chemical phenomena could be the effect of an increase in temperature which later on increases the particulate matter concentration. This is the case for a higher temperature. But in colder temperatures also Particulate matter concentration increases due to the temperature inversion effect. Here in our study in both seasons, we can observe the negative value of the regression coefficient of temperature. Comparing the values of these coefficients we can see both are negative and relatively there is less difference in the average temperature of the winter season (December 2019) and pre-monsoon season (March 2020).

Under high relative humidity conditions (RH $\geq$ 80%) hygroscopic particles (e.g., ammonium nitrate and ammonium sulfate) can grow to 2–10 times their normal size, increasing the light extinction efficiencies of the particle (Ghotbi et al., 2016). Hence, the same AOD value at high relative humidity corresponds to lower particle dry mass compared to the obtained value at low humidity (Liu et al., 1999). In our study, both season regression coefficients of Relative Humidity (RH) are negative which indicates the reverse effect of RH on AOD. Comparatively winter season has more RH compared with the pre-monsoon season. Here correlation coefficient of RH is negative but the regression coefficient of the winter season is more than that of the pre-monsoon season.

Planetary Boundary Layer Height (PBLH) also plays an important role in this analysis. Here PBLH in the pre-monsoon season is more compared to that of the winter season so the regression coefficient of PBLH is positive in the pre-monsoon season and that of the winter season is negative.

We have found that PBLH has a positive coefficient with PM concentrations in the pre-monsoon season and a negative coefficient with PM concentrations in the winter season. The reason for this difference is that the atmospheric conditions during the pre-monsoon and winter seasons are different, which can affect how PBLH influences PM concentrations. In the pre-monsoon season, the higher PBLH values may allow for greater vertical mixing of pollutants and greater dispersion of PM, leading to a positive correlation between PBLH and PM concentrations. In contrast, during the winter season, lower PBLH values may lead to the accumulation of pollutants near the surface, resulting in a negative correlation between PBLH and PM concentrations.

Model Validation

Two different concentrations of estimated $PM_{2.5}$ were obtained from the single variable regression model and multivariable regression model respectively. The estimated $PM_{2.5}$ concentrations of these two models were calculated for two seasons one winter season (December) and the other pre-monsoon season (March). And these estimated $PM_{2.5}$ concentrations were validated by taking reference to the station of Phora Durbar, Kantipath, Kathmandu, Nepal. Both the estimated and station $PM_{2.5}$ data were daily 24-hour average data. Statistical tools like correlation coefficient, Normalized root mean square error, and mean bias were evaluated from Python to validate the model.

$PM_{2.5}$ was estimated using a single variable regression model equation developed in equations

Table 4: Statistical Analysis of Single Variable Regression Model

	Mean Bias (µg/m^3)	Correlation Coefficient	Normalized Root mean square
$PM_{2.5}$ December	-8.94	0.1388	0.234
$PM_{2.5}$ March	2.828	0.3223	0.47

(4) and (5) for pre-monsoon and winter seasons respectively. Table 4 shows the statistical analysis of the single-variable regression model. Here this model underestimated $PM_{2.5}$ for winter while the estimation for pre-monsoon is good. For December, the model showed a lower correlation in both seasons. Comparatively, the correlation coefficient of the pre-monsoon season is higher. From the values of the Normalized Root Mean Square difference, we can see the difference between the observed PM and estimated PM. The scatter plot of observed $PM_{2.5}$ and estimated $PM_{2.5}$ for winter and pre-monsoon season is shown in Figures 6 and 7 respectively.

Here, Figure 8 and Figure 9 show the relationship between observed $PM_{2.5}$ and single variable regression model estimated $PM_{2.5}$ in winter and pre-monsoon season respectively. Here we can see the gaps in both figures which is the result of missing AOD in our model equations due to the cloudy days.

Similarly, as a single variable regression model, $PM_{2.5}$ was estimated using a multivariable regression model equation developed in equations (6) and (7) for pre-monsoon and winter seasons respectively. Table 5 shows the statistical analysis of the multi-variable regression model.

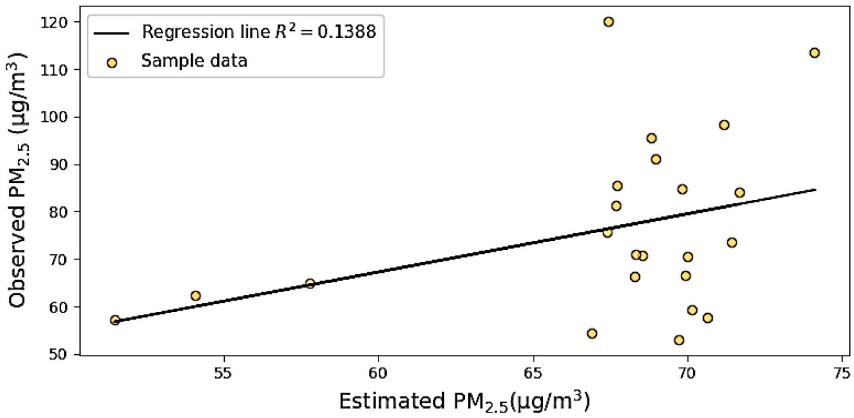

Figure 6: Scatter Plot of Phora Observed $PM_{2.5}$ and Single variable regression model Estimated $PM_{2.5}$ of Winter (December 2019)

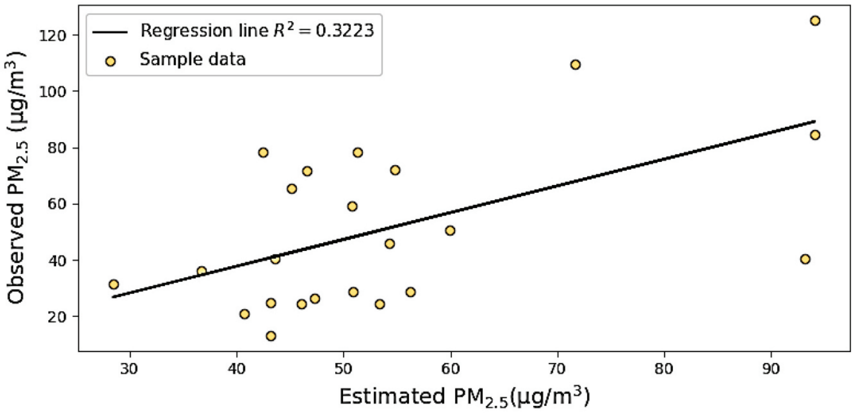

Figure 7: Scatter Plot of Phora Observed $PM_{2.5}$ and Single variable regression model Estimated $PM_{2.5}$ of Pre-monsoon (March 2020)

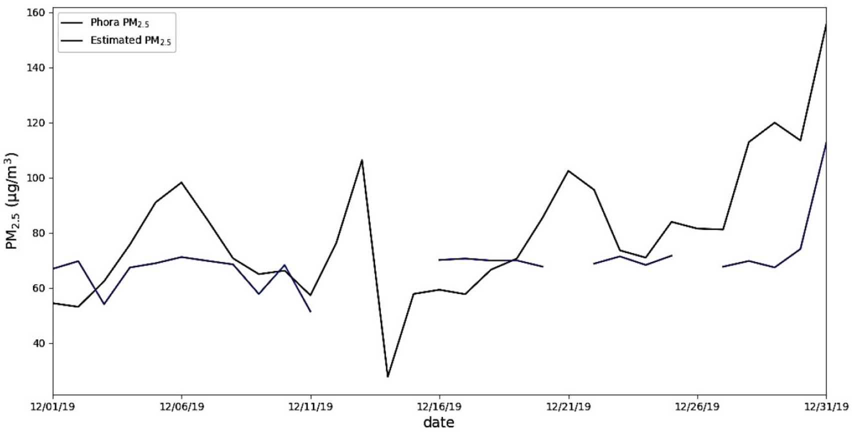

Figure 8: Comparison of Observed $PM_{2.5}$ and Single variable regression model Estimated $PM_{2.5}$ in Winter (December 2019)

Table 5: Statistical Analysis of Multi-Variable Regression Model

	Mean Bias (µg/m^3)	Correlation Coefficient	Normalized Root Mean square
$PM_{2.5}$ December	0.2307	0.4687	0.194
$PM_{2.5}$ March	-0.142	0.7266	0.292

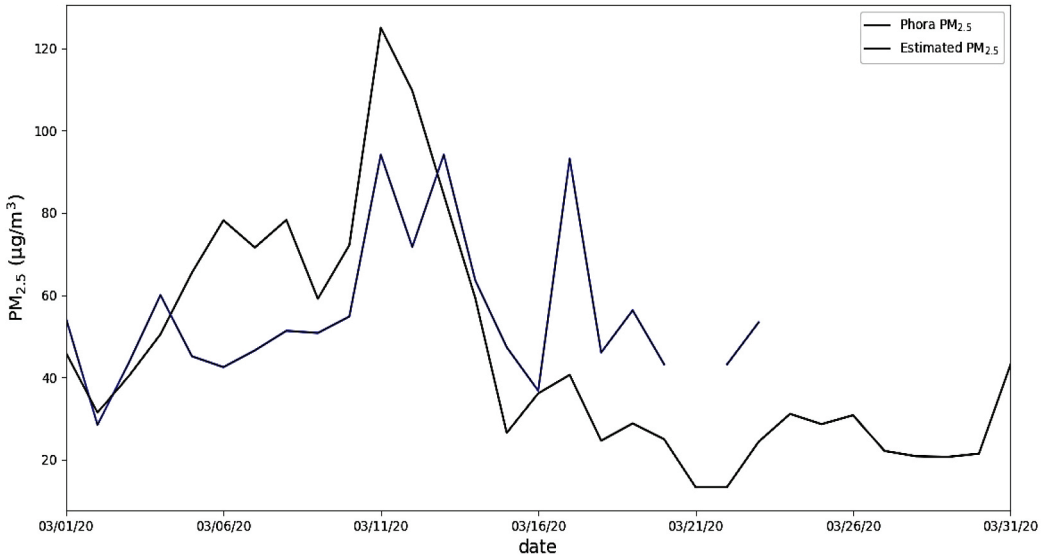

Figure 9: Comparison of Observed $PM_{2.5}$ and Single variable regression model Estimated $PM_{2.5}$ in Pre-monsoon (March 2020)

Here this model shows a good estimation of PM in both seasons. Here meteorological data are adopted from the station of New Road and PBLH from the WRF model. For December, the model showed a good correlation in the pre-monsoon season and a relatively lower correlation in the winter season. From the values of the Normalized Root Mean Square difference, we can see the difference between the observed PM and estimated PM. The scatter plot of observed $PM_{2.5}$ and estimated $PM_{2.5}$ for winter and pre-monsoon season is shown in Figures 10 and 11 respectively. Comparing the results with the single variable regression model has better results.

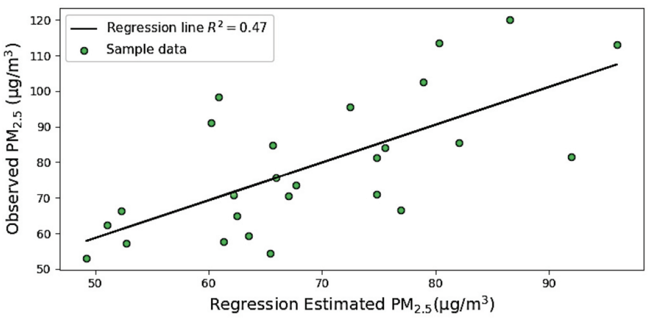

Figure 11: Scatter plot of Observed $PM_{2.5}$ and multi-variable regression model Estimated $PM_{2.5}$ of Winter (December 2019)

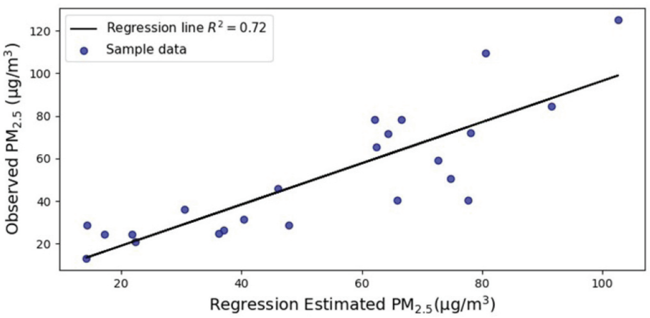

Figure 10: Scatter plot of Observed $PM_{2.5}$ and multi-variable regression model Estimated $PM_{2.5}$ of Premonsoon (March 2020)

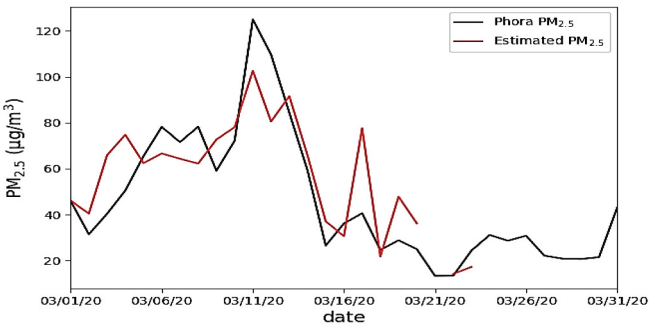

Figure 12: Comparison of Observed $PM_{2.5}$ and multi-variable regression model Estimated $PM_{2.5}$ in Premonsoon (March 2020)

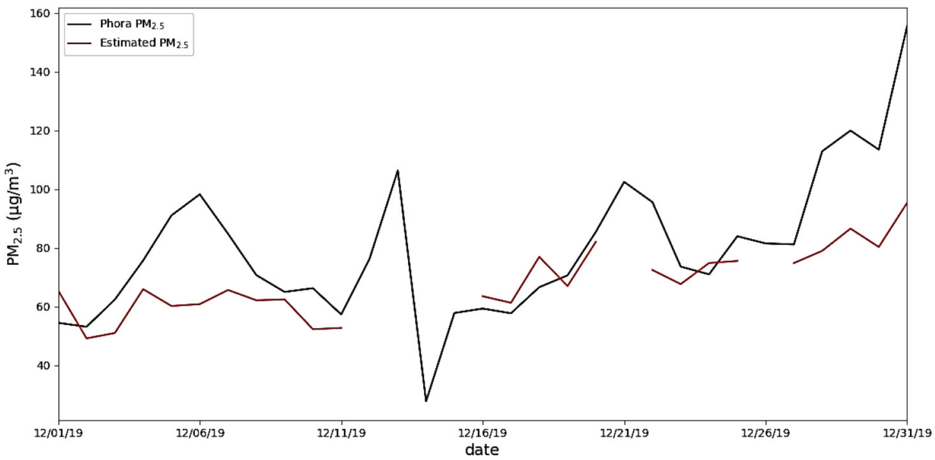

Figure 13: Comparison of Observed $PM_{2.5}$ and multi-variable regression model Estimated $PM_{2.5}$ in Winter (December 2019)

Similarly, Figure 12 and Figure 13 show the relationship between observed $PM_{2.5}$ and single variable regression model estimated $PM_{2.5}$ in winter and pre-monsoon season respectively. Here we can see the gaps in both figures which is the result of missing AOD in our model equations due to the cloudy days. To minimize the error caused by missing AOD during statistical analysis, masking is used in Python which helps to handle the missing or unwanted data.

In summary, the multivariable regression model gives better results in estimating particulate matter than the single-variable regression model. Here in the winter season correlation coefficients are quite low compared to the pre-monsoon season, it is due to the higher quality of pixels present during the extraction of AOD in the pre-monsoon season due to the availability of more clear days. During the estimation of PM, other meteorological variables also play an important role along with the AOD.

Conclusion

The present study estimates the ground-level particulate matter concentration using satellite-based MODIS AOD and derives the relation between particulate matter and AOD. AOD has been used as an input to a single variable regression model and AOD along with other meteorological factors (RH, wind speed, temperature, and PBLH) are used as an input to multivariable regression model developing model equations. MODIS Aqua and Terra retrieved AOD measurements and were employed to derive the correlation coefficient between $PM_{2.5}$ concentrations and AOD during December 2019 and March 2020. The AOD retrievals demonstrate geographical and seasonal variations in their relation to $PM_{2.5}$. Here two regression models were employed to estimate the particulate matter in two seasons (winter and pre-monsoon). A good correlation coefficient was observed in the pre-monsoon season using a multivariate Regression model which generated a higher coefficient of determination ($R^2 = 0.47$ in the winter season and $R^2 = 0.73$ in the pre-monsoon season). While the single variable regression model had a lower coefficient of determination ($R^2 = 0.13$ in the winter season and $R^2 = 0.45$ in the pre-monsoon season). So, the multivariable regression model can derive good relation between particulate matter and AOD. Comparing the seasonal results of each regression model pre-monsoon season has good results compared to the winter season. After comparing with ground station $PM_{2.5}$ concentrations, it can be concluded that $PM_{2.5}$ concentrations predicted by the multivariable regression model nearly followed a similar trend as $PM_{2.5}$ concentrations measured by ground stations. The multivariable Regression model generated the best performance among the two models. So, multivariable regression models can be valuable to conduct research related to air pollution and public health perspectives soon to estimate PM from satellite AOD.

Acknowledgement

Firstly, we would like to express our sincere gratitude to the Department of Environmental Science and Engineering, Kathmandu University for his overwhelming enthusiasm, continuous support, and supervision throughout our research. His ideas, knowledge, and efforts are heavily embedded in this research project. Above all, our deepest gratitude to our parents and family members for their unremitting support and encouragement throughout life.

References

Arvani, B., Pierce, R. B., Lyapustin, A. I., Wang, Y., Ghermandi, G., & Teggi, S. (2015). High spatial resolution aerosol retrievals used for daily particulate matter monitoring over Po valley, northern Italy. *Atmospheric Chemistry and Physics Discussions*, *15*(15), 123–155. https://doi.org/10.5194/acpd-15-123-2015

Chen, J., Xin, J., An, J., Wang, Y., Liu, Z., Chao, N., & Meng, Z. (2014). Observation of aerosol optical properties and particulate pollution at background station in the Pearl River Delta region. *Atmospheric Research*, *143*, 216–227. https://doi.org/10.1016/j.atmosres.2014.02.011

Christopher, S. A., & Gupta, P. (2010). Satellite remote sensing of particulate matter air quality: The cloud-cover problem. *Journal of the Air and Waste Management Association*, *60*(5), 596–602. https://doi.org/10.3155/1047-3289.60.5.596

Duncan, B. N., Prados, A. I., Lamsal, L. N., Liu, Y., Streets, D. G., Gupta, P., Hilsenrath, E., Kahn, R. A., Nielsen, J. E., Beyersdorf, A. J., Burton, S. P., Fiore, A. M., Fishman, J., Henze, D. K., Hostetler, C. A., Krotkov, N. A., Lee, P., Lin, M., Pawson, S., … Ziemba, L. D. (2014). Satellite data of atmospheric pollution for U.S. air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid. *Atmospheric Environment*, *94*, 647–662. https://doi.org/10.1016/j.atmosenv.2014.05.061

Embassy, U. (2020). *Air Quality Monitor - U.S. Embassy in Nepal*. US Embassy in Nepal. https://np.usembassy.gov/air-quality-monitor/

Ferrero, L., Riccio, A., Ferrini, B. S., D'Angelo, L., Rovelli, G., Casati, M., Angelini, F., Barnaba, F., Gobbi, G. P., Cataldi, M., & Bolzacchini, E. (2019). Satellite AOD conversion into ground PM10, PM2.5 and PM1 over the Po valley (Milan, Italy) exploiting information on aerosol vertical profiles, chemistry, hygroscopicity and meteorology. *Atmospheric Pollution Research*, *10*(6), 1895–1912. https://doi.org/10.1016/j.apr.2019.08.003

Ghotbi, S., Sotoudeheian, S., & Arhami, M. (2016). Estimating urban ground-level PM10 using MODIS 3km AOD product and meteorological parameters from WRF model. *Atmospheric Environment*, *141*, 333–346. https://doi.org/10.1016/j.atmosenv.2016.06.057

Goldberg, D. L., Gupta, P., Wang, K., Jena, C., Zhang, Y., Lu, Z., & Streets, D. G. (2019). Using gap-filled MAIAC AOD and WRF-Chem to estimate daily PM2.5 concentrations at 1/ km resolution in the Eastern United States. *Atmospheric Environment*, *199*, 443–452. https://doi.org/10.1016/j.atmosenv.2018.11.049

Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *Journal of Geophysical Research Atmospheres*, *114*(14), 1–13. https://doi.org/10.1029/2008JD011496

Lennartson, E. M., Wang, J., Gu, J., Castro Garcia, L., Ge, C., Gao, M., Choi, M., Saide, P. E., Carmichael, G. R., Kim, J., & Janz, S. J. (2018). Diurnal variation of aerosol optical depth and PM2.5 in South Korea: A synthesis from AERONET, satellite (GOCI), KORUS-AQ observation, and the WRF-Chem model. *Atmospheric Chemistry and Physics*, *18*(20), 15125–15144. https://doi.org/10.5194/acp-18-15125-2018

Liu, R. Y., Parelius, J. M., & Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Annals of Statistics*, *27*(3), 783–858.

Mirzaei, M., Bertazzon, S., Couloigner, I., Farjad, B., & Ngom, R. (2020). Estimation of local daily PM2.5 concentration during wildfire episodes: integrating MODIS AOD with multivariate linear mixed effect (LME) models. *Air Quality, Atmosphere and Health*, *13*(2), 173–185. https://doi.org/10.1007/s11869-019-00780-y

Mues, A., Rupakheti, M., Münkel, C., Lauer, A., Bozem, H., Hoor, P., Butler, T., & Lawrence, M. G. (2017). Investigation of the mixing layer height derived from ceilometer measurements in the Kathmandu Valley and implications for local air quality. *Atmospheric Chemistry and Physics*, *17*(13), 8157–8176. https://doi.org/10.5194/acp-17-8157-2017

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., Peckham, S. E., Grell, G. A., Michalakes, J., Trahan, S., Benjamin, S. G., Alexander, C. R., Dimego, G. J., Wang, W., Schwartz, C. S., Romine, G. S., … Duda, M. G. (2017). The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, *98*(8), 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1

Song, W., Jia, H., Huang, J., & Zhang, Y. (2014). A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China. *Remote Sensing of Environment*, *154*, 1–7. https://doi.org/10.1016/j.rse.2014.08.008

Sorek-Hamer, M., Kloog, I., Koutrakis, P., Strawa, A. W., Chatfield, R., Cohen, A., Ridgway, W. L., & Broday, D. M. (2015). Assessment of PM2.5 concentrations over bright surfaces using MODIS satellite observations. *Remote Sensing of Environment*, *163*, 180–185. https://doi.org/10.1016/j.rse.2015.03.014

Stirnberg, R., Cermak, J., & Andersen, H. (2018). An analysis of factors influencing the relationship between satellite-derived AOD and ground-level PM10. *Remote Sensing*, *10*(9). https://doi.org/10.3390/rs10091353

Tian, J., & Chen, D. (2010). A semi-empirical model for predicting hourly ground-level fine particulate matter (PM2.5) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sensing of Environment*, *114*(2), 221–229. https://doi.org/10.1016/j.rse.2009.09.011

Tsai, T. C., Jeng, Y. J., Chu, D. A., Chen, J. P., & Chang, S. C. (2011). Analysis of the relationship between MODIS aerosol optical depth and particulate matter from 2006 to 2008. *Atmospheric Environment*, *45*(27), 4777–4788. https://doi.org/10.1016/j.atmosenv.2009.10.006

You, W., Zang, Z., Zhang, L., Li, Z., Chen, D., & Zhang, G. (2015). Estimating ground-level PM10 concentration in northwestern China using geographically weighted regression based on satellite AOD combined with CALIPSO and MODIS fire count. *Remote Sensing of Environment*, *168*, 276–285. https://doi.org/10.1016/j.rse.2015.07.020

Zhao, R., Gu, X., Xue, B., Zhang, J., & Ren, W. (2018). Short period PM2.5 prediction based on multivariate linear regression model. *PLoS ONE*, *13*(7), 1–15. https://doi.org/10.1371/journal.pone.0201011