Estimating ground-level PM$_{2.5}$ over Bangkok Metropolitan Region in Thailand using aerosol optical depth retrieved by MODIS

Bussayaporn Peng-in$^{1}$ · Peeyaporn Sanitluea$^{1}$ · Pimnapat Monjatturat$^{1}$ · Pattaraporn Boonkerd$^{1}$ · Arthit Phosri$^{1,2}$

Received: 28 January 2022 / Accepted: 15 August 2022 / Published online: 26 August 2022 © The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

A number of previous studies have shown that statistical model with a combination of satellite-derived aerosol optical depth (AOD) and PM$_{2.5}$ measured by the monitoring stations could be applied to predict spatial ground-level PM$_{2.5}$ concentration, but few studies have been conducted in Thailand. This study aimed to estimate ground-level PM$_{2.5}$ over the Bangkok Metropolitan Region in 2020 using linear regression model that incorporates the Moderate Resolution Imaging Spectroradiometer (MODIS) AOD measurements and other air pollutants, as well as various meteorological factors and greenness indicators into the model. The 12-fold cross-validation technique was used to examine the accuracy of model performance. The annual mean (standard deviation) concentration of observed PM$_{2.5}$ was 22.37 (± 12.55) µg/m$^3$ and the mean (standard deviation) of PM$_{2.5}$ during summer, winter, and rainy season was 18.36 (± 7.14) µg/m$^3$, 33.60 (± 14.48) µg/m$^3$, and 15.30 (± 4.78) µg/m$^3$, respectively. The cross-validation yielded $R^2$ of 0.48, 0.55, 0.21, and 0.52 with the average of predicted PM$_{2.5}$ concentration of 22.25 (± 9.97) µg/m$^3$, 21.68 (± 9.14) µg/m$^3$, 29.43 (± 9.45) µg/m$^3$, and 15.74 (± 5.68) µg/m$^3$ for the year round, summer, winter, and rainy season over the Bangkok Metropolitan Region. We also observed that integrating NO$_2$ and O$_3$ into the regression model improved the prediction accuracy significantly for a year round, summer, winter, and rainy season over the Bangkok Metropolitan Region.

In conclusion, estimating ground-level PM$_{2.5}$ concentration from the MODIS AOD measurement using linear regression model provided the satisfactory model performance when incorporating many possible predictor variables that would affect the association between MODIS AOD and PM$_{2.5}$ concentration.

Keywords PM$_{2.5}$ · Aerosol optical depth · MODIS · Satellite · Bangkok Metropolitan Region

Introduction

A number of previous epidemiological studies have revealed that exposure to particulate matter with aerodynamic diameter less than or equal to 2.5 µm (PM$_{2.5}$), both short-term and long-term exposure, was associated with an increased risk of morbidity and mortality for particularly cardiovascular and respiratory diseases (Alexeeff et al. 2021; Atkinson et al. 2014; Fan et al. 2016; Farhadi et al. 2020; Sangkharat et al. 2019; Thongphunchung et al. 2021; Wang et al. 2020). Over the recent years, PM$_{2.5}$ has been considered one of the most serious environmental and public health problems in Thailand, especially in Bangkok Metropolitan Region (BMR) (i.e., Bangkok and its five surrounding provinces, including Nakhon Pathom, Nonthaburi, Pathum Thani, Samut Prakan, and Samut Sakhon), which is the most densely populated region of Thailand. Moreover, the rapid economic development with many commercial activities in BMR has exacerbated in favors of high PM$_{2.5}$ emissions, leading to increase ambient PM$_{2.5}$ concentrations where daily average concentration has sometimes higher than the 24-h ambient standard of Thailand, especially during winter (i.e., November to February) (Thongphunchung et al. 2021). Because monitoring ground-level PM$_{2.5}$ concentration in Thailand, including in the BMR, has been operated by the Pollution Control Department (PCD) with sparsely distributed numbers of monitoring stations across the BMR, the concentrations of PM$_{2.5}$ in unmonitored locations have not been captured. Therefore, there is necessary to develop the
possibly alternative techniques to estimate the concentration of PM$_{2.5}$ in the areas with no monitoring station in the BMR that can be further applied for informing PM$_{2.5}$ situation in the region.

Satellite remote sensing with different algorithms is one of the powerful tools used to estimate ground-level PM$_{2.5}$. Specifically, the Deep Blue (DB) algorithm has been developed to calculate aerosol optical depth (AOD), which is a parameter indicating the extinction of light due to the suspension of aerosols in the atmosphere. The AOD data is generally captured by the instruments aboard the satellites, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua satellites. This data has been extensively applied along with different predictor variables (i.e., meteorological factors, greenness indicators, and planetary boundary layer) to estimate PM$_{2.5}$ concentration at different spatial and temporal scales (He et al. 2020; Kong et al. 2016; Lv et al. 2017; Zheng et al. 2016). The linear relationship between AOD and PM$_{2.5}$ has been considered in earlier previous studies taken into account many spatiotemporal predictors that might be affecting the number of atmospheric aerosols (Liu et al. 2005; Ma et al. 2014). In recent years, more advanced statistical models have been utilized to describe the complex relationship between AOD and ground-level PM$_{2.5}$ concentration such as linear mixed-effects model (Wang et al. 2019; Xie et al. 2015; Zheng et al. 2016), the Bayesian-based downscaling framework (Lv et al. 2017), and linear regression with geographically and temporally weighted regression (GTWR) model (Chen et al. 2020; He et al. 2020).

The different MODIS AOD data products have been used to estimate ground-level PM$_{2.5}$ concentration in many previous studies. In particular, a spatial resolution of 10 km MODIS AOD has long been used to derive ambient concentration of PM$_{2.5}$ in the regional scales (Kong et al. 2016; Ma et al. 2014), but a 3-km MODIS AOD have been applied recently to gain insight into finer details of spatial information in urban levels contributing to population-based PM$_{2.5}$ exposure and its health effect studies (Lv et al. 2017; Xie et al. 2015). Although satellite-derived AOD products have widely used to estimate surface PM$_{2.5}$ concentration in many parts of the world, especially in mainland China, with various statistical models (Hu et al. 2014; Liu et al. 2005; Ma et al. 2014; Wang et al. 2019; Xie et al. 2015), very few studies have been conducted in Thailand where there is a limited number of air quality monitoring stations. Therefore, ambient concentration of PM$_{2.5}$ in an area with no monitoring station could not be captured; meanwhile, using PM$_{2.5}$ concentration from a fixed site monitoring station representing the same exposure level for all residents throughout the region may bias the results of epidemiological studies. Hence, this study aimed to estimate daily PM$_{2.5}$ concentration in the BMR, the most polluted region of Thailand, using 3-km resolution MODIS AOD dataset from January to December 2020 for exposure estimation of epidemiological studies. The multiple-linear regression with various temporally meteorological variables and vegetation index was established to examine the relationship between satellite-derived AOD and PM$_{2.5}$ concentration. The accuracy of estimated PM$_{2.5}$ concentration was then explored using cross-validation technique.

**Methods**

The Bangkok Metropolitan Region (BMR) is consisting of Bangkok Metropolitan Area and its five surrounding provinces (i.e., Nakhon Pathom, Nonthaburi, Pathum Thani, Samut Prakan, and Samut Sakhon). This region is located in the central part of Thailand with 10.9 million populations in 2020, representing 16.5% of Thailand’s total populations (Thailand National Statistical Office 2020). The BMR has been considered the most economic development area in Thailand that often experiences heavy concentration of PM$_{2.5}$ generated mainly by vehicle emission, biomass burning, and industrial activities, respectively (Narita et al. 2019). The BMR covers an area of approximately 8000 km$^2$, and 23 air quality monitoring stations have been operated by the Pollution Control Department throughout the BMR as shown in Fig. 1.

**MODIS AOD data product**

The MODIS AOD data was obtained from the Atmosphere Archive and Distribution System (LAADS) Distributed Active Archive Center (DAAC) (https://ladsweb.modaps. eosdis.nasa.gov/search/order/1), which is one of the DAACs under the National Aeronautics and Space Administration (NASA)’s Earth Observing System (EOS). In particular, the AOD data in this study was acquired from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard the Terra and Aqua satellites Collection 6.1, Level 2 Deep Blue aerosol product (Optical_Depth_Land_And_Ocean at the 550 nm wavelength) with a resolution of 3 km×3 km (MODO4_3K and MYD04_3K) during January through December 2020 and then resampled the data into fixed grid that covers spatial resolution of 1 km×1 km using bilinear interpolation technique. The daily average AOD value was calculated by taking the average of daily AOD data from the Terra and Aqua satellites, overlapping the study area respectively at approximately 10:30 and 13:30 local time, because a previous study revealed that the MODIS AOD data from Terra and Aqua satellites may provide good estimates of daily AOD values (Zhang et al. 2012). The AOD values were then averaged throughout the BMR by date during the study period, and the missing AOD
data was imputed using the expectation–maximization (EM) algorithm under the assumptions of a multivariate normal distribution (Junger and Ponce de Leon 2015).

**Ground-level PM$_{2.5}$ and meteorological data**

Hourly concentration of air pollution (i.e., PM$_{2.5}$ (µg/m$^3$), NO$_2$ (ppb), and O$_3$ (ppb)) and meteorological variables (i.e., ambient temperature (°C), relative humidity (%), and wind speed (m/s)) were obtained from the air quality monitoring stations operated by the Pollution Control Department of Thailand that situated across the BMR as presented in Fig. 1 during January through December 2020. Because MODIS AOD data are monitored at a single point in time when the satellites are overpassing the study area (i.e., 10:30 and 13:30 local times for Terra and Aqua satellites, respectively), the 7-h average concentrations of every single air pollutant and meteorological variable from 9:00 through 15:00 local times each day were computed that covers the time for Terra and Aqua satellites overpassing the study area. In the case that there were missing data more than 2 h out of 7 h, data of that particular day was assigned as missing. To obtain daily average concentrations of individual air pollutants and meteorological variables, the 7-h average concentrations obtained from all monitoring stations located in the BMR were then averaged. The missing air pollutant and meteorological data was imputed using the EM algorithm (Junger and Ponce de Leon 2015).

**Greenness indicator data**

The 16-day interval greenness data during January through December 2020 was acquired from the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard the NASA’s Terra satellite with a spatial resolution of 0.1°×0.1° (https://neo.gsfc.nasa.gov/view.php?datasetId=MOD_NDVI_M). The satellite-based greenness data have been measured as the Normalized Difference Vegetation Index (NDVI), which is calculated from the ratio of the difference between reflected near-infrared light and infrared visible light to the addition of near-infrared and infrared visible light reflectance. The value of NDVI has no unit, ranging from −1.0 to 1.0, in which higher value of NDVI indicates the larger green vegetation space (Denpetkul and Phosri 2021). The daily average NDVI value was then computed by taking the average of all 0.1°×0.1° grids for every single day over the BMR.

**Statistical analysis**

The multiple-linear regression model was constructed in this study to examine the relationship between daily variation of AOD and daily concentrations of ground-level PM$_{2.5}$ taken into account many potential confounders that may affect
the relationship of daily varying AOD value with ground-level PM$_{2.5}$. In particular, the linear regression model was tested whether a number of predictor variables increase the model performance using the forward stepwise selection method, where the multiple-linear regression model in this study was ultimately accounted for surface temperature, relative humidity, wind speed, NO$_2$, O$_3$, and NDVI. Daily surface temperature, relative humidity, and wind speed were included in the model because previous study indicated that increased levels of temperature, relative humidity, and wind speed were associated with reduced concentration of PM$_{2.5}$ (Sritong-anon et al. 2021). Furthermore, NDVI was incorporated into the model because PM$_{2.5}$ concentration might be explained by the variation of NDVI. Specifically, higher vegetation coverage may associate with lower concentration of PM$_{2.5}$ (Zhou et al. 2021). Daily average concentrations of NO$_2$ and O$_3$ were also incorporated into the model because NO$_2$ and O$_3$ may be associated with the variations of PM$_{2.5}$ concentration. In particular, high O$_3$ concentration as strong oxidative pollutant during summer months is able to introduce the formation of secondary particles, thereby increasing PM$_{2.5}$ concentration. On the other hands, high concentration of PM$_{2.5}$ during winter suppresses solar radiation, resulting in lower O$_3$ production and concentration (Jia et al. 2017). Moreover, chemical reaction of NO$_2$ in the air promotes the formation of secondary PM$_{2.5}$ (Balamarugan et al. 2022; Wu et al. 2016). Because of COVID-19 pandemic in 2020, Thai’s government has imposed the lockdown measure to prevent the transmission of SARS-CoV-2. This policy might influence the levels of AOD and ground-level PM$_{2.5}$ concentration. Thus, indicator variable of COVID-19 lockdown period was also incorporated into the model. The time series linear regression model was applied in this study because we hypothesized that the association between AOD and ground-level PM$_{2.5}$ concentration varies intensively daily, but minimally spatially over the specific region (Lee et al. 2011). The final model used to examine the relationship between AOD and ground-level PM$_{2.5}$ is represented by Eq. (1).

$$
\text{Log}(\text{PM}_{2.5}) = \beta_1 \text{AOD}_t + \beta_2 \text{Temp}_t + \beta_3 \text{Humid}_t + \beta_4 \text{Wind}_t + \beta_5 \text{NO}_2 + \beta_6 \text{O}_3 + \beta_7 \text{NDVI}_t + \beta_8 \text{COVID}_t
$$

(1)

wherein Log(PM$_{2.5}$) indicates the logarithmic scale of ground-level PM$_{2.5}$ concentration at day $t$ and AOD$_t$, Temp$_t$, Humid$_t$, Wind$_t$, NO$_2$, O$_3$, and NDVI$_t$ are the level of AOD, surface temperature, relative humidity, wind speed, nitrogen dioxide, ozone, and NDVI at day $t$, respectively. COVID$_t$ is the dummy variable for COVID-19 lockdown period at day $t$, whereas $\beta_{1,2,3,\ldots,8}$ indicate the vector of regression coefficients of AOD, surface temperature, humidity, wind speed, nitrogen dioxide, ozone, NDVI, and COVID-19 lockdown, respectively. The sensitivity analysis was also performed by excluding NO$_2$ and O$_3$ from regression model to explore the robustness of the model performance.

The 12-fold cross-validation method was then used to assess the model performance. Specifically, we stratified the observed PM$_{2.5}$ concentration into 12 groups ($N=12$) by calendar month and selected a group of month as the test set and the remaining $N-1$ months were utilized as the training set. The prediction of a test set was generated from the model output of training set using Eq. (1). This procedure was repeated for 12 times until the prediction was completed in all 12 months. The simple linear regression model was then fitted to explore the correlation between observed and predicted PM$_{2.5}$ concentrations where $R^2$ value, as well as the root mean square error (RMSE) and mean absolute error (MAE), was utilized to evaluate the model performance. This model performance was also separately explored in different seasons, including winter (i.e., November to February), summer (March to June), and rainy reason (i.e., July to October). All statistical analyses were performed by R (version 4.1.0).

The annual and season-specific average concentrations of PM$_{2.5}$ were calculated by averaging daily predicted PM$_{2.5}$ concentration for the whole year and for season-specific, respectively. Because of missing AOD data due to cloud cover and precipitation (Levy et al. 2007), the bias-corrected annual or season-specific PM$_{2.5}$ concentration was applied by calculating the ratio of annual or season-specific average of predicted PM$_{2.5}$ concentration for all grid cells to annual or season-specific average of daily predicted PM$_{2.5}$ concentration obtained from regression model as shown in Eq. (2). The output was then used to compute the bias-corrected annual or season-specific average PM$_{2.5}$ concentration by multiplying with predicted annual or season-specific average concentration of PM$_{2.5}$ for each grid cell (Zheng et al. 2016). All maps were produced by ArcMap (version 10.8).

$$
\text{Corr}_{\text{bias}} = \frac{\text{PM}_{2.5}\text{grid}}{\text{PM}_{2.5}\text{predicted}}
$$

(2)

wherein Corr$_{\text{bias}}$ is the correction factor; PM$_{2.5}\text{grid}$ defines annual/season-specific average of predicted PM$_{2.5}$ for all grid cells; and PM$_{2.5}\text{predicted}$ is annual/season-specific average of daily predicted PM$_{2.5}$ level.

### Results

The descriptive statistics of ground-level PM$_{2.5}$, AOD, and other predictor variables in the BMR during the study period (i.e., January through December 2020) are shown in Table 1. The average of ground-level PM$_{2.5}$ concentration (minimum–maximum) throughout the study period was 22.37 $\mu$g/m$^3$ (7.23–86.67 $\mu$g/m$^3$), which peaked during
winter (November to February). The average of AOD value was 0.51 (0.03–2.28), where the highest level was observed between March and April (Fig. 2). The average of surface temperature, relative humidity, wind speed, NO2, and O3 was respectively 29.05 °C (23.59–32.87), 70.16% (41.41–93.35), 0.62 m/s (0.30–1.53), 13.50 ppb (3.71–50.07), and 36.65 ppb (12.01–78.31), whereas that of NDVI was 0.51 (0.43–0.60) during the study period. Figure 2 indicates time series variations of PM2.5 concentration and AOD during the study period, where PM2.5 concentration was peaked during winter (November to February) and the highest level of AOD was observed between March and April. The spatial distribution of AOD over the BMR is shown in Supplementary Fig. S1.

Table 2 illustrates regression coefficient and its standard error (SE) of the relationship between MODIS AOD and concentration of PM2.5 controlling for many possible confounders as described in Eq. (1). In particular, a 1 unit increase of AOD was positively associated with 41.46% [(1.4146 − 1)*100]±5.75% increase of ground-level PM2.5 concentration and 1 ppb increase of NO2 and O3 was associated with 2.12 ±0.25% and 1.81 ±0.14% increase of PM2.5 concentration, respectively, whereas increasing each 1 °C of temperature, 1% of humidity, 1 m/s of wind speed, and 1 unit of NDVI was associated with 2.12 ±0.25% and 1.81 ±0.14% increase of PM2.5 concentration, respectively. Moreover, PM2.5 concentration during the time of COVID-19 lockdown was lower by 24.59 ±5.87% compared to that during no imposed lockdown period. The multicollinearity is not likely to be a problem of the regression model in this study since the correlation among predictor variables was not very high (Supplementary Table S1). The multiple R2 of regression model is 0.71, indicating that all independent variables (i.e., MODIS AOD, surface temperature, relative humidity, wind speed, NO2, O3, NDVI, and COVID-19 lockdown period) could be able to explain 71% of the variability in ground-level PM2.5 concentration in BMR significantly.

| Variables | Mean | SD  | Min  | 25th | 50th | 75th | Max   |
|-----------|------|-----|------|------|------|------|-------|
| PM2.5 (µg/m³) | 22.37 | 12.55 | 7.23 | 13.79 | 18.16 | 27.07 | 86.67 |
| AOD       | 0.51 | 0.31 | 0.03 | 0.32 | 0.43 | 0.59 | 0.64  |
| Temperature (°C) | 29.05 | 1.61 | 23.59 | 28.19 | 29.06 | 30.19 | 32.87 |
| Relative humidity (%) | 70.16 | 8.87 | 41.41 | 66.00 | 69.65 | 75.64 | 93.35 |
| Wind speed (m/s) | 0.62 | 0.18 | 0.30 | 0.48 | 0.59 | 0.73 | 1.53  |
| NO2 (ppb) | 13.50 | 7.00 | 3.71 | 7.98 | 11.97 | 17.76 | 50.07 |
| O3 (ppb)  | 36.65 | 13.40 | 12.01 | 25.53 | 34.02 | 46.73 | 78.31 |
| NDVI      | 0.51 | 0.05 | 0.43 | 0.47 | 0.51 | 0.55 | 0.60  |

SD is standard deviation; Min and Max are minimum and maximum values, respectively; 25th, 50th, and 75th define 25th, 50th, and 75th percentiles, respectively.

Fig. 2 The concentration of PM2.5 and AOD over time during the study period, spanning from January through December 2020.
There is high correlation between observed PM$_{2.5}$ and predicted PM$_{2.5}$ concentrations obtained from cross-validation for the whole year and for season-specific with Pearson’s correlation coefficient of 0.82 and $p$-value $< 0.05$ (Fig. 3A). The daily variation of observed PM$_{2.5}$ was highly correlated with predicted PM$_{2.5}$ concentration (Fig. 3B), indicating that the prediction accuracy was improved when NO$_2$, O$_3$, temperature, relative humidity, NDVI, and dummy variables of COVID-19 lockdown period were integrated into the model.

The results of 12-fold cross-validation are depicted in Fig. 4. Specifically, the association between observed PM$_{2.5}$ concentration and predicted PM$_{2.5}$ concentration obtained from cross-validation for the whole year and for season-specific was indicated in terms of correlation plot, $R^2$, MAE, and RMSE. Based on correlation plot, the closer the points are, the higher model accuracy is. Moreover, higher $R^2$ and lower MAE and RMSE indicate higher model accuracy, and vice versa. The $R^2$ of cross-validation for the whole year period, as well as for summer, winter, and rainy season was 0.48, 0.55, 0.21, and 0.52, respectively, which are lower than the $R^2$ from the regression model (i.e., 0.71). The RMSE and MAE of cross-validation in winter (13.72 µg/m$^3$ and 6.64 µg/m$^3$) were greater compared to those in summer (6.97 µg/m$^3$ and 4.24 µg/m$^3$) and rainy season (4.01 µg/m$^3$ and 2.94 µg/m$^3$), respectively. Therefore, the prediction accuracy during summer and rainy season was higher than that in winter and over the whole year period. Moreover, the performance of regression model yielded reasonable prediction for the whole year and for season-specific as indicated by the slopes, intercepts, and $R^2$ values. In particular, the model underestimates PM$_{2.5}$ concentration by 0.5% and 12.4% for the whole year period and in winter, respectively, while overestimates PM$_{2.5}$ concentration in summer and rainy season by 18.1% and 2.9%, respectively (Table 3). Figure 5 indicates the spatial distribution of

![Table 2: Regression coefficient and its standard error (SE) of the relationship between MODIS AOD and concentration of PM$_{2.5}$ controlling for many possible confounders](image)

| Variables          | $\beta$ coefficients | Exp($\beta$) | Standard error (SE) | Exp(SE) |
|--------------------|-----------------------|--------------|---------------------|---------|
| Intercept          | 4.3454                | 77.1193      | 0.4462              | 1.5624  |
| AOD                | 0.3469                | 1.4146       | 0.0559              | 1.0575  |
| Temperature        | -0.0183               | 0.9818       | 0.0107              | 1.0108  |
| Relative humidity  | -0.0020               | 0.9980       | 0.0021              | 1.0021  |
| Wind speed         | -0.0364               | 0.9642       | 0.0895              | 1.0936  |
| NO$_2$             | 0.0209                | 1.0212       | 0.0025              | 1.0025  |
| O$_3$              | 0.0179                | 1.0181       | 0.0014              | 1.0014  |
| NDVI               | -3.4171               | 0.0328       | 0.3759              | 1.4563  |
| COVID-19 lockdown  | -0.2822               | 0.7541       | 0.0570              | 1.0587  |

Regression coefficient is interpreted per 1 unit change of AOD and NDVI, 1 °C change of temperature, 1% change of humidity, 1 m/s change of wind speed, and 1 ppb change of NO$_2$ and O$_3$, whereas that of COVID-19 lockdown is interpreted by comparing between with and without lockdown period.

![Fig 3: The correlation between observed and predicted PM$_{2.5}$ concentrations (A) and time series plot of observed and predicted PM$_{2.5}$ concentrations (B) obtained from the regression model](image)
the bias-corrected annual and season-specific average PM$_{2.5}$ concentration over the BMR, where dark blue and red indicate lower and higher concentrations of PM$_{2.5}$, respectively. The annual average concentration of PM$_{2.5}$ was higher in the downtown area of Bangkok compared to the rest of the study area (Fig. 5A). The average concentration of PM$_{2.5}$ during winter (Fig. 5B) was also higher in the city center of Bangkok and that was higher than that during summer (Fig. 5C) and rainy season (Fig. 5D). The additional information of bias-corrected annual and

![Graphs of PM$_{2.5}$ concentrations over different seasons](image)

**Fig. 4** The simple linear regression between observed and predicted PM$_{2.5}$ concentrations obtained from 12-fold cross-validation for a year round, winter, summer, and rainy season. The solid lines indicate the regression line, whereas dashed lines represent the 0:1 (intercept:slope) line

| Table 3 | The 12-fold cross-validation results on ground-level PM$_{2.5}$ prediction |
|---------|-----------------------------|----------------|----------|-----------|-----------|
| Variables | Intercept | Slope  | R$^2$  | RMSE | PM$_{2.5}$ (mean ± SD) (µg/m$^3$) |
|          |           |        |        |     | Predicted | Observed |
| Full model | 9.98  | 0.55 | 0.48 | 9.16 | 22.25 ± 9.97 | 22.37 ± 12.55 |
| Whole year | 4.27  | 0.95 | 0.55 | 6.97 | 21.68 ± 9.14 | 18.36 ± 7.14 |
| Summer | 19.29 | 0.30 | 0.21 | 13.72 | 29.43 ± 9.45 | 33.60 ± 14.48 |
| Winter | 2.64  | 0.86 | 0.52 | 4.01 | 15.74 ± 5.68 | 15.30 ± 4.78 |
| Rainy | no-NO$_2$ and O$_3$ model | 14.98 | 0.29 | 0.19 | 11.63 | 21.39 ± 8.22 | 22.37 ± 12.55 |
| Whole year | 12.72 | 0.51 | 0.19 | 8.98 | 22.04 ± 8.27 | 18.36 ± 7.14 |
| Summer | 24.59 | 0.05 | 0.01 | 17.49 | 26.40 ± 8.31 | 33.60 ± 14.48 |
| Winter | 11.99 | 0.25 | 0.14 | 4.63 | 15.83 ± 3.16 | 15.30 ± 4.78 |

Full model is including the following variables into the model as showed in Eq. (1): AOD, surface temperature, relative humidity, wind speed, NO$_2$, O$_3$, NDVI, and indicator of COVID-19 lockdown period.
season-specific average of PM$_{2.5}$ concentration over the BMR is shown in Supplementary Table S2.

Figure 6 illustrates the results of 12-fold cross-validation by means of excluding NO$_2$ and O$_3$ from the model. Finding indicates that the accuracy of model performance was lower than that of the model with NO$_2$ and O$_3$ (as indicated in Fig. 4), in which $R^2$ of the model without NO$_2$ and O$_3$ was lower and RMSE was greater for both whole year and season-specific periods. Specifically, incorporating NO$_2$ and O$_3$ into the model considerably improved the model performance, where the slope increased by 0.26 and intercept declined by 5.00 for the whole year period. This finding was also observed for season-specific analysis, where including NO$_2$ and O$_3$ into the model substantially increased the slope by 0.44, 0.25, and 0.61, and the corresponding intercept decreased by 8.45, 5.30, and 9.35 in summer, winter, and rainy season, respectively (Table 3).

**Discussion**

Many previous studies have revealed that regression model with the MODIS AOD measurement can be used as predictor variable to estimate the spatial ground-level PM$_{2.5}$ concentration, taken into account many potential confounders (Guo et al. 2021; Hu et al. 2014; Xu and Zhang 2020; Zheng et al. 2016). This method would benefit for exposure assessment for epidemiological research, especially in the areas with no monitoring station network that is commonly used to explore the association of PM$_{2.5}$ with morbidity and mortality. In this study, we found that linear regression model would

![Fig. 5 The spatial distribution of annual average bias-corrected PM$_{2.5}$ concentration (A), and the mean during winter (B), summer (C), and rainy season (D) over the BMR. The black dots define air quality monitoring stations situated in the BMR.](image-url)
dramatically improve the prediction accuracy when air pollutants, meteorological variables, greenness indicator, and dummy variable for COVID-19 lockdown period were integrated into the model with the overall $R^2$ of 0.71, but cross-validation $R^2$ was 0.48, suggesting that the overfitting was found in the model. This finding is similar to many previous studies (Guo et al. 2021; Li et al. 2015; Wang et al. 2019; Zheng et al. 2016), implying that the overfitting is commonly existed in the model. The model performance on investigating the association between MODIS AOD and ground-level PM$_{2.5}$ concentration varied by season (Xin et al. 2014). Specifically, the prediction accuracy observed during winter was lower than that during summer, rainy season, and a year round, where the $R^2$ from 12-fold cross-validation was 0.21, 0.55, 0.52, and 0.48 for winter, summer, rainy season, and a year round, respectively. The RMSE from 12-fold cross-validation during winter (13.72 µg/m$^3$) was also higher than that in summer (6.97 µg/m$^3$), rainy season (4.01 µg/m$^3$), and the whole year (9.16 µg/m$^3$). This finding might be contributed from the substantially higher concentration of PM$_{2.5}$ observed during winter, where multiple-linear regression model may not be able to predict the high concentration of PM$_{2.5}$ accurately (Zheng et al. 2016). In addition, the concentration of PM$_{2.5}$ was the lowest in rainy season with some missing grid cells as we found in Fig. 5D. This finding was caused by missing MODIS AOD data due to substantial cloud cover observed during rainy season. The slope of simple linear regression between observed PM$_{2.5}$ and predicted PM$_{2.5}$ concentrations obtained from 12-fold cross-validation for the whole year, winter, summer, and rainy season was respectively 0.55, 0.30, 0.95, and 0.86, indicating that the predicted PM$_{2.5}$ concentration from linear regression model during summer was highly accordant with the observed concentration since the slope is close to 1. This result suggests that the model performance during summer is more accurate, to some extent, compared to that during other seasons, although regression model was overfitted the data.

We also found that incorporating various meteorological variables and greenness indicator into the model would improve the model performance on the prediction of ground-level PM$_{2.5}$ concentration, which is in agreement
with previous studies (Kloog et al. 2011; Liu et al. 2007; Ma et al. 2014; Wang et al. 2019; Zheng et al. 2016). Furthermore, we compared the prediction performance of regression model between a model with and without NO₂ and O₃, where incorporating NO₂ and O₃ into the model substantially improved the prediction accuracy, in which the slopes of cross-validation were increased and the intercepts were dramatically decreased. This finding is similar to a previous study (Zheng et al. 2016). Therefore, the local air pollutants might also influence the model performance on investigating the association between MODIS AOD and ground-level PM₂.₅ concentration.

The principal limitation of this study is a lacking of MODIS AOD data in a certain day during the study period that is commonly caused by high cloud cover or high surface reflectance (Hu et al. 2014). However, the main objective of this study was to develop the statistical model to predict ground-level PM₂.₅ concentration using temporal AOD data, controlling for other potential confounders, because we assumed a priori that the association between AOD and ground-level PM₂.₅ varies intensively daily, but minimally spatially over the BMR. Therefore, the availability of daily AOD value could be applied to estimate ground-level PM₂.₅ concentration, to some extent. Moreover, we replaced the missing AOD data using EM algorithm, where the additional measurement error might be introduced, but a previous study revealed that this method is appropriate for time series data that is fitted to the data of this study (Junger and Ponce de Leon 2015). We also interpolated AOD data from 3- to 1-km resolution using the bilinear interpolation technique, which might also introduce the measurement error. Therefore, further studies are needed to estimate the finer spatial scales of PM₂.₅ concentration using the finer resolution of AOD.

AOD products have been widely used to estimate ground-level PM₂.₅ concentration in many previous studies using linear mixed effect model and geographically weighted regression (GWR) models (Lee et al., 2011; Ma et al., 2014; Zheng et al., 2016). However, integration of the MODIS AOD, air pollutants, meteorological variables, and greenness indicators in the current study provides strength to accurately estimate ground-level concentration of PM₂.₅ using multiple-linear regression model. This finding indicates that although the complicated models or algorithms are not utilized, linear regression model with appropriate predictor variables can be used to accurately estimate surface concentration of PM₂.₅. Moreover, this study is the first study that comprehensively estimates ground-level PM₂.₅ concentration in Bangkok by integrating the MODIS AOD, air pollutants, meteorological variables, and greenness indicator using linear regression, and the prediction accuracy was improved, to some extent. Therefore, model coefficients of various predictor variables generated from this study can be utilized to estimate the concentration of PM₂.₅ in an area with no air quality monitoring station in Bangkok, Thailand.

Conclusion

The multiple-linear regression model was used to estimate ground-level PM₂.₅ concentration using the MODIS AOD as predictor, controlling for various meteorological variables, greenness indicator, local air pollutants, and dummy variable of COVID-19 lockdown period. The annual average concentration of predicted PM₂.₅ (22.25 ± 9.97 µg/m³) was similar to that of observed values (22.37 ± 12.55 µg/m³), indicating the accurate results. The accuracy of model performance was varied by seasons, where the most accurate finding was observed during summer with R² and RMSE of the cross-validation of 0.55 and 6.97 µg/m³, respectively. The model performed better when the daily average ambient NO₂ and O₃ concentration was incorporated into the model. Finding from this study suggests that linear regression model with appropriate predictor variables could be used to estimate ground-level PM₂.₅ concentration that might be applied further for epidemiological studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11869-022-01238-4.

Acknowledgements The authors are grateful to the Pollution Control Department of Thailand for providing air pollution and meteorological variables for the analyses in this study. The authors are also grateful to the NASA’s Earth Observation for the provision of AOD (https://ladsweb.modaps.eosdis.nasa.gov/search/order/1) and NDVI (https://neo.gsfc.nasa.gov/view.php?datasetId=MOD_NDVI_M) data applied in this study. We also gratefully acknowledged the effort of Mr. Thomas McManamon for professional English language editing.

Author contribution Bussayaporn Peng-in: data curation, methodology, writing—original draft. Peeyaporn Sanitlua: data curation, methodology, writing—original draft. Pimnapat Monjatturat: data curation, methodology, writing—original draft. Pattaranap Boonkerd: data curation, methodology, writing—original draft. Arthit Phosri: conceptualization, supervision, writing—review and editing.

Data availability The data are available in electronic supplementary materials and from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.
References

Alexeeff SE, Liao NS, Liu X, Van Den Eeden SK, Sidney S (2021) Long-term PM2.5 exposure and risks of ischemic heart disease and stroke events: review and meta-analysis. J Am Heart Assoc 10:1–22. https://doi.org/10.1161/JAHA.120.016890

Atkinson RW, Kang S, Anderson HR, Mills IC, Walton HA (2014) Epidemiological time series studies of PM2.5 and daily mortality and hospital admissions: a systematic review and meta-analysis. Thorax 69:660–665. https://doi.org/10.1136/thoraxjnl-2013-204492

Balamurugan V, Chen J, Qu Z, Wang J, Li, Geng D, Chen J et al (2022) Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing. Environ Sci Technol 56:9015–9026. https://doi.org/10.1021/acs.est.2c01670

Farhadi Z, Abulghasem Gorgi H, Shabaninejad H, Aghajani Delavar Denpetkul T, Phosri A (2021) Daily ambient temperature and mortality in Thailand: estimated effects, attributable risks, and effect modifications by greenness. Sci Total Environ 791:148373. https://doi.org/10.1016/j.scitotenv.2021.141093

Denpetkul T, Phosri A (2021) Daily ambient temperature and mortality in Thailand: estimated effects, attributable risks, and effect modifications by greenness. Sci Total Environ 791:148373. https://doi.org/10.1016/j.scitotenv.2021.141093

Fan J, Li S, Fan C, Bai Z, Yang K (2016) The impact of PM2.5 on asthma emergency department visits: a systematic review and meta-analysis. Environ Sci Pollut Res 23:843–850. https://doi.org/10.1007/s11356-015-5321-x

Farhadi Z, Abulghasem Gorgi H, Shabaninejad H, Aghajani Delavar Denpetkul T, Phosri A (2021) Daily ambient temperature and mortality in Thailand: estimated effects, attributable risks, and effect modifications by greenness. Sci Total Environ 791:148373. https://doi.org/10.1016/j.scitotenv.2021.141093

He Q, Gu Y, Zhang M (2020) Spatiotemporal trends of PM25 concentrations in central China from 2003 to 2018 based on MAIAC-derived high-resolution data. Environ Int 137:105536. https://doi.org/10.1016/j.envint.2020.105536

Hu X, Waller LA, Lu X, Yasar A, Wang Y, Al-Hamdan MZ, Crosson WL et al (2014) Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. Remote Sens Environ 140:220–232. https://doi.org/10.1016/j.rse.2013.08.032

Jia M, Zhao T, Cheng X, Gong S, Zhang X, Tang L et al (2017) Inverse Relations of PM25 and O3 in air compound pollutions by greenness. Sci Total Environ 791:148373. https://doi.org/10.1016/j.scitotenv.2021.141093

Junger WL, Ponce de Leon A (2015) Imputation of missing data in time series for air pollutants. Atmos Environ 102:96–104. https://doi.org/10.1016/j.atmosenv.2014.11.049

Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J (2015) Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmos Environ 45:6267–6275. https://doi.org/10.1016/j.atmosenv.2011.08.066

Kong L, Xin J, Zhang W, Wang Y (2016) The empirical correlations between PM2.5, PM10 and AOD in the Beijing metropolitan region and the PM2.5, PM10 distributions retrieved by MODIS. Environ Pollut 216:350–360. https://doi.org/10.1016/j.envpol.2016.05.085

Lee HJ, Liu Y, Coull BA, Schwartz J, Koutrakis P (2011) A novel calibration approach of MODIS AOD data to predict PM2.5 concentrations. Atmos Chem Phys 11:7991–8002. https://doi.org/10.5194/acp-11-7991-2011

Levy RC, Remer LA, Matoo S, Vermote EF, Kaufman YJ (2007) Second-generation operational algorithm: retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. J Geophys Res Atmos 112:1–21. https://doi.org/10.1029/2006JD007811

Li R, Gong J, Chen L, Wang Z (2015) Estimating ground-level PM2.5 using fine-resolution satellite data in the megacity of Beijing, China Aerosol Air Qual Res 15:1347–1356. https://doi.org/10.4209/aaqr.2015.01.0009

Liu Y, Sarnat JA, Kilaru V, Jacob DI, Koutrakis P (2005) Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing. Environ Sci Technol 39:3269–3278. https://doi.org/10.1021/es049352m

Liu Y, Franklin M, Kahn R, Koutrakis P (2007) Using aerosol optical thickness to predict ground-level PM2.5 concentrations in the St. Louis area: a comparison between MISR and MODIS. Remote Sens Environ 107:33–44. https://doi.org/10.1016/j.rse.2006.05.022

Lv B, Hu Y, Chang HH, Russell AG, Cai J, Xu B et al (2017) Daily estimation of ground-level PM2.5 concentrations at 4 km resolution over Beijing-Tianjin-Hebei by fusing MODIS AOD and ground observations. Sci Total Environ 580:235–244. https://doi.org/10.1016/j.scitotenv.2016.12.049

Ma Z, Xu X, Huang L, Bi J, Liu Y (2014) Estimating ground-level PM2.5 in China using satellite remote sensing. Environ Sci Technol 48:7436–7444. https://doi.org/10.1021/es509399

Narita D, Kim Oanh NT, Sato K, Huo M, Permadi DA, Chi NNH et al (2019) Pollution characteristics and policy actions on fine particulate matter in a growing Asian economy: the case of Bangkok Metropolitan Region. Atmosphere 10:1–18. https://doi.org/10.3390/atmos10050227

Sangkharat K, Fisher P, Thomas GN, Thornes J, Pope FD (2019) The impact of air pollutants on ambulance dispatches: a systematic review and meta-analysis of acute effects. Environ Pollut 254:112769. https://doi.org/10.1016/j.envpol.2019.06.065

Sritong-aon C, Thomya J, Kertpromphhan C, Phosri A (2021) Estimated effects of meteorological factors and fire hotspots on ambient particulate matter in the northern region of Thailand. Air Qual Atmos Health 14(1):1857–1868. https://doi.org/10.1007/s11869-021-01059-x

Thailand National Statistical Office (2020) Population size and structure. Retrieved December 5, 2021, from http://statbbs.nso.go.th/statireport/page/sector/en/01.aspx

Thongphunchong K, Phosri A, Sihabut T, Patthanaissaranukroh W (2021) Short-term effects of particulate matter on outpatient department visits for respiratory diseases among children in Bangkok Metropolitan Region: a case - crossover study. Air Qual Atmos Health 14:1785–1795. https://doi.org/10.1007/s11869-021-01053-3

Wang B, Eum KD, Kazemiparkouhi F, Li C, Manjourides J, Pavlu V et al (2020) The impact of long-term PM2.5 exposure on specific causes of death: exposure-response curves and effect modification among 53 million U.S. Medicare Beneficiaries Environ Health 19:1–12. https://doi.org/10.1186/s12940-020-00575-0

Wang Q, Zeng Q, Tao J, Sun L, Zhang L, Gu T et al (2019) Estimating PM2.5 concentrations based on MODIS AOD and NAQPMs data over Beijing–Tianjin–Hebei. Sensors 19. https://doi.org/10.3390/s19051207
Wu J, Xu C, Wang Q, Cheng W (2016) Potential sources and formations of the PM2.5 pollution in urban Hangzhou. Atmosphere 7(8):1–15. https://doi.org/10.3390/atmos7080100

Xie Y, Wang Y, Zhang K, Dong W, Lv B, Bai Y (2015) Daily estimation of ground-level PM2.5 concentrations over Beijing using 3 km resolution MODIS AOD. Environ Sci Technol 49:12280–12288. https://doi.org/10.1021/acs.est.5b01413

Xin J, Zhang Q, Wang L, Gong C, Wang Y, Liu Z et al (2014) The empirical relationship between the PM2.5 concentration and aerosol optical depth over the background of North China from 2009 to 2011. Atmos Res 138:179–188. https://doi.org/10.1016/j.atmosres.2013.11.001

Xu X, Zhang C (2020) Estimation of ground-level PM2.5 concentration using MODIS AOD and corrected regression model over Beijing. China Plos One 15:1–15. https://doi.org/10.1371/journal.pone.0240430

Zhang Y, Yu H, Eck TF, Smirnov A, Chin M, Remer LA et al (2012) Aerosol daytime variations over North and South America derived from multiyear AERONET measurements. J Geophys Res Atmos 117:1–13. https://doi.org/10.1029/2011JD017242

Zheng Y, Zhang Q, Liu Y, Geng G, He K (2016) Estimating ground-level PM2.5 concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. Atmos Environ 124:232–242. https://doi.org/10.1016/j.atmosenv.2015.06.046

Zhou M, Huang Y, Li G (2021) Changes in the concentration of air pollutants before and after the COVID-19 blockade period and their correlation with vegetation coverage. Environ Sci Pollut Res 28(18):23405–23419. https://doi.org/10.1007/s11356-020-12164-2

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.