Journal of Current Science and Technology, July - September 2025 Vol. 15 No. 3, Article 117

ISSN 2630-0656 (Online)

Cite this article: Ruktamatakul, S., Kuntamanee, N., Puengsaychol, P., Ruktametakul, S., & Yimprayoon, P. (2025). Comparison of multiple linear regression and periodic models for estimating $PM_{2.5}$ and PM_{10} concentrations in Bangkok. *Journal of Current Science and Technology*, *15*(3), Article 117. https://doi.org/10.59796/jcst.V15N3.2025.117

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

Comparison of Multiple Linear Regression and Periodic Models for Estimating $PM_{2.5}$ and PM_{10} Concentrations in Bangkok

Sittipong Ruktamatakul[1], Natchaya Kuntamanee[1], Preeyada Puengsaychol[1],
Sarisa Ruktametakul[2], and Pornpis Yimprayoon[1,*]

[1]Department of Computational Science and Digital Technology, Faculty of Liberal Arts and Science,
Kasetsart University, Kamphaeng Saen Campus, Nakhonpathom 73140, Thailand
[2]Triam Udom Suksa School, Bangkok 10330, Thailand

[*]Corresponding author; E-mail: faasppy@ku.ac.th

Received 2 December 2024; Revised 27 January 2025; Accepted 18 March 2025; Published online 15 June 2025

Abstract

This study compares the performance of Multiple Linear Regression (MLR) and Periodic Models in estimating $PM_{2.5}$ and PM_{10} concentrations in Bangkok using a 60-month dataset (2019–2023). Eight independent variables, including air temperature, rainfall, air pressure, wind speed, ozone concentrations, nitrogen dioxide concentrations, the number of vehicles, and the number of factories, were analyzed to determine their influence on $PM_{2.5}$ and PM_{10} levels. Model accuracy was assessed using Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The results revealed that the Periodic Model more accurately predicted $PM_{2.5}$ (MAE = 4.65, MAPE = 12.69), while the MLR model performed better for PM_{10} (MAE = 6.93, MAPE = 10.54). These findings highlight the complementary strengths of the two modeling approaches: Periodic Models effectively capture seasonal trends, while MLR reveals specific influencing factors. These findings provide valuable insights into the strengths and limitations of each model, offering guidance for developing targeted and efficient measures to control $PM_{2.5}$ and PM_{10} levels in Bangkok, ultimately enhancing public health and urban living conditions.

Keywords: *Bangkok; multiple linear regression; periodic model;* PM_{10}*;* $PM_{2.5}$

1. Introduction

Particulate matter ($PM_{2.5}$ and PM_{10}) is a critical air pollutant that raises significant public health concerns due to its severe impact on human well-being. The World Health Organization (WHO) has established stringent guidelines for $PM_{2.5}$ and PM_{10} to mitigate the health risks associated with air pollution (World Health Organization, 2021). In Thailand, the Pollution Control Department has implemented air quality standards to address rising $PM_{2.5}$ and PM_{10} concentrations, particularly in urban areas such as Bangkok. However, despite these measures, PM concentrations in Bangkok frequently exceed the recommended levels, posing serious health threats to the population (Sooktawee et al., 2023).

As one of Asia's largest and most densely populated cities, Bangkok has experienced persistent air quality challenges over the past several decades. The primary sources of $PM_{2.5}$ and PM_{10} pollution include traffic emissions, biomass burning, and industrial activities, which collectively contribute to elevated PM levels (Tesfaldet, & Chanpiwat, 2023). Prolonged exposure to high PM levels has been associated with increased risks of cardiovascular and respiratory diseases, leading to higher hospital admissions and mortality rates, especially among vulnerable populations (Pengjan et al., 2019). Seasonal variations further exacerbate this issue, with higher PM levels typically observed during the dry season (Kanchanasuta et al., 2020). Recent studies have also highlighted the importance of effective communication tools, such as the Air Quality Health Index (AQHI), in conveying health risks of PM exposure in Bangkok and supporting public health interventions (Kanchanasuta et al., 2024).

To address the persistent issue of elevated PM concentrations, statistical modeling techniques have been extensively applied in air pollution research. Among these techniques, Multiple Linear Regression (MLR) models are commonly used to examine the relationship between independent variables and particulate matter concentrations, providing valuable insights into the factors influencing $PM_{2.5}$ and PM_{10} levels (Sirisumpun et al., 2023). Periodic models, on the other hand, are particularly effective in capturing seasonal and temporal patterns in PM concentrations, making them well-suited for time-series analysis. However, both models have limitations: MLR assumes linear relationships between variables, whereas periodic models lack causal explanations for observed patterns. Recent studies have increasingly highlighted the potential of advanced modeling techniques in air pollution forecasting. For instance, video-based spatiotemporal models have demonstrated high accuracy in weekly $PM_{2.5}$ predictions (Minsan et al., 2024; Pranonsatit et al., 2025). Furthermore, the application of hybrid deep learning architectures has been shown to enhance predictive precision in visual estimation tasks (Laohakiat et al., 2024). In addition, real-time monitoring and forecasting of $PM_{2.5}$ concentrations can be effectively implemented through AIoT-based systems, offering timely insights for public health interventions (An, 2025).

Although MLR and periodic models have been used in previous studies to estimate PM concentrations, limited research has directly compared their performance in urban contexts such as Bangkok. This study aims to address this gap by evaluating the accuracy and effectiveness of MLR and Periodic Models in estimating $PM_{2.5}$ and PM_{10} levels in Bangkok. The findings are expected to provide valuable insights into the strengths and limitations of these models, thereby offering guidance for effective air quality management and policy development.

2. Objectives

This study aims to compare the performance of Multiple Linear Regression and periodic models in estimating $PM_{2.5}$ and PM_{10} concentrations in Bangkok. It focuses on analyzing how eight independent variables air temperature, rainfall, air pressure, wind speed, ozone concentration, nitrogen dioxide levels, number of vehicles, and number of factories, collectively influence $PM_{2.5}$ and PM_{10} levels.

3. Materials and Methods

3.1 Data Collection

Data on $PM_{2.5}$, PM_{10}, ozone concentrations, and nitrogen dioxide concentrations were collected from the Pollution Control Department (2024); air temperature and rainfall from the Climate Center, Thai Meteorological Department (2024); air pressure and wind speed from Meteostat (2024); the number of vehicles from the Transport Statistics Group, Planning Division, Department of Land Transport (2024); and the number of factories from the Department of Industrial Works (2024). The dataset, covering 60 months from January 2019 to December 2023, was compiled by aligning data from all sources using consistent timestamps to ensure accuracy and uniformity. Additionally, all units were standardized to maintain consistency across variables. The collected data, originally recorded as daily averages, were aggregated into monthly averages to reduce variability and better reflect long-term trends.

3.2 Multiple Linear Regression Analysis (MLR)

According to Annette (1990), if a dependent variable Y depends on multiple independent variables $X_1,X_2,\dots,X_k$, the multiple linear regression equation is expressed as

$$\widehat{Y}=a+b_1X_1+b_2X_2+b_3X_3+\dots+b_kX_k.$$

Here, $\widehat{Y}$ represents the mean value of Y at a given point defined by $X_1,X_2,\dots,X_k$. This equation implies that the coefficient b_1 represents the expected change in Y when X_1 increases by one unit, assuming that $X_2,X_3,\dots,X_k$ remain constant. Similarly, the coefficients $b_2,b_3,\dots,b_k$ represent the partial regression coefficients of Y with respect to each corresponding variable.

The least squares method (Wang, & Liu, 2019) is used to calculate the constants $a,b_1,b_2,\dots,b_k$ by minimizing the sum of squared differences between the observed values (Y_i) and the predicted values ($\widehat{Y}_i$). The parameters b are determined using the equation

$$b=\left(X^TX\right)^{-1}\left(X^TY\right),$$

where $X=\begin{bmatrix} 1 & X_{11} & \cdots & X_{k1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & \cdots & X_{kn} \end{bmatrix}$ is the design matrix and Y is the vector of observed values.

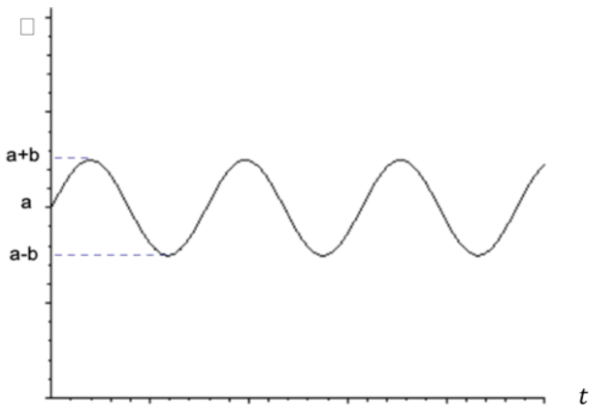

Figure 1 Graph of Sine Function with the Equation $Y=a+b\sin(\Omega(t-t_0))$

After obtaining the initial regression model, each variable is assessed based on its *p*-value to evaluate its significance. Variables with *p*-values exceeding the specified significance level are considered non-significant and are sequentially removed from the model. This iterative process continues until only variables with statistically significant effects on the dependent variable remain in the final model.

3.3 Periodic Models

Lippman, & Rasmussen (2022) explained that periodic models are mathematical models designed to capture recurring patterns over regular time intervals. These models are particularly effective for analysing data that exhibits cyclical trends. A sine function, commonly used in periodic models, represents such recurring patterns, as illustrated in Figure 1.

In Figure 1, the sine function is described by the equation $Y=a+b\sin(\Omega(t-t_0))$. The parameters of the equation are defined as follows:

a : The mean value of Y, representing the midpoint between the maximum and minimum values.

b : The amplitude of the graph, defined as half the vertical distance between the maximum and minimum points.

Ω : The angular frequency, measured in radians per second, which determines the frequency of the sine wave. The period of the wave is the time interval between two consecutive peaks and is given by $T=\frac{2\pi}{\Omega}$.

t_0 : The phase shift, derived from real-world data, which adjusts the starting point of the function to align with observed data.

Alternatively, the equation can be expressed as: $Y=a+b\sin(\Omega t+\phi)$ where $\phi=-\Omega t_0$ represents the phase angle. The frequency, which measures the number of cycles per unit time, is inversely proportional to the period.

This model is well-suited for analysing seasonal and temporal variations in $PM_{2.5}$ and PM_{10} concentrations, providing insights into cyclical patterns and their effects on air quality.

3.4 Comparison of Forecasting Method Performance

Botchkarev (2019) emphasized that the performance of forecasting methods can be effectively evaluated using mean absolute error (MAE) and mean absolute percentage error (MAPE), which are widely adopted due to their simplicity and capacity to quantify prediction accuracy. Models with lower MAE and MAPE values indicate higher accuracy in capturing the observed data. The formulas for calculating MAE and MAPE are as follows:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|e_t|$$

and

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{e_t}{Y_t}\right|$$

where $e_t = Y_t-\hat{Y}_t$, Y_t is the observed value at time t, $\hat{Y}$ is the predicted value at time t, and n represents the total number of observations.

MAE and MAPE have distinct advantages and limitations. MAE provides an absolute measure of average prediction error in the same units as the dependent variable, making it straightforward to interpret. MAPE expresses errors as a percentage of observed values, allowing for standardized comparisons across datasets with different scales. However, MAE does not account for the magnitude of values, which may underestimate errors in datasets with high variability. MAPE, however, can be disproportionately affected by very small observed values, resulting in inflated percentage errors. Despite these limitations, MAE and MAPE remain practical and commonly used metrics for assessing model performance.

4. Results

In this study, at time i, the variables were defined as follows:

X_{1i} represented air temperature (°C),
X_{2i} represented rainfall (mm),
X_{3i} represented air pressure (mbar),
X_{4i} represented wind speed (km/h),
X_{5i} represented ozone concentrations (ppb),
X_{6i} represented nitrogen dioxide concentrations (ppb),
X_{7i} represented the number of vehicles,
X_{8i} represented the number of factories,
Y_{1i} represented $PM_{2.5}$ concentrations (μg/m³),
Y_{2i} represented PM_{10} concentrations (μg/m³),
$\hat{Y}_{1im}$ represented the predicted $PM_{2.5}$ concentration using the multiple linear regression model,
$\hat{Y}_{2im}$ represented the predicted PM_{10} concentration using the multiple linear regression model,
$\hat{Y}_{1ip}$ represented the predicted $PM_{2.5}$ concentration using the periodic model, and
$\hat{Y}_{2ip}$ represented the predicted PM_{10} concentration using the periodic model.

4.1 Multiple Linear Regression

The general form of the multiple linear regression (MLR) model is given by:

$$\hat{Y}_i = b_0+b_1X_{1i}+b_2X_{2i}+b_3X_{3i}+b_4X_{4i}+b_5X_{5i}+b_6X_{6i}+b_7X_{7i}+b_8X_{8i}.$$

The final MLR model for $PM_{2.5}$ was derived by initially considering all eight independent variables (X_{1i}, X_{2i}, X_{3i},…, X_{8i}). Using the least squares method, the initial model coefficients were calculated, resulting in the equation:

$$\hat{Y}_{1i} = -1322.4313-0.0835X_{1i}-0.0126X_{2i}+1.2403X_{3i}-2.9579X_{4i}+0.6950X_{5i}+0.1946X_{6i}+0.0002X_{7i}+0.0049X_{8i}$$

To refine the model, each variable's significance was assessed using p-values, with a threshold of $\alpha=0.05$. Variables with p-values exceeding this threshold were removed step-by-step, starting with the least significant variable. In Step 1, air temperature (X_{1i}) was excluded due to $p_{1i}=0.9426$. In Step 2, the number of factories (X_{8i}) was removed with $p_{8i}=0.8214$. In Step 3, nitrogen dioxide concentrations were excluded due to $p_{6i}=0.2439$, and in Step 4, rainfall (X_{2i}) was removed with $p_{2i}=0.2084$. After these iterations, only statistically significant variables ($p < 0.05$) were retained in the final model, as shown in Table 1, resulting in the following equation:

$$\hat{Y}_{1i}=-1973.0527+1.9629X_{3i}-3.2490X_{4i}+0.7432X_{5i}+0.0002X_{7i}.$$

The retained variables (air pressure (p_{3i}= 0.0004), wind speed (p_{4i}= 0.0492), ozone concentrations (p_{5i}= 0.0000), and the number of vehicles (p_{7i}= 0.0057)) demonstrated statistical significance in predicting $PM_{2.5}$ concentrations. This model indicates that the $PM_{2.5}$ concentrations are significantly influenced by X_{3i}, X_{4i}, X_{5i} and X_{7i}.

Similarly, the final MLR model for PM_{10} was developed by initially including all eight independent variables (X_{1i}, X_{2i}, X_{3i},…, X_{8i}). Using the least squares method, the initial model coefficients were computed, resulting in the equation:

$$\hat{Y}_{2i}= -2715.8217-0.4720X_{1i}-0.0251X_{2i}+2.6375X_{3i}-3.9840X_{4i}+0.4017X_{5i}+0.4798X_{6i}+0.0002X_{7i}+0.0069X_{8i}$$

Each variable's statistical significance was then assessed using p-values, with a threshold of $\alpha = 0.05$. Variables with p-values exceeding this threshold were removed in a stepwise manner, starting with the least significant variable. In Step 1, the number of factories (X_{8i}) was excluded due to P_{8i}=0.8259. In Step 2, air temperature (X_{1i}) was removed with P_{1i}=0.7399. In Step 3, the number of vehicles (X_{7i}) was excluded due to P_{7i}=0.1332. In Step 4, wind speed (X_{4i}) was removed with P_{4i}=0.2211, followed by ozone concentrations (X_{5i}) in Step 5, with P_{5i}=0.2573. Finally, in Step 6, rainfall (X_{2i}) was excluded with $P_{2i} = 0.0160$. After these iterations, only statistically significant variables ($p < 0.05$) were retained in the final model, as shown in Table 2, resulting in the following equation:

$$\hat{Y}_{2i} = -3162.5008-0.0262X_{2i}+3.1898X_{3i}+0.6986X_{6i}.$$

The retained variables (rainfall (p_{2i}= 0.0160), air pressure (p_{3i}= 0.0005), and nitrogen dioxide concentrations (p_{6i}= 0.0007) demonstrated statistical significance in predicting PM_{10} concentrations. This model highlights that PM_{10} concentrations are significantly influenced by X_{2i}, X_{3i} and X_{6i}. The predicted values based on this final model are presented in Table 3.

Table 1 Stepwise Regression Results for $PM_{2.5}$ Using Multiple Linear Regression Method

Models	Maximum *p*-value
$\hat{Y}_{1i}$=-1322.4313-0.0835X_{1i}-0.0126X_{2i}+1.2403X_{3i}-2.9579X_{4i}+0.6950X_{5i}+0.1946X_{6i}+0.0002X_{7i}+0.0049X_{8i} (p_{1i}= 0.9426, p_{2i}= 0.2133, p_{3i}= 0.1118, p_{4i}= 0.1153, p_{5i}= 0.0008, p_{6i}= 0.2814, p_{7i}= 0.0219, p_{8i}= 0.8318)	p_{1i}=0.9426
$\hat{Y}_{1i}$=-1355.8255-0.0123X_{2i}+1.2673X_{3i}-2.9841X_{4i}+0.6083X_{5i}+0.1985X_{6i}+0.0002X_{7i}+0.0051X_{8i} (p_{2i}= 0.1626, p_{3i}= 0.0612, p_{4i}= 0.1021, p_{5i}= 0.0007, p_{6i}= 0.2453, p_{7i}= 0.0205, p_{8i}= 0.8214)	p_{8i}=0.8214
$\hat{Y}_{1i}$=-1297.3994-0.0123X_{2i}+1.2951X_{3i}-3.0485X_{4i}+0.6092X_{5i}+0.1854X_{6i}+0.0002X_{7i} (p_{2i}= 0.1583, p_{3i}= 0.0498, p_{4i}= 0.0881, p_{5i}= 0.0006, p_{6i}= 0.2439, p_{7i}= 0.0114)	p_{6i}=0.2439
$\hat{Y}_{1i}$=-1645.2975-0.0109X_{2i}+1.6418X_{3i}-3.7551X_{4i}+0.6886X_{5i}+0.0002X_{7i} (p_{2i}= 0.2084, p_{3i}= 0.0062, p_{4i}= 0.0273, p_{5i}= 0.0000, p_{7i}= 0.0050)	p_{2i}=0.2084
$\hat{Y}_{1i}$=-1973.0527+1.9629X_{3i}+3.2490X_{4i}+0.7432X_{5i}+0.0002X_{7i} (p_{3i}= 0.0004, p_{4i}= 0.0492, p_{5i}= 0.0000p_{7i}= 0.0057,)	p_{4i}=0.0492*

Note: * significant at $p<0.05$

Table 2 Stepwise Regression Results for PM_{10} Using Multiple Linear Regression Method

Models	Maximum *p*-value
$\hat{Y}_{2i}$=-2715.8217-0.4720X_{1i}-0.0251X_{2i}+2.6375X_{3i}-3.9840X_{4i} +0.4017X_{5i}+0.4798X_{6i}+0.0002X_{7i}+0.0069X_{8i} (p_{1i}= 0. 7659, p_{2i}= 0. 0725, p_{3i}= 0. 0149, p_{4i}= 0.1200, p_{5i}= 0.0909, p_{6i}= 0.0547, p_{7i}= 0.1803, p_{8i}= 0.8259)	p_{8i}=0.8259
$\hat{Y}_{2i}$=-2620.0072-0.5170X_{1i}-0.0253X_{2i}+2.6599X_{3i}-4.0558X_{4i}+0.4036X_{5i}+0.4603X_{6i}+0.0002X_{7i} (p_{1i}= 0. 7399, p_{2i}= 0. 0669, p_{3i}= 0.0128, p_{4i}= 0.1074, p_{5i}= 0.0862, p_{6i}= 0.0462, p_{7i}= 0.1324)	p_{1i}=0.7399
$\hat{Y}_{2i}$=-2811.5904-0.0231X_{2i}+2.8347X_{3i}-4.2343X_{4i}+0.3965X_{5i}+0.4808X_{6i}+0.0002X_{7i} (p_{2i}= 0. 0541, p_{3i}= 0. 0022, p_{4i}= 0.0831, p_{5i}= 0.0877, p_{6i}= 0.0295, p_{7i}= 0.1332)	p_{7i}=0.1332
$\hat{Y}_{2i}$=-2808.2067-0.0232X_{2i}+2.8395X_{3i}-2.7420X_{4i}+0.3574X_{5i}+0.5498X_{6i} (p_{2i}= 0. 0564, p_{3i}= 0. 0024, p_{4i}= 0.2211, p_{5i}= 0.1248, p_{6i}= 0.0125)	p_{4i}=0.2211
$\hat{Y}_{2i}$=-3083.1920-0.0203X_{2i}+3.1064X_{3i}+0.2400X_{5i}+0.6253X_{6i} (p_{2i}= 0. 0883, p_{3i}= 0. 0007, p_{5i}= 0.2573, p_{6i}= 0.0035)	p_{5i}=0.2573
$\hat{Y}_{2i}$=-3162.5008-0.0262X_{2i}+3.1898X_{3i}+0.6986X_{6i} (p_{2i}= 0. 0160, p_{3i}= 0. 0005, p_{6i}= 0.0007)	p_{2i}=0.0160*

Note: * significant at $p<0.05$

Table 3 Observed and Predicted Values of $PM_{2.5}$ and PM_{10} using multiple linear regression and periodic models (January 2019 – December 2023)

Month year	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	Y_{1i}	Y_{2i}	$\hat{Y}_{1im}$	$\hat{Y}_{2im}$	$\hat{Y}_{1ip}$	$\hat{Y}_{2ip}$
Jan. 2019	28.00	1.10	1013.00	1.50	22.44	39.30	91828	17146	62.00	102.00	47.39	96.17	49.20	91.00
Feb. 2019	29.50	0.00	1012.30	2.80	21.89	16.56	87966	17142	37.00	61.00	40.54	78.08	50.00	84.80
Mar. 2019	30.20	9.20	1012.30	2.80	20.89	16.44	96702	17124	38.00	69.00	41.72	77.76	41.20	71.00
Apr. 2019	31.70	38.80	1008.50	2.50	19.13	14.33	73750	17108	34.00	69.00	28.87	63.39	38.20	68.60
May. 2019	31.30	89.30	1006.90	2.10	18.67	17.73	91043	17090	35.00	73.00	30.50	59.33	28.40	59.80
Jun. 2019	30.10	151.30	1006.80	1.80	13.33	14.82	82424	17090	27.00	67.00	25.41	55.35	23.40	52.40
Jul. 2019	29.50	146.40	1006.80	1.80	12.00	14.09	82376	17091	29.00	70.00	24.41	54.97	23.20	51.40
Aug. 2019	29.00	97.80	1006.30	2.20	10.71	13.82	86165	17092	28.00	56.00	22.00	54.46	22.80	48.20
Sep. 2019	28.60	361.40	1009.30	1.50	15.43	22.70	76508	17088	32.00	69.00	31.55	63.33	25.20	53.60
Oct. 2019	29.50	156.10	1010.10	1.60	20.14	27.30	81317	17072	34.00	73.00	37.35	74.47	30.60	64.80
Nov. 2019	28.80	6.70	1010.90	1.60	22.43	32.36	72221	17041	34.00	80.00	38.62	84.48	34.40	76.00
Dec. 2019	27.50	0.00	1012.90	1.80	22.50	33.18	52904	17024	44.00	84.00	37.69	91.61	41.00	85.40
Jan. 2020	29.10	22.70	1011.70	1.80	25.71	27.45	85815	17014	52.00	92.00	44.97	83.18	49.20	91.00
Feb. 2020	29.10	37.00	1012.70	2.60	27.57	22.55	82254	17007	50.00	92.00	44.93	82.57	50.00	84.80
Mar. 2020	30.20	5.00	1009.90	3.20	20.71	13.00	89836	17004	32.00	62.00	34.06	67.81	41.20	71.00
Apr. 2020	30.70	42.50	1010.10	2.50	27.75	12.50	56200	17000	30.00	61.00	34.55	67.11	38.20	68.60
May. 2020	31.70	78.90	1007.80	2.70	21.00	12.25	52661	16998	26.00	53.00	23.59	58.65	28.40	59.80
Jun. 2020	30.10	183.90	1007.10	2.10	16.50	13.38	62481	17001	21.00	46.00	22.98	54.45	23.40	52.40
Jul. 2020	29.90	243.40	1006.80	1.80	13.00	17.17	64348	16997	21.00	42.00	21.18	54.58	23.20	51.40
Aug. 2020	29.40	288.20	1006.30	2.10	12.33	14.43	70376	16990	21.00	42.00	20.05	49.90	22.80	48.20

Table 3 Cont.

Month year	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	Y_{1i}	Y_{2i}	$\hat{Y}_{1im}$	$\hat{Y}_{2im}$	$\hat{Y}_{1ip}$	$\hat{Y}_{2ip}$
Sep. 2020	29.40	342.30	1007.20	1.90	18.33	15.00	67948	16992	22.00	44.00	26.39	51.75	25.20	53.60
Oct. 2020	27.30	373.70	1007.60	1.70	17.67	25.20	72402	16993	28.00	59.00	28.32	59.33	30.60	64.80
Nov. 2020	28.70	86.60	1011.20	1.50	21.67	32.75	70028	17005	37.00	72.00	38.48	83.61	34.40	76.00
Dec. 2020	27.50	0.00	1011.50	1.80	25.88	37.67	60486	17010	47.00	88.00	39.13	90.28	41.00	85.40
Jan. 2021	26.40	0.00	1012.00	1.80	33.38	37.00	79012	17021	57.00	106.00	49.76	91.41	49.20	91.00
Feb. 2021	28.60	80.70	1011.10	2.40	34.50	28.80	71153	17016	57.00	115.00	45.15	80.69	50.00	84.80
Mar. 2021	30.00	16.90	1009.70	3.10	27.56	15.17	85774	17028	40.00	77.00	38.19	68.37	41.20	71.00
Apr. 2021	29.80	191.90	1008.90	2.40	21.78	16.67	64042	17051	32.00	63.00	29.81	62.28	38.20	68.60
May. 2021	30.60	175.70	1007.40	2.30	24.22	13.50	75618	17071	24.00	53.00	31.55	55.71	28.40	59.80
Jun. 2021	30.40	94.50	1007.30	2.50	16.78	10.33	73741	17079	23.00	52.00	24.76	55.31	23.40	52.40
Jul. 2021	29.40	209.90	1006.60	2.10	13.56	9.67	42067	17097	21.00	45.00	15.32	49.58	23.20	51.40
Aug. 2021	29.50	262.20	1007.60	1.80	14.00	9.33	60584	17106	20.00	44.00	22.66	51.17	22.80	48.20
Sep. 2021	28.30	346.10	1007.70	1.60	13.50	12.50	63749	17122	24.00	50.00	23.83	51.50	25.20	53.60
Oct. 2021	28.50	270.40	1008.70	1.50	19.25	14.80	57992	17121	27.00	55.00	29.13	58.28	30.60	64.80
Nov. 2021	28.60	89.00	1010.20	1.60	23.50	20.00	72217	17121	31.00	69.00	38.04	71.45	34.40	76.00
Dec. 2021	27.40	0.00	1013.10	2.00	31.63	25.60	63794	17121	41.00	87.00	46.62	86.95	41.00	85.40
Jan. 2022	28.70	34.40	1011.50	2.30	26.86	22.75	73663	17116	37.00	81.00	41.13	78.95	49.20	91.00
Feb. 2022	28.50	115.30	1010.50	2.60	22.43	18.75	79063	17114	39.00	78.00	36.09	70.85	50.00	84.80
Mar. 2022	30.20	74.80	1008.50	3.40	23.67	12.50	93742	17107	34.00	67.00	33.72	61.16	41.20	71.00
Apr. 2022	30.60	20.50	1008.40	3.00	33.80	16.00	65636	17101	39.00	73.00	36.16	64.71	38.20	68.60
May. 2022	29.50	182.80	1006.50	3.50	21.20	14.00	85993	17097	25.00	54.00	25.93	53.00	28.40	59.80
Jun. 2022	30.10	231.30	1006.90	4.40	16.40	11.50	85201	17085	23.00	51.00	20.05	51.26	23.40	52.40
Jul. 2022	29.30	403.40	1005.60	2.00	13.00	10.50	63348	17077	21.00	48.00	17.95	41.90	23.20	51.40
Aug. 2022	28.60	304.70	1006.40	2.00	11.50	11.00	81308	17083	23.00	50.00	22.37	47.39	22.80	48.20
Sep. 2022	27.80	687.00	1007.30	1.80	12.25	13.00	88358	17080	26.00	54.00	26.89	41.63	25.20	53.60
Oct. 2022	28.20	372.40	1009.80	1.90	24.00	20.00	68966	17080	33.00	67.00	35.93	62.75	30.60	64.80
Nov. 2022	28.70	117.90	1009.70	1.50	27.00	22.00	75474	17076	36.00	71.00	40.70	70.50	34.40	76.00
Dec. 2022	27.30	18.30	1011.50	2.10	34.00	24.00	67940	17074	35.00	75.00	45.83	80.25	41.00	85.40
Jan. 2023	27.20	0.00	1012.20	2.30	39.00	32.00	82002	17070	38.00	74.00	53.37	88.55	49.20	91.00
Feb. 2023	28.60	45.30	1011.30	3.00	41.00	33.00	91394	17070	67.00	78.00	52.88	85.19	50.00	84.80
Mar. 2023	29.80	4.70	1010.70	3.20	41.00	30.00	98272	17067	62.00	80.00	52.57	82.25	41.20	71.00
Apr. 2023	31.50	25.40	1007.50	2.70	40.00	21.00	66128	17067	56.00	77.00	40.09	65.21	38.20	68.60
May. 2023	32.00	73.80	1007.30	2.80	37.00	23.00	92640	17067	32.00	66.00	42.98	64.70	28.40	59.80
Jun. 2023	30.70	97.40	1006.60	3.30	28.00	18.00	92260	17067	23.00	46.00	33.21	58.35	23.40	52.40
Jul. 2023	30.20	223.20	1006.80	3.70	20.00	19.00	72881	17065	24.00	52.00	22.09	56.39	23.20	51.40
Aug. 2023	29.90	198.00	1007.20	3.40	18.00	18.00	85659	17050	22.00	49.00	25.18	57.63	22.80	48.20
Sep. 2023	29.00	380.10	1007.20	2.70	19.00	26.00	88358	17042	22.00	51.00	28.79	58.44	25.20	53.60
Oct. 2023	29.00	258.00	1010.30	2.30	24.00	36.00	68966	17039	31.00	70.00	35.62	78.52	30.60	64.80
Nov. 2023	28.70	92.90	1011.70	2.50	30.00	43.00	75474	17037	34.00	88.00	43.61	92.20	34.40	76.00
Dec. 2023	28.90	11.90	1012.60	2.90	34.00	24.00	67940	17038	38.00	93.00	45.39	83.93	41.00	85.40

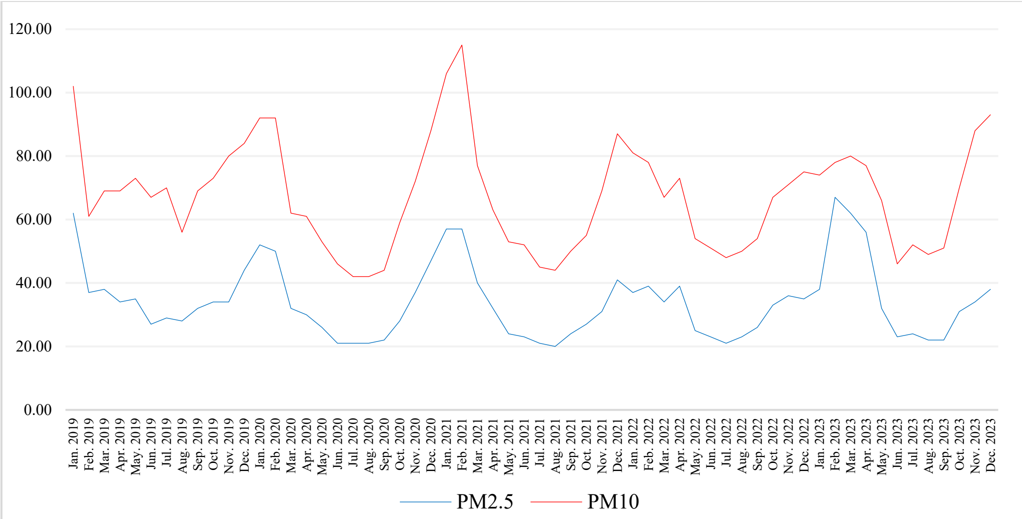

Figure 2 Monthly average concentrations of $PM_{2.5}$ and PM_{10} in Bangkok from January 2019 to December 2023, demonstrating seasonal and cyclical variation

Table 4 Monthly Average Data for $PM_{2.5}$ and PM_{10} in Bangkok (aggregated over 2019-2023)

Month (t)	**Monthly average**	
	$PM_{2.5}$ ($\bar{Y}$)	**PM_{10} ($\bar{Y}$)**
January	49.2	91.0
February	50.0	84.8
March	41.2	71.0
April	38.2	68.6
May	28.4	59.8
June	23.4	52.4
July	23.2	51.4
August	22.8	48.2
September	25.2	53.6
October	30.6	64.8
November	34.4	76.0
December	41.0	85.4

Table 5 Forecasting performance (MAE and MAPE) of multiple linear regression and periodic models for $PM_{2.5}$ and PM_{10} in Bangkok

PM	**Models**	**MAE**	**MAPE**
$PM_{2.5}$	MLR	4.98	14.17
	Periodic	4.65	12.69
PM_{10}	MLR	6.93	10.54
	Periodic	7.18	10.70

Figure 2 shows that the graphs of particulate matter exhibit a periodic pattern, supporting the application of a periodic model described by the equation:

$$Y=a+b\sin(\Omega(t-t_0)).$$

Using 60 months of $PM_{2.5}$ and PM_{10} data, monthly averages were calculatedto obtain $\bar{Y}$, as shown in Table 4. The parameters a and b were calculated as follows:

$$a=\frac{\bar{Y}_{max}-\bar{Y}_{min}}{2}$$

and

$$b=\bar{Y}_{max}-a$$

from $T=\frac{2\pi}{\Omega}$, where T is the period (12 months), Ω was calculated using the equation: $\Omega=\frac{2\pi}{T}=\frac{2\pi}{12}$. To determine t_0, the equation $t_0=\sin(\Omega(t\text{-}t_0))$ was used. The value of $\bar{Y}$ reached its minimum when $\sin(\Omega(t\text{-}t_0))=\sin\left(-\frac{\pi}{2}\right)$. Therefore $\Omega(t\text{-}t_0)=-\frac{\pi}{2}$. Substituting the values of Ω and t, where t corresponded to the month with the minimum $\bar{Y}$ (August), t=8 was set. Substituting these values, the periodic model for $PM_{2.5}$ in Bangkok was defined as: $\widehat{Y}_{1i}= 36.4+13.6\sin((0.52356)(t\text{-}11))$. Similarly, the periodic model for PM_{10} in Bangkok was $\widehat{Y}_{2i} = 69.6+21.4\sin((0.52356)(t\text{-}11))$. The predicted values based on these models are presented in Table 4.

Forecasting performance was evaluated using MAE and MAPE for both $PM_{2.5}$ and PM_{10} models, as shown in Table 5.

The results of the performance comparison of the statistical models influencing $PM_{2.5}$ and PM_{10} are presented in Table 5. The periodic model yielded the most accurate predictions for $PM_{2.5}$ concentrations, with MAE = 4.65 and MAPE = 12.69. In contrast, the multiple linear regression model provided the most accurate predictions for PM_{10} concentrations, with MAE = 6.93 and MAPE = 10.54.

The periodic models used in this study were developed based on the observation that $PM_{2.5}$ and PM_{10} concentrations exhibited recurring pattern over time, as shown in Figure 2. This behaviour, identified from 60 months of monthly averages, justified the use of a sine-based periodic model to capture the cyclical nature of the data. The parameters of the periodic model were derived mathematically from the data, to ensure alignment with observed trends. Although the periodic model does not directly analyze the physical or chemical mechanisms underlying the relationship between variables, it effectively captures seasonal and temporal variations in particulate matter concentrations. This approach is particularly useful for highlighting cyclical behaviours not readily apparent in traditional regression models. As shown in Table 5, the periodic model demonstrated superior accuracy (lower MAE and MAPE) in predicting $PM_{2.5}$ concentrations, emphasizing its utility in forecasting periodic trends. However, for PM_{10} concentrations, the multiple linear regression model was found to be more effective, likely due to the stronger influence of the independent variables on PM_{10} levels compared to its periodicity. This complementary use of periodic and regression models highlights their respective strengths: capturing seasonal trends versus analyzing the effects of specific variables. It is recognized that the periodic model is designed to align with observed cyclical patterns but is limited in its ability to establish causal relationships between independent and dependent variables. Future research could integrate periodic modeling with physical or chemical analyses to explore the mechanisms underlying PM concentrations, offering a more comprehensive understanding of the observed trends.

5. Discussion

The analysis of statistical models for identifying factors influencing $PM_{2.5}$ and PM_{10} concentrations in Bangkok highlights the complexity of urban air pollution. The coefficients of the final multiple linear regression (MLR) models for $PM_{2.5}$ and PM_{10} provide valuable insights into the significance and magnitude of each variable's effect. For $PM_{2.5}$, air pressure (X_{3i} = 1.9629), wind speed (X_{4i} = -3.2490), ozone concentration (X_{5i} = 0.7432), and the number of vehicles (X_{7i} = 0.0002) were identified as the most influential factors. However, the coefficient for the number of vehicles was relatively small, suggesting a limited direct influence compared to other variables such as wind speed, air pressure, and ozone concentration, which demonstrated stronger effects on $PM_{2.5}$ concentrations. This underscores the importance of considering both statistical and practical significance when interpreting model outcomes.

The retained variables for PM_{10} include rainfall, air pressure, and nitrogen dioxide concentration, which demonstrate the multifactorial nature of its behavior. The periodic model was more effective for $PM_{2.5}$ due to its cyclical patterns, whereas MLR was more suitable for PM_{10}, likely reflecting its sensitivity to specific environmental and anthropogenic influences. This difference may be attributed to PM_{10} concentrations are affected more strongly by short-term, localized events rather than consistent seasonal patterns.

The study further emphasizes the importance of supporting findings with evidence from existing literature. Previous research has shown that air pressure and wind speed influence particulate matter levels by affecting atmospheric stability and dispersion. Ozone concentrations contribute to secondary PM formation, while vehicle emissions are a primary source of particulate matter in urban environments. Referencing these mechanisms reinforces the validity of the findings and supports their relevance to real-world air quality management.

Nevertheless, the limitations of both models must be acknowledged. The periodic model effectively captures seasonal trends but cannot establish causal relationships. In contrast, MLR relies on observed data and assumes linear relationships, which may oversimplify complex interactions among variables. Future studies could integrate periodic models with advanced techniques, such as machine learning or hybrid models, to better capture non-linear relationships and enhance predictive accuracy. Additionally, incorporating physical or chemical analyses could provide a more comprehensive understanding of the dynamics influencing particulate matter concentrations.

6. Conclusion

This study developed and evaluated statistical models to investigate factors influencing $PM_{2.5}$ and PM_{10} concentrations in Bangkok. By applying the periodic model to $PM_{2.5}$ and the multiple linear regression model to PM_{10}, the study underscores the importance of tailoring modeling techniques to the specific characteristics of each pollutant. The results revealed that air pressure, wind speed, ozone concentration, and the number of vehicles significantly influence $PM_{2.5}$ concentrations, whereas rainfall, air pressure, and nitrogen dioxide concentration are the primary factors for PM_{10}. These findings highlight the multifactorial nature of urban air pollution, driven by both anthropogenic and natural factors. The insights from this study provide actionable recommendations for policymakers, including promoting cleaner transportation, regulating industrial emissions, and improving urban planning. When combined with adaptive air quality management strategies that consider seasonal and temporal variations, these measures could help reduce pollutant levels and improve public health outcomes in Bangkok. While the models effectively predict pollutant concentrations, their limitations must be recognized. The periodic model captures cyclical trends but lacks the ability to establish causal relationships, whereas MLR assumes linearity, which may oversimplify complex variable interactions. Future research should address these limitations by incorporating advanced techniques such as hybrid models or machine learning and expanding the set of variables to include factors like relative humidity to improve predictive accuracy. Additionally, integrating physical and chemical analyses may offer a more comprehensive understanding of the mechanisms influencing particulate matter concentrations. Building on these findings, future efforts can support the development of more effective and adaptive air quality management strategies in Bangkok.

7. References

An, A. (2025). An AIoT-based air quality monitoring system for real-time PM2.5 prediction in urban environments. ASEAN *Journal of Scientific and Technological Reports*, *28*(1), Article e255168. https://doi.org/10.55164/ajstr.v28i1.255168

Annette, J. D. (1990). *An introduction to generalized linear models* (5th ed.). New South Wales, USA: Chapman & Hall.

Botchkarev, A. (2019). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *Interdisciplinary Journal of Information, Knowledge, and Management, 14*, 45-76. https://doi.org/10.48550/arXiv.1809.03006

Climate Center, Thai Meteorological Department. (2024). *Weather summary (monthly)*. Retrieved from https://www.tmd.go.th/en/climate/summarymonthly

Department of Industrial Works. (2024). Industrial factory statistics. Retrieved from https://www.diw.go.th/webdiw/static-fac/

Kanchanasuta, S., Ingviya, T., Dumavibhat, N., Wongrathanandha, C., Sansanayudh, N., Chalongviriyalert, P., ... & Bunplod, N. (2024). Constructing an AQHI as a health risk communication tool for Bangkok, Thailand. *Environmental Challenges*, *16*, Article 100991. https://doi.org/10.1016/j.envc.2024.100991

Kanchanasuta, S., Sooktawee, S., Patpai, A., & Vatanasomboon, P. (2020). Temporal variations and potential source areas of fine particulate matter in Bangkok, Thailand. *Air, Soil and Water Research*, *13*, Article 1178622120978203. https://doi.org/10.1177/1178622120978203

Laohakiat, S., Klerkkidakan, S., & Wiwatwattana, N. (2024). Visually estimating and forecasting PM2.5 levels using hybrid architecture deep neural network. *Current Applied Science and Technology*, *24*(3), Article e0258074. https://doi.org/10.55003/cast.2023.258074

Lippman, D., & Rasmussen, M. (2022). *Precalculus: An investigation of functions* (2nd ed.). USA: Creative Commons.

Meteostat. (2024). *Weather data*. Retrieved from https://meteostat.net/en/station/48455?t=2019-01-01/2019-01-01

Minsan, W., Minsan, P., & Panichkitkosolkul, W. (2024). Enhancing decomposition and Holt-winters weekly forecasting of PM2.5 concentrations in Thailand's eight northern provinces using the Cuckoo Search algorithm.

Thailand Statistician, *22*(4), 963–985. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/256084

Pengjan, S., Fan, C., Bonnet, S., & Garivait, S. (2019). Assessment of the PM2.5/PM10 ratio in the Bangkok Metropolitan Region during. *Journal of Sustainable Energy & Environment, 10*, 75-84.

Pollution Control Department. (2024). *Regional air quality and situation reports*. Retrieved from http://air4thai.pcd.go.th/webV3/#/History

Pranonsatit, J., Wongwailikhit, K., Painmanakul, P., Vateekul, P. (2025). Enhancing PM2.5 forecasting using video-based spatiotemporal models and cyclical encoding. In Nguyen, N.T., et al. Recent Challenges in Intelligent Information and Database Systems. ACIIDS 2025. *Communications in Computer and Information Science*, *2494*. Springer, Singapore. https://doi.org/10.1007/978-981-96-5884-8_26

Sirisumpun, N., Wongwailikhit, K., Painmanakul, P., & Vateekul, P. (2023). Spatio-Temporal PM2. 5 Forecasting in Thailand Using Encoder-Decoder Networks. *IEEE Access*, *11*, 69601-69613. https://doi.org/10.1109/ACCESS.2023.3293398

Sooktawee, S., Kanchanasuta, S., & Bunplod, N. (2023). Assessment of 24-h moving average PM2. 5 concentrations in Bangkok, Thailand against WHO guidelines. *Sustainable Environment Research*, *33*(1), Article 3. https://doi.org/10.1186/s42834-023-00165-y

Tesfaldet, Y. T., & Chanpiwat, P. (2023). The effects of meteorology and biomass burning on urban air quality: The case of Bangkok. *Urban Climate*, *49*, Article 101441. https://doi.org/10.1016/j.uclim.2023.101441

Transport Statistics Group, Planning Division, Department of Land Transport. (2024). *Number of new registered cars*. Retrieved from https://web.dlt.go.th/statistics/

Wang, X., & Liu, Y. (2019). Research on the application of improved least square method in linear fitting. *IOP Conference Series: Earth and Environmental Science*, 252, 052158. https://doi.org/10.1088/1755-1315/252/5/052158

World Health Organization (WHO). (2021). *WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide, and carbon monoxide*. Retrieved from https://apps.who.int/iris/handle/10665/345329