

Spatiotemporal Cokriging for Modeling and Prediction of PM_{2.5} Concentrations in Jakarta

Fariza Alamanda Putri, I Gede Nyoman Mindra Jaya, Bertho Tantular

Abstract— Monitoring PM_{2.5} concentrations is indeed a crucial thing to be concerned about due to its significant impact on environmental quality and public health measurement. PM_{2.5} in its high levels is considered a hazardous air pollutant that can lead to respiratory and cardiovascular diseases, making reliable estimations of the pollutant essential for policymaking and public awareness. Despite how important it is, the sparse distribution of monitoring stations often leads to data gaps, particularly in urban areas like Jakarta, Indonesia, where air pollution is a serious matter. Traditional interpolation methods such as Inverse Distance Weighting (IDW) and kriging are acknowledged to have limitations in capturing the complex spatiotemporal dynamics of PM_{2.5} concentrations and fail to incorporate secondary variables that influence air pollution. Therefore, this study proposes a spatiotemporal cokriging model to estimate PM_{2.5} concentrations in Jakarta by integrating temporal variations and environmental factors, including rainfall, relative humidity, wind speed, and nitrogen dioxide (NO₂) concentrations. Using data from 29 monitoring stations (January–December 2023), the sum-metric semivariogram model was the best fit for spatiotemporal variability due to its ability to show more precise cokriging weight calculations. The proposed model achieved high predictive accuracy, with a Mean Absolute Percentage Error (MAPE) of 0.659% and a Root Mean Square Error (RMSE) of 0.384. Results revealed seasonal trends, with higher PM_{2.5} levels during the dry season and lower levels during the rainy season. Spatially, the northern and southern Jakarta exhibit elevated concentrations due to industrial activities and vehicular emissions. Thus, these findings underscore the model's effectiveness in improving air pollution predictions and provide a robust framework for policymakers to design targeted mitigation strategies in urban settings.

Index Terms— Air Quality, Jakarta, PM_{2.5}, Spatiotemporal Cokriging

I. INTRODUCTION

AIR pollution is a significant environmental issue faced by densely populated cities, particularly those that serve as centers for economic and industrial activities, such

as Jakarta, the capital of Indonesia. Air pollution in Jakarta largely stems from energy consumption, fossil fuel use, household activities, and industrial processes [1]. Among air pollutants, particulate matter 2.5 (PM_{2.5}) poses a critical threat to human health due to its small size (≤ 2.5 micrometer, abbreviated as μm), which allows it to penetrate deep into the respiratory system [2]. Recognizing the danger, the Government of Indonesia has set the healthy threshold for PM_{2.5} at 50 $\mu\text{g}/\text{m}^3$ (micrograms per cubic meter) [3]. Therefore, reliable information on PM_{2.5} concentrations is essential for policymakers to devise mitigation strategies and for the public to protect themselves from exposure.

Despite its importance, monitoring PM_{2.5} concentrations is often limited by the availability of the monitoring equipment, resulting in data gaps across regions. To address this, interpolation methods are needed to estimate pollutant concentrations at unmonitored locations. The commonly used interpolation methods include Inverse Distance Weighting (IDW) and kriging.

IDW estimates values at unknown points based solely on their distances from known locations, assigning higher weights to closer points. However, IDW does not account for spatial dependencies or underlying variability patterns, and it cannot incorporate additional predictor variables, which limits its ability to provide accurate estimates in complex environmental systems.

Kriging improves upon IDW by utilizing spatial dependencies through modeling a semivariogram, which captures the relationship between variability and distance among observed locations. This results in more accurate predictions than IDW, particularly in regions with significant spatial autocorrelation. However, like IDW, kriging does not allow for including predictor variables and does not explicitly consider temporal dependencies.

Air pollution, however, is influenced by a combination of spatial and temporal factors and secondary variables such as rainfall, relative humidity, wind speed, and NO₂ concentrations. To overcome these limitations, spatiotemporal cokriging is employed as it integrates both spatial and temporal dependencies into the model while allowing for the inclusion of secondary variables as predictors. This approach enables a more comprehensive analysis of pollutant distribution patterns and improves prediction accuracy by incorporating the influence of additional environmental factors alongside PM_{2.5} concentrations.

Given this context, incorporating spatial-temporal

Manuscript received December 2, 2024; revised May 10, 2025.

This work was supported by Universitas Padjadjaran through research grant No: 1893/UN6.3.1/PT.00/2024.

Fariza Alamanda Putri is a graduate of the Master's Program in Applied Statistics at Universitas Padjadjaran, West Java, 45363, Indonesia. (e-mail: fariza20001@mail.unpad.ac.id).

*I Gede Nyoman Mindra Jaya is a lecturer in Department of Statistics Universitas Padjadjaran, West Java, 45363, Indonesia. (e-mail: mindra@unpad.ac.id).

Bertho Tantular is a lecturer in Department of Statistics Universitas Padjadjaran, West Java, 45363, Indonesia. (e-mail: bertho@unpad.ac.id).

dependencies and additional environmental variables into a cokriging model is expected to enhance the accuracy of PM_{2.5} concentrations estimates. This study aims to develop a spatiotemporal cokriging model to provide more precise predictions of PM_{2.5} concentrations in Jakarta. By including rainfall, relative humidity, wind speed, and NO₂ concentrations as secondary variables, the proposed model seeks to capture spatial and temporal variations more effectively, ultimately delivering more accurate estimations that reflect pollution dynamics.

II. DATA

This study was carried out in the Jakarta Province area, which includes the administrative areas of Central, North, West, South, and East Jakarta City (Fig. 1).



Fig. 1. Study Area. The map highlights the administrative boundaries of Jakarta, the focus region of the study, along with its surrounding areas, including Tangerang, Depok, and Bekasi

As the capital of the country, Jakarta is experiencing rapid industrialization and urbanization [1]. This contributes to an increase in the number of motor vehicles, industrial activities, and infrastructure development. The dense activity and mobilization cause an increase in air pollutant emissions. In August 2024, the concentrations of pollutants in Jakarta will be the highest in the world [2]. IQAir also noted that Jakarta is often one of the cities with the worst air quality in the world [3].

TABLE I
THE VARIABLES USED IN SPATIOTEMPORAL MODEL

Variable	Data Source	Description	Unit ^a
PM _{2.5} concentration	PT Nafas Aplikasi Indonesia Jakarta	The concentrations of PM _{2.5} particles in the air	µg/m ³
NO ₂ concentration	Environment Agency (https://satudata.jakarta.go.id/)	The concentrations of nitrogen dioxide in the air	µg/m ³
Rainfall	BMKG (https://dataonline.bmkg.go.id/)	Indicates the total amount of rainfall over a one-month period	mm
Relative Humidity	BMKG (https://dataonline.bmkg.go.id/)	The percentage of water vapor content in the air	%
Wind Speed	BMKG (https://dataonline.bmkg.go.id/)	The speed of air movement	m/s

^aThe standard unit based on the data source

Notes: µg/m³ = micrometers per cubic meter; mm = millimeter; m/s = meters per second

This study uses PM_{2.5} data and secondary variables, including rainfall, relative humidity, wind speed, and NO₂ concentrations (Table I). Rainfall, relative humidity, and wind speed were chosen because several studies [4], [5], [6], [7] prove that these variables affect PM_{2.5} concentration. Meanwhile, the variable NO₂ was chosen as a proxy for population density because it generally comes from human activities such as industry, power generation, and the use of motor vehicles [8], [9].

Data were collected in 12 monthly periods throughout 2023, with a total of 29 observation points in the study area with the distribution of points (Fig. 2). From a total sample of 29 point; the data will be divided into 80% training data (23 points) and test data as much as 20% (6 points).

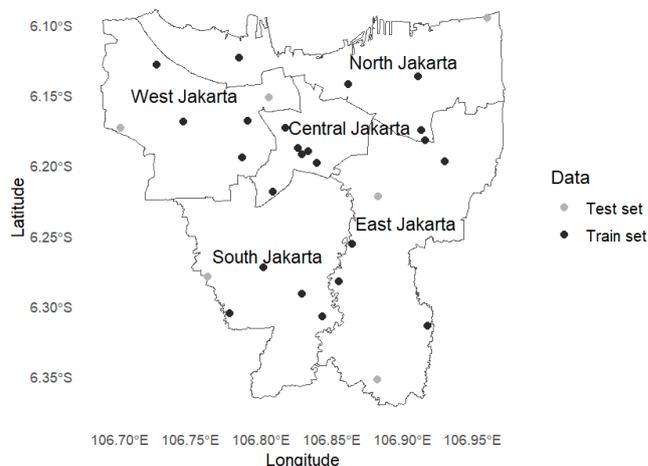


Fig. 2. Sample Point Distribution

III. METHODS

A. Variogram

A variogram is a function that describes dependencies in a regional variable. Variograms are used to show the differences between random variables, while the covariance function is used to show similarities between [10]. Meanwhile, a semivariogram is half the quantity of a variogram. Empirical semivariogram refer to semivariogram calculated from sample data. An empirical semivariogram is calculated by (1).

$$\gamma(\mathbf{h}, u) = \frac{1}{2\#N(\mathbf{h}, u)} \sum_{(s_i, t), (s_j, t') \in N(\mathbf{h}, u)} (Z(s_i, t) - Z(s_j, t'))^2 \quad (1)$$

Refer to (1), $\gamma(\mathbf{h}, u)$ expresses a spatiotemporal semivariogram as well as $N(\mathbf{h}, u) = \{(s_i, t)(s_j, t') : s_i - s_j = \mathbf{h} \text{ and } |t - t'| = u\}$ is a set of pairs in spatial distance \mathbf{h} and temporal lag u and $\#$ is a notation that expresses the cardinality of the set or the number of elements in the set. Next, (s_i, t) and (s_j, t') is a pair of points with spatial distance \mathbf{h} and temporal lag u for $i, j = 1, \dots, n$ and $t, t' = 1, \dots, 12$ because the data consist of 12 monthly periods, where n represents the number of spatial locations in the training data.

Empirical semivariogram cannot be used directly to obtain kriging predictions because they are only defined for a set of lag distances calculated based on sample data. To overcome these limitations, empirical semivariogram need to be fitted to a continuous mathematical model, which is then called a theoretical semivariogram. A theoretical semivariogram is needed as a function that generates the value of the semivariogram in the calculation of kriging prediction. The function of the theoretical semivariogram allows for continuous and consistent depiction of dependency patterns at various distances between observation points. This function also simplifies the random fluctuations or variations that may be encountered in empirical semivariogram. In spatiotemporal analysis, the covariance function model can be categorized as separable and non-separable. The separable model assumes that the spatial and temporal components of the covariance or semivariogram function can be separated and modeled independently [11]. In contrast, the non-separable model assumes that the spatial and temporal components in a variogram model or covariance function must be modeled together. The following is the spatiotemporal covariance model used in this study [12].

1) Product covariance model

This model assumes that the function of spatiotemporal covariance can be represented by the multiplication of the spatial and temporal covariance functions.

$$\gamma_{sep}(\mathbf{h}, u) = C_s(\mathbf{0})\gamma_t(u) + C_t(\mathbf{0})\gamma_s(\mathbf{h}) - \gamma_s(\mathbf{h})\gamma_t(u) \quad (2)$$

Refer to (2), $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ and $u = |t - t'|$ represent distances in space and time, respectively, with \mathbf{s}_i and \mathbf{s}_j being the coordinates of the locations, and t and t' being the observation periods. Consequently, $\gamma_{sep}(\mathbf{h}, u)$ denotes the semivariogram of the product model, $C_s(\mathbf{0})$ and $C_t(\mathbf{0})$ is the sill of the spatial and temporal components whose values are known, $\gamma_s(\mathbf{h})$ and $\gamma_t(u)$ consecutively is the spatial and temporal semivariogram.

2) Product-sum covariance model

This model is defined in (3).

$$\gamma_{ps}(\mathbf{h}, u) = \gamma_s(\mathbf{h})(1 + c.C_t(\mathbf{0})) + \gamma_t(u)(1 + c.C_s(\mathbf{0})) - c.\gamma_s(\mathbf{h})\gamma_t(u) \quad (3)$$

Refer to (3), $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ and $u = |t - t'|$ represent distances in space and time, respectively, with \mathbf{s}_i and \mathbf{s}_j being the coordinates of the locations, and t and t' being the observation periods. Consequently, $\gamma_{ps}(\mathbf{h}, u)$ denotes the semivariogram of the product-sum model, $C_s(\mathbf{0})$ and $C_t(\mathbf{0})$ each of which is sequentially a sill of $\gamma_s(\mathbf{h})$ and $\gamma_t(u)$ which is consecutively a spatial and temporal semivariogram, and $c > 0$ is a constant that guarantees the positive definite properties of the semivariogram.

3) Metric covariance model

The metric model is also referred to as the combined distance model. The hypothesis underlying this model is that the spatiotemporal covariance has a uniform structure in both the spatial domain and the temporal domain.

$$\begin{aligned} \gamma_m(\mathbf{h}, u) &= \gamma\left(\sqrt{\|\mathbf{h}\|^2 + \kappa^2 u^2}\right) \\ &= C(\mathbf{0}, 0) - C\left(\sqrt{\|\mathbf{h}\|^2 + \kappa^2 u^2}\right) \end{aligned} \quad (4)$$

Referring to (4), $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ and $u = |t - t'|$ represent distances in space and time, respectively, with \mathbf{s}_i and \mathbf{s}_j being the coordinates of the locations, t and t' being the observation periods, $\gamma_m(\mathbf{h}, u)$ is the semivariogram of the metric model, and $C(\mathbf{0}, 0)$ is the sill whose value is known.

Then, $\|\mathbf{h}\|^2 = \mathbf{h}^T \mathbf{h}$ represents the squared norm of the spatial distance vector, and κ is the anisotropy correction coefficient used to align the scales of the spatial and temporal dimensions.

4) Sum-metric covariance model

Sum-metric models are formed by summing the spatial covariance function, the temporal covariance function and the metric covariance function.

$$\begin{aligned} \gamma_{sm}(\mathbf{h}, u) &= \gamma_s(\mathbf{h}) + \gamma_t(u) \\ &+ \gamma\left(\sqrt{\|\mathbf{h}\|^2 + \kappa^2 u^2}\right) \end{aligned} \quad (5)$$

Refer to (5), $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ and $u = |t - t'|$ represent distances in space and time, respectively, with \mathbf{s}_i and \mathbf{s}_j being the coordinates of the locations, t and t' being the observation periods, $\gamma_{sm}(\mathbf{h}, u)$ stating the sum-metric model spatiotemporal semivariogram, $\gamma_s(\mathbf{h})$ is a spatial semivariogram, $\gamma_t(u)$ is a temporal semivariogram. Then, $\|\mathbf{h}\|^2 = \mathbf{h}^T \mathbf{h}$ represents the squared norm of the spatial distance vector, and κ is the anisotropy correction coefficient used to align the scales of the spatial and temporal dimensions.

5) Simple sum-metric covariance model

The simple-sum metric model adds a spatiotemporal nugget component so that it will have a semivariogram formulated as follows.

$$\begin{aligned} \gamma_{ssm}(\mathbf{h}, u) &= nug.1_{\|\mathbf{h}\| > 0} + \gamma_s(\mathbf{h}) \\ &+ \gamma_t(u) + \gamma\left(\sqrt{\|\mathbf{h}\|^2 + \kappa^2 u^2}\right) \end{aligned} \quad (6)$$

In (6), $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ and $u = |t - t'|$ represent distances in space and time, respectively, with \mathbf{s}_i and \mathbf{s}_j being the coordinates of the locations, t and t' being the observation periods. Therefore, $nug.1_{\|\mathbf{h}\| > 0}$ means that the nugget

effect will be included in the semivariogram calculation when the spatial distance $\|\mathbf{h}\| = \sqrt{\mathbf{h}^T \mathbf{h}} > 0$, and κ is the anisotropy correction coefficient.

In the theoretical semivariogram models, the functions of the spatial and temporal variograms $\gamma_s(\mathbf{h})$ and $\gamma_t(u)$ are respectively referred as marginal semivariograms. The spatial and temporal marginal semivariogram models do not have to be identical. Marginal semivariograms can be constructed from three frequently used function models, namely spherical, exponential, and Gaussian semivariogram functions.

B. Cross-Variogram

In contrast to a semivariogram, a cross-variogram refers to a variogram that is used to describe the relationship between the main variable and the secondary variable. In this study, $\{\mathbf{Z}(\mathbf{s}_i, t) = Z_1(\mathbf{s}_i, t), \dots, Z_5(\mathbf{s}_i, t)\}$ with $i = 1, \dots, n$ and $t = 1, \dots, 12$, which is the realization of a multivariable spatial process, the cross-variograms can be estimated by (7) [13].

$$\gamma_{ab}(\mathbf{h}, u) = \frac{1}{2\#N(\mathbf{h}, u)} \sum_{(\mathbf{s}_i, t)(\mathbf{s}_j, t') \in N(\mathbf{h}, u)} [Z_a(\mathbf{s}_i, t) - Z_a(\mathbf{s}_j, t')][Z_b(\mathbf{s}_i, t) - Z_b(\mathbf{s}_j, t')] \quad (7)$$

Where $\gamma_{ab}(\mathbf{h}, u)$ is the cross-variogram between the variable Z_a and Z_b where, $a, b = 1, \dots, 5$ and $N(\mathbf{h}, u) = \{(\mathbf{s}_i, t)(\mathbf{s}_j, t') : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h} \text{ and } |t - t'| = u\}$ with $i, j = 1, \dots, n$ and $t, t' = 1, \dots, 12$, is the set of pairs of points in distance \mathbf{h} and time interval u . The # notation expresses the cardinality of the set or the number of elements in the set.

C. Spatiotemporal Cokriging

Once the semivariogram and cross-variograms models are obtained, the next step is to form a spatiotemporal cokriging model. To make it easier to mention, starting from this section onwards, the spatiotemporal location (\mathbf{s}_i, t) will be referred to as x_{it} , so that the unobserved location (\mathbf{s}_0, t') will be referred to as $x_{0t'}$. Furthermore, Z_1, Z_2, Z_3, Z_4 and Z_5 represent the concentrations of PM_{2.5}, NO₂, rainfall, relative humidity, and wind speed, respectively. Among these, Z_1 serves as the primary variable to be predicted, while Z_1, Z_2, Z_3, Z_4 and Z_5 are used as secondary variables.

$$\begin{aligned} Z_1(x_{0t'}) &= \sum_{i=1}^n \sum_{t=1}^{12} \lambda_{i,t,1}^1 Z_1(x_{i,t}) + \dots \\ &\quad + \sum_{i=1}^n \sum_{t=1}^{12} \lambda_{i,t,5}^1 Z_5(x_{i,t}) \quad (8) \\ &= \sum_{m=1}^5 \sum_{i=1}^n \sum_{t=1}^{12} \lambda_{i,t,m}^1 Z_m(x_{i,t}) \end{aligned}$$

The cokriging predictor for $Z_1(x_{0t'})$ is given by (8). By $Z_1(x_{0t'})$ being a prediction of cokriging for the concentrations of PM_{2.5} the location $x_{0t'}$, $\lambda_{i,t,m}^1$ is the cokriging weight for the m -th variable at the spatiotemporal location $x_{i,t}$ by taking into account the influence of the concentrations of PM_{2.5} (Z_1), and $Z_m(x_{i,t})$ is the observation value of the m -th variable at the spatiotemporal location $x_{i,t}$. Equation (8), in its matrix form, is written as (9) [14].

$$Z_1(x_0) = (\lambda_{1,1,1}^1 \dots \lambda_{n,12,1}^1 \dots \lambda_{1,1,5}^1 \dots \lambda_{n,12,5}^1) \begin{pmatrix} Z_1(x_{1,1}) \\ \vdots \\ Z_1(x_{n,12}) \\ \vdots \\ Z_5(x_1) \\ \vdots \\ Z_5(x_{n,12}) \end{pmatrix} \quad (9)$$

The Λ matrix, which contains the cokriging weights $\lambda_{i,t,m}^k$, is obtained by solving (10) [15].

$$\begin{pmatrix} \Gamma & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Lambda \\ \Delta \end{pmatrix} = \begin{pmatrix} \mathbf{G} \\ \mathbf{I} \end{pmatrix} \quad (10)$$

With

$$\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} & \Gamma_{13} & \Gamma_{14} & \Gamma_{15} \\ \Gamma_{21} & \Gamma_{22} & \Gamma_{23} & \Gamma_{24} & \Gamma_{25} \\ \Gamma_{31} & \Gamma_{32} & \Gamma_{33} & \Gamma_{34} & \Gamma_{35} \\ \Gamma_{41} & \Gamma_{42} & \Gamma_{43} & \Gamma_{44} & \Gamma_{45} \\ \Gamma_{51} & \Gamma_{52} & \Gamma_{53} & \Gamma_{54} & \Gamma_{55} \end{pmatrix},$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{1} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & & & \vdots \\ \mathbf{0} & & \mathbf{1} & & \mathbf{0} \\ \vdots & & & \ddots & \vdots \\ \mathbf{1} & \dots & \mathbf{0} & \dots & \mathbf{1} \end{pmatrix}, \mathbf{G} = \begin{pmatrix} \gamma_1^1 & \gamma_1^2 & \dots & \gamma_1^m \\ \gamma_2^1 & \gamma_2^2 & \dots & \gamma_2^m \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_m^1 & \gamma_m^2 & \dots & \gamma_m^m \end{pmatrix},$$

$$\Lambda = \begin{pmatrix} \lambda_{1,1,1}^1 & \dots & \lambda_{n,12,1}^1 & \dots & \lambda_{1,1,5}^1 & \dots & \lambda_{n,12,5}^1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \lambda_{1,1,1}^5 & \dots & \lambda_{n,12,1}^5 & \dots & \lambda_{1,1,5}^5 & \dots & \lambda_{n,12,5}^5 \end{pmatrix}$$

With Γ_{m_1, m_2} of size $(n \times 12) \times (n \times 12)$, containing the semivariogram values if $m_1 = m_2$, which represents the variogram between the same variables at observed locations, or the cross-variogram between variables m_1 and m_2 at observed locations in the training data (if $m_1 \neq m_2$). Subsequently, $\mathbf{1} = (1, \dots, 1)^T$ is a vector of size $(n \times 12) \times 1$, Δ is a matrix of size (5×5) containing the Lagrange multiplier, and \mathbf{I} is the identity matrix. The semivariogram

vector for variable m_1 (if $m_1 = m_2$) or the cross-variogram between variables m_1 and m_2 (if $m_1 \neq m_2$) is defines as $\gamma_{m_1}^{m_2} = (\gamma_{1,1,m_1}^{m_2}, \dots, \gamma_{1,12,m_1}^{m_2})^T$, where $\gamma_{i,t,m_1}^{m_2}$ represents the variogram between location $x_{0,t'}$ and the observed location i for time period t , with $i = 1, \dots, n$, $t, t' = 1, \dots, 12$, and $m_1, m_2 = 1, \dots, 5$.

D. Prediction Evaluation

After obtaining weights Λ for the PM_{2.5} concentrations cokriging prediction model, the next step is to evaluate the accuracy of the model in predicting PM_{2.5} concentrations at unobserved regional points. Model evaluation will be conducted using test data by treating it as unobserved data. In this study, two measures of error will be used, namely RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error). RMSE and MAPE values can be calculated using the following formula.

$$RMSE = \sqrt{\frac{1}{n_{test} \times 12} \sum_{t=1}^{12} \sum_{i=1}^{n_{test}} (Z(x_{i,t}) - \hat{Z}(x_{i,t}))^2} \quad (11)$$

$$MAPE = \frac{1}{n_{test} \times 12} \sum_{t=1}^{12} \sum_{i=1}^{n_{test}} \frac{|Z(x_{i,t}) - \hat{Z}(x_{i,t})|}{Z(x_{i,t})} \times 100\% \quad (12)$$

Where n_{test} represents the number of points used as test data, $Z_1(x_{i,t})$ is the actual value of PM_{2.5} concentrations at the spatiotemporal location $x_{i,t}$, and $\hat{Z}_1(x_{i,t})$ is the prediction value of PM_{2.5} concentrations for the spatiotemporal location $x_{i,t}$.

IV. RESULTS

A. Descriptive Statistics

First, a descriptive statistical analysis was conducted on the research variables without considering the time period. The results of the maximum, minimum, mean, and standard deviation calculations are presented in Table II.

TABLE II
DESCRIPTIVE STATISTICS

Variable	Min	Max	Mean	SD
PM _{2.5} concentration	14.56	68.83	39.57	11.69
NO ₂ concentration	3.90	29.97	19.37	5.83
Rainfall	0	558.83	139.82	137.78
Relative Humidity	67.70	89.79	78.40	4.82
Wind Speed	2.14	6.22	3.79	0.87

Notes: Min = Minimum; Max = Maximum; SD = standard deviation

The PM_{2.5} concentrations exhibited variation throughout the year, as indicated by a standard deviation of 11.69 $\mu\text{g}/\text{m}^3$. This value suggests that the PM_{2.5} concentrations deviate by 11.69 $\mu\text{g}/\text{m}^3$ from the mean. The minimum PM_{2.5} concentrations occurred in February, with a value of 14.56 $\mu\text{g}/\text{m}^3$, while the maximum concentrations was recorded in August, reaching 68.83 $\mu\text{g}/\text{m}^3$. The lowest

NO₂ concentrations was 3.9 $\mu\text{g}/\text{m}^3$, which occurred in December, whereas the highest concentration, also recorded in December but at a different location, was 29.97 $\mu\text{g}/\text{m}^3$. Rainfall demonstrated higher variability compared to the other variables, as reflected in a standard deviation of 137.78 mm. The minimum rainfall in Jakarta during the 2023 period was 0 mm, indicating that no rain fell for an entire month, whereas the maximum rainfall reached 558.83 mm. The relative humidity and wind speed variables exhibited relatively homogeneous values across locations and time periods, with standard deviations of 4.82% and 0.87 m/s, respectively.

Table III presents the minimum, mean, maximum, and standard deviation of PM_{2.5} concentrations recorded across the study area over 12 months. The minimum concentrations of 14.56 $\mu\text{g}/\text{m}^3$ was observed in February, while the maximum concentrations of 68.83 $\mu\text{g}/\text{m}^3$ occurred in August. The overall mean concentrations ranged from 21.59 $\mu\text{g}/\text{m}^3$ (in February) to 50.15 $\mu\text{g}/\text{m}^3$ (in October). The highest standard deviation was recorded in May (8.40 $\mu\text{g}/\text{m}^3$), indicating significant variability in PM_{2.5} concentrations during this period. Notably, the dry season months of July and August exhibited elevated mean concentrations of 49.27 $\mu\text{g}/\text{m}^3$ and 49.82 $\mu\text{g}/\text{m}^3$, respectively.

TABLE III
TEMPORAL CHARACTERISTICS OF THE MONTHLY PM_{2.5} CONCENTRATION

Month	Min	Max	Mean	SD
January	17.01	38.08	24.83	5.12
February	14.56	29.68	21.59	3.71
March	23.83	48.57	33.85	6.56
April	22.79	41.02	29.20	4.77
May	32.01	64.90	45.99	8.40
June	34.65	61.65	45.12	7.09
July	37.95	67.43	49.27	7.89
August	37.54	68.83	49.82	8.37
September	35.40	62.53	45.43	7.10
October	40.12	68.15	50.15	7.75
November	30.90	54.38	41.93	6.49
December	27.71	51.49	37.72	5.96

In the cokriging analysis to be carried out, the value of a variable will be estimated using information from other variables. Two approaches are employed to examine the relationships between variables: trend visualization and Pearson's correlation coefficients. Fig. 3 visualizes the trends in the mean values of PM_{2.5}, relative humidity, precipitation, NO₂, and wind speed throughout the year, providing insights into their temporal relationships through parallel or contrasting patterns of movement. Meanwhile, the Pearson correlation coefficients, shown in Table IV, quantify the linear relationships between these variables.

Fig. 3 shows trends from the mean of PM_{2.5} concentrations values, relative humidity, rainfall, NO₂, and wind speed throughout the year. Fig. 3 shows that PM_{2.5} and NO₂ tend to increase from January to August and decrease from September to December. The lowest PM_{2.5} concentrations occurred in February, while the highest value occurred in August. In contrast, the variables of relative humidity, rainfall, and wind speed tend to decrease from March to September and increase from October to February. The highest values of rainfall and relative humidity are achieved in February, while the lowest values are in

September.

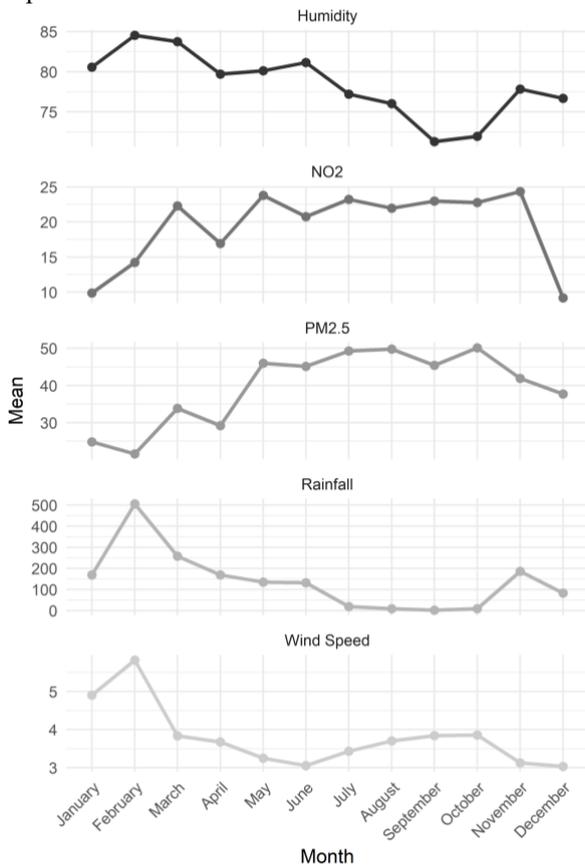


Fig. 3. Mean Trend of The Variables

This trend indicates an inverse relationship between $PM_{2.5}$ and NO_2 concentrations with meteorological variables such as rainfall, relative humidity, and wind speed. During the rainy season (January–February and October–

December), higher rainfall and relative humidity levels contribute to pollutant dispersion and removal, resulting in lower $PM_{2.5}$ concentration. In contrast, during the dry season (March–September), reduced rainfall and lower relative humidity levels allow for pollutant accumulation, which is further exacerbated by lower wind speeds, limiting air circulation.

Table IV below shows that there is a negative correlation between rainfall, relative humidity, and wind speed to $PM_{2.5}$ concentration. When relative humidity, wind speed, and rainfall increase, $PM_{2.5}$ concentrations tend to fall. In contrast, NO_2 has a positive correlation with $PM_{2.5}$. This means that $PM_{2.5}$ concentrations will also increase as NO_2 increases.

TABLE IV
PEARSON CORRELATION COEFFICIENT

	$PM_{2.5}$	NO_2	Rainfall	Relative Humidity	Wind Speed
$PM_{2.5}$	1				
NO_2	0.31	1			
Rainfall	-0.58	-0.34	1		
Relative Humidity	-0.13	-0.43	0.73	1	
Wind Speed	-0.75	-0.19	0.54	0.04	1

B. Spatiotemporal Semivariogram

Fig. 4 illustrates the empirical semivariogram alongside the theoretical semivariogram model. The empirical semivariogram was generated using training data comprising 23 location points over a 12-month period. The spatial component of the empirical semivariogram shows minimal variation in gamma values as spatial distance increases, indicating limited spatial variability within the observed range. Conversely, gamma values exhibit a clear

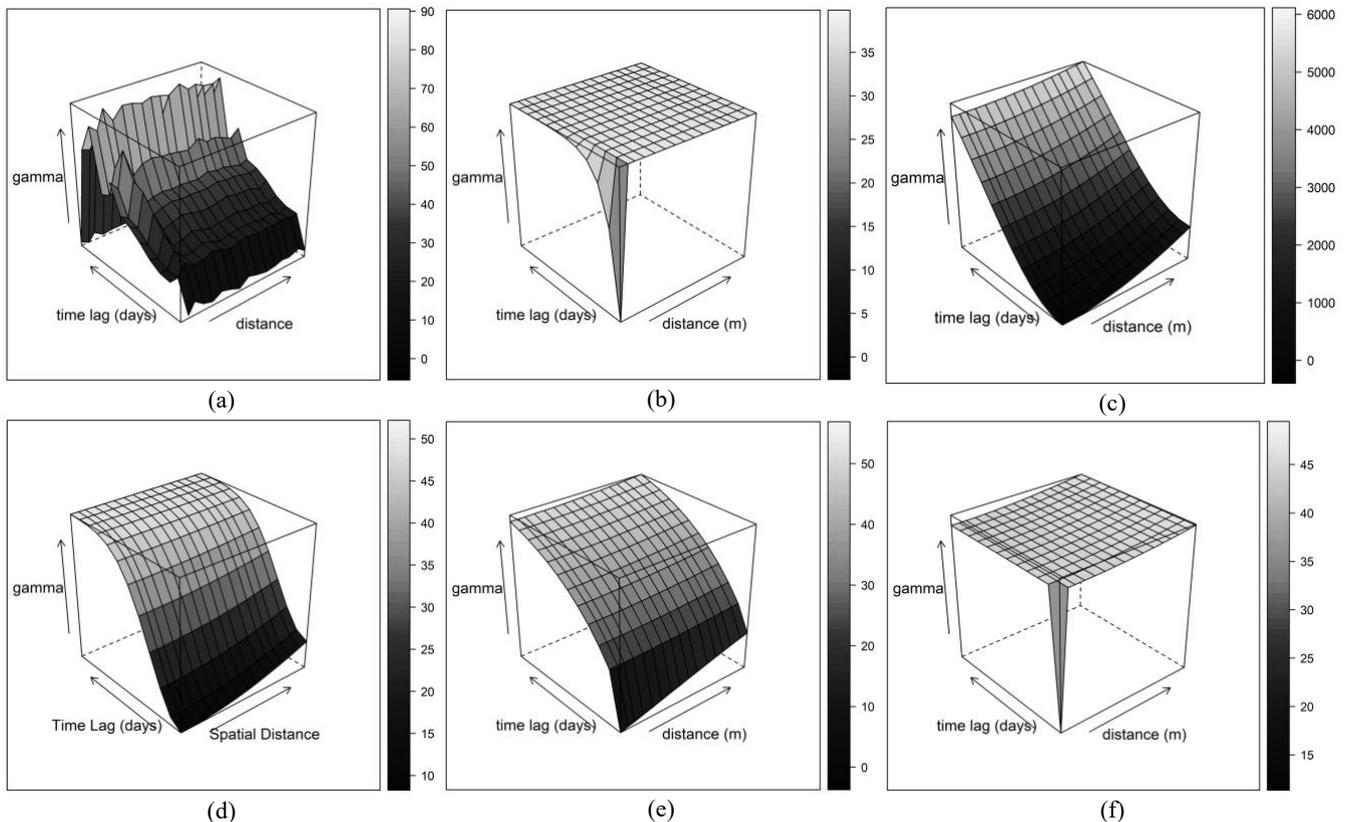


Fig. 4. Empirical Semivariogram and Theoretical Semivariograms Model, where (a): Empirical Semivariogram; (b): Separable; (c): Product-Sum; (d): Metric; (e): Sum-Metric; (f): Simple Sum-Metric

increase as the time lag grows, reflecting greater variability over time. Interestingly, a decline in gamma values is observed at the largest time lag, suggesting potential smoothing effects or reduced temporal correlation at longer intervals.

The empirical semivariogram is further modeled using various theoretical semivariograms. By combining three marginal semivariogram models (spherical, exponential, and Gaussian) for spatial, temporal, and *joint* (spatial) components, a total of 9 *separable* models, 9 *product-sum* models, 3 *metric* models, 27 *sum-metric* models, and 27 *simple sum-metric* models were evaluated. Fig. 4 illustrates the spatiotemporal semivariogram shapes derived from the best-performing combination in each theoretical model category. This modeling process aims to determine which theoretical semivariogram most accurately captures the spatiotemporal variability patterns observed in the empirical semivariogram. The color gradient in each plot represents the semivariance (gamma) values, which describe the variability between data points as a function of spatial distance and time lag. In all models, lower semivariance values are represented by darker colors (blue), indicating a strong correlation between data points over shorter spatial and temporal distances.

Table V summarizes the best combination of theoretical models based on the mean squared error (MSE) criterion, providing insight into the most suitable model for representing the data. The theoretical semivariogram with the lowest MSE indicates its ability to explain the variability pattern observed in the empirical semivariogram. Therefore, the theoretical semivariogram with the lowest MSE will be used to generate the gamma matrix in the cokriging prediction equation.

TABLE V
COMPARISONS OF MSE VALUES FOR EACH THEORETICAL SEMIVARIOGRAMS

Model	Spatial	Temporal	Joint	MSE
Separable	Exp	Exp	-	465.59
Product-Sum	Gau	Gau	-	8,906.309
Metric	-	-	Gau	300.29
Sum-Metric	Gau	Exp	Sph	292.15
Simple Sum-Metric	Gau	Gau	Exp	538.85

Notes: Exp is Exponential; Gau is Gaussian; Sph is Spherical

Based on the results in Table V above, the theoretical model with the lowest MSE is the *sum-metric* model with an MSE of 292.15. In addition, it is also evident that the *sum-metric* semivariogram (Fig. 4e) exhibits the pattern most similar to the empirical semivariogram (Fig. 4a). Thus, the *sum-metric* model will be used to generate the gamma matrix (Γ) for the subsequent calculation of the cokriging weight.

A. Semivariogram Parameter Estimation

Furthermore, fitting of the empirical semivariogram using the model will be carried out. This fitting process will produce semivariogram parameters for spatial, temporal, and joint (spatiotemporal) dimensions. Table VI displays the parameters of the semivariogram of the sum-metric model fitting results against the empirical semivariogram.

The sill indicates the maximum variability that can be explained by a semivariogram. In the spatial dimension, when the distance between points reaches 30,762.69 m

(spatial range), the variability that occurs is 13.76 (sill). This indicates that the spatial dependency between the data is still significant up to such a distance, while the distance greater than that no longer shows any dependency.

TABLE VI
SEMIVARIOGRAM PARAMETERS ESTIMATES

Component	Parameter	Estimate
Spatial	Sill	13.76
	Range	30,762.69 m
	Nugget	0.00
Temporal	Sill	43.72
	Range	177.95 days
	Nugget	0.00
Joint	Sill	13.67
	Range	19,662.36 m
	Nugget	0.00

Meanwhile, in the temporal dimension, when the time lag between observations reaches 177.95 days (temporal range), the sill is 43.72. This shows that the temporal relationship between the data up to a time interval of 6 months, while the time lag greater than that no longer shows a temporal relationship. A zero nugget value in a semivariogram indicates that the variability in the data is fully explained by the semivariogram model through its parameters, such as sill and range. This suggests that there is no unexplained variability or random noise at very short distances or time lags, and the data transitions smoothly without abrupt changes.

B. Cross-variogram Parameter Estimation

The fitting process is also applied to the cross-variogram between PM_{2.5} and predictor variables following the same steps as the semivariogram fitting. This process aims to estimate the cross-variogram parameters by minimizing the difference between empirical and theoretical values. The result is a continuous function that serves as the basis for generating cross-variogram values.

1) Cross-variogram of PM_{2.5} and NO₂

In the cross-variogram fitting process for PM_{2.5} and NO₂ concentrations, the sum-metric model yielded the lowest mean squared error (MSE) compared to other models. In this fitting, the spatial marginal semivariogram function is represented by the Gaussian model, the temporal marginal function by the spherical model, and the joint component by the exponential function.

TABLE VII
ESTIMATION OF CROSS-VARIOGRAM PARAMETERS BETWEEN PM_{2.5} AND NO₂

Component	Parameter	Estimate
Spatial	Sill	13.76
	Range	30,762.69 m
	Nugget	0
Temporal	Sill	43.72
	Range	280.97 days
	Nugget	2.83
Joint	Sill	13.67
	Range	21,954.45 m
	Nugget	0.56

The sum-metric model, with its specified arrangement of marginal functions, provides parameter estimates for the cross-variogram between PM_{2.5} and NO₂ concentrations, as presented in Table VII.

Fig. 5 presents a comparison between the empirical cross-variogram of PM_{2.5} and NO₂ concentrations and the theoretical cross-variogram based on the sum-metric model.

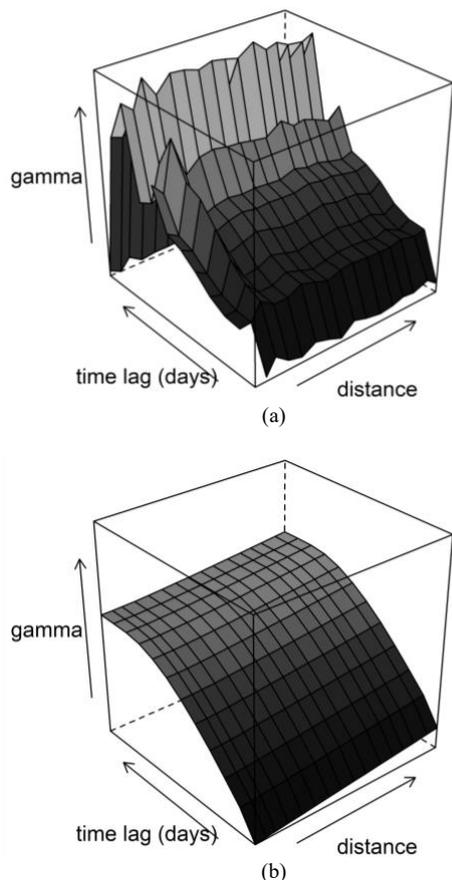


Fig. 5. Comparison of (a) the empirical cross-variogram between PM_{2.5} and NO₂ and (b) the theoretical cross-variogram based on the sum-metric model.

2) Cross-variogram of PM_{2.5} and Rainfall

In the fitting process of the cross-variogram between PM_{2.5} concentrations and rainfall, the simple sum-metric model achieved the lowest mean squared error (MSE) compared to other models.

This model employs a Gaussian function for the spatial component, a spherical function for the temporal component, and a Gaussian function for the joint component.

The parameter estimation results indicate a spatial sill value of 914.19, a temporal sill of 22,763.12, and a joint sill of 6,105.81. Additionally, Fig. 6 shows that gamma reaches approximately 40,000, which is significantly higher than the sill values observed in the semivariogram and cross-variogram of other variables. This discrepancy arises due to the substantially wider range of rainfall values compared to other variables.

Although the cross-variogram exhibits extreme values, it does not affect the unbiased nature of PM_{2.5} concentrations estimation. The extreme values only influence the magnitude of the Lagrange multiplier, which does not contribute to cokriging estimation. Consequently, the cokriging estimation results remain unbiased despite the presence of extreme cross-variogram values.

The simple sum-metric model provides parameter estimates for the cross-variogram between PM_{2.5} concentrations and rainfall variables, as shown in Table VIII.

TABLE VIII
ESTIMATION OF CROSS-VARIOGRAM PARAMETERS BETWEEN PM_{2.5} AND RAINFALL

Component	Parameter	Estimate
Spatial	Sill	914.19
	Range	15,879.99 m
Temporal	Sill	22,763.12
	Range	120.76 days
Joint	Sill	6,105.81
	Range	31,469.09 m
	Nugget	13.67

Fig. 6 presents a comparison between the empirical cross-variogram of PM_{2.5} concentrations and rainfall with the theoretical cross-variogram obtained from the simple sum-metric model.

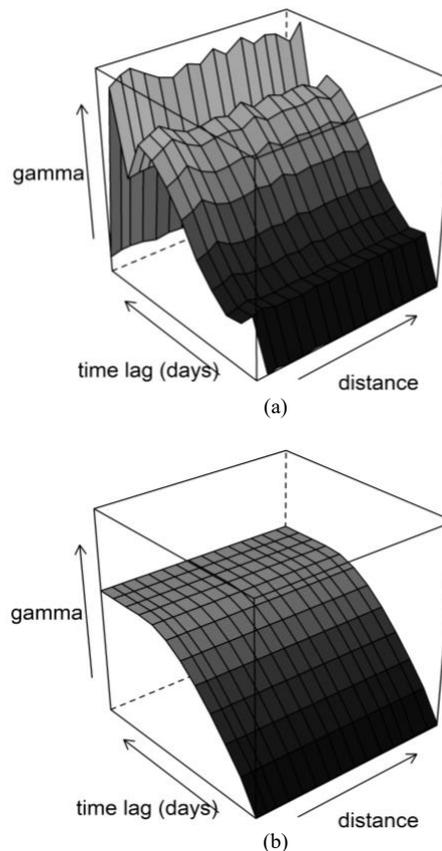


Fig. 6. Comparison of (a) the empirical cross-variogram between PM_{2.5} and rainfall and (b) the theoretical cross-variogram based on the simple sum-metric model.

3) Cross-variogram of PM_{2.5} and Humidity

The cross-variogram fitting results for PM_{2.5} concentrations and humidity indicate that the metric model with a Gaussian function achieves the lowest mean squared error (MSE) compared to other models. The estimated semivariogram parameters for this model are presented in Table IX.

TABLE IX
ESTIMATION OF CROSS-VARIOGRAM PARAMETERS BETWEEN PM_{2.5} AND HUMIDITY

Component	Parameter	Estimate
Joint	Sill	24.43
	Range	22,603.55 m
	Nugget	6.31

Fig. 7 presents a comparison between the empirical cross-variogram of PM_{2.5} concentration and humidity and the theoretical cross-variogram based on the metric model.

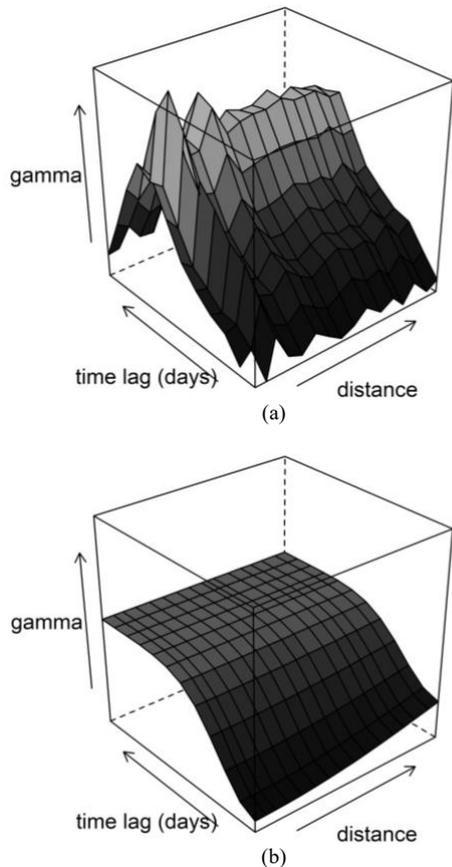


Fig. 7. Comparison of (a) the empirical cross-variogram between PM_{2.5} concentrations and humidity and (b) the theoretical cross-variogram based on the metric model.

4) Cross-variogram of PM_{2.5} and Wind Speed

The cross-variogram fitting between PM_{2.5} concentration and wind speed identified the sum-metric model as the one achieving the lowest mean squared error (MSE) compared to other models. This model employs the Gaussian function for the spatial, temporal, and joint components. The estimated parameters for the cross-variogram between PM_{2.5} concentration and wind speed are presented in Table X.

TABLE X
ESTIMATION OF CROSS-VARIOGRAM PARAMETERS
BETWEEN PM_{2.5} AND WIND SPEED

Component	Parameter	Estimate
Spatial	Sill	1
	Range	30,762.69 m
	Nugget	0
Temporal	Sill	1.88
	Range	341 days
	Nugget	0.22
Joint	Sill	1
	Range	26,843.34 m
	Nugget	0

The parameter estimation of the cross-variogram for wind speed yields sill values for all three components. It is evident that these sill values are significantly lower than those observed in the semivariogram and cross-variogram of other variables. This difference is also illustrated in Figure 8, where the gamma values range from 0 to approximately 2.5. This occurs because the range of wind speed data is considerably smaller compared to other variables.

Similar to the case of rainfall, the sill and gamma values in this cross-variogram do not affect the kriging weights or the estimation results of PM_{2.5} concentration.

Fig. 8 presents a comparison between the empirical cross-variogram of PM_{2.5} concentrations and wind speed and the theoretical cross-variogram based on the sum-metric model.

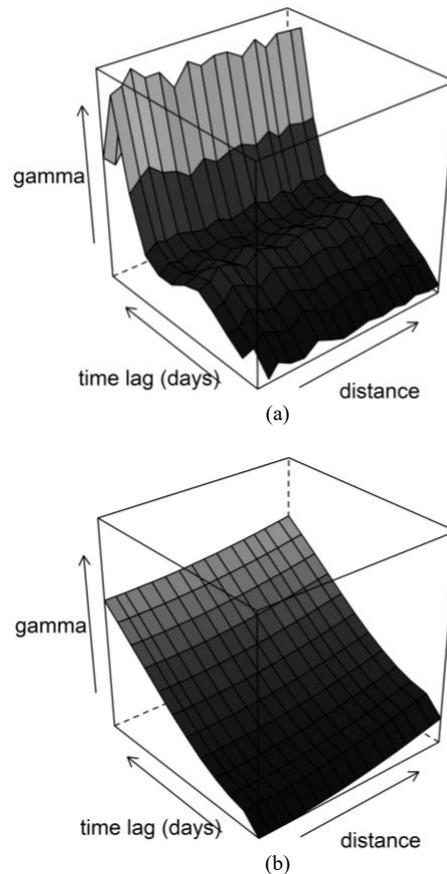


Fig. 8. Comparison of (a) the empirical cross-variogram between PM_{2.5} concentrations and wind speed and (b) the theoretical cross-variogram based on the sum-metric model.

C. Comparison of Spatiotemporal Ordinary Kriging and Spatiotemporal Cokriging

As a baseline comparison, we implement spatiotemporal (ST) ordinary kriging, which does not incorporate additional predictor information. This method relies solely on the spatial and temporal correlations of the primary variable without utilizing secondary data sources. By comparing these two approaches, we aim to highlight the potential advantages of incorporating auxiliary information in spatiotemporal modeling and assess the extent to which cokriging enhances prediction accuracy.

To quantitatively evaluate the differences between these methods, Table XI presents a comparison of the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE).

TABLE XI
COMPARISON OF ERROR ANALYSIS BETWEEN SPATIOTEMPORAL ORDINARY KRIGING AND SPATIOTEMPORAL COKRIGING

Model	RMSE	MAPE
ST Ordinary Kriging	9.63	18.31%
ST Cokriging	0.38	0.66%

The results presented in Table XI indicate a significant improvement in prediction accuracy when using spatiotemporal cokriging compared to spatiotemporal ordinary kriging. The RMSE for ordinary kriging is 9.63, whereas for cokriging, it is substantially lower at 0.38,

demonstrating that cokriging yields predictions much closer to the actual values. Similarly, the MAPE for ordinary kriging is 18.31%, while for cokriging, it is only 0.66%, highlighting a substantial reduction in relative error.

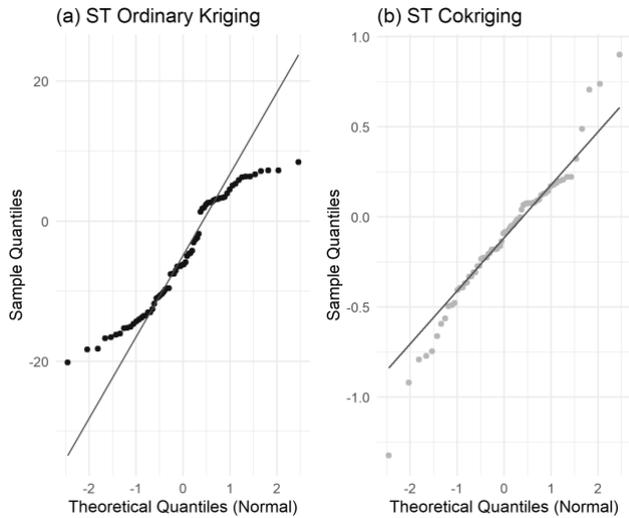


Fig. 9. Residual distributions of (a) the spatiotemporal ordinary kriging model and (b) the spatiotemporal cokriging model compared to the normal distribution.

In Fig. 9a, which represents the residuals from spatiotemporal ordinary kriging, substantial deviations from the reference line indicate non-normality in the residuals. Specifically, the presence of heavy tails at both ends suggests that the model struggles to capture extreme values, leading to skewed predictions and potential underestimation or overestimation in certain regions.

In contrast, Fig. 9b, which illustrates the residuals from spatiotemporal cokriging, shows a much closer alignment with the reference line, particularly in the middle quantiles. The overall pattern suggests that cokriging produces residuals that are more normally distributed and less prone to extreme prediction errors. This improved normality implies a better model fit and greater predictive reliability, reinforcing the advantage of incorporating auxiliary variables in spatiotemporal interpolation.

These findings indicate that integrating additional predictor variables in cokriging leads to a notable enhancement in model performance. Unlike ordinary kriging, which relies solely on the spatial and temporal correlations of the primary variable, cokriging utilizes auxiliary information, resulting in significantly improved accuracy. This demonstrates the advantage of leveraging secondary data sources in spatiotemporal modeling. The substantial reduction in error suggests that spatiotemporal cokriging is a more reliable method when relevant secondary variables are available, making it a preferred choice over ordinary kriging for applications requiring enhanced predictive accuracy.

D. Prediction

The spatiotemporal cokriging method was employed to estimate five variables: PM_{2.5} concentration, NO₂ concentration, relative humidity, rainfall, and wind speed. The estimation process utilized cokriging weights, which were derived (10) based on the Γ matrix. The Γ matrix was generated using the sum-metric semivariogram function with parameters specified in Table III.

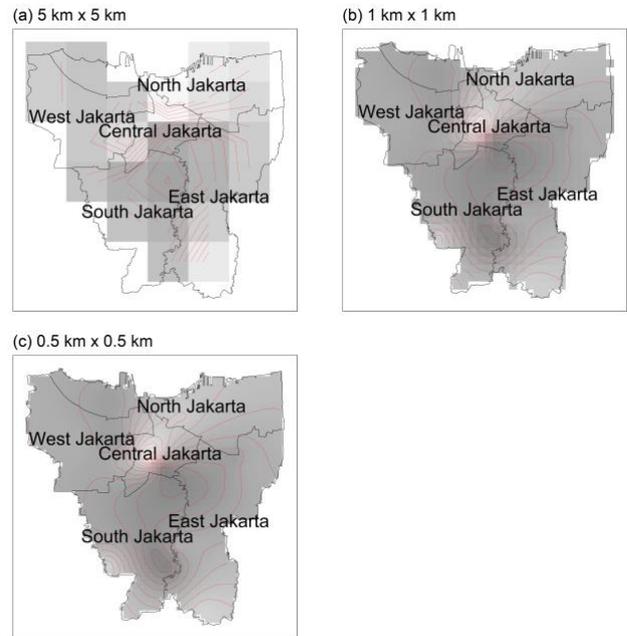


Fig. 10. Comparison of prediction maps at different resolutions: (a) 5 km, (b) 1 km, (c) 0.5 km.

In spatiotemporal cokriging, resolution primarily influences the smoothness of the prediction map rather than the accuracy of the estimates. Fig. 10a shows the predicted PM_{2.5} concentrations at a coarse resolution of 5 km × 5 km, where the spatial pattern appears more fragmented. In contrast, Fig. 10c presents predictions at a finer resolution of 0.5 km × 0.5 km, offering slightly more detail but without substantial differences compared to the 1 km × 1 km resolution in Fig. 10b.

The 1 km × 1 km resolution provides a smoother representation of PM_{2.5} distribution compared to 5 km × 5 km, while maintaining a level of detail comparable to the 0.5 km × 0.5 km resolution. Therefore, 1 km × 1 km was selected as the optimal resolution to balance smoothness and computational efficiency.

Using the same variogram model, PM_{2.5} concentrations were mapped to analyze their spatial distribution across Jakarta. Fig. 11 presents the predicted PM_{2.5} concentrations for 646 grids, each measuring 1 km × 1 km, regularly distributed across the DKI Jakarta region from January to December 2023.

The grayscale gradient represents PM_{2.5} levels, where darker shades indicates lower concentrations (20–30 μg/m³), medium gray shades represents moderate concentrations (40–50 μg/m³), and lighter shades indicates high concentrations (above 50 μg/m³). The map reveals lower PM_{2.5} concentrations during the beginning of the year (January–April) and the end of the year (November–December), coinciding with the rainy season. In contrast, higher PM_{2.5} concentrations are observed during the mid-year period (May–October), corresponding to the dry season.

The map also highlights regional differences in PM_{2.5} concentrations. The northern and southern parts of Jakarta consistently exhibit higher PM_{2.5} levels compared to other areas, likely due to a combination of high transportation activity, industrial emissions, and population density. Additionally, parts of East Jakarta show elevated PM_{2.5} concentrations during specific periods, reflecting localized

or temporal factors that contribute to air pollution in this region.

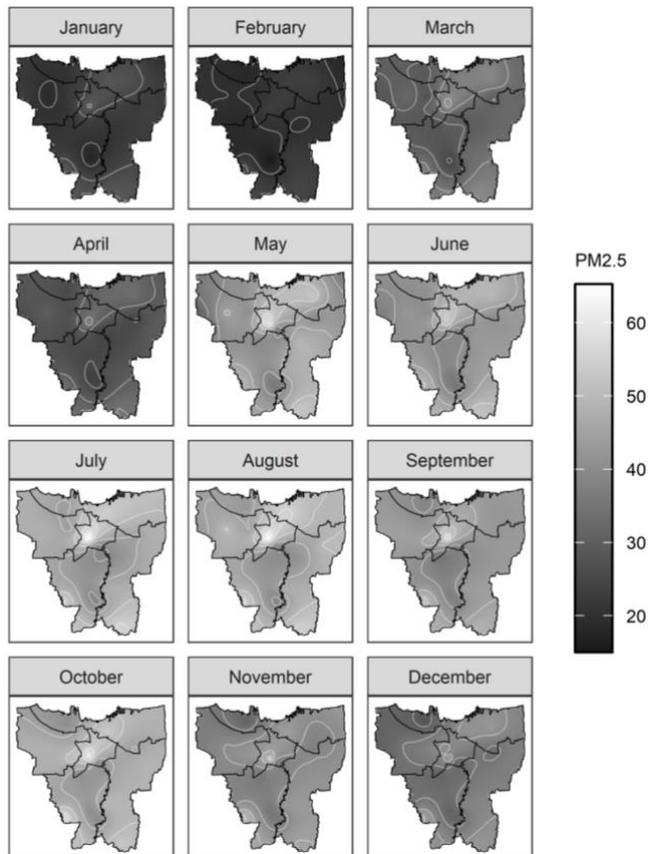


Fig. 11. Prediction Map of Monthly PM_{2.5} Concentrations in Jakarta (2023)

V. DISCUSSIONS

This study aims to estimate PM_{2.5} concentrations in Jakarta using the spatiotemporal cokriging method, which integrates spatial and temporal dependencies in prediction calculations while incorporating rainfall, relative humidity, wind speed, and NO₂ concentrations as additional variables. Predictions are made by accounting for the correlations between variables as well as spatial and temporal dependencies between locations. The weights assigned to each variable and location are calculated based on the spatiotemporal semivariogram and cross-variogram, which represent the variability structure and relationships across spatial and temporal dimensions.

The empirical semivariogram reveals spatiotemporal variability in the sample data. As shown in Fig. 4a, the empirical semivariogram for PM_{2.5} concentrations in Jakarta throughout 2023 indicates that temporal variability dominates over spatial variability. Gamma values consistently increase from the lag of 1 to 9 months, followed by a decrease between lags 10 and 12 months. This decrease corresponds to seasonal patterns, where PM_{2.5} concentrations from October to December exhibit similarities with January and February, contrasting with mid-year months such as June and July.

During the semivariogram fitting process, the sum-metric model was identified as the best fit for the empirical semivariogram pattern, achieving a Mean Squared Error (MSE) of 292.15. This model integrates a Gaussian function for spatial variability, an exponential function for temporal variability, and a spherical function for the spatiotemporal

dimension. The selection of these functions is justified by their compatibility with the observed variability characteristics in the data: the Gaussian function effectively captures gradual spatial changes, the exponential function reflects temporal fluctuations, and the spherical function describes linear changes over short spatiotemporal distances [16], [17], [18].

The sum-metric semivariogram model produced a gamma matrix that accurately reflects the data's variability structure, enabling the calculation of optimal cokriging weights. Using this model, the cokriging predictor achieved high predictive accuracy, with a Mean Absolute Percentage Error (MAPE) of 0.66% and a Root Mean Square Error (RMSE) of 0.38. These results are significantly superior to those obtained using spatiotemporal ordinary kriging, demonstrating that the spatiotemporal cokriging approach, incorporating additional variables, effectively captures the complex spatiotemporal dynamics of PM_{2.5} concentrations in Jakarta.

Fig. 11 illustrates the estimated PM_{2.5} concentrations across Jakarta from January to December 2023. The spatial distribution shows lower PM_{2.5} levels at the beginning (January–April) and end (November–December) of the year, with significant increases during the dry season (May–October). This pattern corresponds to meteorological factors, such as increased rainfall during the rainy season (late and early in the year), which dissolves and removes pollutants, and lower rainfall during the dry season, which limits air purification. Additionally, lower wind speeds during the dry season contribute to pollutant accumulation in localized areas.

High PM_{2.5} concentrations are predominantly observed in North, Central, and parts of South Jakarta, influenced by various factors specific to each region. In North and Central Jakarta, the high population density, extensive industrial activities, and significant transportation emissions are the primary contributors to elevated PM_{2.5} levels. Fossil fuel combustion from motor vehicles and industrial processes also increases NO₂ concentrations, compounding the air pollution in these areas. Meanwhile, high PM_{2.5} levels in South Jakarta are driven by both local and regional sources. The region's proximity to Bogor and Depok, which experience substantial urbanization and transportation activities, facilitates the transport of pollutants into South Jakarta. This phenomenon is particularly evident during the dry season, when atmospheric conditions favor the dispersion of pollutants from neighboring areas.

VI. CONCLUSION

The findings of this study provide valuable insights into the distribution of PM_{2.5} across Jakarta, including regions without monitoring equipment. These results can enhance public awareness and encourage preventive measures to mitigate the health risks associated with PM_{2.5} exposure. For policymakers, the study offers a data-driven foundation for air quality management initiatives, such as expanding urban green spaces and implementing effective waste management policies to address industrial emissions in high-risk areas.

The lack of green spaces in North, Central, and parts of South Jakarta exacerbates the accumulation of pollutants, as these areas lack natural mechanisms to absorb or mitigate air pollution. The interplay between meteorological factors and human activities underscores the necessity for a comprehensive air quality management strategy. Such a strategy should include controlling transportation emissions,

regulating industrial activities, and increasing the availability of green spaces in urban areas.

Future research could focus on developing more flexible temporal semivariogram functions to account for seasonal patterns. Additionally, incorporating wind direction data could improve prediction accuracy, as pollutant distribution is not uniform across all directions but is influenced by wind patterns. Another critical area for improvement lies in addressing the issue of coarse spatial and temporal data granularity. If observation points are spaced too far apart, localized pollution sources or microclimate effects may not be captured. Similarly, if data is collected at long intervals, short-term fluctuations could remain undetected. Enhancing the spatial and temporal resolution of data collection would allow for a more accurate representation of micro-scale variability and improve the robustness of semivariogram modeling.

ACKNOWLEDGMENT

This research was supported by PT Nafas Aplikasi Indonesia through the provision of data used in this study. The opinions presented in this paper do not constitute an endorsement or approval by PT Nafas Aplikasi Indonesia and reflect only the personal views of the authors.

REFERENCES

[1] A. B. Pratomo, H. N. Muthmainah, N. Kristiono, and G. C. Setyawan, 'Implementation of Internet of Things (IoT) Technology in Air Pollution Monitoring in Jakarta: Quantitative Analysis of the Influence of Air Quality Change and Its Impact on Public Health in Jakarta', *West Sci. Nat. Technol.*, vol. 1, no. 01, pp. 40–47, Sep. 2023, doi: 10.58812/wsnt.v1i01.225.

[2] OHCE UGM, 'Jakarta's Pollution Ranks 1st in the World, What Are the Health Impacts? (Polusi Jakarta Peringkat 1 di Dunia, Bagaimana Dampaknya pada Kesehatan?)', OHCE UGM. Accessed: Nov. 08, 2024. [Online]. Available: <https://ohce.wg.ugm.ac.id/polusi-jakarta-peringkat-1-di-dunia-bagaimana-dampaknya-pada-kesehatan/>

[3] AFP, 'Jakarta becomes world's most polluted major city: IQAir - Jakarta', *The Jakarta Post*. Accessed: Nov. 20, 2024. [Online]. Available: <https://www.thejakartapost.com/indonesia/2023/08/10/jakarta-becomes-worlds-most-polluted-major-city-iqair.html>

[4] Y. Zhang and W. Jiang, 'Pollution characteristics and influencing factors of atmospheric particulate matter (PM_{2.5}) in Chang-Zhu-Tan area', *IOP Conf. Ser. Earth Environ. Sci.*, vol. 108, p. 042047, Jan. 2018, doi: 10.1088/1755-1315/108/4/042047.

[5] Z. Chen *et al.*, 'Understanding meteorological influences on PM_{2.5} concentrations across China: a temporal and spatial perspective', *Atmospheric Chem. Phys.*, vol. 18, no. 8, pp. 5343–5358, Apr. 2018, doi: 10.5194/acp-18-5343-2018.

[6] G. Xu, X. Ren, K. Xiong, L. Li, X. Bi, and Q. Wu, 'Analysis of the driving factors of PM_{2.5} concentration in the air: A case study of the Yangtze River Delta, China', *Ecol. Indic.*, vol. 110, p. 105889, Mar. 2020, doi: 10.1016/j.ecolind.2019.105889.

[7] I. G. N. M. Jaya and H. Folmer, 'High-Resolution Spatiotemporal Forecasting with Missing Observations Including an Application to Daily Particulate Matter 2.5 Concentrations in Jakarta Province, Indonesia', *Mathematics*, vol. 12, no. 18, p. 2899, Sep. 2024, doi: 10.3390/math12182899.

[8] M. Rendana *et al.*, 'Spatiotemporal analysis of the link between VCD_{trop} NO₂ and population size for provinces in Sumatra Island during 2012-2020', *Environ. Pollut. Bioavailab.*, vol. 36, no. 1, p. 2298570, Dec. 2024, doi: 10.1080/26395940.2023.2298570.

[9] L. N. Lamsal, R. V. Martin, D. D. Parrish, and N. A. Krotkov, 'Scaling Relationship for NO₂ Pollution and Urban Population Size: A Satellite Perspective', *Environ. Sci. Technol.*, vol. 47, no. 14, pp. 7855–7861, Jul. 2013, doi: 10.1021/es400744g.

[10] A. Xu, Q. Wang, L. Hu, and H. Shu, 'Computing spatiotemporal variogram and covariance model', in *2011 Seventh International*

Conference on Natural Computation, Shanghai, China: IEEE, Jul. 2011, pp. 2032–2036, doi: 10.1109/ICNC.2011.6022553.

[11] J. M. Montero, G. Fernández-Avilés, and J. Mateu, *Spatial and spatio-temporal geostatistical modeling and kriging*, in *Wiley series in probability and statistics*. Chichester, West Sussex: Wiley, 2015.

[12] B. Gräler, E. Pebesma, and G. Heuvelink, 'Spatio-Temporal Interpolation using gstat', *R J.*, vol. 8, no. 1, p. 204, 2016, doi: 10.32614/RJ-2016-014.

[13] H. Wackernagel, *Multivariate Geostatistics: An Introduction with Applications*. Berlin Heidelberg: Springer, 1995. doi: 10.1007/978-3-662-03098-1.

[14] R. Giraldo, J. Mateu, and P. Delicado, 'Cokriging and multivariate kriging methods based on data of a functional random field', *Comun. En Estadística*, vol. 10, no. 2, pp. 315–344, 2017, doi: 10.15332/2422474x.3645.

[15] J. M. Ver Hoef and R. P. Barry, 'Constructing and fitting models for cokriging and multivariable spatial prediction', *J. Stat. Plan. Inference*, vol. 69, no. 2, pp. 275–294, Jun. 1998, doi: 10.1016/S0378-3758(97)00162-6.

[16] Rr. K. N. Sari and U. S. Pasaribu, 'The comparison of isotropic and anisotropic semivariogram for Gauss model', presented at the 4TH INTERNATIONAL CONFERENCE ON MATHEMATICS AND NATURAL SCIENCES (ICMNS 2012): Science for Health, Food and Sustainable Energy, Bandung, Indonesia, 2014, pp. 508–511. doi: 10.1063/1.4868855.

[17] P. I. Brooker and M. A. Stewart, 'A comparative study of simulation techniques for two dimensional data honouring specified exponential semivariograms', *J. Aust. Math. Soc. Ser. B Appl. Math.*, vol. 36, no. 3, pp. 249–260, Jan. 1995, doi: 10.1017/S0334270000010420.

[18] T. B. Wkm, Diah Chaerani, and B. N. Ruchjana, 'Exploration of R Software for Fitting Spherical Semivariogram Using Linear Programming and Sensitivity Analysis Test (Eksplorasi Software R Untuk Fitting Semivariogram Spherical Menggunakan Pemrograman Linear dan Uji Analisis Sensitivitas)', *J. Mat. Integratif*, vol. 12, no. 2, p. 75, Jul. 2017, doi: 10.24198/jmi.v12.n2.11918.75-82.

Fariza Alamanda Putri received her B.S. degree in Statistics from the Department of Statistics, Universitas Padjadjaran, in 2024 and completed her Master degree in Applied Statistics at the same department in 2025.

I Gede Nyoman Mindra Jaya received his B.S. degree (S.Si) from the Department of Statistics, Universitas Padjadjaran, in 2003, and his Master of Science degree from Institut Pertanian Bogor in 2009. He completed his Doctor of Philosophy degree at the Faculty of Spatial Sciences, University of Groningen, Netherlands, in 2022 with cum laude honors. His research focuses on spatial and Bayesian methodologies.

Bertho Tantular received the B.S. degree (S.Si) from Department of Statistics Universitas Padjadjaran in 1998 and Master of Science from Institut Pertanian Bogor in 2009. He completed his Doctor of Mathematics at Faculty of Mathematics and Natural Science, Universitas Padjadjaran. He is active in Longitudinal Data and Multilevel Analysis researches.