

## Research Article

# Prediction of PM<sub>2.5</sub> in an Urban Area of Northern Thailand Using Multivariate Linear Regression Model

Teerachai Amnuaylojaroen <sup>1,2</sup>

<sup>1</sup>Department of Environmental Science, School of Energy and Environment, University of Phayao, Phayao 56000, Thailand

<sup>2</sup>Atmospheric Pollution and Climate Change Research Unit, School of Energy and Environment, University of Phayao, Phayao 56000, Thailand

Correspondence should be addressed to Teerachai Amnuaylojaroen; [teerachai4@gmail.com](mailto:teerachai4@gmail.com)

Received 24 November 2021; Revised 20 January 2022; Accepted 5 February 2022; Published 22 February 2022

Academic Editor: Ilan Levy

Copyright © 2022 Teerachai Amnuaylojaroen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As a result of considerable changes in rural areas in Northern Thailand, the frequency and intensity of haze outbreaks from particulate pollution, particularly fine particulate matter (PM<sub>2.5</sub>), has increased in this region. To supplement ground-based monitoring where PM<sub>2.5</sub> observation is limited, this study applied a multivariate linear regression model to predict PM<sub>2.5</sub> concentrations in 2020 using aerosol optical depth (AOD); meteorological parameters of wind velocity, temperature, and relative humidity; and gaseous pollutants such as SO<sub>2</sub>, NO<sub>2</sub>, CO, and O<sub>3</sub> from ground-based measurements at three locations: Chiang Mai, Lampang, and Nan provinces in Northern Thailand. Two multivariate linear regression models were conducted in this study. The first model (model 1) is a generic model with meteorological parameters of aerosol optical depth (AOD), temperature, relative humidity, and wind speed. The second model (model 2) includes meteorological parameters and several gaseous pollutants, such as SO<sub>2</sub>, NO<sub>2</sub>, CO, and O<sub>3</sub>. In general, the regression model, which used hourly data from 2020 of the three provinces, adequately characterized the PM<sub>2.5</sub> concentrations. The performance of model 2 was good for the prediction of PM<sub>2.5</sub> concentrations at Chiang Mai ( $R^2 = 0.52$ ) and Lampang ( $R^2 = 0.60$ ). Model 2 improved the prediction of PM<sub>2.5</sub> concentration compared to model 1 for both wet and dry seasons. However, model uncertainties were also present, which lays a foundation for further study.

## 1. Introduction

Northern Thailand is a region with frequent air pollution problems from biomass burning, particularly at the beginning of the year, from January to April [1, 2]. Moreover, severe haze events in this region caused by particulate pollution have become more intense and frequent in recent years. Widespread biomass-burning emissions combined with human activities that emit particulate pollutants are important contributors to air quality problems in this region [3–5]. In addition, certain meteorological and topographic effects also produce favorable conditions that contribute to the air pollution problem in Northern Thailand.

As a result of the changing settlement in rural areas under urban pressure in Northern Thailand, the frequency and intensity of haze outbreaks from particulate pollution in this

region have increased, particularly fine particulate matter (PM<sub>2.5</sub>) [6, 7]. Pollution and smoke are exacerbated by weather conditions such as in the Chiang Mai basin that favor particulate matter accumulation, burning of stubble in preparation for upcoming rain and planting of rice, and narrow valleys creating appropriate basins for air pollution [8]. On the other hand, air pollution in Northern Thailand originates not only from domestic sources but also from long-distance pollutants brought by meteorological factors such as wind, temperature, and humidity [2]. According to AirVisual data, at the beginning of 2020, the measured PM<sub>2.5</sub> in Chiang Mai, one of the largest cities in Northern Thailand, reached a staggering level of 330 over one weekend, while the airborne pollution levels across the region varied between 100 and 390.

Collecting PM<sub>2.5</sub> data is the first stage of air quality monitoring, by using a professional sampler with a filter

membrane to measure precise  $PM_{2.5}$  concentrations in space. Many  $PM_{2.5}$  sampling and monitoring facilities, however, only examine critical sample data points, which cannot accurately represent data throughout the full research area. The lack of ground-monitoring stations and sampling locations for spatial interpolation may cause mistakes in the monitoring data [9]. To fill this information gap, the method of estimating particulate matter at a specific scale at locations with fewer monitors is an important approach that is now being widely investigated. Approaches to estimating  $PM_{2.5}$  are divided into two types: those that rely on ground-level monitor-based estimation and those that rely on satellite-based estimation [10, 11]. The general methods that incorporate ground-level monitor-based approximation include (a) land use regression (LUR) model, (b) generalized additive mixed model, (c) hierarchical model, and (d) geostatistical interpolation. The methods that rely on satellite-based (monitor-free) estimation explicitly refer to remote sensing techniques [11]. Satellite-based remote sensing data on aerosol properties, such as moderate-resolution imaging spectroradiometer (MODIS) and aerosol optical depth (AOD), are gradually being applied to air quality monitoring [12]. AOD is related to radiation wavelength, aerosol size, vertical profile, and particle size distribution. In the visible and near-infrared bands, particle sizes corresponding to AOD inversion in the visible and near-infrared bands vary from 0.1 to  $2\ \mu\text{m}$ , which is very close to the particle size range of  $PM_{2.5}$ .

Meanwhile, many methods, including simple linear regression [13] multivariable statistical methods [14], mixed-effects [15], and chemical transport models, have been developed to estimate surface  $PM_{2.5}$  from AOD values based on satellite-measurement data. The goal is to establish a measurable relationship between satellite-measured AOD and  $PM_{2.5}$  concentrations that enables continuous and real-time monitoring of air quality in specific locations [16]. Previously, the relationship between  $PM_{2.5}$  and AOD was shown using a simple linear regression model if the two values were stable over time. Van Donkelaar et al. [17] proposed using the  $PM_{2.5}/\text{AOD}$  ratio to predict  $PM_{2.5}$ , but that was fraught with uncertainty due to a lack of  $PM_{2.5}$  data from ground monitoring. According to research by Wang et al. [18] and Rogula-Kozłowska et al. [19], some pollutants, including organic carbides, ammonium nitrate, nitrogen dioxide, carbon monoxide, and sulfate, are important components contributing to  $PM_{2.5}$ , so the effects of AOD, meteorological conditions, and gaseous pollutants have been widely studied in relation to variations in  $PM_{2.5}$  concentrations. For example, Zhao et al. [20] proposed a multivariate linear regression model with the main parameters of aerosol optical depth obtained through remote sensing, meteorological factors from ground monitoring (wind velocity, temperature, and relative humidity), and gaseous pollutants ( $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{CO}$ , and  $\text{O}_3$ ) to predict  $PM_{2.5}$  concentrations in Beijing. Their results indicated that the regression model based on annual data had good performance in estimating  $PM_{2.5}$  concentration. The relationship of particulate matter ( $PM_{10}$  and  $PM_{2.5}$ ) with gaseous pollutants and several meteorological parameters was investigated by

Liu et al. [21], who revealed that pollution showed high levels of  $PM_{2.5}$  and  $PM_{10}$  concentrations during the 9-year observation period (2004–2012) in Beijing. They found significant positive correlations between the particulates and  $\text{CO}$  and  $\text{NO}_x$ , which suggests that both traffic-related emissions and combustion sources were major contributors to the particulates. In addition, the wind profile appeared to be a key factor in the variation of particles.

To supplement ground-based monitoring, this study aimed to establish a quantitative model for estimating the continuous monitoring of  $PM_{2.5}$  in Northern Thailand. This research utilized multivariate linear regression, using  $PM_{2.5}$  concentration as the dependent variable and AOD data; meteorological characteristics such as wind speed, temperature, and relative humidity; and gaseous pollutants such as  $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{CO}$ , and  $\text{O}_3$  as independent variables. To begin, three provinces in upper Northern Thailand were chosen as typical case studies based on the locations of Thai Pollution Control Department monitoring stations in Chiang Mai, Lampang, and Nan, and all necessary data for the model from 2020 were obtained.

## 2. Methodology

In this study, two linear multivariate linear regression models were created: a generic model (model 1) that includes meteorological parameters such as aerosol optical depth (AOD), temperature, relative humidity, and wind speed and a generic model (model 2) that includes chemical species such as nitrogen dioxide ( $\text{NO}_2$ ), sulfur dioxide ( $\text{SO}_2$ ), ozone ( $\text{O}_3$ ), and carbon monoxide ( $\text{CO}$ ). To demonstrate whether the multivariate linear regression model improved the prediction of  $PM_{2.5}$  concentration, the model performance was compared with that of a generic model [22] as follows:

$$PM_{2.5} = (\alpha + \varepsilon_1) + (\beta_1 + \varepsilon_1) \times \text{AOD} + (\beta_2 + \varepsilon_3) \times \text{Temp} + (\beta_3 + \varepsilon_4) \times \text{RH} + (\beta_4 + \varepsilon_5) \times \text{WSPD}, \quad (1)$$

where  $PM_{2.5}$  is given as mass concentration at the ground level ( $\mu\text{g}/\text{m}^3$ ), AOD is derived from MODIS (dimensionless), Temp is temperature ( $^\circ\text{C}$ ), RH is relative humidity (%), and WSPD is wind velocity (m/s).

A generic model including gaseous pollutants, based on Zhao et al. [20], applied a multivariate linear regression model to estimate  $PM_{2.5}$  in Beijing city. That study used  $PM_{2.5}$  data from stations in Beijing from 2015 and AOD product data from the Aqua-MODIS 550 nm Collection 6. The equation is as follows:

$$PM_{2.5} = (\alpha + \varepsilon_1) + (\beta_1 + \varepsilon_1) \times \text{AOD} + (\beta_2 + \varepsilon_3) \times \text{Temp} + (\beta_3 + \varepsilon_4) \times \text{RH} + (\beta_4 + \varepsilon_5) \times \text{WSPD} + (\beta_5 + \varepsilon_6) \times \text{CO} + (\beta_6 + \varepsilon_7) \times \text{NO}_2 + (\beta_7 + \varepsilon_8) \times \text{SO}_2 + (\beta_8 + \varepsilon_9) \times \text{O}_3, \quad (2)$$

where  $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{CO}$ , and  $\text{O}_3$  are the concentrations of the four pollutants at the ground level;  $\beta_1, \beta_2, \dots, \beta_8$  are the

slopes corresponding to the respective variables; and  $(\alpha + \varepsilon_1)$  is the intercept.

This study used measurement data from the Pollution Control Department in Northern Thailand. Figure 1 shows the three ground-monitoring stations in urban areas selected as case studies: the Municipal Office in Nan, the Provincial Government Center in Chiang Mai, and the Meteorological Station in Lampang. This area of upper Northern Thailand is influenced by southwest and northeast monsoons ([https://www.tmd.go.th/en/archive/thailand\\_climate.pdf](https://www.tmd.go.th/en/archive/thailand_climate.pdf)). This was demonstrated by separating the PM<sub>2.5</sub> monitoring data into 2 seasons (wet season: May to October; dry season: November, December, January to April). As demonstrated in Table 1, the mean PM<sub>2.5</sub> concentration in the dry season exceeded that of the wet season. Throughout 2020, data from these stations covering hourly mass concentrations of PM<sub>2.5</sub> and the four principal air pollutants (NO<sub>2</sub>, CO, SO<sub>2</sub>, and O<sub>3</sub>) as well as meteorological data (temperature, wind speed, and relative humidity) were gathered. The quality assurance/quality control (QA/QC) procedures were based on protocols of the US Environmental Protection Agency (EPA) [23]. The performance criterion of sampling was to ensure quantified data for all PM<sub>2.5</sub> exposures and microenvironmental concentrations. QA was based on the following principles: (1) all procedures should be carefully planned, tested, and performed according to standard operation procedures (SOPs) approved by the study director; (2) every piece of data must be fully traceable; and (3) any deviations or irregularities must be recorded. Missing data in PCD measurements occurred on a regular basis but only 15% of the time [3].

For this study, AOD product data from the Aqua-MODIS 550 nm Collection 6 were also retrieved. MODIS is a medium-resolution image spectrometer carried by the US Earth Observing System's Terra and Aqua satellites that provides daily aerosol data to scientists worldwide [24]. The standard MODIS level-2 (L2) product has a spatial resolution of 10 km, whereas the recently released MODIS Collection 6 (C6) product (MYD04 3K) has a resolution of 3 km. Furthermore, MODIS C6 has been improved in a number of areas, including sensor calibration, cloud detection, lookup table structure, radiation transmission computation, and gas absorption corrections [25]. However, Munchak et al. [25] also suggested low performance of data production over urban surfaces. It is definitely a limitation in terms of air quality applications. Despite this limitation, the product provides additional capabilities for researching aerosol on a local scale, such as resolving small-scale AOD gradients and point sources, retrieving aerosols in patchy cloud fields, and retrieving data closer to coastlines. In addition, the performance of the AOD data was verified with 10 stations located in various regions: Beijing, XiangHe, SACOL, and Taihu, China; Taiwan; Moscow, Russia; Minsk, Belarus; Moldova; Kyiv, Ukraine; and Belsk, Poland. The summary of AOD MODIS Collection 6 was indicated by an  $R$  value in the range of 0.769–0.948, RMSE in the range of 0.148–0.324, and MAE in the range of 0.102–0.3 [26].

The coefficient of determination ( $R^2$ ), root mean square error (RMSE), and standard error (SE) were used to assess the accuracy and precision of the two models:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^N (\text{model}_i - \text{observation}_i)^2}{N}}, \\ \text{SE} &= \frac{\sigma}{\sqrt{N}}, \\ \sigma &= \sqrt{\frac{\sum[(\text{observation} - \text{model}) - (\overline{\text{observation}} - \overline{\text{model}})]^2}{N}}, \end{aligned} \quad (3)$$

where  $N$  = number of samples.

### 3. Results and Discussion

Figures 2–4 show the scatter distributions for the fitting of models 1 and 2 to the annual, wet season, and dry season PM<sub>2.5</sub> data from Chiang Mai, Lampang, and Nan for 2020. The fitted line is based on the least-squares method to estimate the linear trend with the best fit among the scattered points. Chiang Mai, Lampang, and Nan are in upper Northern Thailand and are influenced by the monsoon climate, which is dominated by southwest and northeast monsoons ([https://www.tmd.go.th/en/archive/thailand\\_climate.pdf](https://www.tmd.go.th/en/archive/thailand_climate.pdf)). The seasonality of the data became clear after the PM<sub>2.5</sub> monitoring data were further classified into two seasons (wet season: May to October; dry season: November, December, January to April). As shown in Table 1, the mean PM<sub>2.5</sub> concentrations during the dry season substantially exceeded those during the wet season. In general, the dry season (November to April) is characterized by heavy pollution and the wet season (May to October) by light pollution in Northern Thailand, as indicated in Table 1.

Table 2 shows the  $R^2$  and error measure for models 1 and 2 fitting results in Chiang Mai, Lampang, and Nan. The  $R^2$  value of annual data from model 1 was in the range of 0.02–0.36, RMSE was 25.22–35.11  $\mu\text{g}/\text{m}^3$ , and SE was 21.72–30.27  $\mu\text{g}/\text{m}^3$ . Model 2 outperformed model 1, with  $R^2$  values of 0.18–0.26 and 0.13–0.42 for the wet and dry seasons, respectively; for model 1, the values were 0.09–0.22 and 0.03–0.17, respectively. RMSE and SE of model 2 were 31.47–42.10 and 6.56–7.08  $\mu\text{g}/\text{m}^3$ , respectively, during the wet season and 55.06–90.44 and 26.96–34.07  $\mu\text{g}/\text{m}^3$  during the dry season. In general, the dry season is characterized by heavy pollution and the wet season by light pollution in Northern Thailand, as indicated in Table 1. Model 2 was better able to capture the trend of PM<sub>2.5</sub> in the dry season than the wet season, as demonstrated by higher  $R^2$  values, particularly for Chiang Mai and Lampang, which were 0.40 and 0.32, respectively. However, for model 2 in the dry season, RMSE and SE were greater than model 1 by 76.91 and 34.07 for Chiang Mai and 90.44 and 25.81 for Lampang, respectively.

A higher  $R^2$  result for model 2 compared to model 1 is mainly due to the addition of gaseous air pollutants, which is inherent in the respective datasets for the two seasons utilized for model building. Furthermore, overall PM<sub>2.5</sub> changes across seasons were so substantial that the random selection of test data could not fully guarantee that the data distribution would be proportional to the circumstances of the two seasons. The uncertainty of land surface would affect the accuracy of the MODIS product, which may be a limitation in terms of air

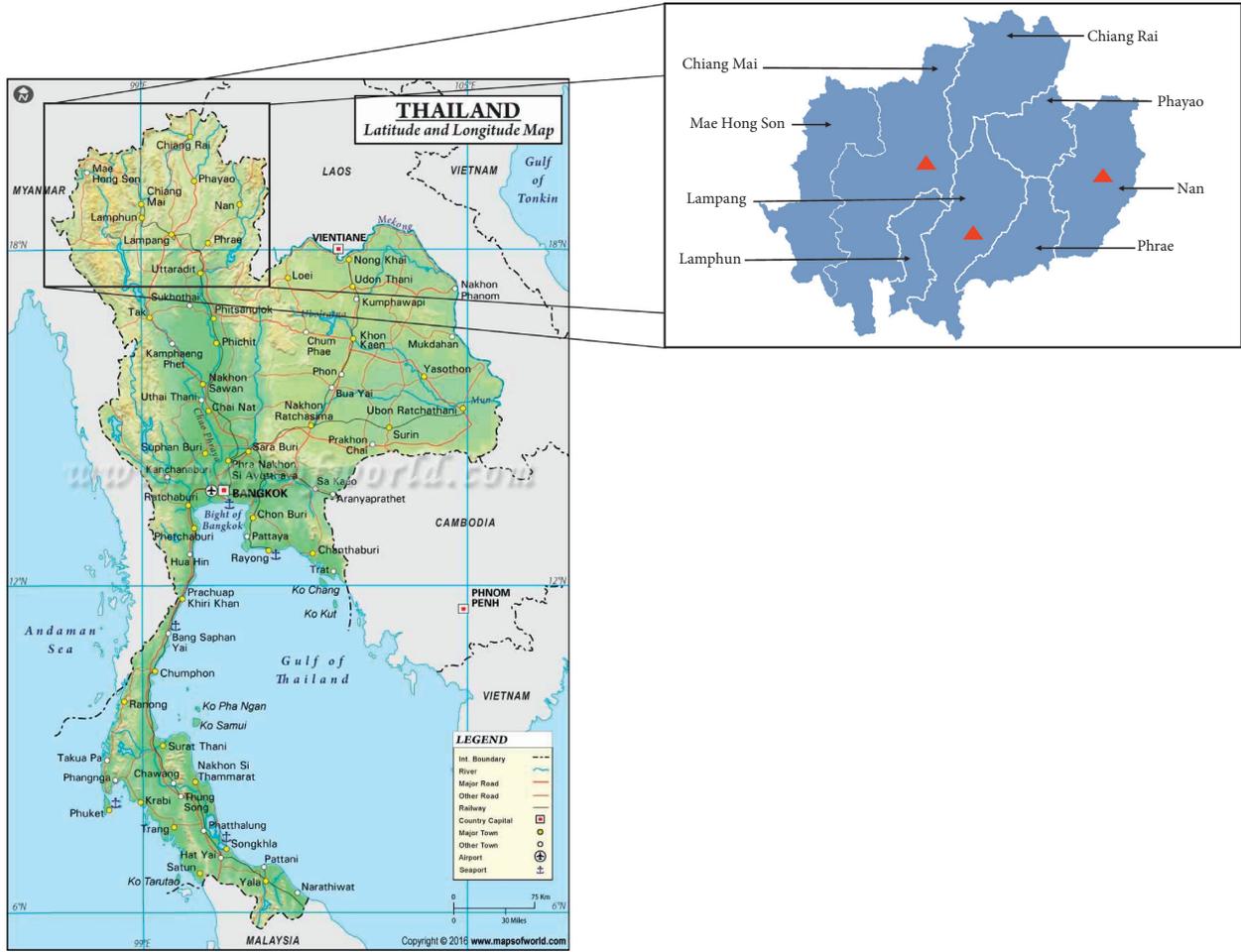


FIGURE 1: Map of Thailand; highlighted region shows the location of PM<sub>2.5</sub> monitoring stations (red triangles).

TABLE 1: Descriptive statistics and summary of PM<sub>2.5</sub> observations in Northern Thailand.

	Mean ( $\mu\text{g}/\text{m}^3$ )	SD ( $\mu\text{g}/\text{m}^3$ )	Min ( $\mu\text{g}/\text{m}^3$ )	Max ( $\mu\text{g}/\text{m}^3$ )
<i>Chiang Mai (lat: 18.83, long: 98.97)</i>				
Annual	33	32.19	3	468
Wet season	16	7.25	3	63
Dry season	63	32.19	10	468
<i>Lampang (lat: 18.28, long: 99.50)</i>				
Annual	30	30.61	1	214
Wet season	10	7.46	1	52
Dry season	63	31.23	2	214
<i>Nan (lat: 18.78, long: 100.77)</i>				
Annual	29	27.05	1	276
Wet season	13	8.22	1	65
Dry season	57	28.84	3	276

quality applications [25]. Moreover, the overestimation of modeled PM<sub>2.5</sub> is similar to the results of Zhao et al. [20], who applied a multivariate linear regression model to achieve short-term prediction of PM<sub>2.5</sub> in Beijing. PM<sub>2.5</sub> ground-monitoring stations were predominantly located in cities and metropolitan areas, with only a few in suburbs. Systematic variations intrinsic to the tapered element oscillating microbalance (TEOM) PM<sub>2.5</sub> measuring technology also highlighted uncertainties in the PM<sub>2.5</sub> data sources [27]. As a result, the

estimation accuracy was reduced [25]. Of concern, the proposed model overlooks the complicated mechanism of PM<sub>2.5</sub> formation, which would affect its performance.

Sulfate ( $\text{SO}_4^{2-}$ ) and nitrate ( $\text{NO}_3^-$ ) were reported to be the two most sensitive parameters contributing to PM<sub>2.5</sub> in the atmosphere [28].  $\text{SO}_4^{2-}$  can be generated by oxidizing  $\text{SO}_2$  with OH in the gas phase or with dissolved  $\text{H}_2\text{O}_2$  or  $\text{O}_2$  in clouds [29]. During the day,  $\text{NO}_2$  with hydroxyl radical (OH), followed by condensation, dominates nitrate production, while at

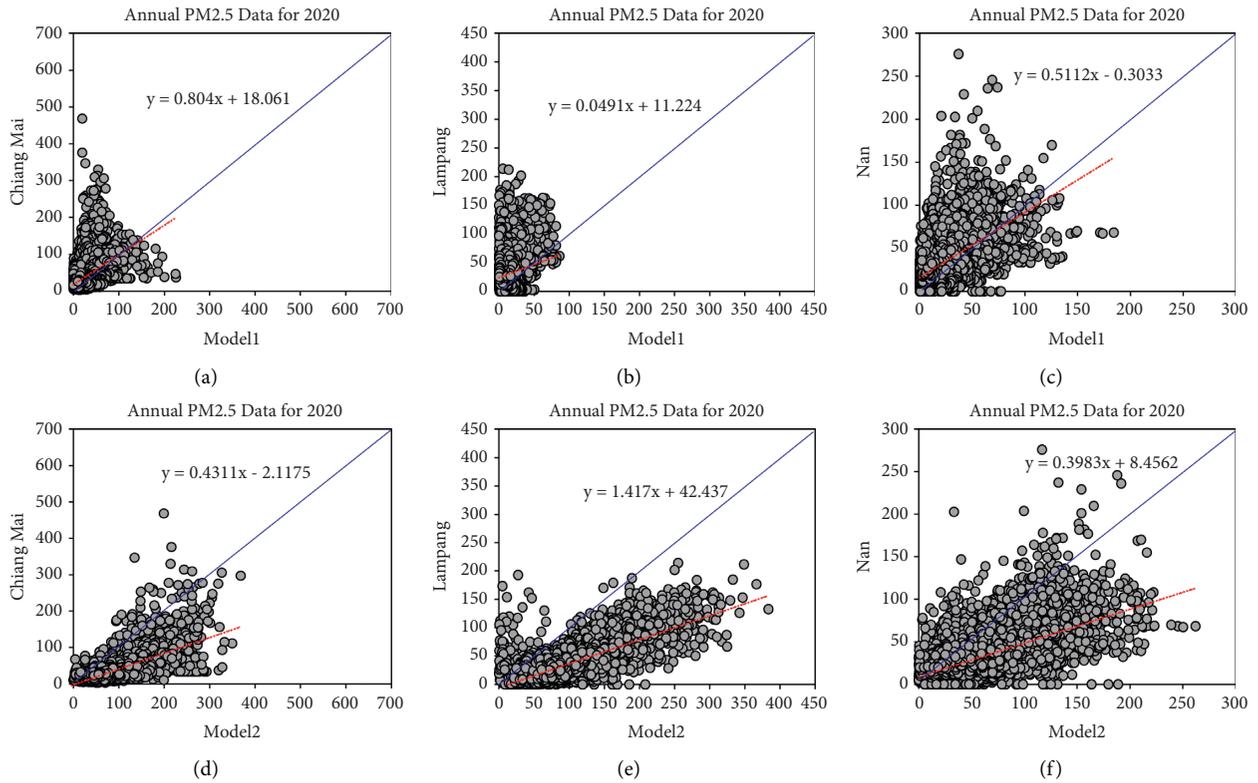


FIGURE 2: Fitting results of annual PM<sub>2.5</sub> data for (a–c) model 1 (generic) and (d–f) model 2 (multivariate linear regression).

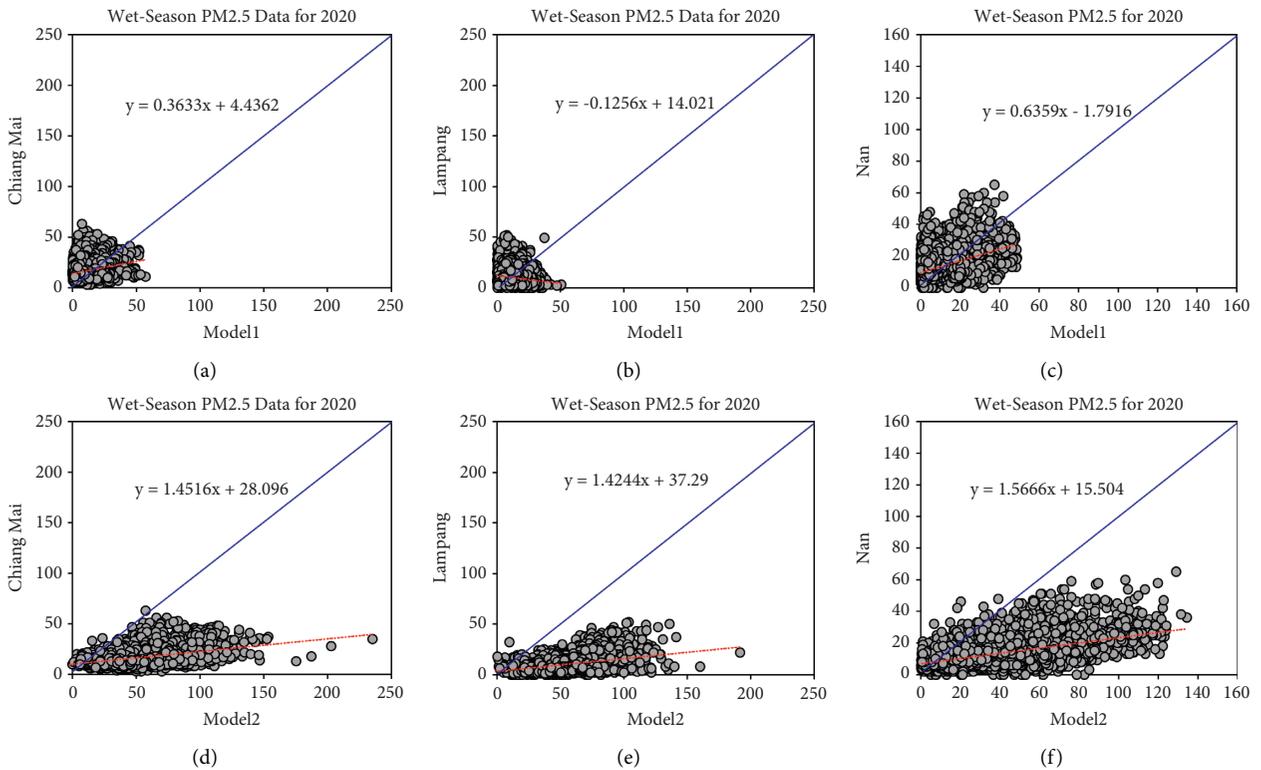


FIGURE 3: Fitting results of wet-season PM<sub>2.5</sub> data for (a–c) model 1 (generic) and (d–f) model 2 (multivariate linear regression).

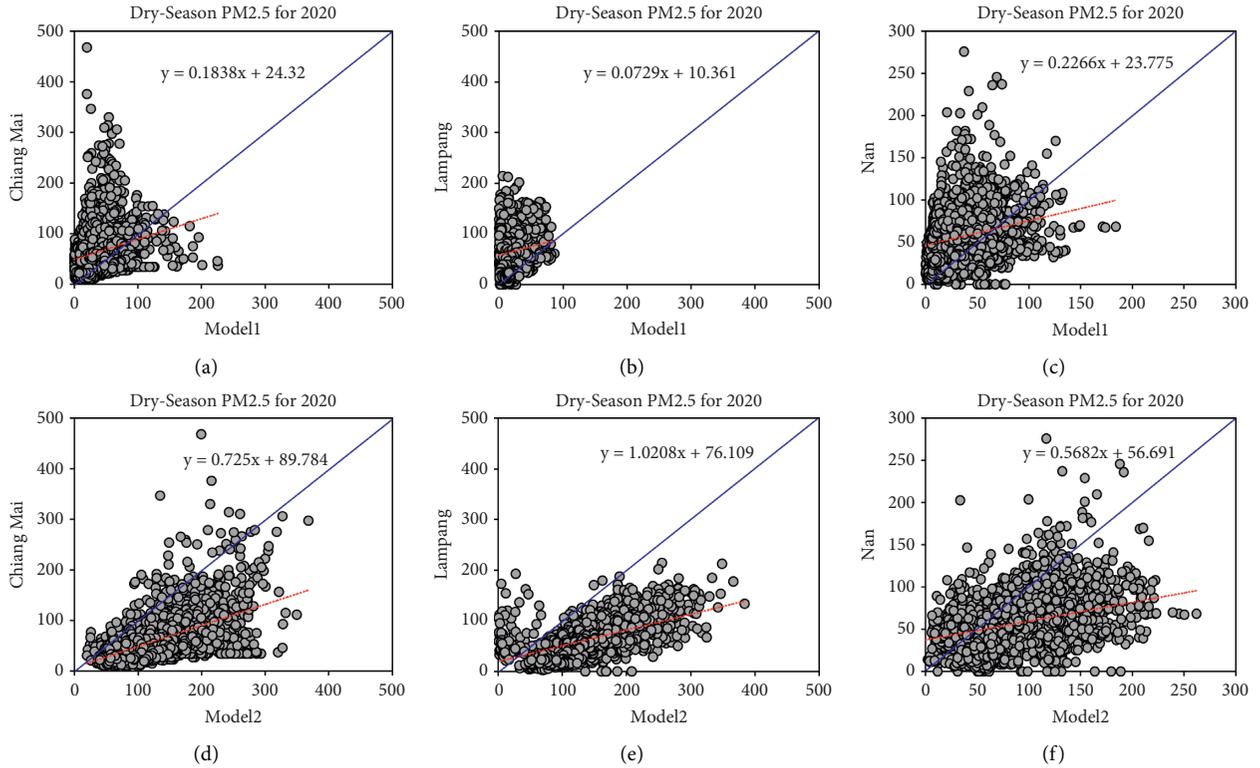


FIGURE 4: Fitting results of dry-season  $PM_{2.5}$  data for (a–c) model 1 (generic) and (d–f) model 2 (multivariate linear regression).

TABLE 2: Summary of  $R^2$  and error measure for fitting results.

	$R^2$		Annual RMSE		SE		Mean	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Chiang Mai	0.28	0.52	31.1	61.91	27.33	22.36	19.07	82.38
Lampang	0.02	0.60	35.11	66.48	30.27	19.37	12.69	84.48
Nan	0.36	0.39	25.22	40.23	21.72	21.05	17.36	51.89
Mean	0.22	0.50	30.48	56.21	26.44	20.93	16.37	72.92
<i>Wet Season</i>								
Chiang Mai	0.09	0.18	11.43	42.1	6.93	6.56	10.42	51.99
Lampang	0.02	0.18	11.2	47.37	7.39	6.77	12.78	51.34
Nan	0.22	0.26	10.46	31.47	7.27	7.08	8.49	35.28
Mean	0.11	0.21	11.03	40.31	7.20	6.80	10.56	46.20
<i>Dry Season</i>								
Chiang Mai	0.17	0.4	42.58	76.91	39.19	34.07	35.94	135.62
Lampang	0.03	0.32	57.67	90.44	30.85	25.81	14.93	140.15
Nan	0.06	0.13	38.82	55.06	27.9	26.96	36.74	88.95
Mean	0.09	0.28	46.36	74.14	32.65	28.95	29.20	121.58

night, heterogeneous reactions of nitrate radical ( $NO_3$ ) or dinitrogen pentoxide ( $N_2O_5$ ) predominate [30]. A comparison can be made with other methodologies, for example, 3-dimensional (3D) chemical model simulation. In a simulation using the community multiscale air quality (CMAQ) modeling system conducted by Pimonsree et al. [31] in Northern Thailand, the average estimated  $PM_{2.5}$  concentration from model 2 was  $121.58 \mu g/m^3$ , while a 3D chemical model simulation with Fire Inventory from NCAR (FINN) and modified FINN with fire radiative power (FRP) estimated 238 and  $92 \mu g/m^3$ , respectively, in the dry season. Model 2 underestimated by

50% and overestimated by 31% in the FINN (IOA = 0.25) and modified FINN with FRP (IOA = 0.55) simulations of the high pollution episode (March 2012).

$PM_{2.5}$  prediction is a challenging issue influenced by several variables, including weather conditions and environmental seasonality. Various factors influence the regression findings used to predict  $PM_{2.5}$ . Aside from weather, there may be some correlation between  $PM_{2.5}$ ,  $SO_2$ ,  $NO_2$ ,  $CO$ , and  $O_3$ . Table 3 shows that  $PM_{2.5}$  correlates with other measurements and factors. During the study period, mean  $PM_{2.5}$  concentrations correlated strongly with  $CO$ ,  $NO_2$ , and  $SO_2$  concentrations and

TABLE 3: Pearson correlation between  $PM_{2.5}$  and various parameters.

Variables	Chiang Mai	Lampang	Nan
Temp	0.06	-0.17	0.18
RH	-0.46	-0.25	-0.57
AOD	0.48	0.47	0.53
WSPD	0.02	-0.17	-0.06
$NO_2$	0.48	0.56	0.44
$SO_2$	0.35	0.39	0.01
$O_3$	0.30	0.27	0.32
CO	0.58	0.76	0.34

AOD. The  $PM_{2.5}$  and AOD relationship is seen to be spatial and temporal [32]. AOD has been widely used to retrieve ground  $PM_{2.5}$  concentrations because of its wide coverage and continuous spatial distribution. While  $NO_2$  and  $SO_2$  directly affect  $PM_{2.5}$ , CO has an indirect effect. The hydroxyl radical combines with CO, and the hydrogen atom reacts with  $O_2$  to form  $HO_2$ . When NO is present, the most important atmospheric reaction that  $HO_2$  undergoes is with NO. The results reflect that the variables employed in the regression model are reasonable. The model's predictive validity improved considerably after integrating the chemical species of gases to capture the impact of gaseous contaminants on  $PM_{2.5}$  production, especially during the dry season when biomass burning is a significant source of CO and NO in this region. However, this study estimated  $PM_{2.5}$  for only three locations and disregarded regional variations in the atmosphere. To improve prediction on a large scale, the spatial heterogeneity of  $PM_{2.5}$  data should be obtained, as suggested by Ma et al. [33, 34]. In addition, the latest version of AOD product data from the Aqua-MODIS 550 nm Collection 6.1, which improves the retrieval algorithms to decrease uncertainty [26], should be obtained for improved prediction.

Even though this model surpassed earlier models, there are still some uncertainties. The sources of  $PM_{2.5}$  measurements are suspect for many reasons. However, systematic errors in the TEOM approach for assessing  $PM_{2.5}$  concentrations were also revealed. Considering the high-quality AOD products at 3 km resolution, there were uncertainties in the AOD data that reduced the prediction accuracy. Third, the model as a whole has certain unsolved questions. The mechanism by which  $PM_{2.5}$  is formed was not taken into account in this study, which would have reduced the model accuracy. Spatial and temporal heterogeneity can cause uncertainty. Localization and time of day can affect  $PM_{2.5}$  concentrations. The model described here provides accurate predictions for a small area. If the model can be used to predict  $PM_{2.5}$  concentrations across a broader geographical and temporal range, it merits further investigation. This could be a useful technique for supplementing ground-based monitoring; however, additional development, including the addition of spatial heterogeneity, should be conducted in the future.

#### 4. Conclusion

This study applied a multivariate linear regression model to predict  $PM_{2.5}$  concentrations using data from remote sensing of aerosol optical depth (AOD), meteorological elements from ground monitoring (wind velocity, temperature, and

relative humidity), and gaseous contaminants ( $SO_2$ ,  $NO_2$ , CO, and  $O_3$ ) at three locations in 2020: Chiang Mai, Lampang, and Nan Provinces in Northern Thailand. In this investigation, two multivariate linear regression models were used. Model 1, a general model, contains meteorological factors such as AOD, temperature, relative humidity, and wind speed. Model 2 incorporates meteorological characteristics as well as gaseous contaminants such as  $SO_2$ ,  $NO_2$ , CO, and  $O_3$ . The performance of the proposed model was examined by several statistical indicators: coefficient of determination ( $R^2$ ), root mean square error (RMSE), and standard error (SE). The findings show that the model performed well at capturing annual  $PM_{2.5}$  concentrations in Northern Thailand with  $R^2$  values in the range of 0.39–0.60. By including gaseous pollutants in the multivariate linear regression model, model 2 showed improved characterization of  $PM_{2.5}$  concentrations compared to model 1. However, the model found a large discrepancy, and to improve its efficiency, the nonlinear processes of  $PM_{2.5}$  formation and the spatial heterogeneity of  $PM_{2.5}$  data should be considered for prediction on a large scale to supplement ground-based measurements. Additionally, for improved prediction, AOD product data from the Aqua-MODIS 550 nm Collection 6.1, which enhances retrieval techniques to reduce uncertainty, should be acquired.

#### Data Availability

The data used in this study were provided with permission by Thailand's Pollution Control Department (PCD) and hence cannot be made freely available. Access to these data can be obtained by contacting the PCD at <http://www.pcd.go.th/>.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Authors' Contributions

Teerachai Amnuaylojaroen conceived and designed the experiments; performed the experiments; analysed and interpreted the data; contributed the reagents, materials, analysis tools, and data; and wrote the manuscript.

#### Acknowledgments

This research was funded by the National Research Council of Thailand and the University of Phayao, Thailand. The author thank the Thai Pollution Control

Department for supporting ground-based measurement data and the National Research Council of Thailand for financial support.

## References

- [1] S. Yin, X. Wang, X. Zhang, M. Guo, M. Miura, and Y. Xiao, "Influence of biomass burning on local air pollution in mainland Southeast Asia from 2001 to 2016," *Environmental Pollution*, vol. 254, Article ID 112949, 2019.
- [2] T. Amnuaylojaroen, J. Inkom, R. Janta, and V. Surapipith, "Long range transport of southeast asian PM2.5 pollution to northern Thailand during high biomass burning episodes," *Sustainability*, vol. 12, no. 23, p. 10049, 2020.
- [3] T. Amnuaylojaroen, M. C. Barth, L. K. Emmons et al., "Effect of different emission inventories on modeled ozone and carbon monoxide in Southeast Asia," *Atmospheric Chemistry and Physics*, vol. 14, no. 23, pp. 12983–13012, 2014.
- [4] H.-H. Lee, O. Iraqui, Y. Gu et al., "Impacts of air pollutants from fire and non-fire emissions on the regional air quality in Southeast Asia," *Atmospheric Chemistry and Physics*, vol. 18, no. 9, pp. 6141–6156, 2018.
- [5] H.-H. Lee, O. Iraqui, and C. Wang, "The impact of future fuel consumption on regional air quality in southeast asia," *Scientific Reports*, vol. 9, no. 1, p. 2648, 2019.
- [6] T. Amnuaylojaroen, R. C. Macatangay, and S. Khodmanee, "Modeling the effect of VOCs from biomass burning emissions on ozone pollution in upper Southeast Asia," *Heliyon*, vol. 5, no. 10, Article ID e02661, 2019.
- [7] S. Khodmanee and T. Amnuaylojaroen, "Impact of biomass burning on Ozone, Carbon monoxide and Nitrogen dioxide in northern Thailand," *Frontiers in Environmental Science*, vol. 9, p. 27, 2021.
- [8] T. Amnuaylojaroen and J. Kreasuwun, "Investigation of fine and coarse particulate matter from burning areas in Chiang Mai, Thailand using the WRF/CALPUFF," *Chiang Mai Journal of Science*, vol. 39, no. 2, pp. 311–326, 2012.
- [9] L. Qu, H. Xiao, N. Zheng, Z. Zhang, and Y. Xu, "Comparison of four methods for spatial interpolation of estimated atmospheric nitrogen deposition in South China," *Environmental Science and Pollution Research*, vol. 24, no. 3, pp. 2578–2588, 2017.
- [10] S.-J. Lee, M. L. Serre, A. van Donkelaar, R. V. Martin, R. T. Burnett, and M. Jerrett, "Comparison of geostatistical interpolation and remote sensing techniques for estimating long-term exposure to ambient PM 2.5 concentrations across the continental United States," *Environmental Health Perspectives*, vol. 120, no. 12, pp. 1727–1732, 2012.
- [11] G. Zhang, X. Rui, and Y. Fan, "Critical review of methods to estimate PM2.5 concentrations within specified research region," *ISPRS International Journal of Geo-Information*, vol. 7, no. 9, p. 368, 2018.
- [12] R. M. Hoff and S. A. Christopher, "Remote sensing of particulate pollution from space: have we reached the promised land?" *Journal of the Air & Waste Management Association*, vol. 59, no. 6, pp. 645–675, 2009.
- [13] H. Zhang, R. M. Hoff, and J. A. Engel-Cox, "The relation between moderate resolution imaging spectroradiometer (MODIS) aerosol optical depth and PM2.5 over the United States: a geographical comparison by U.S. Environmental protection agency regions," *Journal of the Air & Waste Management Association*, vol. 59, no. 11, pp. 1358–1369, 2009.
- [14] P. Gupta and S. A. Christopher, "Seven year particulate matter air quality assessment from surface and satellite measurements," *Atmospheric Chemistry and Physics*, vol. 8, no. 12, pp. 3311–3324, 2008.
- [15] I. Kloog, F. Nordio, B. A. Coull, and J. Schwartz, "Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the Mid-Atlantic states," *Environmental Science & Technology*, vol. 46, no. 21, pp. 11913–11921, 2012.
- [16] X. Hu, L. A. Waller, A. Lyapustin, Y. Wang, M. Z. Al-Hamdan, and W. L. Crosson, "Estimating ground-level PM2.5 concentrations in the southeastern United States using MAIAC AOD retrievals and a two-stage model," *Remote Sensing of Environment*, vol. 140, pp. 220–232, 2014.
- [17] A. van Donkelaar, R. V. Martin, M. Brauer et al., "Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application," *Environmental Health Perspectives*, vol. 118, no. 6, p. 847, 2010.
- [18] J. Wang, Y. Song, J. Zuo, and H. Wu, "Compositions and pollutant sources of haze in Beijing urban sites," *Environmental Science and Pollution Research*, vol. 23, no. 9, pp. 8827–8836, 2016.
- [19] W. Rogula-Kozłowska, K. Klejnowski, P. Rogula-Kopiec et al., "Spatial and seasonal variability of the mass concentration and chemical composition of PM 2.5 in Poland," *Air Quality, Atmosphere & Health*, vol. 7, no. 1, pp. 41–58, 2014.
- [20] R. Zhao, X. Gu, B. Xue, J. Zhang, and W. Ren, "Short period PM2.5 prediction based on multivariate linear regression model," *PLoS One*, vol. 13, no. 7, Article ID e0201011, 2018.
- [21] Z. Liu, B. Hu, L. Wang, F. Wu, W. Gao, and Y. Wang, "Seasonal and diurnal variation in particulate matter (PM 10 and PM 2.5) at an urban site of Beijing: analyses from a 9-year study," *Environmental Science and Pollution Research*, vol. 22, no. 1, pp. 627–642, 2015.
- [22] W. Song, H. Jia, J. Huang, and Y. Zhang, "A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China," *Remote Sensing of Environment*, vol. 154, pp. 1–7, 2014.
- [23] K. Sen, G. S. Fout, R. Haugland, C. Moulton, A. C. Grimm, and G. DiGiovanni, *Quality Assurance/quality Control Guidance for Laboratories Performing PCR Analyses on Environmental Samples*, Publication as an EPA document, USA, 2004.
- [24] R. C. Levy, L. A. Remer, R. G. Kleidman et al., "Global evaluation of the Collection 5 MODIS dark-target aerosol products over land," *Atmospheric Chemistry and Physics*, vol. 10, no. 21, pp. 10399–10420, 2010.
- [25] L. A. Munchak, R. C. Levy, S. Mattoo, L. A. Remer, B. N. Holben, and J. S. Schafer, "MODIS 3 km aerosol product: applications over land in an urban/suburban region," *Atmospheric Measurement Techniques*, vol. 6, no. 7, pp. 1747–1759, 2013.
- [26] M. Filonchyk and V. Hurynovich, "Validation of MODIS aerosol products with AERONET measurements of different land cover types in areas over Eastern Europe and China," *Journal of Geovisualization and Spatial Analysis*, vol. 4, no. 1, pp. 1–11, 2020.
- [27] J. Engel-Cox, N. T. K. Oanh, A. van Donkelaar, R. V. Martin, and E. Zell, "Toward the next generation of air quality monitoring: particulate matter," *Atmospheric Environment*, vol. 80, pp. 584–590, 2013.
- [28] Z. Shi, L. Huang, J. Li, Q. Ying, H. Zhang, and J. Hu, "Sensitivity analysis of the surface ozone and fine particulate matter to meteorological parameters in China," *Atmospheric Chemistry and Physics*, vol. 20, no. 21, pp. 13455–13466, 2020.

- [29] J. H. Seinfeld and S. N. Pandis, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, Wiley, New York, NY, USA, 2nd edition, 2006.
- [30] F. Jiang, F. Liu, Q. Lin, Y. Fu, and Y. Yang, "Characteristics and formation mechanisms of sulfate and nitrate in size-segregated atmospheric particles from urban Guangzhou, China," *Aerosol and Air Quality Research*, vol. 19, no. 6, pp. 1284–1293, 2019.
- [31] S. Pimonsree, P. Vongruang, and S. Sumitsawan, "Modified biomass burning emission in modeling system with fire radiative power: simulation of particulate matter in Mainland Southeast Asia during smog episode," *Atmospheric Pollution Research*, vol. 9, no. 1, pp. 133–145, 2018.
- [32] Q. Yang, Q. Yuan, L. Yue, T. Li, H. Shen, and L. Zhang, "The relationships between PM<sub>2.5</sub> and AOD in China: about and behind spatiotemporal variations," 2018, <https://arxiv.org/abs/1808.05729>.
- [33] Z. Ma, X. Hu, L. Huang, J. Bi, and Y. Liu, "Estimating ground-level PM<sub>2.5</sub> in China using satellite remote sensing," *Environmental science & technology*, vol. 48, no. 13, pp. 7436–7444, 2014.
- [34] Y. Xie, B. Zhao, L. Zhang, and R. Luo, "Spatiotemporal variations of PM<sub>2.5</sub> and PM<sub>10</sub> concentrations between 31 Chinese cities and their relationships with SO<sub>2</sub>, NO<sub>2</sub>, CO and O<sub>3</sub>," *Particuology*, vol. 20, pp. 141–149, 2015.