Estimating Ground-Level Hourly $PM_{2.5}$ Concentrations in Thailand Using Satellite Data: A Log-Linear Model With Sum Contrast Analysis

Suhaimee Buya, *Member, IEEE*, Hideomi Gokon, Hieu Chi Dam, Sasiporn Usanavasin, and Jessada Karnjana

***Abstract*—This study introduces a novel method for estimating hourly concentrations of particulate matter 2.5 μm ($PM_{2.5}$) using satellite data. The pollution control department of Thailand collected hourly $PM_{2.5}$ data nationwide in 2020. NASA's Earth Observing System Data and Information System encompasses all moderate resolution imaging spectroradiometer satellite data. We employed aerosol optical depth (AOD), land surface temperature (LST), normalized difference vegetation index (NDVI), and elevation (EV) in our analysis. The approach incorporates a weighted sum contrast log-linear regression model that integrates satellite data, allowing for the examination of small-scale hourly variations in $PM_{2.5}$ concentrations. The results reveal a high correlation between hourly $PM_{2.5}$ levels and AOD, LST, NDVI, EV, time, and week of the year in terms of spatial distribution, with an R^2 value of 53.8%. The mean hourly $PM_{2.5}$ concentration was 23.1 μg/m^3, displaying elevated concentrations during the dry season (November to March) and peak hours (8 to 11 A.M. and 8 to 12 P.M.). Positive correlations between AOD and $PM_{2.5}$, especially when AOD exceeded 0.52, and between LST and $PM_{2.5}$, particularly when LST exceeded 33.9 °C, along with NDVI ranging from −0.08 to 0.18 and EV above 67.9 m, resulted in higher $PM_{2.5}$ levels than the overall mean. The proposed model proved valuable for interpretation and practical application, offering comparable estimated hourly $PM_{2.5}$ concentrations at a 1-km resolution with monitoring stations. This suggests that researchers or policymakers may use the model to understand hourly $PM_{2.5}$ fluctuations and their impact on human health and the environment.**

***Index Terms*—Hourly $PM_{2.5}$, log-linear regression, satellite data, weighted sum contrast.**

Manuscript received 4 October 2023; revised 18 January 2024 and 1 March 2024; accepted 1 April 2024. Date of publication 4 April 2024; date of current version 17 April 2024. This work was supported by the JSPS KAKENHI under Grant 23H00256. *(Corresponding author: Sasiporn Usanavasin.)*

Suhaimee Buya is with the SIIT, Thammasat University, Pathum Thani 12120, Thailand, and also with the School of Knowledge Science, JAIST, Nomi 923-1211, Japan (e-mail: suhaimee.buy@dome.tu.ac.th).

Hideomi Gokon is with the Division of Advanced Science and Technology, JAIST, Nomi 923-1211, Japan (e-mail: gokon@jaist.ac.jp).

Hieu Chi Dam is with the Director of Co-Creative Intelligence Research Area, Knowledge Science, Co-Creative Intelligence, Research Centre for Interpretable AI, International Research Center for Materials Informatics (Excellent Core), JAIST, Nomi 923-1211, Japan (e-mail: dam@jaist.ac.jp).

Sasiporn Usanavasin is with the School of ICT at Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani 12120, Thailand (e-mail: sasiporn.us@siit.tu.ac.th).

Jessada Karnjana is with the Artificial Intelligence Research Group, National Electronics and Computer Technology Center, National Science and Technology Development Agency, Pathum Thani 12120, Thailand (e-mail: jessada.karnjana@nectec.or.th).

Digital Object Identifier 10.1109/JSTARS.2024.3384964

I. Introduction

Suspended particulate matter from human and natural sources significantly impacts the climate and environment [1], [2]. Epidemiological studies indicate that atmospheric particles with an aerodynamic diameter of less than 2.5 μm ($PM_{2.5}$) can cause severe harm to human health, resulting in increased deaths due to cardiovascular and respiratory diseases and lung cancer [3], [4], [5]. Due to the uneven distribution of monitoring sites, analyzing the spatial and temporal distribution of $PM_{2.5}$ particles through traditional ground-based monitoring is restricted. Hence, it is crucial to acquire precise and dependable estimates of $PM_{2.5}$ concentrations with high coverage and resolution to evaluate the impact of air quality on public health.

Various studies have revealed a substantial correlation between surface $PM_{2.5}$ concentrations and the aerosol optical depth (AOD) obtained from satellite data. However, to enhance the accuracy of the modeling, other variable factors are frequently incorporated [6], [7]. With advancements in satellite technology, it is now possible to estimate ground-level $PM_{2.5}$ concentrations using satellite data products. This approach is expected to provide accurate and reliable estimates of $PM_{2.5}$ concentrations with high coverage and resolution, overcoming the limitations of conventional ground-based monitoring systems. While the previously mentioned studies primarily used AOD data from the moderate resolution imaging spectroradiometer (MODIS), other satellite sensors, including the visible infrared imaging radiometer suite, multiangle imaging spectro radiometer, and Korean geostationary ocean color imager, also offer AOD products [8], [9], [10], [11], [12]. However, these sensors, situated on polar orbit satellites, provide only one observation per day, leading to a limited representation of aerosol distribution during the day [13], [14].

Studies have established a strong correlation between AOD and $PM_{2.5}$, with statistical models developed to relate the two [15], [16], [17]. By combining satellite remote sensing AOD with ground-based $PM_{2.5}$ observations and employing statistical methods, it is possible to generate $PM_{2.5}$ estimates with extensive spatial coverage [18]. In a review on predicting ground $PM_{2.5}$ concentrations using satellite AOD, multiple linear regression (MLR) (25 articles), mixed-effect model (MEM) (23 articles), chemical transport model (16 articles), and geographically weighted regression (10 articles) were widely employed [6]. However, no clear "best" model was identified, as each method has its strengths and limitations.

The MLR model has been extensively employed since 2005 to predict $PM_{2.5}$ levels using satellite AOD data. In this model, AOD is the independent variable, while ground-level $PM_{2.5}$ is the dependent variable. Previous research utilized the MLR model to forecast $PM_{2.5}$ concentrations in various regions, including cities, suburbs, and the countryside in the eastern United States during 2001 [19]. They observed significant variations in coefficients among regions, resulting in low R-squared (R^2) values of 0.420, 0.490, 0.590, and 0.430 in cities, suburban areas, the countryside, and the entire region, respectively. Recent studies have explored covariate factors within the MLR model to enhance the model's performance across different circumstances [20], [21], [22], [23], [24].

To estimate $PM_{2.5}$ using satellite data analysis with MLR, it is necessary to utilize treatment contrasts when the independent variables are categorical. Treatment contrasts in linear regression involve a linear combination of predictors whose coefficients add up to zero, facilitating the comparison of various treatments [25]. Initially, this method was developed to compare one or more treatment groups with a control group by setting the control group's parameter to zero and allowing the treatment parameters to reflect their respective treatment effects. Nonetheless, the propose an alternative approach for constructing 95% confidence intervals (CI) to compare means without selecting a reference group [26]. This approach provides informative 95% CIs for comparing each mean with the overall mean and employs different contrasts referred to as "sum" contrasts when a control group is absent. Sum contrasts restrict the parameters associated with each factor level, indicating the difference between that level and the overall mean outcome. It is worth noting that this technique has not yet been applied to estimate $PM_{2.5}$ using satellite data.

Weighted sum contrasts have been commonly utilized in previous studies employing linear and logistic regression models. For instance, utilized weighted sum contrasts in linear regression to evaluate the influence of land-cover transformation and elevation (EV) on decadal changes in land surface temperature (LST) by comparing the adjusted mean of all factors [27]. Conducted a comparison of blood lead levels among children in the Pattani River region of Thailand [28], while the examined HIV mortality by age group and gender in Thailand between 2014 and 2015 [29]. In addition, employed weighted sum contrasts logistic regression to investigate land-use change in Thailand [30], [31], [32]. Furthermore, utilized the same technique to explore the increase in LST in Bali, Indonesia, from 2001 to 2020 [33].

$PM_{2.5}$ concentration exhibits temporal and spatial variations that require high-resolution monitoring, typically not met by polar-orbiting satellites [34]. To address this issue, geostationary meteorological satellites have been increasingly used to estimate $PM_{2.5}$ [35]. The $PM_{2.5}$ data collected at ground station sites are typically recorded hourly, necessitating the development of accurate methods to estimate hourly levels to cover all areas of study. No study has been done for the estimation of ground-level hourly $PM_{2.5}$ in Thailand. Some have been done for the estimation of daily $PM_{2.5}$ [36], [37]. Our previous research also mentioned the satellite data that can be used to estimate $PM_{2.5}$ in Thailand and to model daily $PM_{2.5}$ in Thailand [38], [39]. It begins with AOD as a base factor and then adds other variables to improve accuracy in estimating $PM_{2.5}$ levels in Thailand. Specifically, we have selected LST, normalized difference vegetation index (NDVI), and EV data to represent land use and cover, as well as time and week of the year (WOY) as time and seasoning factors. Therefore, this study uses satellite data from our previous research to estimate ground-level hourly $PM_{2.5}$. We aim to develop an easily interpretable method for estimating hourly $PM_{2.5}$ levels using a weighted sum contrasts linear regression approach. Our proposed method will be valuable for researchers and policymakers in understanding the hourly fluctuations of $PM_{2.5}$ and their impact on human health and the environment.

II. Materials and Methods

A. Study Area

Thailand, a dynamic nation in Southeast Asia, boasts a diverse geography that encompasses the Andaman Sea and the Gulf of Thailand. With a population of approximately 70 million, Thailand is traditionally divided into four regions: central, north, northeast, and south. The four-region system is the administrative classification developed by the Ministry of Interior and used for statistical or academic purposes. It holds significant regional importance, consisting of 77 provinces and covering an extensive area of 513 120 km^2. The pollution control department (PCD) is a legally recognized government agency in Thailand that collects data on air pollution parameters throughout the country. We utilized hourly $PM_{2.5}$ data from 67 stations in 2020, as displayed in Fig. 1.

B. Satellite Data

In this study, we utilized AOD, LST, NDVI, and EV data obtained from the satellite products of MODIS. All the data used in this research were retrieved from the Distributed Active Archive Center, made available through the National Aeronautics and Space Administration (NASA) Earth Observing System Data and Information System.

We processed data obtained from the MCD19A2 product, including AOD (AOD at 045 Microns) retrieved from Terra and Aqua satellites. The AOD data are collected twice daily at 10:30 A.M. and 1:30 P.M. local standard time, with a spatial resolution of 1 kilometer (km) per pixel. The LST data from Terra (MOD11A1 product) and Aqua (MYD11A1 product) satellites were combined to calculate the daily average LST values, taking the arithmetic mean when data from both satellites were available or using the data from a single satellite if only one was operational on a specific day.

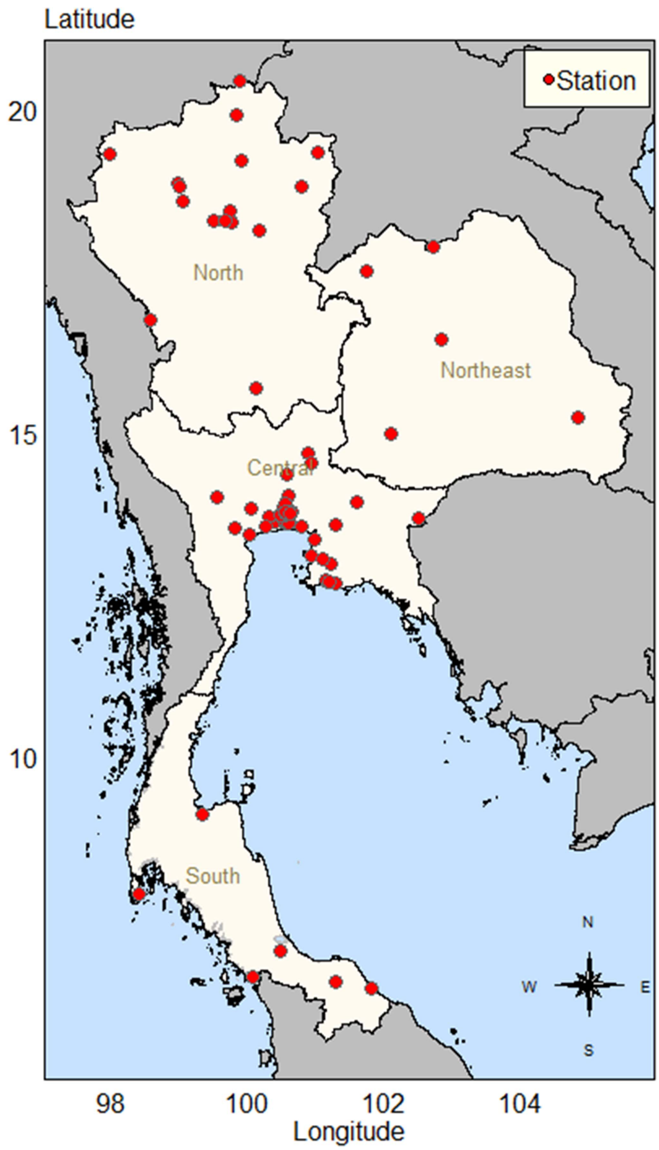

Fig. 1. $PM_{2.5}$ station in 2020.

In addition, the MOD13A1 NDVI product, with a temporal resolution of 16 days and a spatial resolution of 500 m, was employed to depict land cover changes and monitor global vegetation conditions. This dataset provides valuable insights for modeling biogeochemical and hydrologic processes, understanding climates at global and regional scales, and characterizing various biophysical features and processes on the ground surface. Lastly, the EV data from the "Land Digital Elevation Model (MODDEM1KM) - Land/sea mask and digital elevation model" with a spatial resolution of 1 km were utilized.

C. Data Analysis

1) Study Variables: We examine the association between hourly $PM_{2.5}$ concentrations and satellite data in 2020. To achieve this, we matched the $PM_{2.5}$ concentrations for each station with the average satellite data within a 5 km radius. In addition, we employed data imputation techniques, specifically using the nearest date and pixel. This matching process was accomplished using station latitude-longitude and date variables. We then analyzed the relationship between hourly $PM_{2.5}$ concentrations and various satellite variables, including AOD, LST, NDVI, and EV, as well as the time and WOY variables. The satellite data were grouped into ten levels using the ten quantiles to facilitate our analysis. It is important to note that the time variable was measured over 24 h, while the WOY variable was measured over 53 weeks. The summary of the data used in this study, including variables, units, temporal resolution, spatial resolution, and source, is displayed in Table I.

TABLE I
SUMMARY OF THE DATA USED IN THIS STUDY

Variable	Unit	Temporal resolution	Spatial resolution	Source
$PM_{2.5}$	$\mu g/m^3$	1 h	Site	PCD
AOD	Unitless	1 day	1 km	MODIS
LST	°C	1 day	1 km	MODIS
NDVI	Unitless	16 days	500 m	MODIS
EV	Meter (m)	–	1 km	MODIS
Time	hour	1 h	–	–
WOY	week	1 week	–	–

2) Linear Regression Based Weighted Sum Contrasts: This study focused on hourly $PM_{2.5}$ values as the outcome of interest, along with a group of AOD, LST, NDVI, and EV, as well as categories of WOY and time variables as determinants. When a categorical variable is used as a determinant in a regression model, it is referred to as a factor. The model formula features a set of *k–1* parameters, where *k* refers to the number of distinct categories. The linear regression model formulated as $y = a + factor(x_1) + factor(x_2) + factor(x_3) + factor(x_4) + factor(x_5) + factor(x_6)$, where y is the hourly $PM_{2.5}$, a is the constant term, x_1 is AOD group, x_2 is LST group, x_3 is NDVI group, x_4 is EV group, x_5 is time, and x_6 is WOY. Since the factors only have a *k-1* parameter, then the model was formulated as follows:

$$y = a + \sum_{i=2}^{k=10} b_i x_{1i} + \sum_{i=2}^{k=10} c_i x_{2i} + \sum_{i=2}^{k=10} d_i x_{3i} + \sum_{i=2}^{k=10} e_i x_{4i} + \sum_{i=2}^{k=24} f_i x_{5i} + \sum_{i=2}^{k=53} g_i x_{6i} \quad (1)$$

where b_i, c_i, d_i, e_i, f_i, and g_i were the coefficients of x_1, x_2, x_3, x_4, x_5, and x_6 at identity *i*, respectively, and b_1, c_1, d_1, e_1, f_1, and g_1 were equal to 0 in the case that no contrast option was selected when specifying the model. Usually, without a specified contrast option, the parameters are alphabetically assigned to each factor level, with the first parameter being set to 0. Termed "treatment" contrasts, these contrasts were initially employed in experiments that compared one or more treatment groups to a control group. By setting the parameter corresponding to the control group to 0, these contrasts ensure that the parameters associated with the treatments accurately capture the real treatment effects. However, given that our research did not incorporate a control group, we excluded six categories from our analysis of determinants and covariates.

Our study aimed to compare the mean of all relevant factors to evaluate their impact on hourly $PM_{2.5}$ levels. In order to achieve our objective, we implemented the weighted sum contrasts [26]. This approach recommends constructing 95% CI to compare means without the need for selecting a reference group, thereby offering informative intervals for comparing each mean with the overall mean. When there is no control group available, we employed "sum" contrasts, which restrict the parameters linked to each factor level, allowing us to measure the difference between that level and the overall mean of the outcome. The formulation bears a resemblance to that of treatment contrasts but includes extra terms (b_1, c_1, d_1, e_1, f_1, and g_1), that is

$$y = a + b_1 x_{11} + \sum_{i=2}^{k=10} b_i x_{1i} + c_1 x_{21} + \sum_{i=2}^{k=10} c_i x_{2i} + d_1 x_{31} + \sum_{i=2}^{k=10} d_i x_{3i} + e_1 x_{41} + \sum_{i=2}^{k=10} e_i x_{4i} + f_1 x_{51} + \sum_{i=2}^{k=24} f_i x_{5i} + g_1 x_{61} + \sum_{i=2}^{k=53} g_i x_{6i}. \quad (2)$$

During the linear regression using the weighted sum contrasts method, several statistical model parameters were calculated, including the overall mean (the average hourly $PM_{2.5}$), overall adjusted R^2 (derived from the regression model without contrast option), crude mean (the $PM_{2.5}$ mean in each factor category), and adjusted R^2 and p-value for each factor. These parameters were compared and investigated. Furthermore, 95% CI were computed for each factor category using the "democratic" approach to evaluate the mean variation. To compare means effectively, 95% CI for the difference between means were employed, representing the difference between each mean and the overall mean. These 95% CI align with the p-value and do not necessitate the selection of a control group for comparison, rendering them "democratic" in nature.

III. Results

A. Normal Distribution Test

The left-hand plot of Fig. 2 displays the quantile-quantile (Q-Q) plot of $PM_{2.5}$, revealing a highly skewed distribution. In contrast, the right-hand plot depicts the distribution after log-transforming $PM_{2.5}$, which appears to follow a normal distribution. Consequently, we will utilize the log-transformed $PM_{2.5}$ outcome in this study, as it aligns with the assumption of the linear regression model.

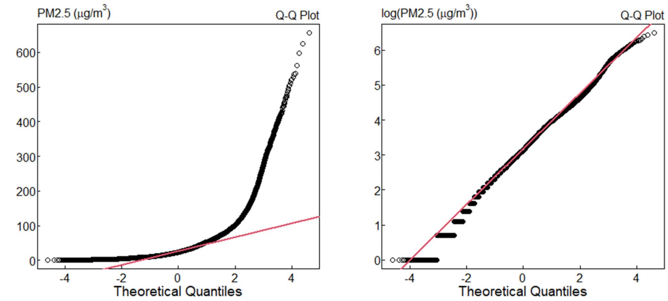

Fig. 2. Normal Q-Q plots of $PM_{2.5}$ (left) and log-transformed $PM_{2.5}$ (right).

TABLE II
Variable Selection in a Stepwise Regression Model

Response: log($PM_{2.5}$)			
Factor variable	**R^2**	**RMSE**	**P-value***
AOD+Time	20 (+20)	26.3	< 0.001
AOD+LST+Time	21.7 (+1.7)	26.2 (–0.1)	< 0.001
AOD+LST+NDVI+Time	24.8 (+3.1)	25.8 (–0.4)	< 0.001
AOD+LST+NDVI+EV+Time	31.4 (+6.6)	24.3 (–1.5)	< 0.001
AOD+LST+NDVI+EV+Time+WOY	53.8 (+22.4)	22.2 (–2.1)	< 0.001

Note: *All factor variable has p-value < 0.001.

B. Select Variables

We generated a log-linear regression model by iteratively adding and removing predictor variables based on their p-values, starting from a set of potential predictor variables. The final model, displayed in Table II, incorporates all the variables, resulting in an R^2 value of 53.8%, supported by significantly low p-values (< 0.001). Notably, the factor variables WOY (+22.4) and AOD (+20) contributed the most to the model's accuracy improvement.

C. Log-Linear Model With Sum Contrast Analysis Results

Fig. 3 displays the estimated mean of hourly $PM_{2.5}$ and the comparative 95% CI (plus sign) after adjusting for various factors, including AOD, LST, NDVI, EV, time, and WOY, relative to the overall mean for each factor. The horizontal red lines represent the simple average hourly $PM_{2.5}$ from all stations, while the blue dots indicate the crude means for each factor group. The "r-sq:" label indicates the adjusted R^2 value obtained from regression fitting with separated determinants. The "Overall r-sq:" label denotes the adjusted R^2 value obtained from the regression model adjusting for all factors.

The overall mean of hourly $PM_{2.5}$ in 2020 was 23.1 μg/m^3, falling below the National Ambient Air Quality Standards (NAAQS) of 25 μg/m^3 but still exceeding the World Health Organization (WHO) guidelines of 5 μg/m^3. The overall R^2 value was 53.8%, and all factors were significant, with p-values < 0.001. The individual R^2 values were ranked in descending order, with WOY accounting for 33.8%, followed by AOD (18.7%), NDVI (9.8%), EV (6.4%), LST (4.8%), and time (1.2%).

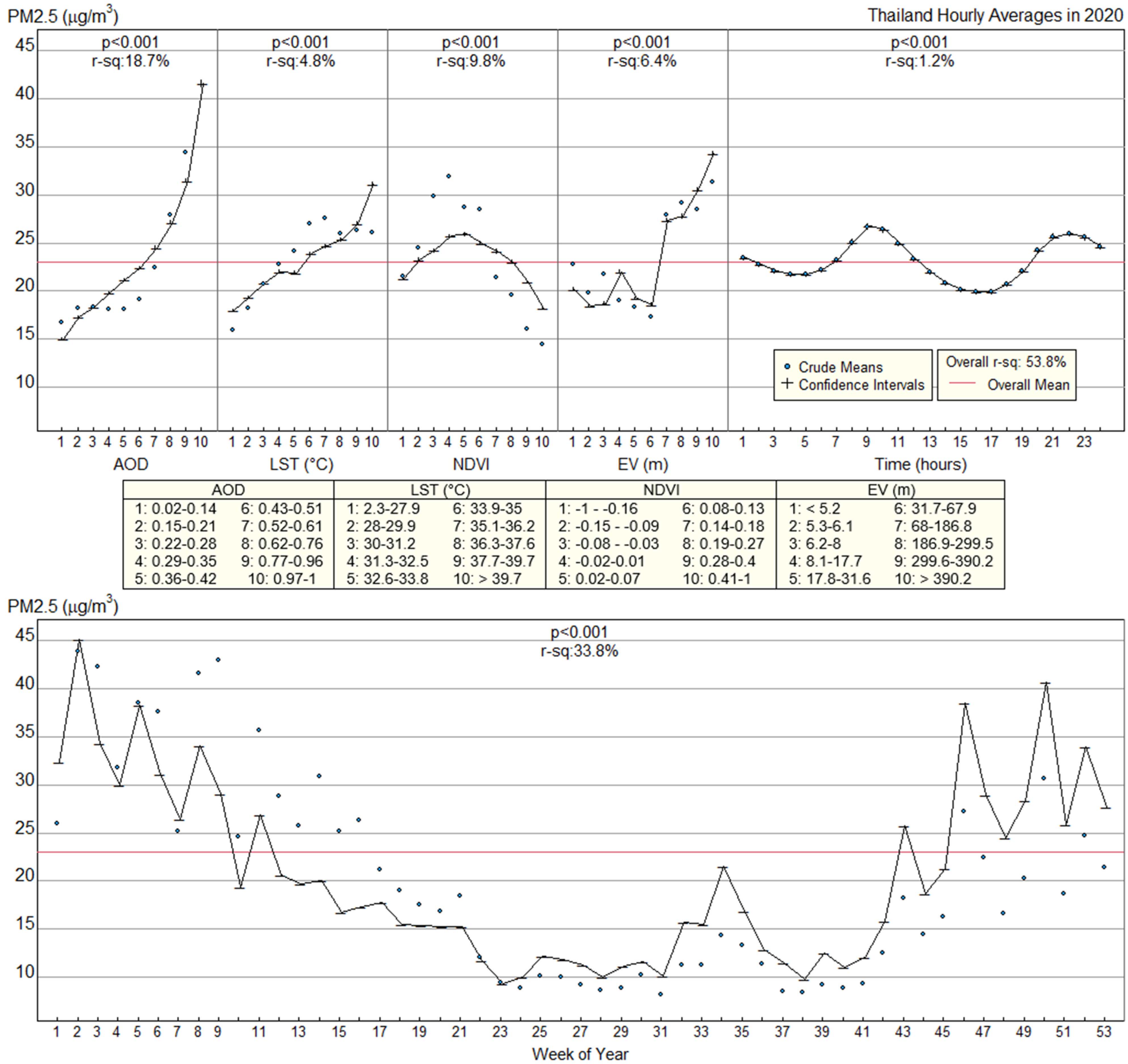

Fig. 3. Mean and confidence intervals of hourly $PM_{2.5}$ based on a log-linear regression model fitting using weighted sum contrast with satellite and time variables as determinants.

The analysis revealed that the WOY factor had higher than the overall mean and NAAQS hourly $PM_{2.5}$ levels from January to March (1–9) and November to December (46–53). The AOD factor was positively correlated with $PM_{2.5}$, particularly when it exceeded 0.52, resulting in $PM_{2.5}$ levels higher than the overall mean. Similarly, LST was positively associated with $PM_{2.5}$, particularly when it exceeded 33.9 °C, leading to $PM_{2.5}$ levels higher than the overall mean. For NDVI, $PM_{2.5}$ levels were higher than the overall mean within the range of −0.08–0.18, indicating areas likely with no green leaves and possibly urbanized. Furthermore, the EV factor showed higher $PM_{2.5}$ levels than the overall mean in areas above 67.9 m. Lastly, the time variable indicated that $PM_{2.5}$ levels were higher than the overall mean during 8–11 A.M. and 20–24 P.M.

The 95% CI plot in Fig. 3 demonstrates that it is relatively easy to estimate the hourly concentration of $PM_{2.5}$ based on each factor. For instance, if we are aware of the values of AOD (0.55), LST (35.5 °C), NDVI (0.15), EV (300 m), time (12 A.M.), and WOY (14), we can examine the graph for each factor to determine the corresponding $PM_{2.5}$ values (24.5, 24.7, 24.2, 30.5, 23.3, and 20, respectively (see actual values in Table III). We can estimate hourly $PM_{2.5}$ as follows:

$$\text{where the constant term is} - 15.69,$$

$$y = \log(24.5) + \log(24.7) + \log(24.2) + \log(30.5)$$

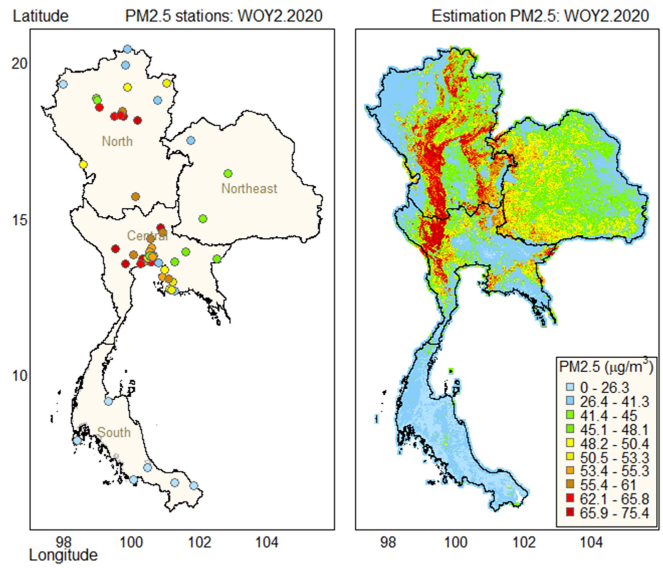

Fig. 4. Hourly $PM_{2.5}$ at stations (left) and estimated hourly $PM_{2.5}$ in each pixel 1 km resolution (right) in WOY 2 in Thailand 2020.

$$+ \log(23.3) + \log(20) - 15.69 = 3.465$$

$$\text{Hourly PM}_{2.5} = \exp(3.465) = 32 \mu\text{g/m}^3$$

Consequently, we can infer that the hourly $PM_{2.5}$ concentration in this area is around 32 μg/m^3. This straightforward approach lets us quickly estimate the hourly $PM_{2.5}$ concentration by hand and easily interpret the results.

D. Estimation Hourly $PM_{2.5}$ at 1 km Resolution

Fig. 4 displays the hourly $PM_{2.5}$ levels recorded by monitoring stations on the left and the estimated levels at a 1 km resolution on the right. The estimation used a log-linear regression model with a weighted sum contrast in WOY 2 (6–12 January) of 2020. Both the hourly $PM_{2.5}$ levels at monitoring stations and the estimated levels show comparable values. This method enables us to obtain $PM_{2.5}$ data for every small area since 2000, when MODIS satellite data became available for download.

Fig. 5 displays the estimation of hourly $PM_{2.5}$ levels, categorized by region. The R^2 values in the northeast (46.3%), central (44.1%), and north (36.2%) regions are higher than those in the south (11.1%). In the north, the average hourly $PM_{2.5}$ is 42.7 μg/m^3, a level that may impact health. On the other hand, the northeast and central regions have average hourly $PM_{2.5}$ levels of 33.3 and 27.5 μg/m^3, respectively, which represent median range values. Only the south region exhibits an average $PM_{2.5}$ lower than 15 μg/m^3, indicating good air quality in this area. The standard levels for 24 h of NAAQS are 37.5 μg/m^3, and WHO recommends 15 μg/m^3.

IV. Discussion

The log-linear regression model we propose, incorporating weighted sum contrasts, enhances the understanding of the relationship between satellite data and $PM_{2.5}$, enabling the estimation of hourly $PM_{2.5}$ concentrations in Thailand with a high spatial resolution of 1 km. This model facilitates the examination of spatiotemporal variations in hourly $PM_{2.5}$ concentrations at fine scales, offering valuable insights for epidemiological research and empowering individuals to make informed decisions regarding air pollution.

Although we used a linear regression model, assuming that the response variables follow a normal distribution, the Q-Q plot applied to the test revealed that log-transformation was more appropriate for the hourly $PM_{2.5}$ outcome, given its skewed distribution. Previous studies have also shown that log-transformation can reduce skewness in the distribution of $PM_{2.5}$, increasing the accuracy of estimation [40], [41], [42], [43]. This finding is consistent with others who similarly found log-transformation to be appropriate for their study on the hospital cost of Chronic-Disease Patients in Southern Thailand, as demonstrated by the Q-Q plot [44]. In statistics, a Q-Q plot is a graphical representation of the differences between observed and expected values based on the assumption of a normal distribution. A horizontal band close to zero without any discernible pattern indicates that the observed scores are normally distributed [45].

Using the log-linear regression model, we performed an iterative process of adding and removing predictor variables based on their p-values. The variables included satellite variables (AOD, LST, NDVI, and EV), time, and WOY. This process resulted in a final R^2 value of 53.8%. Notably, the WOY and AOD factor variables contributed the most to the model's accuracy improvement. This finding is consistent with previous studies in China, highlighting the significant contribution of AOD to $PM_{2.5}$ modeling [46], [47]. Moreover, time variables, such as year, month, WOY, and time, explain the seasonal variation in $PM_{2.5}$ levels and the trend over time [48].

This study has identified various factors associated with elevated levels of $PM_{2.5}$. Specifically, the analysis has shown that $PM_{2.5}$ levels were higher than the overall mean and NAAQS hourly $PM_{2.5}$ levels during particular months, namely January to March and November to December, as indicated by the WOY factor. In addition, the AOD factor was found to be positively correlated with $PM_{2.5}$ levels, particularly when exceeding 0.52, while LST was positively associated with $PM_{2.5}$ levels exceeding 33.9 °C. According to research in northern Thailand, the dry season occurring from November to April experiences high levels of ground PM_{10} and $PM_{2.5}$ concentrations due to AODs [49]. These high levels result from extensive agricultural field burning and open-air biomass burning in the region and neighboring countries. The study also found that NDVI within the range of –0.08 to 0.18 and EV in areas above 67.9 m were associated with higher $PM_{2.5}$ levels than the overall mean, indicating a relationship with the urbanization area in the central part and higher areas in the northern part of Thailand.

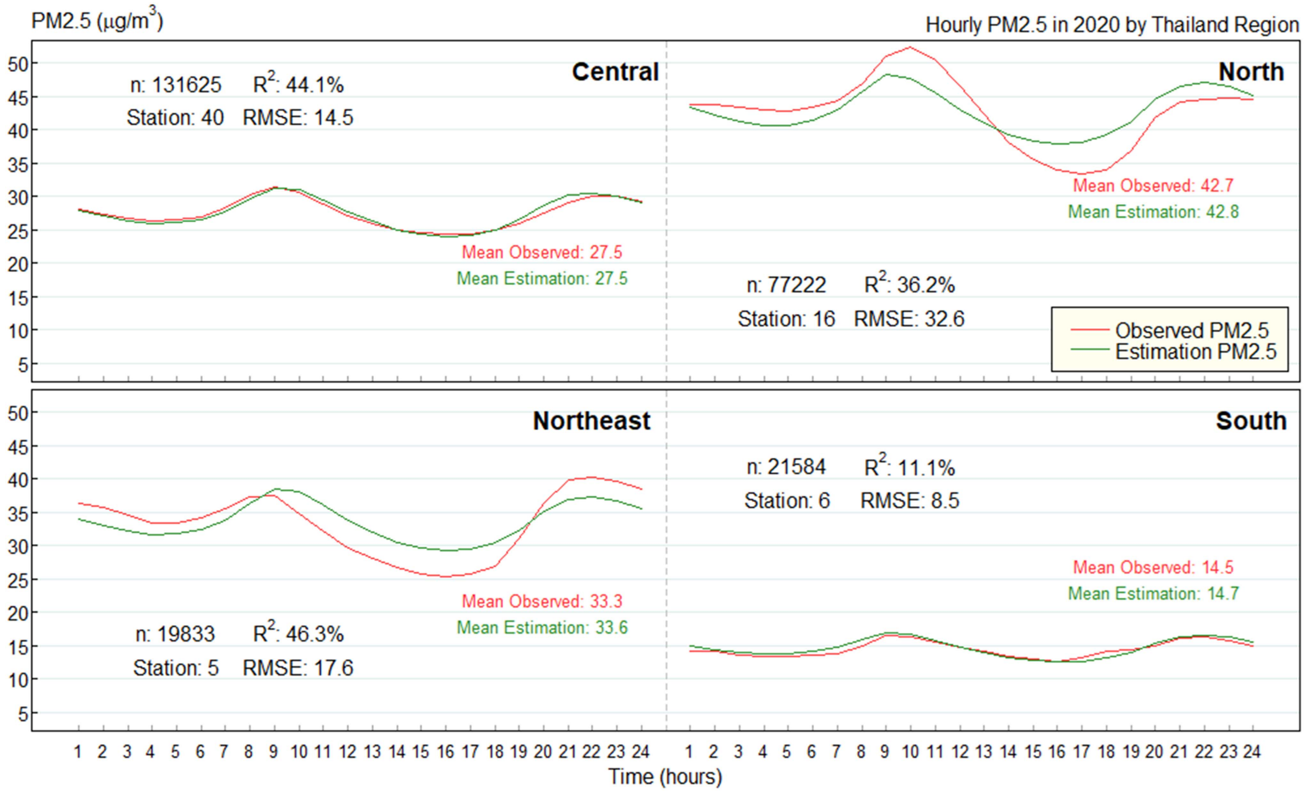

Fig. 5. Estimation of hourly $PM_{2.5}$ levels separated by region.

The time variable revealed higher $PM_{2.5}$ levels during specific hours, specifically 8–11 A.M. and 20–24 P.M., which may be related to working hours in the Bangkok Metropolitan Region (BMR), the capital of central Thailand. In the BMR, automobiles, road dust, biomass burning, and meat cooking are significant sources of $PM_{2.5}$, contributing to the elevated levels observed in the region [50], [51]. The rapid urbanization of Southeast Asia in recent years has increased the likelihood of air pollution from vehicles, industries, and construction activities [52]. Also, the study in China described that $PM_{2.5}$ had greater effects on mortality in urban cities than rural areas [53].

The weighted sum contrast linear regression model helps estimate hourly $PM_{2.5}$ levels and understand the relationship between determinants and $PM_{2.5}$. Although some studies indicate that machine learning models may offer more precise estimates than linear regression [46], [47], the latter is still valuable due to its simplicity and practicality. In contrast, machine learning models may need more interpretability, require expensive computing resources, and rely on high-resolution satellite images, which can be challenging for researchers with limited budgetary support for analyzing satellite data. Thus, the weighted sum contrast linear regression model remains viable to overcome these limitations. Statistics make population inferences from a sample, while machine learning identifies generalizable predictive patterns [54].

This method produces hourly $PM_{2.5}$ estimates at a resolution of 1 km that are similar to those obtained from monitoring stations. By leveraging MODIS satellite data, which has been accessible for download since 2000, this approach allows for the estimation of $PM_{2.5}$ data in small areas, facilitating the application of the weighted sum contrast linear regression model for hourly $PM_{2.5}$ estimation. This enables policymakers to acquire data on short-term and long-term $PM_{2.5}$ fluctuations and their effects on human health and the environment.

V. Conclusion

This study utilized a weighted sum contrast log-linear regression model to examine the impact of various factors on hourly $PM_{2.5}$ concentrations in Thailand. The model included six predictor variables: AOD, LST, NDVI, EV, time, and WOY, resulting in an R^2 of 53.8%. The WOY and AOD factors contributed the most to the model's accuracy improvement. The analysis also revealed specific periods and thresholds for each factor correlated with higher hourly $PM_{2.5}$ levels. The 95% CI plot demonstrated that it is easy to estimate hourly $PM_{2.5}$ concentrations based on each factor. The map revealed that the spatial distribution of $PM_{2.5}$ levels was comparable to those observed at the monitoring stations. Overall, the study provides valuable insights into the factors influencing hourly $PM_{2.5}$ concentrations in Thailand and the potential for utilizing satellite data to estimate hourly $PM_{2.5}$ levels in small areas.

Appendix

TABLE III
Weighed Sum Contrasts Log-Linear Regression Model Results

Variables	Crude Means (μg/m³)	Adjusted Means (μg/m³)	95% CI (μg/m³)
AOD			
01: 0.02–0.14	16.8	15.0	14.9–15.1
02: 0.15–0.21	18.3	17.2	17.1–17.4
03: 0.22–0.28	18.4	18.3	18.2–18.4
04: 0.29–0.35	18.2	19.8	19.7–19.9
05: 0.36–0.42	18.1	21.2	21.1–21.3
06: 0.43–0.51	19.2	22.5	22.3–22.6
07: 0.52–0.61	22.4	24.5	24.4–24.7
08: 0.62–0.76	27.9	27.2	27.0–27.3
09: 0.77–0.96	34.4	31.5	31.3–31.7
10: 0.97–1	51.8	41.6	41.3–41.9
LST (°C)			
01: 2.3–27.9	15.9	18.0	17.8–18.1
02: 28–29.9	18.3	19.4	19.2–19.5
03: 30–31.2	20.7	20.8	20.7–21.0
04: 31.3–32.5	22.8	22.0	21.9–22.2
05: 32.6–33.8	24.1	21.9	21.8–22.1
06: 33.9–35	27.0	23.9	23.7–24.0
07: 35.1–36.2	27.6	24.7	24.5–24.9
08: 36.3–37.6	25.9	25.4	25.2–25.6
09: 37.7–39.7	26.3	27.0	26.8–27.2
10: > 39.7	26.0	31.1	30.8–31.3
NDVI			
01: -0.33–-0.16	21.6	21.3	21.2–21.5
02: -0.15–-0.09	24.5	23.3	23.1–23.5
03: -0.08–-0.03	29.9	24.2	24.1–24.4
04: -0.02–0.01	31.9	25.7	25.5–25.9
05: 0.02–0.07	28.7	25.9	25.8–26.1
06: 0.08–0.13	28.5	25.0	24.8–25.1
07: 0.14–0.18	21.4	24.2	24.0–24.3
08: 0.19–0.27	19.6	23.1	22.9–23.2
09: 0.28–0.4	16.1	21.0	20.8–21.1
10: 0.42–1	14.5	18.2	18.1–18.4
EV (m)			
01: 3–5.2	22.7	20.2	20.1–20.3
02: 5.3–6.1	19.9	18.4	18.3–18.5
03: 6.2–8	21.7	18.7	18.6–18.9
04: 8.1–17.7	19.0	22.0	21.9–22.2
05: 17.8–31.6	18.4	19.3	19.1–19.4
06: 31.7–67.9	17.3	18.6	18.5–18.7
07: 68–186.8	28.0	27.4	27.2–27.5
08: 186.9–299.5	29.1	27.8	27.6–28.0
09: 299.6–390.2	28.5	30.5	30.3–30.7
10: > 390.2	31.3	34.3	34.0–34.5
Time			
1	23.5	23.5	23.3–23.7
2	22.8	22.8	22.5–23.0
3	22.1	22.1	21.8–22.3
4	21.7	21.7	21.5–21.9
5	21.8	21.8	21.6–22.0
6	22.2	22.2	22.0–22.5
7	23.3	23.3	23.0–23.5
8	25.1	25.1	24.9–25.4
9	26.7	26.7	26.5–27.0
10	26.4	26.4	26.1–26.7
11	24.9	24.9	24.7–25.2
12	23.3	23.3	23.1–23.6
13	22.0	22.0	21.8–22.2
14	20.8	20.8	20.6–21.0
15	20.2	20.2	20.0–20.4
16	19.9	19.9	19.7–20.1
17	19.9	19.9	19.7–20.1
18	20.7	20.7	20.5–20.9
19	22.1	22.1	21.9–22.4
20	24.2	24.2	24.0–24.5
21	25.7	25.7	25.4–25.9
22	26.0	26.0	25.7–26.2
23	25.6	25.6	25.3–25.9
24	24.7	24.7	24.4–24.9
WOY			
1	26.0	32.3	31.9–32.8
2	43.9	45.1	44.6–45.6
3	42.3	34.3	33.9–34.7
4	31.8	29.9	29.6–30.3
5	38.5	38.3	37.9–38.8
6	37.6	31.1	30.7–31.4
7	25.1	26.5	26.2–26.8
8	41.6	34.1	33.7–34.4
9	43.0	29.1	28.7–29.5
10	24.6	19.3	19.1–19.6
11	35.7	26.9	26.6–27.2
12	28.8	20.6	20.3–20.9
13	25.8	19.7	19.5–20.0
14	30.9	20.0	19.7–20.2
15	25.1	16.8	16.5–17.0
16	26.3	17.3	17.0–17.6
17	21.2	17.7	17.5–18.0
18	19.0	15.5	15.2–15.8
19	17.6	15.3	15.1–15.5
20	16.9	15.3	15.1–15.5
21	18.4	15.3	14.8–15.8
22	12.0	11.7	11.5–11.9
23	9.5	9.4	9.1–9.6
24	8.8	10.0	9.8–10.2
25	10.1	12.2	11.9–12.5
26	10.0	11.9	11.6–12.1
27	9.2	11.2	10.9–11.6
28	8.6	10.0	9.8–10.2
29	8.8	11.1	10.9–11.3
30	10.2	11.6	11.4–11.9
31	8.2	10.1	9.8–10.4
32	11.3	15.7	15.0–16.4
33	11.3	15.5	14.7–16.4
34	14.3	21.5	20.8–22.4
35	13.3	16.9	16.6–17.2
36	11.3	12.8	12.5–13.1
37	8.5	11.5	11.2–11.8
38	8.3	9.8	9.5–10.2
39	9.1	12.6	12.3–12.8
40	8.8	11.1	10.8–11.3
41	9.3	12.0	11.6–12.4
42	12.5	15.8	15.4–16.3
43	18.2	25.7	25.2–26.2
44	14.5	18.7	18.3–19.1
45	16.3	21.3	20.9–21.7
46	27.3	38.5	38.0–39.0
47	22.5	29.0	28.6–29.3
48	16.6	24.4	24.1–24.8
49	20.2	28.4	28.0–28.8
50	30.7	40.7	40.3–41.2
51	18.6	25.8	25.4–26.3
52	24.7	34.0	33.4–34.5
53	21.4	27.7	27.2–28.2

Acknowledgment

The authors would like to express their gratitude to Sirindhorn International Institute of Technology (SIIT) Thammasat University and JAIST for research support. The authors are grateful for the Pollution Control Department for providing the $PM_{2.5}$ data, which has been instrumental in this study. The authors also extend their appreciation to Prof. D. McNeil for his invaluable guidance and expertise throughout this research. The satellite data can be assessed at the Earth Science Data Systems https://ladsweb.modaps.eosdis.nasa.gov (accessed on 18 May 2023).

References

[1] F. Mao, M. Duan, Q. Min, W. Gong, Z. Pan, and G. Liu, "Investigating the impact of haze on MODIS cloud detection," *J. Geophysical Res.*, vol. 120, no. 23, pp. 12212–237247, 2015, doi: 10.1002/2015JD023555.

[2] Z. Pan, F. Mao, W. Wang, B. Zhu, X. Lu, and W. Gong, "Impacts of 3D aerosol, cloud, and water vapor variations on the recent brightening during the South Asian monsoon season," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 651, doi: 10.3390/rs10040651.

[3] F. Dominici et al., "Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases," *Jpn Automobile Manufacturers Assoc.*, vol. 295, no. 10, pp. 1127–1134, 2006, doi: 10.1001/jama.295.10.1127.

[4] W. J. Gauderman et al., "The effect of air pollution on lung development from 10 to 18 years of age," *New England J. Med.*, vol. 351, no. 11, pp. 1057–1067, 2004, doi: 10.1056/NEJMoa040610.

[5] J. Cai et al., "Association of developmental coordination disorder with early-life exposure to fine particulate matter in Chinese preschoolers," *Innovation*, vol. 4, no. 1, 2023, Art. no. 100347, doi: 10.1016/j.xinn.2022.100347.

[6] Y. Chu et al., "A review on predicting ground $PM_{2.5}$ concentration using satellite aerosol optical depth," *Atmosphere*, vol. 7, no. 10, 2016, Art. no. 129, doi: 10.3390/atmos7100129.

[7] X. Hu, L. A. Waller, A. Lyapustin, Y. Wang, and Y. Liu, "10-year spatial and temporal trends of $PM_{2.5}$ concentrations in the southeastern US estimated using high-resolution satellite data," *Atmos. Chem. Phys.*, vol. 14, no. 12, pp. 6301–6314, 2014, doi: 10.5194/acp-14-6301-2014.

[8] X. Meng, M. J. Garay, D. J. Diner, O. V. Kalashnikova, J. Xu, and Y. Liu, "Estimating PM2.5 speciation concentrations using prototype 4.4 km-resolution MISR aerosol properties over Southern California," *Atmos. Environ.*, vol. 181, pp. 70–81, 2018, doi: 10.1016/j.atmosenv.2018.03.019.

[9] W. Wang, F. Mao, L. Du, Z. Pan, W. Gong, and S. Fang, "Deriving hourly PM2.5 concentrations from himawari-8 AODs over Beijing–Tianjin–Hebei in China," *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 858, doi: 10.3390/rs9080858.

[10] J. Wu, F. Yao, W. Li, and M. Si, "VIIRS-based remote sensing estimation of ground-level PM2.5 concentrations in Beijing–Tianjin–Hebei: A spatiotemporal statistical model," *Remote Sens. Environ.*, vol. 184, pp. 316–328, 2016, doi: 10.1016/j.rse.2016.07.015.

[11] J. W. Xu et al., "Estimating ground-level $PM_{2.5}$ in eastern China using aerosol optical depth determined from the GOCI satellite instrument," *Atmos. Chem. Phys.*, vol. 15, no. 22, pp. 13133–13144, 2015, doi: 10.5194/acp-15-13133-2015.

[12] L. Zang, F. Mao, J. Guo, W. Gong, W. Wang, and Z. Pan, "Estimating hourly PM1 concentrations from himawari-8 aerosol optical depth in China," *Environ. Pollut.*, vol. 241, pp. 654–663, 2018, doi: 10.1016/j.envpol.2018.05.100.

[13] J. Guo et al., "Impact of diurnal variability and meteorological factors on the PM2.5 - AOD relationship: Implications for PM2.5 remote sensing," *Environ. Pollut.*, vol. 221, pp. 94–104, 2017, doi: 10.1016/j.envpol.2016.11.043.

[14] M. Feiyue et al., "Assimilating moderate resolution imaging spectroradiometer radiance with the weather research and forecasting data assimilation system," *J. Appl. Remote Sens.*, vol. 11, no. 3, 2017, Art. no. 36002, doi: 10.1117/1.JRS.11.036002.

[15] X. Ma, J. Wang, F. Yu, H. Jia, and Y. Hu, "Can MODIS AOD be employed to derive PM2.5 in Beijing-Tianjin-Hebei over China?," *Atmos. Res.*, vol. 181, pp. 250–256, 2016, doi: 10.1016/j.atmosres.2016.06.018.

[16] Q. Xu, X. Chen, S. Yang, L. Tang, and J. Dong, "Spatiotemporal relationship between Himawari-8 hourly columnar aerosol optical depth (AOD) and ground-level $PM_{2.5}$ mass concentration in mainland China," *Sci. Total Environ.*, vol. 765, 2021, Art. no. 144241, doi: 10.1016/j.scitotenv.2020.144241.

[17] Q. Yang, Q. Yuan, L. Yue, T. Li, H. Shen, and L. Zhang, "The relationships between PM2.5 and aerosol optical depth (AOD) in mainland China: About and behind the spatio-temporal variations," *Environ. Pollut.*, vol. 248, pp. 526–535, 2019, doi: 10.1016/j.envpol.2019.02.071.

[18] Q. Xiao et al., "Full-coverage high-resolution daily $PM_{2.5}$ estimation using MAIAC AOD in the Yangtze river delta of China," *Remote Sens. Environ.*, vol. 199, pp. 437–446, 2017, doi: 10.1016/j.rse.2017.07.023.

[19] Y. Liu, J. A. Sarnat, V. Kilaru, D. J. Jacob, and P. Koutrakis, "Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing," *Environ. Sci. Technol.*, vol. 39, no. 9, pp. 3269–3278, 2005, doi: 10.1021/es049352m.

[20] M. Green, S. Kondragunta, P. Ciren, and C. Xu, "Comparison of GOES and MODIS aerosol optical depth (AOD) to aerosol robotic network (AERONET) AOD and IMPROVE PM2.5 mass at Bondville, Illinois," *J. Air Waste Manage. Assoc.*, vol. 59, no. 9, pp. 1082–1091, 2009, doi: 10.3155/1047-3289.59.9.1082.

[21] N. Kumar, D. Liang, A. Comellas, A. D. Chu, and T. Abrams, "Satellite-based PM concentrations and their application to COPD in cleveland, OH," *J. Exposure Sci. Environ. Epidemiol.*, vol. 23, no. 6, pp. 637–646, 2013, doi: 10.1038/jes.2013.52.

[22] L. Jean-Francois, L. Cathy, G. - L. Corinne, D. Thierno, and C. Hélène, "Monitoring of ambient fine particulate matter concentrations from space: Application to European and African cities," *Proc. SPIE*, vol. 7826, 2010, Art. no. 78262A, doi: 10.1117/12.864954.

[23] M. Schaap, A. Apituley, R. M. A. Timmermans, R. B. A. Koelemeijer, and G. de Leeuw, "Exploring the relation between aerosol optical depth and $PM_{2.5}$ at Cabauw, The Netherlands," *Atmos. Chem. Phys.*, vol. 9, no. 3, pp. 909–925, 2009, doi: 10.5194/acp-9-909-2009.

[24] J. Tao et al., "A method to estimate concentrations of surface-level particulate matter using satellite-based aerosol optical thickness," *Sci. China Earth Sci.*, vol. 56, no. 8, pp. 1422–1433, 2013, doi: 10.1007/s11430-012-4503-3.

[25] G. Casella, *Statistical Design*. New York, NY, USA: Springer, 2008, doi: 10.1007/978-0-387-75965-4.

[26] P. Tongkumchum and D. McNeil, "Confidence intervals using contrasts for regression model," *Songklanakarin J. Sci. Technol.*, vol. 31, no. 2, pp. 151–156, 2009.

[27] S. Abdulmana, A. Lim, S. Wongsai, and N. Wongsai, "Effect of land cover change and elevation on decadal trend of land surface temperature: A linear model with sum contrast analysis," *Theor. Appl. Climatol.*, vol. 149, no. 1, pp. 425–436, 2022, doi: 10.1007/s00704-022-04038-z.

[28] G. Af, "Blood lead levels among school children living in the pattani river Basin: Two contamination scenarios?," *J. Environ. Med.*, vol. 2, pp. 11–16, 2000.

[29] H. D. Tulu, A. Lim, A. Ma-A-Lee, K. Bundhamcharoen, and N. Makka, "Prediction of HIV mortality in Thailand using three data sets from the national AIDS program database," *Sains Malaysiana*, vol. 49, no. 1, pp. 155–160, 2020, doi: 10.17576/jsm-2020-4901-19.

[30] S. Buya, P. Tongkumchum, and B. E. Owusu, "Modelling of land-use change in Thailand using binary logistic regression and multinomial logistic regression," *Arabian J. Geosciences*, vol. 13, no. 12, 2020, Art. no. 437. [Online]. Available: https://library.tu.ac.th/eds/detail/edssjs_edssjs.76F9F511

[31] S. Buya, P. Tongkumchum, K. Rittiboon, and S. Chaimontree, "Logistic regression model of built-up land based on grid-digitized data structure: A case study of Krabi, Thailand," *J. Indian Soc. Remote Sens.*, vol. 50, no. 5, 2022, Art. no. 909. [Online]. Available: https://library.tu.ac.th/eds/detail/edssjs_edssjs.E501BA54

[32] P. Chuangchang, O. Thinnukool, and P. Tongkumchum, "Modelling urban growth over time using grid-digitized method with variance inflation factors applied to spatial correlation," *Arabian J. Geosciences*, vol. 9, no. 5, 2016, Art. no. 342, doi: 10.1007/s12517-016-2375-0.

[33] S. Buya, P. Chuangchang, and B. A. Owusu, "Analysis of land surface temperature with land use and land cover and elevation from NASA MODIS satellite data: A case study of Bali, Indonesia," *Environ. Monit. Assess.*, vol. 194, no. 8, 2022, Art. no. 566, doi: 10.1007/s10661-022-10252-z.

[34] L. Zhang, J. P. Wilson, B. MacDonald, W. Zhang, and T. Yu, "The changing $PM_{2.5}$ dynamics of global megacities based on long-term remotely sensed observations," *Environ. Int.*, vol. 142, 2020, Art. no. 105862, doi: 10.1016/j.envint.2020.105862.

[35] T. Zhang, L. Zang, Y. Wan, W. Wang, and Y. Zhang, "Ground-level $PM_{2.5}$ estimation over urban agglomerations in China with high spatiotemporal resolution based on Himawari-8," *Sci. Total Environ.*, vol. 676, pp. 535–544, 2019, doi: 10.1016/j.scitotenv.2019.04.299.

[36] B. Peng-in, P. Sanitluea, P. Monjatturat, P. Boonkerd, and A. Phosri, "Estimating ground-level PM2.5 over bangkok metropolitan region in Thailand using aerosol optical depth retrieved by MODIS," *Air Qual., Atmos. Health*, vol. 15, no. 11, pp. 2091–2102, 2022, doi: 10.1007/s11869-022-01238-4.

[37] P. Wongnakae, P. Chitchum, R. Sripramong, and A. Phosri, "Application of satellite remote sensing data and random forest approach to estimate ground-level $PM_{2.5}$ concentration in northern region of Thailand," *Environ. Sci. Pollut. Res.*, vol. 30, pp. 88905–88917, 2023, doi: 10.1007/s11356-023-28698-0.

[38] B. Suhaimee, U. Sasiporn, G. Hideomi, and K. Jessada, "An estimation of daily $PM_{2.5}$ concentration in Thailand using satellite data at 1-kilometer resolution," *Sustainability*, vol. 15, no. 10024, 2023, Art. no. 10024. [Online]. Available: https://library.tu.ac.th/eds/detail/edsdoj_edsdoj.03d29fa989304607bcc040fc9c0da03a

[39] S. Buya, S. Usanavasin, G. Hideomi, and J. Karnjana, "The association of satellite data with $PM_{2.5}$ data from ground monitoring stations in Thailand," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 1533–1536, doi: 10.1109/IGARSS52108.2023.10282670.

[40] Z. - Y. Chen, T. - H. Zhang, R. Zhang, Z. - M. Zhu, C. - Q. Ou, and Y. Guo, "Estimating $PM_{2.5}$ concentrations based on non-linear exposure-lag-response associations with aerosol optical depth and meteorological measures," *Atmos. Environ.*, vol. 173, pp. 30–37, 2018, doi: 10.1016/j.atmosenv.2017.10.055.

[41] Y. Liu, M. Franklin, R. Kahn, and P. Koutrakis, "Using aerosol optical thickness to predict ground-level $PM_{2.5}$ concentrations in the St. Louis area: A comparison between MISR and MODIS," *Remote Sens. Environ.*, vol. 107, no. 1, pp. 33–44, 2007, doi: 10.1016/j.rse.2006.05.022.

[42] J. Tian and D. Chen, "A semi-empirical model for predicting hourly ground-level fine particulate matter ($PM_{2.5}$) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements," *Remote Sens. Environ.*, vol. 114, no. 2, pp. 221–229, 2010, doi: 10.1016/j.rse.2009.09.011.

[43] W. Song, H. Jia, J. Huang, and Y. Zhang, "A satellite-based geographically weighted regression model for regional $PM_{2.5}$ estimation over the Pearl River Delta region in China," *Remote Sens. Environ.*, vol. 154, pp. 1–7, 2014, doi: 10.1016/j.rse.2014.08.008.

[44] W. Thongpeth, A. Lim, S. Kraonual, A. Wongpairin, and T. Thongpeth, "Determinants of hospital costs for management of chronic-disease patients in southern Thailand," *J. Health Sci. Med. Res.*, vol. 39, no. 4, pp. 313–320, 2021, doi: 10.31584/jhsmr.2021787.

[45] K. Rani Das, "A brief review of tests for normality," *Amer. J. Theor. Appl. Statist.*, vol. 5, no. 1, 2016, Art. no. 5, doi: 10.11648/j.ajtas.20160501.12.

[46] Y. Xu et al., "Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level $PM_{2.5}$," *Environ. Pollut.*, vol. 242, pp. 1417–1426, 2018, doi: 10.1016/j.envpol.2018.08.029.

[47] J. Wei et al., "Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach," *Remote Sens. Environ.*, vol. 231, 2019, Art. no. 111221, doi: 10.1016/j.rse.2019.111221.

[48] M. D. Yazdi et al., "Predicting fine particulate matter ($PM_{2.5}$) in the greater london area: An ensemble approach using machine learning methods," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 914, doi: 10.3390/rs12060914.

[49] P. Phuengsamran and P. Lalitaporn, "Estimating particulate matter concentrations in central Thailand using satellite data," *Thai Environ. Eng. J.*, vol. 35, no. 3, pp. 1–11, 2021.

[50] C. ChooChuay et al., "Impacts of $PM_{2.5}$ sources on variations in particulate chemical compounds in ambient air of Bangkok, Thailand," *Atmospheric Pollut. Res.*, vol. 11, no. 9, pp. 1657–1667, 2020, doi: 10.1016/j.apr.2020.06.030.

[51] N. Chuersuwan, S. Nimrat, S. Lekphet, and T. Kerdkumrai, "Levels and major sources of $PM_{2.5}$ and PM_{10} in Bangkok metropolitan region," *Environ. Int.*, vol. 34, no. 5, pp. 671–677, 2008, doi: 10.1016/j.envint.2007.12.018.

[52] D. Yang, C. Ye, X. Wang, D. Lu, J. Xu, and H. Yang, "Global distribution and evolvement of urbanization and $PM_{2.5}$ (1998–2015)," *Atmos. Environ.*, vol. 182, pp. 171–178, 2018, doi: 10.1016/j.atmosenv.2018.03.053.

[53] T. Liu et al., "Urban-rural disparity of the short-term association of $PM_{2.5}$ with mortality and its attributable burden," *Innovation*, vol. 2, no. 4, Nov. 2021, Art. no. 100171, doi: 10.1016/j.xinn.2021.100171.

[54] D. Bzdok, N. Altman, and M. Krzywinski, "Statistics versus machine learning," *Nature Methods*, vol. 15, no. 4, pp. 233–234, 2018, doi: 10.1038/nmeth.4642.

Suhaimee Buya (Member, IEEE) received the Bachelor of Science degree in applied mathematics and the Master of Science degree in research methodology from the Prince of Songkla University, Hat Yai, Thailand, in 2012 and 2016. He is currently working toward the double degree in a Doctoral program in the School of Information, Computer, and Communication Technology, Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani, Thailand, and the School of Knowledge Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan.

He also received an Erasmus Mundus Mobility with Asia (EMMA) scholarship, which enabled him to spend two semesters as an exchange student at Lucian Blaga University of Sibiu in Romania. He has demonstrated his expertise in applied statistics and data analytics, focusing on geoinformatics, statistics, remote sensing, and health data. He has authored or coauthored several papers, including 13 in top-ranking journals, with four as the first and corresponding author.

Hideomi Gokon received the Doctor of Engineering degree from the Graduate School of Engineering, Tohoku University, Sendai, Japan, in 2015.

He served as an Assistant Professor with the Institute of Industrial Science, The University of Tokyo, from 2015 to 2019. Since 2019, he has been an Associate Professor with the Japan Advanced Institute of Science and Technology, Nomi, Japan. Among his notable publications, *Detecting Urban Floods with Small and Large Scale Analysis of ALOS-2/PALSAR-2 Data* was featured in Remote Sensing, Volume 15, Issue 2, on January 16, 2023.

Hieu Chi Dam received the B.Sc. degree in physics from the University of Tokyo, Tokyo, Japan, in 1998, and the M.Sc. and Ph.D. degrees in computational materials from the Japan Advanced Institute of Science and Technology, Nomi, Japan, in 2000 and 2003, respectively.

His current research interests include machine learning and data mining techniques with measurement data, propelling the evolution of data-driven analysis methods across various fields.

Sasiporn Usanavasin received the B.Sc. (Hons.) degree in information technology from Sirindhorn International Institute of Technology (SIIT), Thammasat University, Bangkok, Thailand, in 1999, and the M.Eng. and Ph.D. degrees in software engineering from Keio University, Japan, in 2003 and 2006, respectively.

She is an Assistant Professor with the School of Information, Computer, and Communication Technology, SIIT.

Jessada Karnjana received the B.Eng. degree in electronics engineering from the King's Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 1999, the M.Eng. degree in microelectronics from the Asian Institute of Technology, Khlong Nueng, Thailand, in 2006, the Ph.D. degree in information science from the Japan Advanced Institute of Technology, Nomi, Japan, in 2016, and the Ph.D. degree in engineering and technology from the Sirindhorn International Institute of Technology, Thammasat University, Bangkok, in 2017.

He has worked as a Researcher with the National Electronics and Computer Technology Center, Pathumthani, Thailand, since 1999. His research interests include data analysis, signal analysis, signal processing, wireless sensor network, and artificial intelligence.