A Meteorology Based Particulate Matter Prediction Model for Megacity Dhaka

Sadia Afrin1,2†, Mohammad Maksimul Islam1,2†, Tanvir Ahmed1

1 Department of Civil Engineering, Bangladesh University of Engineering and Technology, Dhaka - 1000, Bangladesh
2 Department of Civil, Construction, and Environmental Engineering, North Carolina State University, 2501 Stinson Dr, Raleigh, NC 27695, USA

ABSTRACT

Dhaka, the capital of Bangladesh, is one of the megacities in the world with the worst air quality. In this study, we develop statistical models for predicting particulate matter (PM) concentration in ambient air of Dhaka using meteorological and air quality data from 2002 to 2004 of a continuous air quality monitoring station (CAMS). Model for finer fraction of PM (PM2.5) explains up to 57% variability of daily PM2.5 concentration, whereas model for coarser fraction (PM2.5-10) explains up to 35% of its variability, indicating that PM2.5 is influenced more by meteorology than PM2.5-10. Temperature, wind speed, and wind direction account for 94% of total PM2.5 variability explained by the model, while relative humidity contributes to 75% of total PM2.5 variability. Inclusion of PM lag effect increases models’ predictive power by 4–16%. In general, our developed models show promising performance in capturing the seasonal variability of Dhaka’s PM concentration, although overestimate the low concentrations during wet season (April–September). We validate these models using a recent dataset (2013–2017) from the same monitoring site, in which modeled PM show strong positive correlations with observed concentrations (r = 0.81 and 0.76 for PM2.5 and PM2.5-10 respectively). Models also exhibit strong predictive power in forecasting PM levels of two other CAMSs in Dhaka. Thus, the developed models have potentials to explain the temporal and spatial variability of daily PM within Dhaka. These models can be helpful to policymakers as they can predict daily PM at any location of Dhaka with reasonable accuracy if daily meteorological data and previous day’s PM concentration are available. The effect of climate change scenarios on air pollution dynamics of Dhaka can also be assessed using these models.

Keywords: Air quality, Particulate matter, Regression model, Seasonal variability

1 INTRODUCTION

Air pollution is the dominant driver of global mortality according to the ‘Global Burden of Disease’ study (Cohen et al., 2017). Every year, outdoor and indoor air pollution attributes to around 8 million premature deaths globally (WHO, 2016). Especially particulate matter (PM)—the most health-damaging air pollutant—has been found to be strongly associated with negative health outcomes (Bae and Hong, 2018). PM shows a strong positive association with all-cause mortality (Choi et al., 2009), respiratory morbidity, and possibly cardiovascular mortality and morbidity (Thurston et al., 2016). PM2.5 (particles with an aerodynamic diameter less than 2.5 μm) is also a big concern in ambient air in Bangladesh as evidenced by various observations spanning over several decades (Rahman et al., 2019). Observations since 1990 show that the population of this country has been exposed to PM2.5 levels above all three interim targets of the World Health Organization (WHO), making it one the countries with the highest PM2.5 exposures (Health Effects Institute, 2019). Thus, ambient PM2.5 has become one of the top risk factors affecting cardiovascular morbidity and mortality in Bangladesh (Islam et al., 2017). Each year, air pollution causes 195,000 premature deaths here (Rahman et al., 2019). PM2.5-attributed annual...
mortality in Bangladesh is also increasing over time and has been increased by 52% in the last two decades (Health Effect Institute, 2017).

The air pollution problem in Bangladesh is mostly concentrated in the urban cities of the country (Hossain et al., 2019). Among the major urban cities, air pollution in Dhaka, the capital of Bangladesh, is most concerning since it is one of the most densely populated cities in the world. Dhaka is ranked as the third most polluted megacities (Rahman et al., 2019) and the second most polluted capital in terms of poor air quality (IAQAir, 2018). The huge population of the city (~9 million) is constantly exposed to high concentration of PM$_{2.5}$ and PM$_{10}$ (particles with an aerodynamic diameter less than 10 µm) for many years (Begum et al., 2013; Majumder et al., 2019). Both PM$_{2.5}$ and PM$_{10}$ in Dhaka frequently exceed the Bangladesh National Ambient Air Quality Standard (NAAQS) and the U.S. NAAQS (Hossain and Easa, 2012), causing a great health concern. A recent study by Tasmin et al. (2019) found an association of short-term exposure to PM with worsening of lung function of school children in Dhaka. Maji et al. (2018) also estimated 9051 PM$_{2.5}$ attributed deaths in Dhaka during the year 2016. The estimated economic costs of air pollution attributed to sickness and deaths in Dhaka by the World Bank are valued approximately US$ 200–800 million per year (Afrin et al., 2012).

Brick kiln operations surrounding the city, fossil fuel combustion, industrial operation, vehicular emission, re-suspension of dust from unpaved roads and soil, construction works, and metal smelters are the primary sources of elevated PM concentration in Dhaka (Begum et al., 2010). Studies so far explored different aspects of air pollution in this city. Some studies focused on trend analysis of air pollutants (Rana et al., 2016a; Begum and Hopke, 2018), some explored the factors affecting seasonal variability of the air pollutants (Islam and Afrin, 2014; Islam et al., 2015), while several studies assessed the effectiveness of specific policies including the banning of “Two-Stroke Engines” and introduction of “Unleaded Gasoline” (Biswa et al., 2003; Begum et al., 2006). Some papers identified local sources by applying source apportionment techniques (Salam et al., 2008; Begum et al., 2010), while others highlighted the effect of transboundary sources (Rana et al., 2016a, b).

The elevated level of ambient PM concentration in Dhaka warrants initiatives to minimize the air pollution problem of the city. According to a World Bank report (World Bank, 2006), annually $169–492 million could be saved by a 20% reduction in PM$_{10}$ concentration in Dhaka. A recent report by “State of Global Air” (Health Effects Institute, 2019) shows that Bangladesh will gain the highest benefit in terms of life expectancy if it can meet the WHO guideline of 10 µg m$^{-3}$ of PM$_{2.5}$. An air quality prediction model could serve as an important policy tool to reduce the air pollution problem (Liang et al., 2015; Tong et al., 2019). Two basic types of models are used for air quality prediction—deterministic or chemical transport model (CTM) and empirical model. CTMs consider fundamental physics and chemistry of all possible emissions, and pollutant transport and transformation processes associated with air pollution. However, because of the huge complexities of CTM, uncertainties might arise from the assumptions considered while simulating pollutant concentration (Cobourn, 2010). Empirical or statistical models have also been used in some studies for air quality forecasting, assessing the effectiveness of adopted measures, and proposing possible measures of pollution reduction (Cobourn, 2010; Liu et al., 2012; Liang et al., 2015). Konovalov et al. (2009) found better performance of the empirical statistical model compared to CTM while forecasting PM$_{10}$ concentration.

In Dhaka, we do not have any unified, high-resolution, and time-varied emission information. Additionally, owing to the complexity and level of details required by a CTM, it is not always feasible to employ this modeling approach to develop a PM prediction model for Dhaka. Only one study so far has applied CTM to simulate pollutant concentrations in Dhaka (Azkar et al., 2012). However, owing to the unavailability of fine resolution emission inventory, the model greatly underestimated the observed pollutant concentrations. Since meteorology has been proved to be a significant driver of air quality worldwide (Tai et al., 2010; Pearce et al., 2011; Leung et al., 2017; Wang et al., 2019), a meteorology based statistical model (Cobourn, 2010) can be envisaged to forecast pollutant concentration in Dhaka. To our knowledge, no statistical/empirical model has been developed for Dhaka yet to predict PM concentration from meteorological inputs. Some studies explored the meteorology-air quality dynamics in Dhaka to different extents (Afrin et al., 2014; Islam et al., 2015, 2018; Kayes et al., 2019). Kayes et al. (2019) applied multiple linear and non-linear regression models to explore the role of three meteorological variables—temperature,
relative humidity, and precipitation—on particulate and gaseous air pollutants. Islam et al. (2015) determined bivariate “Pearson correlation coefficient” between meteorological parameters and air pollutants. That study considered solar radiation and wind speed in addition to the other variables considered by Kayes et al. (2019). In addition, Islam et al. (2018) applied multilinear regression and a cross-correlation approach to assess the effect of meteorological parameters on seasonal variation of PM concentration in Dhaka. Although studies have shown that wind coming from the northwest direction of Dhaka during the winter months increases PM concentration (Islam et al., 2015; Mahmood et al., 2019), wind direction was not included as a predictor in the multilinear regression or correlation analysis in the above-mentioned studies.

In addition to the meteorological variables, the previous day’s PM concentration (lag PM) has been identified as a significant predictor in the PM prediction model for different countries. Some studies considered the lag effect of PM concentration along with different matrices of fire smoke for estimating smoke-related PM$_{2.5}$ (Price et al., 2012; Yao and Henderson, 2013), while Marsha and Larkin (2019) incorporated previous day’s PM concentration for the development of statistical models for predicting ground-level PM$_{2.5}$ prediction. Yao and Henderson (2013) identified lag PM$_{2.5}$ as the most influencing predictor in their model predicting smoke-related PM$_{2.5}$. Given the significance of PM lag effect, a PM regression model having the previous day PM as a predictor might have significant potential in explaining the daily variability of monitored PM concentration of Dhaka. However, modeling studies for Dhaka have not assessed the potential of this predictor in improving overall model performance.

Given the influence of meteorology over the air quality in Dhaka and the potential of meteorological parameters to predict pollutant concentration, we apply multivariate regression to develop a statistical predictive model for ambient PM concentration of Dhaka city. We include wind direction components as predictors in the model in addition to temperature, rainfall or precipitation, relative humidity, solar radiation, and wind speed. We develop two sets of models—one to predict PM$_{2.5}$ concentration and the other to predict PM$_{2.5-10}$ (particles with an aerodynamic diameter in between 2.5 and 10 µm) concentration—using the available data of an air quality monitoring station in Dhaka. We also evaluate the model performance by incorporating first-order interaction terms between the meteorological parameters and the previous day’s PM concentrations as potential predictors. Additionally, we quantify the relative contribution of significant meteorological parameters and PM lag to the variation in 24-hour average PM concentration. Finally, we validate our prediction models by applying them to a recent dataset of the same monitoring site. We also assess the performance of the model in explaining spatial variability by employing it in two other monitoring stations of Dhaka.

2 METHODS

2.1 The Study Area and Data Source

This study analyzed air quality and meteorology data of Dhaka measured at Continuous Air Monitoring Stations (CAMSs). CAMSs record hourly PM data using Met One Model beta attenuation monitor (BAM)-1020, which is a U.S. Environmental Protection Agency Federal Equivalent Method (FEM) monitor (Mahmood et al., 2019). CAMS also measures meteorological parameters including rainfall, temperature, relative humidity, solar radiation, wind speed, and direction. CAMS data are reported monthly by the Department of Environment (DoE), Bangladesh. Dhaka has three CAMSs (Fig. 1) located at Sangshad Bhaban, Sher-e-Bangla Nagar (CAMS-1); Farmgate (CAMS-2); and Darus-Salam, Mirpur (CAMS-3). CAMS-1 is an urban air monitoring site that started operation in April 2002 and is located close to two major roads with heavy traffic congestion: Rokeya Sharani (~150 meters) and Manik Mia Avenue (~300 meters). The second CAMS is also an urban monitor located on the premises of the Bangladesh Agricultural Research Council (BARC) to capture the pollution from the busy Farmgate area. CAMS-3 is located at around 100m from a traffic-congested Mirpur road and near major brick kiln clusters. CAMS-2 started monitoring data since June 2008, while CAMS-3 started reporting data from November 2012 (Afrin et al., 2012).

We collected the daily PM (PM$_{2.5}$ and PM$_{10}$) and meteorological data of CAMS-1 for the 1st three years (April 2002–May 2004) of its operation, and developed a meteorology based
statistical model of PM using this dataset. We used a recent dataset of CAMS-1 from 2013 to 2017 for model validation. We also used the 2013–2017 data of the other two CAMSs for the models’ spatial validation. For the analysis, we split the total PM$_{10}$ concentrations into two size fractions: the finer fraction (PM$_{2.5}$) and the coarser fraction (PM$_{2.5-10}$). We used daily average atmospheric temperature (T), rainfall (R), relative humidity (RH), and solar radiation (SR), wind speed (WS), and two components of wind direction (WD): North-South (Cosine of WD) and East-West (Sine of WD). The daily variations of PM fractions and meteorological parameters at CAMS-1 for a typical winter month (February 2015) are shown in Fig. S2 and the summary statistics of all three CAMSs are presented in Table S1. PM concentrations and meteorological parameters show day-to-day variation over a month with no consistent weekday versus weekend trend. While the daily variation in PM fractions (PM$_{2.5}$ and PM$_{2.5-10}$) seem to follow a similar trend with peaks and falls at the same time (day) of the month, we do not observe any consistent trend among the meteorological variables (Fig. S2).

2.2 Model Development and Validation

In this study, we used multiple regression to develop generalized linear models of PM fractions (PM$_{2.5}$ and PM$_{2.5-10}$). We applied the traditional regression approach (model having all predictors included) and several stepwise approaches (forward selection, backward elimination, and bidirectional) to our data and calculated adjusted R$^2$ and Akaike information criterion (AIC) values of each model as adjusted R$^2$ and AIC are frequently used to assess the relative performance of models (Wang et al., 2018; Han et al., 2019). In stepwise regression, the choice of predictive variables was carried out considering forward, backward, or both-way selections of variables depending on the approaches. The final model only retained the significant predictors based on the p-value for the F or chi-squared test of the change in the deviance by adding or removing the term. In our analysis, we defined a significance level of p < 0.05. We tested assumptions of normality, independence of errors, and collinearity using the probability plots, “Durbin-Watson” test and “variance inflation factor (VIF)” respectively. Assumptions of linearity and homogeneity of variances were also checked with Residuals-Fits plots.

Initially, we simulated a set of models including only linear terms of the meteorological variables, $x_i$ (Eq. (1)). To assess the influence of the interaction between these meteorological variables in explaining the variability of the PM model, we developed a second set of models including both linear terms ($x_i$) and first-order interactions between these meteorological variables ($x_ix_j$) as potential predictors (Eq. (2)). In addition, we developed another set of models appending the lag PM concentrations to the linear and interaction terms of the meteorological variables (Eq. (3)) to explore the possible lag effect of PM.

$$PM = \beta_0 + \sum_{i=1}^{7} \beta_i x_i$$

(1)

$$PM = \beta_0 + \sum_{i=1}^{7} \beta_i x_i + \sum_{i=1}^{6} \sum_{j=i+1}^{7} \beta_{ij} x_i x_j$$

(2)

$$PM = \beta_0 + \beta_{lag}PM + \sum_{i=1}^{7} \beta_i x_i + \sum_{i=1}^{6} \sum_{j=i+1}^{7} \beta_{ij} x_i x_j$$

(3)

In these equations, PM could be either 24-hour average PM$_{2.5}$ or PM$_{2.5-10}$ during the study period considered, $\beta_0$ is the intercept term, $\beta_i$ is the regression coefficient for lag PM term, and $\beta_i$ and $\beta_{ij}$ are the regression coefficients for meteorological variables ($x$).

Further, we quantified the relative importance of the weather predictors and PM lag in the regression models by applying the average increment method (Tonidandel and LeBreton, 2011). This method follows the principle of dominance analysis (Azen and Budescu, 2003) to calculate the shared variance by individual predictors to the total variance or explaining power of a model. Finally, the PM regression models developed using CAMS-1’s earlier dataset (2002–2004) was validated with its recent dataset (2013–2017). We applied daily meteorological data from 2013...
to 2017 on the models to calculate modeled PM and compare it with observed 24-hour average PM for validation. We also applied the recent dataset of two other CAMSs (2 and 3) in the models to assess their spatial validity. In this validation process, we used the Pearson correlation coefficient or “r-value” to measure how strong the correlation is between observed and modeled PM concentrations. We also calculated normalized mean error (NME) and normalized mean bias (NMB) to assess the performance of the models (Appel et al., 2011). To investigate the influence of seasonality in the model-vs-measurement comparison, we divided the yearly data into two characteristic seasons in Bangladesh based on rainfall pattern: dry-season (October–March) and wet-season (April–September) (Islam et al., 2015). We further divided each of these two seasons into two quarterly periods (dry-period: January–March (JFM), October–December (OND); and wet-period: April–June (AMJ), July–September (JAS)) during the model validation process to discern quarterly trend. In addition, we conducted monthly validation of models and assessed the weekday versus weekend influence in model performance.

3 RESULTS AND DISCUSSION

3.1 PM Regression Models with Linear Terms Only

The significant predictors with coefficients (β) of PM$_{2.5}$ and PM$_{2.5-10}$ models for different multilinear regression approaches along with respective adjusted R$^2$ and AIC values are summarized in Table S2. In general, adjusted R$^2$ and AIC values of models obtained for different approaches are close to each other. For the PM$_{2.5}$ model, a backward elimination approach provides the largest adjusted R$^2$ with the lowest AIC value, while for the PM$_{2.5-10}$ model, all three stepwise approaches yield the same result. Hence, we consider the model from the backward elimination approach for the rest of the discussion in this paper and tabulate the summary statistics of its outcome in Table 1.

As illustrated in Table 1, the PM$_{2.5}$ model can explain 57% variability of daily PM$_{2.5}$ concentrations. Temperature, relative humidity, and wind parameters appear as significant predictors for the PM$_{2.5}$ model. The negative association between PM$_{2.5}$ and RH can be explained by the growth and subsequent wet deposition of particles. Several studies (Islam et al., 2015, 2018; Kayes et al., 2019) also observed a similar effect of RH on PM$_{2.5}$. The inverse relationship between T and PM$_{2.5}$ is common in literature (Wang et al., 2013) and likely indicates the cooling effect of PM due to its net negative radiative forcing (Islam et al., 2015). Temperature was observed to be positively associated with planetary boundary layer (PBL) height and thus can reduce the PM$_{2.5}$ concentrations by increasing the PBL height as well as pollutant dispersion. The T-PM$_{2.5}$ negative association can also be linked to seasonal variation of PM of Dhaka, which shows higher PM levels in dry winter months (October–March) compared to wet warmer months (April–September) (Islam et al., 2015). Wind speed shows a negative relation with PM$_{2.5}$ indicating the probable dispersion and subsequent dilution of particle concentration in the air due to wind speed (Islam et al., 2018). The positive association of PM with the North-South component of wind speed (CosWD) suggests that PM$_{2.5}$ concentrations increase when winds come from north directions.

Table 1. Significant parameters of the PM backward elimination regression models considering only the linear terms of meteorological parameters as predictor variables.

<table>
<thead>
<tr>
<th>Response variables (PM)</th>
<th>Predictor variable (x)</th>
<th>Estimate (β)</th>
<th>Standard Error (SE)</th>
<th>t-statistics</th>
<th>P-value</th>
<th>VIF</th>
<th>Durbin-Watson</th>
<th>Adjusted R$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM$_{2.5}$</td>
<td>Intercept</td>
<td>416.86</td>
<td>34.1</td>
<td>12.23</td>
<td>2E-28</td>
<td></td>
<td>0.95</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>-8.39</td>
<td>0.75</td>
<td>-11.26</td>
<td>6E-25</td>
<td>1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RH</td>
<td>-1.2</td>
<td>0.3</td>
<td>-3.95</td>
<td>1E-04</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WS</td>
<td>-11.41</td>
<td>3.24</td>
<td>-3.52</td>
<td>5E-04</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SinWD</td>
<td>-10.08</td>
<td>5.13</td>
<td>-1.96</td>
<td>5E-02</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CosWD</td>
<td>11.98</td>
<td>4.17</td>
<td>2.87</td>
<td>4E-03</td>
<td>1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM$_{2.5-10}$</td>
<td>Intercept</td>
<td>298.8</td>
<td>18.06</td>
<td>16.55</td>
<td>8E-45</td>
<td>1.1</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>-2.51</td>
<td>0.38</td>
<td>-6.62</td>
<td>2E-10</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RH</td>
<td>-2.27</td>
<td>0.2</td>
<td>-11.57</td>
<td>5E-26</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This finding is in line with earlier literature (Islam et al., 2015; Mahmood et al., 2019) where winds coming from north-west direction were found to transport particles to Dhaka from the major brick-kiln clusters located there. The negative association of PM$_{2.5}$ with the East-West wind direction (SinWD) suggests that winds coming from this direction drive particles away from the city. These findings highlight the potential role of wind direction as a model predictor, which has not been explored by previous air quality studies in Dhaka (Islam et al., 2018; Kayes et al., 2019).

The PM$_{2.5:10}$ regression model shows lower predictive power (Adjusted $R^2 = 0.35$) than that of the PM$_{2.5}$ model (Table 1), suggesting that meteorology has less influence over coarser PM fraction than the finer fraction in Dhaka. Like the PM$_{2.5}$ model, T and RH are the significant predictors of the PM$_{2.5:10}$ model, and their negative associations with PM$_{2.5:10}$ are comparable to other studies (Islam et al., 2015). However, the wind speed and direction do not appear to be significant predictors in the PM$_{2.5:10}$ model indicating that particle transportation by winds to and from Dhaka has less influence on the coarser fraction than the finer one.

3.2 Quantification of the Relative Contribution of Model Parameters

The t-Statistics (absolute value) presented in Table 1 indicates the comparative contribution of different predictors to explain model variability. For the PM$_{2.5}$ model, T has a higher contribution than the other significant predictors, whereas, for the PM$_{2.5:10}$ model, RH has a higher influence. Using dominance analysis, we further quantify the proportional contribution of each predictor to the total variability explained by the models. Fig. 1 represents the total variability as well as the shared variability explained by the significant meteorological parameters influencing the PM concentration. For the PM$_{2.5}$ model, T can explain 58% of the total variability, whereas WS and WD together can explain 38% variability. Although RH has a small influence on PM$_{2.5}$, it explains 72% of the total variability of coarser particles with the rest of the variability explained by temperature.

3.3 Validation of Models

As mentioned earlier, we validate the PM$_{2.5}$ and PM$_{2.5:10}$ regression models of CAMS-1 using its recent dataset. Fig. 2 depicts the outcomes of the validation process. Fig. 2(a) shows the time series plots of 24-hour average observed and modeled PM$_{2.5}$ concentrations at CAMS-1 for the 2013-2017 period. Note that, CAMS-1 does not have observed data for the entire 2016 and early dry months (up to March) of 2017. In addition, data are not available for the 50% days of the remaining period between 2013 and 2017. Fig. 2(a) exhibits a distinct seasonal variation of observed PM$_{2.5}$ with high concentrations during the dry season. Modeled PM$_{2.5}$ follows the observed PM$_{2.5}$ over the seasons indicating the model’s ability to capture the seasonal variability. A strong and significant positive correlation coefficient ($r = 0.81$) between observed and modeled datasets (Fig. 2(b)) also verifies the model’s ability to predict ambient PM$_{2.5}$ concentration from the meteorological variables considered. However, the model does not seem to capture some low concentration events during the wet season.

![Fig. 1. The relative contribution of the considered meteorological variables in explaining the total variability.](https://example.com/fig1.png)
This observation led us to assess the model’s performance at yearly, seasonal, quarterly, and monthly bases. Table S3 summarizes the different performance indicators (e.g., NMB, NME, r-value, and associated p-values) at those time bases. The yearly evaluation shows a consistently strong correlation between the modeled and measured PM$_{2.5}$ concentrations from 2013-2015 ($r > 0.80$). Insufficient sample sizes in 2016 and 2017 (9 and 55 data points respectively) do not allow assessing model performance, which is also reflected by the statistical insignificance (p-value > 0.05) in correlations in those years. In terms of NMB and NME, yearly model evaluations also show promise. For example, NME values for the 2013–2015 time period range from 49% to 68%, which are below 75%—one of the NME thresholds used in model evaluation process (Appel et al., 2011). Similarly, NMB values, ranging from 24% to 65%, are also either below or close to a corresponding NMB threshold of 60% (Appel et al., 2011).

The season-wise validation of models reveals an interesting fact (Table S3). The model seems to predict PM$_{2.5}$ concentrations better in dry season than wet season as evidenced by the NME, NMB, and r-values. While NMB and NME values for the dry season are below the respective thresholds mentioned above, these values are much higher than the thresholds during the wet season. Correlation between measured and modeled PM$_{2.5}$ is also stronger during the dry season ($r = 0.63$) relative to wet season ($r = 0.38$). Quarterly evaluation of model performance also indicates better performance of models during the dry season than that of wet season. The two quarters of the wet season (AMJ and JAS) have higher NME, NMB values than the thresholds, and lower correlation coefficients than those of dry season. This finding is also evident from the time series plot of modeled and observed PM concentrations of a representative year covering two dry quarters and two wet quarters from October 2014 to September 2015 (Fig. 3).

NME and NMB values of monthly evaluation are shown in a soccer plot (Fig. 4) and tabulated...
Fig. 3. Variation of observed and modeled 24-hr average PM concentrations at CAMS-1 with and without the lag effect of previous day’s PM. This figure covers the time series plots of a typical dry (October–March) and a wet (April–September) season from October-2014 to September-2015 for (a) PM$_{2.5}$ concentrations and (b) PM$_{2.5-10}$ concentrations. The three-monthly quarters (JFM, AMJ, JAS, and OND) are also highlighted by alternate grey shades.

Fig. 4. Soccer plot for PM$_{2.5}$ and PM$_{2.5-10}$ models performance evaluation. Each data point is representing the model’s performance evaluation statistics by month. For each month, statistics based on data of CAMS-1 and CAMS-3 are shown separately as rectangular and triangular markers respectively. Again, for a specific CAMS, the filled markers are presenting the statistics when all available data points from 2013 to 2017 under that month are considered in the analysis, while the empty markers are presenting the statistics when data points excluding the weekends are considered. These data points are further colored by their corresponding quarters (JFM, AMJ, JAS, and OND).
in Table S3. NMB and NME values of all dry months (except October) lie within or close to the goal bar enclosed by ±30% NMB and 50% NME values, whereas wet months are even outside the bar enclosed by ±60% NMB and 75% NME. Note that on soccer plots, percentages of NME and NMB close to zero indicate better model performance compared to the bars further away. This observation restates the fact that model predicts PM$_{2.5}$ concentration better in dry months relative to wet months. Interestingly, the correlations between the modeled and measured PM$_{2.5}$ concentration at monthly levels are weaker than those for the previous time bases and we observe a gradual decrease in r-values as we move down from yearly to smaller time bases (e.g., yearly > seasonal > quarterly > monthly). Overall, the positive NMB beyond the considered threshold for the time bases of wet-season illustrates that the model is overestimating PM$_{2.5}$ concentration during wet summer months at CAMS-1. Model performance matrices for CAMS-3 is also evaluated and presented in Fig. 4.

Like PM$_{2.5}$, the time series plot of observed PM$_{2.5-10}$ in CAMS-1 (Fig. 2(c)) shows seasonal variation, and the model seems to capture the variability. The strong positive correlation ($r = 0.76$) between the modeled and observed PM$_{2.5-10}$ exhibits a promising performance of the model in predicting coarser PM fractions. However, similar to PM$_{2.5}$, the PM$_{2.5-10}$ model mostly overestimates the observed values during the wet months, and hence, we validate the model performance on different time bases (e.g., yearly, seasonal, quarterly, and monthly). A summary of this validation process is shown in Table S4. NMB and NME values associated with the yearly evaluation range from 30 to 55% and 41 to 58% respectively (excluding 2016 and 2017 due to insufficient sample size), which are below the 60% NMB and 75% NME thresholds. Yearly correlation coefficients between the observed and modeled PM$_{2.5-10}$ concentrations are also high ($r > 0.70$, $p < 0.05$), suggesting consistent model performance throughout the evaluation period. Seasonal and quarterly evaluations indicate better model performance in the dry season than the wet season, similar to PM$_{2.5}$. The monthly evaluation of PM$_{2.5-10}$ model displays better performance in terms of bias and error than that of the PM$_{2.5}$ model as shown in the soccer plot (Fig. 4). However, for both models (PM$_{2.5}$ and PM$_{2.5-10}$), NMB and NME values of all dry months (except October) lie within the ±30% NMB and 50% NME limit, whereas wet months remain outside the ±60% NMB and 75% NME limit. During the monthly model validation, we also explore the weekday versus weekend effect on model performance by calculating the bias and error values excluding the weekend. The resulting NMB and NME values are also shown in the soccer plots (Fig. 4). We do not observe any substantial and consistent change in model (both PM$_{2.5}$ and PM$_{2.5-10}$) performances due to the weekday-weekend effect, suggesting that differences in activities between the weekdays and weekends do not play a major role in model performance for the study area.

### 3.4 Model Application to Other Monitoring Sites

We apply CAMS-1 PM regression models to CAMS-2 and CAMS-3 to test its spatial validity within Dhaka. Using the meteorological data of CAMS-2 and CAMS-3, we calculate PM concentrations employing the CAMS-1 regression models. A comparison between modeled and observed PM demonstrates its applicability to other locations within Dhaka. Figs. S3(a) and S3(c) show the time series plots of observed and modeled PM$_{2.5}$ concentrations at CAMS-3 and CAMS-2 respectively. CAMS-3 has 64% daily data available between 2013 and 2017, whereas CAMS-2 has representative data in 2013 only. For the rest of the period in CAMS-2, data is either missing or having a recording issue (e.g., PM$_{2.5}$ is higher than PM$_{10}$). Similar to CAMS-1, modeled PM$_{2.5}$ seems to follow the observed PM$_{2.5}$ in both CAMSSs. The positive correlation between the observed and modeled PM$_{2.5}$ in CAMS-3 is as strong ($r = 0.81$) as CAMS-1 (Fig. S3(b)). However, the model seems to start underestimating the observed data at a lower concentration (100 µg m$^{-3}$) in CAMS-3 relative to CAMS-1 (180 µg m$^{-3}$). Monthly NMB and NME values are, in general, lower for CAMS-3 data relative to CAMS-1 data (Fig. 4), indicating the promising performance of the model developed by CAMS-1 data in predicting PM concentrations in the CAMS-3 area. For CAMS-2, the association between observed and modeled PM$_{2.5}$ concentration ($r = 0.61$) is not as strong as the other two CAMSs, possibly due to having limited data points (Figs. S3(c) and S3(d)). The model mostly overestimates the available observed data over the entire PM$_{2.5}$ range. A scattered pattern is also observed for the modeled PM when observed PM$_{2.5}$ concentration is below 50 µg m$^{-3}$. Similar to PM$_{2.5}$, the PM$_{2.5-10}$ model works well for CAMS-3 in capturing the seasonality (Fig. S3(e)). The correlation between observed and modeled dataset is also strong ($r = 0.73$), and comparable to
2.5 PM concentrations. We replace the interaction terms and PM lag are included as covariates. Table 2 summarizes the PM models including interaction terms between considered meteorological parameters and reports the estimated intercept, regression coefficients of significant predictors, and adjusted-R² of models. For the PM_{2.5} model, in addition to linear terms (T, RH, WS), interaction between wind speed and wind direction component appears as a significant predictor and tends to reduce the concentration, consistent with the model without interactions. However, we do not see an increase in adjusted-R² suggesting that model’s predictive power remains the same. For PM_{2.5-10}, the model does not change when interaction terms are included depicting the insignificant influence of meteorological interaction over coarse particles.

3.6 PM Regression Models with Previous Day’s Concentration
The inclusion of previous day’s PM concentration (PM lag effect) substantially increases the predictive power of fine and coarse PM models in explaining variability (Table 2). The PM_{2.5} model with PM lag effect can explain 16% more variability than the initial models with meteorological parameters only. By estimating the relative influence of the previous day’s PM_{2.5} concentration in explaining the overall variability of the PM_{2.5} model (e.g., average increment method), we find that lag PM_{2.5} alone can explain ~60% of the total explaining variability of the model, while meteorological variables explain the rest 40%. Similar to the PM_{2.5} model, addition of previous day’s coarse PM fraction increases the PM_{2.5-10} model’s explanatory power, but only by 4% relative to the model with meteorological variables. Consequently, PM concentrations from the model with lag effect show a stronger positive correlation with observed PM_{2.5} concentrations at CAMSs than that of the model without the lag. For instance, at CAMS-1, the estimated Pearson correlation coefficients between observed and modeled (with lag PM) concentrations between 2013 and 2017 are 0.94 and 0.82 respectively for PM_{2.5} and PM_{2.5-10}. Modeled PM concentrations also seem to capture the peak and fall of the observed PM more representatively relative to the model without the lag (Fig. 3). This is also reflected by the lower NMB and NME values of PM models with lag-effect relative to model without lag PM. For example, NME and NMB values of the PM_{2.5} model with lag are 25% and 17% respectively (2013–2017), which are substantially lower than those for the model without lag (59% and 48% respectively). We also explore the model performance by varying the averaging window of lag PM concentrations. We replace the previous day’s PM with the average of previous two days’ PM concentrations as a predictor in

Table 2. Summary of stepwise (backward elimination) multilinear regression models of PM when interaction terms and PM lag are included as covariates.

<table>
<thead>
<tr>
<th>Predictors*</th>
<th>PM_{2.5}</th>
<th>PM_{2.5-10}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interaction and PM lag</td>
<td>Interaction and PM lag</td>
</tr>
<tr>
<td>Intercept</td>
<td>409.63</td>
<td>298.80</td>
</tr>
<tr>
<td>R</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T</td>
<td>-6.29</td>
<td>-2.51</td>
</tr>
<tr>
<td>RH</td>
<td>-1.14</td>
<td>-2.27</td>
</tr>
<tr>
<td>SR</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WS</td>
<td>-11.53</td>
<td>-11.75</td>
</tr>
<tr>
<td>SinWD</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CosWD</td>
<td>12.14</td>
<td>-</td>
</tr>
<tr>
<td>WS*SinWD</td>
<td>-7.09</td>
<td>-</td>
</tr>
<tr>
<td>PM lag</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.57</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>0.73</td>
<td>0.39</td>
</tr>
</tbody>
</table>

*Table only lists the interaction terms that appeared significant (p-value < 0.05) in the models with no multi-collinearity. Predictors with no coefficients (represented by a horizontal dash: -) indicate that those did not appear significant in the respective model 1.
the model to identify any changes in model performance. This new model does not show any increment in adjusted $R^2$ from the previous model (Table S5). A model with previous three days’ PM concentrations, instead of the previous day’s PM, remains the same too (Table S5), indicating that lag PM concentrations beyond one day do not influence model performance for the study area.

4 CONCLUSIONS

This paper presents a simple statistical modeling framework for PM in Dhaka city. The models show substantial predictive power to forecast PM$_{2.5}$ and PM$_{2.5-10}$ concentrations from meteorological inputs. Important meteorological variables seem to explain 57% variability in PM$_{2.5}$ concentration, whereas for PM$_{2.5-10}$, it is 35%, suggesting that finer PM fraction is influenced more by meteorology than the coarser fraction. Temperature and relative humidity were significant predictors for both fractions of PM. Wind speed and wind direction did not appear to have significant influences on the PM$_{2.5-10}$ model, although wind parameters explained more than one-third of the total variability explained by the PM$_{2.5}$ model. Validation of the models by a dataset, 11 years apart, reflects its robustness to capture temporal variability. In addition, the promising performance of the model even for two different monitoring sites of Dhaka indicates its ability to capture spatial variability within the city. These findings imply that the meteorology based air quality model developed for a monitoring site of Dhaka can also be applied to other sites.

The modeling framework presented here has other implications too. We can use these models to predict the impact of climate change on ambient PM concentrations of Dhaka, given the same emission scenario. For example, the model indicates a 21% decrease in daily PM$_{2.5}$ levels in Dhaka for a 5% increase in considered meteorological parameters (R, T, RH, SR, and WS) due to climate change, if same emission scenario prevails. Note that, 5% increase in all meteorological parameters is unlikely and is used here for comparison purposes only. The models presented here, especially with PM lag effect, can also be used to generate PM forecast in Dhaka from weather prediction. In addition, PM concentrations can be predicted from this model for any location in the city if meteorology is known. We can also have a retrospective analysis of the PM trend of the city where PM concentration data were not reported or were missing for a certain period of interest.

The modeling framework presented here is developed based on daily meteorological and PM concentrations data, aiming to predict daily PM levels in Dhaka. A similar modeling approach can be applied to develop models for other air pollutants (e.g., $O_3$, NO$_x$, SO$_x$). In this study, we were not able to develop hourly PM models due to the unavailability of reported hourly air quality and meteorological data. Hence, we recommend future studies to develop models that can predict hourly PM concentrations upon the availability of hourly data. Note that, the predictive power of the developed PM models is indicated by their respective adjusted $R^2$ values, and the predictions will have uncertainty. Hence, predicted PM values should be used with extreme caution acknowledging the limitations. In addition, the model is based on air quality and meteorological data of Dhaka only, and hence cautions are needed against using it for other geographical locations where emission scenarios and meteorology are different.

Despite having some limitations as any other model, this model has policy implications too. For example, this model quantifies the contribution of wind speed and directions to ambient PM variability and thus restates the harmful effect of brick kiln clusters located in the north-west direction of Dhaka (Begum et al., 2011; Islam et al., 2015; Mahmood et al., 2019). This finding might convince policymakers to think about the relocation of brick kiln clusters and help them find suitable sites for the move out based on wind speeds and direction. This study will also make policymakers think twice before approving any air polluting industries in and around Dhaka city.

SUPPLEMENTARY MATERIAL

Supplementary data associated with this article can be found in the online version at https://doi.org/10.4209/aaqr.2020.07.0371
REFERENCES


Cogburn, W.G. (2010). An enhanced PM_{2.5} air quality forecast model based on nonlinear regression and back-trajectory concentrations. Atmos. Environ. 44, 3015–3023. https://doi.org/10.1016/j.atmosenv.2010.05.009


