Jurnal Matematika Integratif.
Vol. 21, No. 2 (2025), pp. 229–244.
p-ISSN:1412-6184, e-ISSN:2549-903
doi:10.24198/jmi.v21.n2.65239.229-244

Estimation of Labor Force Participation Rate (TPAK) in Java's Data-Scarce Areas Using Ordinary Cokriging

Sofia Angelina, Nurul Gusriani, dan Firdaniza

Departemen Matematika, Fakultas MIPA, Universitas Padjadjaran
Jl. Ir. Soekarno Km 21 Jatinangor Sumedang 45363
Email: sofia20002@mail.unpad.ac.id

Abstract

The quality of the labor force is crucial for economic development, and the Labor Force Participation Rate (TPAK) is a key employment indicator. In 2024, TPAK data collection in Java Island faced gaps in DKI Jakarta and Banten Provinces, limiting comprehensive labor mapping. To address this, this study develops a spatial estimation method using the geostatistical approach of ordinary cokriging to predict TPAK in areas with limited data. This method allows estimation by incorporating information from surrounding regions and secondary variables. The Open Unemployment Rate (TPT), which has a strong inverse relationship with TPAK—where each 1% increase in TPAK is associated with a 14.82% decrease in TPT—was used as the secondary variable. Spatial autocorrelation analysis with Moran's I test confirmed that TPAK and TPT exhibit spatial patterns, are normally distributed based on Jarque-Bera test, and meet stationarity assumptions visual analysis. The best cross semivariogram model was selected using k-fold cross validation, which identified the spherical model with the lowest average RMSE of 4.24. The resulting ordinary cokriging model produced accurate TPAK estimates, achieving a MAPE of 3.25%. These estimates enable spatial visualization of TPAK in previously unobserved areas, contributing to a more complete understanding of labor participation across Java Island.

Keywords: *Ordinary cokriging; TPAK; k-fold cross validation; labor market.*

2000 Mathematics Subject Classification: 86A32, 65D05, 62G05
Received: 10-07-2025, accepted: 28-07-2025.

1. Introduction

Labor is a key driver of economic development, serving as both a productive resource and a contributor to consumption and national output [1]. A high proportion of economically active working-age population enhances productivity, expands employment opportunities, and increases regional income [2]. The Labor Force Participation Rate (Tingkat Partisipasi Angkatan Kerja or TPAK) captures this dynamic by indicating the percentage of the working-age population actively participating in the labor market [3]. This indicator is essential for regional development planning, particularly in densely populated areas such as Java Island, which is home to 55.73% of Indonesia's population [4].

However, TPAK data are not always complete. In 2024, Statistics Indonesia did not publish TPAK values for all cities and regencies in DKI Jakarta and Banten Provinces. This data gap restricts spatial analysis and undermines efforts to design equitable, data-driven employment policies. These provinces are not only highly urbanized and economically active, but also central to national labor dynamics, making the lack of data particularly concerning. While previous studies have linked TPAK with various socio-economic indicators through regression and econometric models [5][6][7][8], they largely overlook regions with incomplete data. Furthermore, none have applied geostatistical methods to estimate TPAK.

Ordinary cokriging is one such method, widely used in estimating unmeasured values by leveraging spatial correlation and secondary variables. It has been effectively applied in fields such as pollution monitoring [9], heavy metal assessment [10], and vegetation studies [11], but rarely in labor statistics. This study uses the Open Unemployment Rate (Tingkat Pengangguran Terbuka or TPT) as a secondary variable due to its strong inverse relationship with TPAK, where a 1% increase in TPAK is associated with a 14.82% decrease in TPT [12]. The integration of TPT enhances estimation accuracy in regions lacking TPAK data.

In addition to spatial modeling, this study implements k-fold cross-validation to select the best-fitting theoretical semivariogram, addressing limitations of conventional goodness-of-fit criteria such as SSE or R^2 [13][14]. It also applies assumption testing on normality, stationarity, and spatial autocorrelation using Moran's I [15], which is often neglected in similar studies.

This study aims to estimate the 2024 Labor Force Participation Rate (TPAK) in the cities/regencies of DKI Jakarta and Banten Provinces using the ordinary cokriging method, with the Open Unemployment Rate (TPT) as a secondary variable. These provinces are economically central and highly urbanized, yet omitted from recent TPAK reporting, creating serious blind spots in labor analysis. Ordinary cokriging, by incorporating spatial and statistical relationships between TPAK and TPT, offers a novel and data-efficient solution for such gaps. The goal is to generate accurate spatial predictions and to support equitable labor policy formulation by addressing existing data gaps through robust geostatistical modelling.

2. Methodologies

2.1. **Exploratory Data Analysis.**

The exploratory analysis of the primary and secondary variables to be involved in geostatistical spatial interpolation processes is an important step to be taken before process implementation. Once the interpolation methods are based on stationarity and unbiased assumptions, among others, estimates are affected by dataset features such as bias, outliers, and clusters [15]. Therefore, this study performed exploratory analyses focusing on normality (using Jarque-Bera test), spatial autocorrelation (using Moran's I test), and stationarity (through 3D scatterplots). These steps were carried out to ensure the data met key assumptions required for reliable spatial prediction.

2.1.1. *Normality.* The Jarque–Bera (JB) test is used to assess whether model residuals follow a normal distribution by evaluating skewness and kurtosis [**?**]. The null hypothesis assumes normality and is rejected if the test statistic satisfies $JB \geq \chi^2_{(\alpha;2)}$, where $\chi^2_{(\alpha;2)}$ is the critical value of the chi-squared distribution with 2 degrees of freedom at significance level α. The JB test statistic is defined as:

$$JB = \frac{n}{6}\left(SK^2 + \frac{(K-3)^2}{4}\right), \tag{1}$$

where n is the number of observations, S is the sample skewness, and K is the sample kurtosis. The skewness SK and kurtosis K are computed as follows:

$$SK = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{e_i - \bar{e}}{\sigma_e}\right)^3, \tag{2}$$

$$K = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{e_i - \bar{e}}{\sigma_e}\right)^4, \tag{3}$$

where e_i are the residuals, $\bar{e}$ is the mean of the residuals, and $\sigma_e^2 = \frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2$ is their variance. In geostatistics, residual normality is desirable because kriging relies on second-order moments (i.e., variances and covariances). Skewed or heavy-tailed residuals may distort kriging estimates. Therefore, data transformations (e.g. Box-Cox) are often applied to normalize the distribution, reduce the influence of outliers, and improve stationarity [17].

2.1.2. *Spatial autocorrelation.* The spatial weight matrix is used to quantify the spatial relationship between observation locations [18]. It incorporates both geographic coordinates (latitude and longitude) and map-based contiguity [19]. One common method is the inverse distance weighting, where weights are calculated using the Euclidean distance between coordinates. The inverse distance weight w_{ik} is given by:

$$w_{ik} = \frac{d_{ik}^{-1}}{\sum_{i=1}^{n}\sum_{k=1}^{n} d_{ik}^{-1}}, \tag{4}$$

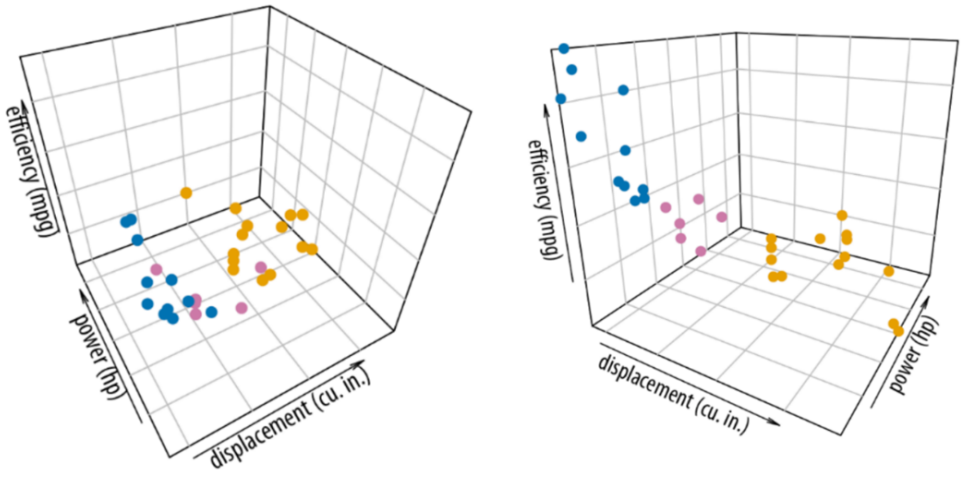
with $d_{ik} = \sqrt{(u_i - u_k)^2 + (v_i - v_k)^2}$, where u_i and v_i represent the geographic coordinates of location i, and similarly for u_k and v_k. These weights are then used to assess spatial autocorrelation, which measures how similar observation values are across space. Spatial autocorrelation is commonly tested using Moran's I index [14], which detects whether values in one location are correlated with neighboring values. The test uses the following hypotheses [18]:
$H_0 : I_M = 0$ (no autocorrelation)
$H_1 : I_M \neq 0$ (existing autocorrelation).
The Moran's I statistic (I_M) and its Z-score are calculated as

$$Z_{\text{hitung}} = \frac{I_M - E(I_M)}{\sqrt{\text{Var}(I_M)}}, \tag{5}$$

where $\text{Var}(I_M)$ is the variance of Moran's I, and $E(I_M) = -\frac{1}{n-1}$. The test rejects the null hypothesis if $Z_{\text{hitung}} \leq -Z_{\alpha/2}$ or $Z_{\text{hitung}} \geq Z_{\alpha/2}$. Moran's I ranges from -1 to 1: positive values indicate positive spatial correlation, negative values indicate negative correlation, and zero means no spatial autocorrelation.

2.1.3. *Stationarity.* A dataset is considered stationary if it shows no trend over time, its fluctuations occur around a constant mean with constant variance [20]. The Augmented Dickey-Fuller (ADF) test is commonly used to assess time series stationarity [21]. Spatial stationarity is the statistical properties of data do not vary across locations. In weak stationarity, the mean, variance, and spatial relationships remain consistent. Strong stationarity extends this to more complex properties. However, many spatial datasets exhibit irregular patterns between locations [22]. 3D plots of coordinate-based observations is suggested to examine spatial stationarity [23]. Figure 1(a) shows a random, patternless distribution indicating stationarity, while Figure 1(b) displays a downward trend, indicating non-stationarity. If spatial data (based on latitude

or longitude) has no clear trend, ordinary kriging can be applied. If a trend is present, ordinary kriging is not suitable [24].



(a) (b)

FIGURE 1. Stationary (a) and non-stationary (b) patterns on 3D scatterplots [25]

2.2. **Crossvariogram.** Variogram modeling is a core part of kriging, used to describe the spatial variation of a regional variable. In cokriging, this concept is extended to the crossvariogram, which incorporates both primary and secondary variables. The crossvariogram $\gamma_{ij}(h)$ is expressed as

$$\gamma_{ij}(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} [Y_i(u_k) - Y_i(u_k + h)] \cdot [X_j(u_k) - X_j(u_k + h)] \tag{6}$$

where $Y_i(u_k)$ and $X_j(u_k)$ are the observed values of the primary and secondary variables in location u_k, respectively, and h is the lag distance between sampled locations. While variogram always yield positive values, crossvariograms can be negative, refelcting a negative spatial correlation between variables [8]. If the crossvariogram reveals spatial dependence, it allows cokriging to provide optimal estimates by accounting for the joint spatial structure of both variables [26].

2.3. **Theoretical Model.** Fitting the experimental variogram to a theoretical model is essential in kriging and cokriging because it directly influences the weight calculation used for spatial interpolation [27]. This process involves matching the curve of the experimental semivariogram or crossvariogram with standard mathematical models such as spherical, exponential, or Gaussian [28] crossvariograms, which capture the spatial correlation between primary and secondary variables, follow the same structure as semivariograms but may take negative values when the variables move in opposite directions [29]. Key parameters in both models include the nugget (micro-scale variation or error), sill (maximum semivariance), and range (distance where spatial correlation becomes negligible) [30]. The most commonly used theoretical model are as follows:

1. Spherical Model

$$\gamma_S^*(h) = \begin{cases} c_0 + c\left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right), & 0 < h \leq a \\ c_0 + c, & h > a \end{cases} \tag{7}$$

2. Exponential Model

$$\gamma_E^*(h) = c_0 + c\left(1 - e^{-h/a}\right), \quad h > 0 \tag{8}$$

3. Gaussian Model

$$\gamma_G^*(h) = c_0 + c\left(1 - e^{-h^2/a^2}\right), \quad h > 0 \tag{9}$$

2.4. **K-Fold Cross Validation.** Cross validation is an evaluation method used to improve model flexibility and performance. Among several types (such as holdout, bootstrapping, and Monte Carlo) k-fold cross validation is widely preferred due to its balanced trade-off between bias and variance. In this method, the dataset is split into k_f equal parts (folds). The model is trained on $k_f - 1$ folds and validated on the remaining fold, repeating the process k_f times so every point is used for both training and testing. This helps prevent overfitting [31] and improves model averaging [32]. Common choices for k_f are 5 or 10 [33]. At each iteration, prediction accuracy is evaluated using the Root Mean Square Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\gamma(h)_i - \gamma_T^*(h)_i\right)^2}, \tag{10}$$

where $\gamma(h)$ is the experimental crossvariogram and $\gamma_T^*(h)$ is the theoretical model [34]. The average error across all folds is then computed as

$$\text{CV}_{k_f} = \frac{1}{k_f}\sum_{i=1}^{k_f}\text{RMSE}_i \tag{11}$$

The best model is selected based on the lowest average RMSE value.

2.5. **Ordinary Cokriging Method.** Matheron describes cokriging as a multivariate geostatistical interpolation method that estimates the primary variable by combining its own experimental data with correlated secondary variable data. It performs well when the correlation between variables is high [35]. The cokriging linear model is written as

$$\hat{y}_0 = \sum_{i=1}^{n} a_i y_i + \sum_{j=1}^{m} b_j^* x_j, \tag{12}$$

where $\hat{y}_0$ is the estimated value of unsampled location, y_i is the primary variable, x_j is the secondary variable, and a_i and b_j^* are the weights for each variables respectively. The cokriging estimator is a weighted sum of both variables, and it satisfies the unbiased condition assuming unknown means. To avoid estimation issues from negative weights, a unique unbiased constraint is used, where the sum of all weights equals one. Ordinary cokriging assumes constant spatial means and determines optimal weights by minimizing the estimation error variance using the Lagrange multiplier method. Covariances needed for weight calculation are derived from theoretical cross semivariograms, expressed as

$$C_{ij}(h) = C_{ij}(0) - \gamma_T^*(h) = C_{ij}(0)\left[1 - \frac{\gamma_T^*(h)}{C_{ij}(0)}\right]. \tag{13}$$

The cokriging weights formula is written as

$$\mathbf{z} = \mathbf{C}^{-1}\mathbf{D}, \tag{14}$$

which can be written in matrix form as

$$\mathbf{z} = \begin{bmatrix} C_{y_1y_1}(h) & \cdots & C_{y_1y_n}(h) & C_{y_1x_1}(h) & \cdots & C_{y_1x_m}(h) & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ C_{y_ny_1}(h) & \cdots & C_{y_ny_n}(h) & C_{y_nx_1}(h) & \cdots & C_{y_nx_m}(h) & 1 & 0 \\ C_{x_1y_1}(h) & \cdots & C_{x_1y_n}(h) & C_{x_1x_1}(h) & \cdots & C_{x_1x_m}(h) & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ C_{x_my_1}(h) & \cdots & C_{x_my_n}(h) & C_{x_mx_1}(h) & \cdots & C_{x_mx_m}(h) & 0 & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{y_0y_1}(h) \\ \vdots \\ C_{y_0y_n}(h) \\ C_{y_0x_1}(h) \\ \vdots \\ C_{y_0x_m}(h) \\ 1 \\ 0 \end{bmatrix}$$

where $\mathbf{C}$ is the covariance matrix, $\mathbf{D}$ is the covariance vector between sampled and unsampled locations, and $\mathbf{z}$ is the vector of weights. Solving this system yields the weights, which are then substituted back into the linear cokriging model in Equation (10) to estimate the primary variable at unsampled locations.

2.6. **Mean Average Percentage Error (MAPE).** The Mean Absolute Percentage Error (MAPE) is an effective measurement for evaluating the accuracy of a predictive model by dividing the difference between the actual value ad the predicted value by the actual value, expressed as a percentage. MAPE can be calculated using the following equation:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{15}$$

where y_i is the i-th actual value and $\hat{y}_i$ is the i-th predicted value. A smaller MAPE value indicates a more accurate prediction model. The scale of judgment for forecast accuracy based on MAPE is presented in Table 1.

Table 1. Scale of judgment of forecast accuracy

MAPE Value	Model Accuracy Level
Less than 10%	Highly accurate
11% to 20%	Good forecast
21% to 50%	Reasonable forecast
51% or more	Inaccurate forecast

Source: [36]

2.7. **Research Flowchart.** To provide a clear overview of the methodological process used in this study, a flowchart is presented in Figure 2 to illustrate the major stages of the research, starting from data preprocessing through to the estimation of TPAK values using the ordinary cokriging method. The flow includes key procedures such as data validation, spatial autocorrelation testing, variogram fitting, model evaluation, and final estimation.

In addition, Figure 3 presents a separate flowchart detailing the k-fold cross-validation process used to select the best-fitting theoretical cross semivariogram model. This procedure involves partitioning the dataset into k folds, performing training and testing iteratively, and assessing performance based on RMSE values. Together, these diagrams enhance the clarity of the research design and illustrate the logical structure of the methodological framework.

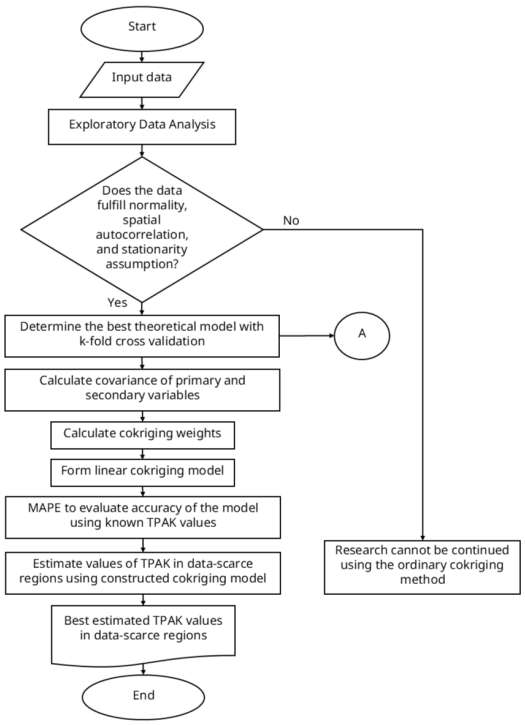


FIGURE 2. Research flowchart from data preprocessing to estimation

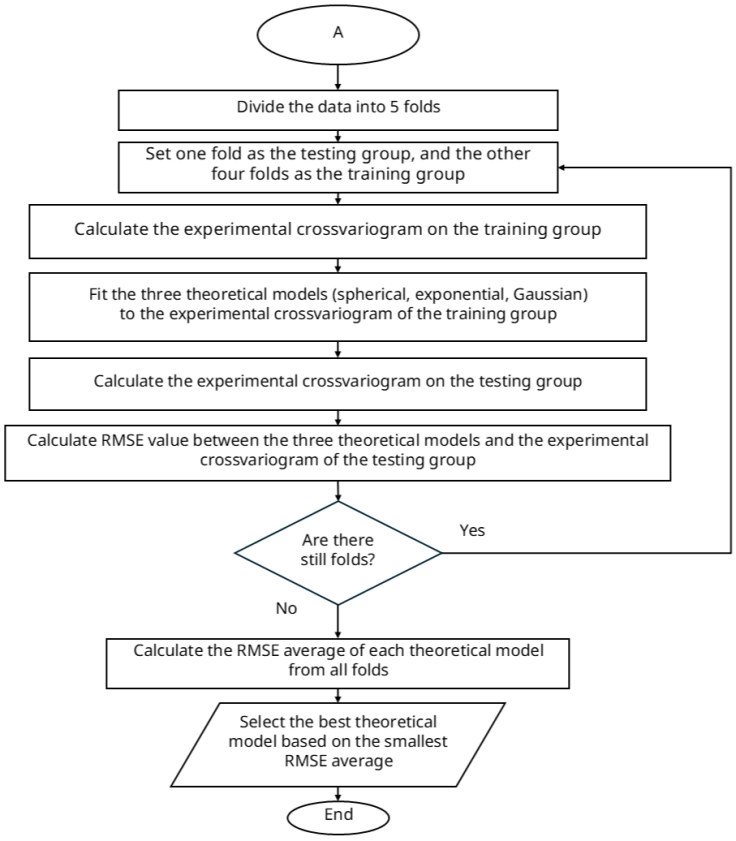


FIGURE 3. K-fold cross validation procedure for model evaluation

3. Result

3.1. **Exploratory Data Analysis.** The Jarque-Bera test uses a 5% significance level and 2 degrees of freedom, resulting in a critical chi-square value of 5.991. This critical value was obtained from the chi-square distribution table. The results of the JB test show a skewness value of 0.1660585 and a kurtosis value of 3.323844, resulting in a JB statistic of 0.9324329. Since this value is less than the critical value of 5.991, the null hypothesis is not rejected, indicating that the residuals are normally distributed for TPAK. This means the cokriging estimation will result in a better prediction as geostatistics analysis generally prefers normally distributed data. After the normality test, we conduct Moran's I test to examine the spatial autocorrelation of both TPAK and TPT datas. The test is done on the coordinates shown in Table 2.

Table 2. Coordinates of cities/regencies in Java Island

Location	City/Regency	Longitude	Latitude
1	Kabupaten Bogor	106.8249650	−6.4796790
2	Kabupaten Sukabumi	106.9266670	−6.9186110
	...		
103	Kota Madiun	111.513702	−7.629714
104	Kota Surabaya	112.734398	−7.289166

Source: [37]

The inverse distance weight of the coordinates is obtained with Equation 2 and presented in Appendix A. The weights are then used to calculate spatial autocorrelation of TPAK and TPT data using Equation 2. Moran's I test result is presented in Table 3.

Table 3. Moran's I test result

Variable	I_M	$E(I_M)$	$\text{Var}(I_M)$	Z_{hitung}
TPAK (Y)	0.22285	−0.00970	0.01966	1.65832
TPT (X)	0.32509	−0.00970	0.01966	2.38736

The result shows that the calculated Z_{hitung} is greater than or equal to $Z_{0.05}$, indicating the presence of spatial autocorrelation in TPAK and TPT values across cities/regencies in Java. This suggests that TPAK and TPT values in a given location tend to be influenced by the values in neighboring areas. The presence of such spatial patterns makes geostatistical approaches like cokriging relevant for use in the estimation process.

Next, a visual analysis was conducted to understand the characteristics of the TPAK data prior to further estimation. A three-dimensional scatterplot was used to evaluate the data distribution patterns and to identify indications of stationarity. This visualization helps reveal whether there are any trends or specific patterns in the data that may influence the modeling process. Figure 2 and Figure 3 presents the three-dimensional scatterplot of TPAK and TPT values across cities/regencies in Java.

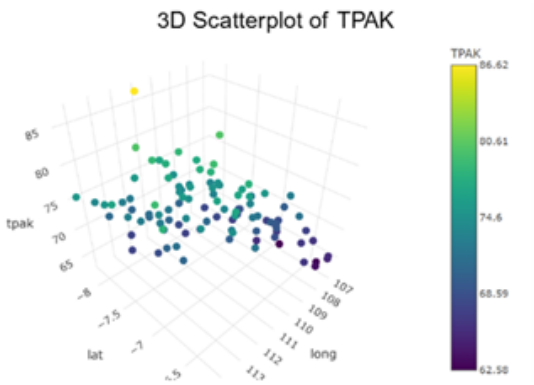


FIGURE 4. 3D scatterplot of TPAK in Java Island

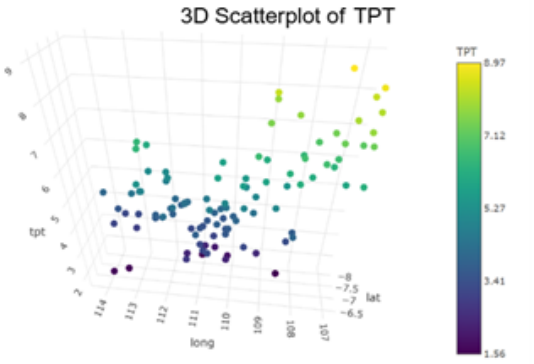


FIGURE 5. 3D scatterplot of TPT in Java Island

The distribution patterns of both TPAK and TPT values appear to be randomly scattered without location-dependent trends, as shown in Figure 1(a). The TPAK data points exhibit no visually significant spatial trend, indicating stationarity across districts/cities in Java, which supports the application of the ordinary cokriging method. Similarly, the visual analysis of TPT data shows a random distribution pattern with no evident spatial tendency, suggesting that the TPT data are also stationary and therefore suitable to be used as a secondary variable in the ordinary cokriging estimation.

3.2. **Theoretical Model Fitting for TPAK and TPT.** Before estimating TPAK values at locations without data, the best theoretical cross semivariogram model was determined using k-fold cross validation with $k_f = 5$. In this study, the data were divided into five folds, with each iteration using four folds as training data and one fold as test data. The model's performance in each iteration was evaluated using RMSE as an indicator of prediction accuracy. The RMSE values from all folds were then averaged to obtain an overall assessment of each theoretical cross semivariogram model's performance. Table 4 presents the RMSE calculations for each fold across all candidate theoretical cross semivariogram models.

TABLE 4. K-fold cross validation result

Fold	$\gamma(h)$	γ_S^*	γ_E^*	γ_G^*	RMSE_S	RMSE_E	RMSE_G
1	8.37529	1.35489	1.21031	1.20403	7.36196	7.49241	7.48861
2	7.83474	1.35852	1.21281	1.22280	7.20362	7.31419	7.28781
3	7.84537	6.60750	5.27225	2.51305	3.21359	3.92629	6.10152
4	5.59761	5.32417	5.32345	5.19987	1.21108	1.19895	1.33524
5	7.56773	7.21266	7.26247	7.18866	2.20680	2.20141	2.23999
Average RMSE (CV_{k_f})					**4.23941**	4.42665	4.89064

The best theoretical semivariogram model was selected based on the lowest average RMSE value. The spherical model yielded the smallest average RMSE of 4.23941, followed by the exponential model with 4.42665, and the Gaussian model with 4.89064. This indicates that the spherical model provides the highest prediction accuracy among the three. Therefore, it was selected as the best theoretical cross semivariogram model to be used in the estimation process using the ordinary cokriging method. A visualization of the spherical model as the best-fit theoretical model is shown in Figure 4.

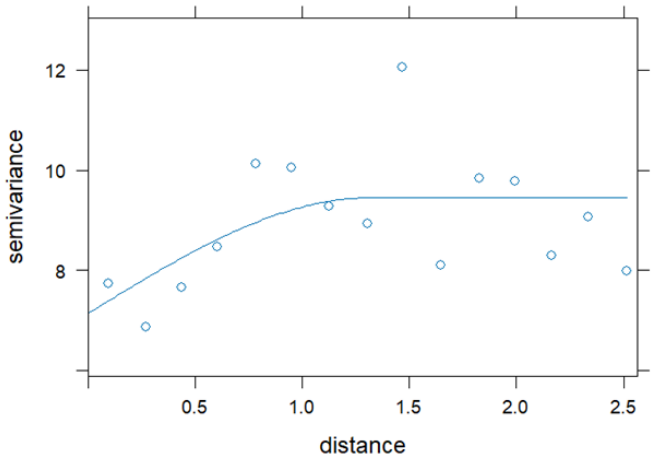


FIGURE 6. Fitting of the spherical model to the experimental crossvariogram

Figure 4 illustrates the fitting of the spherical theoretical cross semivariogram model to the experimental crossvariogram. The dots represent semivariance values at various distances, while the curve depicts the theoretical model pattern, which increases until it reaches a stable level (sill). The alignment between the curve and the distribution of data points supports the selection of the spherical model as the best representation of the spatial structure between the TPAK and TPT variables.

3.3. **Cokriging Weights.** The next step is the weighting of ordinary cokriging using the best theoretical model. Covariance matrix $\mathbf{C}$ entries is calculated using Equation 11, then the matrix is inverted to obtain $\mathbf{C}^{-1}$ as presented in Appendix B. Next, we calculated vector $\mathbf{D}$, the variance vector between the values of the two variables at the sampled locations and the unsampled location. Vector $\mathbf{D}$ is presented in Appendix C. From both matrices, we calculate the cokriging weights vector $\mathbf{z}$ with $\mathbf{z} = \mathbf{C}^{-1}\mathbf{D}$. Vector $\mathbf{z}$ is presented as follows:

$$\mathbf{z} = \begin{bmatrix} 0.03632 \\ -0.00274 \\ -0.02116 \\ \vdots \\ -0.00448 \\ -0.02847 \\ -0.02847 \end{bmatrix}.$$

Weights from vector $\mathbf{z}$ is the foundation to calculate the estimated values. These weights are substituted to the linear cokriging model shown in Equation 10. The model is temporally static, relying solely on 2024 data. Incorporating TPAK values from previous years could reveal trends or structural shifts in labor participation, offering a more dynamic estimation approach.

The constructed cokriging model was first applied to several locations where the actual data was available, but treated as if it were unknown for evaluation purposes. The model's estimated values were then compared to the actual values using MAPE. Based on this evaluation, an overview of the model's predictive accuracy was obtained as shown in Table 5, which served as the basis for determining that the model is suitable for estimating values at locations where data is truly unavailable.

TABLE 5. MAPE calculation of estimated and known TPAK value

Location	City/Regency	Estimated TPAK Value	Known TPAK Value
1	Kabupaten Bogor	70.55253578	66.30
2	Kabupaten Sukabumi	72.16343965	69.75
		...	
103	Kota Madiun	71.18452647	70.60
104	Kota Surabaya	71.43259252	70.49
MAPE			**3.24854%**

The MAPE value is less than 10%, which shows that the model provides highly accurate estimation results and is suitable for predicting TPAK values at locations without data. For further analysis, the comparison graph between the estimated and known TPAK values is shown on Figure 5 to visualize the gap of the values.

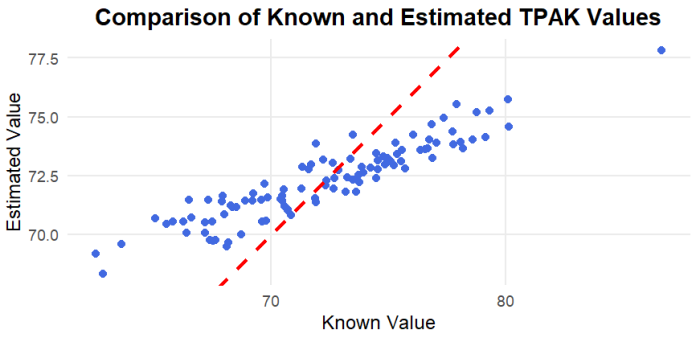


FIGURE 7. Comparison graph of estimated and known TPAK Values

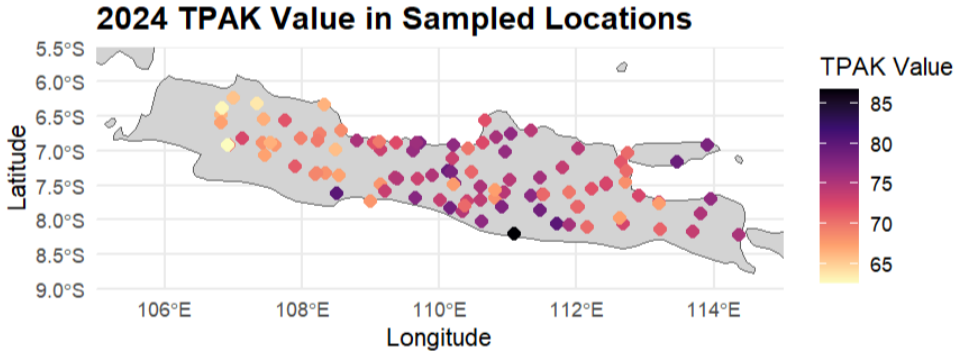
This scatter plot illustrates the relationship between estimated and actual TPAK values. Although several data points exhibit considerable deviation from the actual values, indicating instances of overestimation or underestimation, many estimations align closely with the ideal line ($y = x$), reflecting high precision. This consistency in accurate predictions likely contributes to the notably low MAPE value observed in model evaluation.

3.4. **TPAK Estimation Results with Ordinary Cokriging Method.** Estimation of TPAK value for cities/regencies in Java Island without TPAK value is carried out with the constructed cokriging model and the results are presented in Table 6 below.

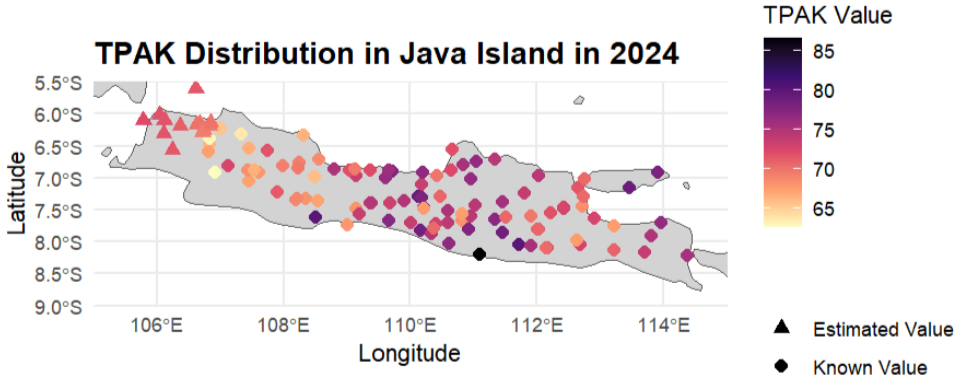
Table 6. Estimated results of TPAK value in targeted locations

Target City/Regency	Known TPT Value	Estimated TPAK Value
Kabupaten Kepulauan Seribu	7.93	71.5290817
Kota Jakarta Selatan	5.22	69.4047539
Kota Jakarta Timur	6.95	69.3942828
Kota Jakarta Pusat	6.24	69.7466378
Kota Jakarta Barat	6.18	69.9647809
Kota Jakarta Utara	6.18	69.9587694
Kabupaten Pandeglang	8.09	71.6250697
Kabupaten Lebak	6.23	71.5165461
Kabupaten Tangerang	6.06	71.0395328
Kabupaten Serang	9.18	72.2063452
Kota Tangerang	5.92	70.0481292
Kota Cilegon	6.08	71.8234746
Kota Serang	7.12	71.6411745
Kota Tangerang Selatan	5.09	69.4048025

The estimated values were then plotted on the TPAK distribution map for cities/regencies in Java Island to provide a comprehensive overview of the TPAK distribution across the island. Previously, the TPAK distribution map of Java appeared empty in the western region of the island due to the absence of data for all cities/regencies in DKI Jakarta and Banten Provinces. After estimating the TPAK values for the cities/regencies in both provinces, the TPAK distribution map of Java Island became more complete, allowing the spatial pattern of TPAK to be more clearly observed, including the distribution patterns of areas with high or low TPAK that were previously unidentifiable due to data limitations. A comparison of both plotted maps is shown in Figure 6.



(a)



(b)

FIGURE 8. Plotted TPAK values in Java Island in 2024: (a) before estimation; (b) after estimation.

As shown in Figure 6(b), most areas fall within the medium TPAK range, which is between 70 and 80. The relatively high TPAK values indicate that the majority of the working-age population in Java Island is economically active. This reflects a high level of labor force participation in overall economic activities across Java Island.

4. Conclusion

This study successfully estimated the 2024 TPAK values in cities and regencies across Java Island, including areas without official data such as DKI Jakarta and Banten Provinces, using the ordinary cokriging method. Among the tested semivariogram models, the spherical model yielded the best performance based on k-fold cross-validation, producing the lowest average RMSE of 4.24. Furthermore, the overall MAPE of 3.25% demonstrates strong predictive accuracy, with many estimated values closely aligning with actual TPAK figures in cities and regencies where data were available, as shown by Figure 5. The resulting TPAK distribution map reveals that estimated regions exhibit relatively moderate to high labor force participation, consistent with patterns observed in adjacent urban and peri-urban areas. These findings confirm that ordinary cokriging, when supported by a well-fitted variogram, is an effective and reliable method for estimating labor force participation in spatially incomplete datasets. This expanded spatial coverage improves labor statistics and supports more equitable policy formulation by illuminating regions previously excluded from analysis. Future studies could extend this approach to other Indonesian islands or enhance the model by incorporating additional secondary variables such as education levels, urbanization rates, or informal employment proportions to further improve estimation robustness and policy relevance.

Appendix A. Inverse distance weight

$$\mathbf{w}_{ik} = \begin{bmatrix} 0 & 49.82608 & \cdots & 313.99004 & \cdots & 659.22007 \\ 49.82608 & 0 & \cdots & 297.43044 & \cdots & 642.88562 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 313.99004 & 297.43044 & \cdots & 0 & \cdots & 345.83113 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 659.22007 & 642.88562 & \cdots & 345.83113 & \cdots & 0 \end{bmatrix}$$

Appendix B. Covariance matrix $\mathbf{C}^{-1}$

$$\mathbf{C}^{-1} = \begin{bmatrix} 49519 & -48076 & \cdots & -49519 & \cdots & -0.00532 \\ -48076 & 49519 & \cdots & 48076 & \cdots & -0.00804 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ -49519 & 48076 & \cdots & 49519 & \cdots & 0.00428 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ -0.00532 & -0.00804 & \cdots & 0.00428 & \cdots & -0.15241 \end{bmatrix}.$$

Appendix C. Variance vector $\mathbf{D}$

$$\mathbf{D} = \begin{bmatrix} 1.74180 \\ 0.71387 \\ 0.70005 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

References

[1] A. E. Ariesti and K. Asmara, "Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Partisipasi Angkatan Kerja (TPAK) di Pulau Jawa," 2023.

[2] R. P. Gautama, "Pengaruh Pengeluaran Perkapita, Jumlah Penduduk Miskin, Tingkat Pengangguran Terbuka (TPT), dan Tingkat Partisipasi Angkatan Kerja (TPAK) Terhadap Pendapatan Pajak Provinsi Jawa Tengah," *Emerging Statistics and Data Science Journal*, vol. 2, no. 1, 2024.

[3] Direktorat Statistik Kependudukan dan Ketenagakerjaan, "Booklet Sakernas Agustus 2024," Aug. 2024. Accessed: Mar. 09, 2025. [Online]. Available: https://www.bps.go.id/id/publication/2024/12/20/145b7ca9b2e159c6c0493290/booklet-sakernas-agustus-2024.html

[4] Badan Pusat Statistik, "Jumlah Penduduk Menurut Provinsi di Indonesia (Ribu Jiwa), 2024," Accessed: Mar. 10, 2025. [Online]. Available: https://sulut.bps.go.id/id/statistics-table/2/OTU4IzI=/jumlah-penduduk-menurut-provinsi-di-indonesia.html

[5] N. Syamsuddin et al., "Pengaruh Tingkat Partisipasi Angkatan Kerja dan Pendidikan terhadap Pertumbuhan Ekonomi di Provinsi Aceh," *Jurnal Sosiohumaniora Kodepena*, vol. 1, no. 2, May 2021. [Online]. Available: http://jsk.kodepena.org/index.php/jsk

[6] G. A. Vanessa, X. Guilin, and D. Jiao, "The Effect of Labor Force Participation Rate (TPAK), Human Development Index (HDI), and Unemployment on Poverty in Central Java in 2022," *International Journal of Noesantara Islamic Studies*, vol. 1, no. 1, 2024, doi: 10.70177/ijnis.v1i1.834

[7] T. Saifudin, A. Faiza, L. Puspasari, and Z. 'Ilmatun Nurrohmah, "Estimating The Concentration of NO_2 with The Cokriging Method in The Capital City of Jakarta," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 4, pp. 1985–1996, Dec. 2023, doi: 10.30598/barekengvol17iss4pp1985-1996

[8] H. Purnomo, S. A. Rande, and R. Prastowo, "Pemetaan Spasial Kadar Kobal pada Endapan Laterit dengan Metode Ordinary Cokriging dan Inverse Distance Weighting," *Jurnal Sains Teknologi & Lingkungan*, vol. 8, no. 1, pp. 73–86, Jun. 2022, doi: 10.29303/jstl.v8i1.317

[9] B. Pavani-Biju, J. G. Borges, S. Marques, and A. C. Teodoro, "Enhancing Forest Site Classification in Northwest Portugal: A Geostatistical Approach Employing Cokriging," *Sustainability (Switzerland)*, vol. 16, no. 15, Aug. 2024, doi: 10.3390/su16156423

[10] N. A. Salsabila, S. Andriani, Mirisda, and D. A. Nohe, "Analisis Pengaruh Tingkat Partisipasi Angkatan Kerja dan Indeks Pembangunan Manusia Terhadap Tingkat Pengangguran Terbuka Menggunakan Regresi Probit Dan Logit," May 2022.

[11] F. D. P. Ramadhani and R. Widianita, "Analisis Dampak Tingkat Partisipasi Angkatan Kerja (TPAK) terhadap Tingkat Kemiskinan di Provinsi Sumatera Barat," *Jurnal Multidisiplin Inovatif*, vol. 8, no. 12, pp. 2246–6110, 2024.

[12] R. Usali, N. Nurwan, F. A. Oroh, and M. R. F. Payu, "Pemodelan Regresi Spasial Dependensi pada Tingkat Partisipasi Angkatan Kerja di Indonesia Tahun 2020," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 15, no. 4, pp. 687–696, Dec. 2021, doi: 10.30598/barekengvol15iss4pp687-696

[13] E. P. Putra, R. Goejantoro, and D. Suyitno, "Penaksiran Kandungan Klorida di Sungai Mahakam Wilayah Samarinda Tahun 2017 dengan Metode Cokriging," *Jurnal EKSPONENSIAL*, vol. 11, no. 2, 2020.

[14] R. Mailanda, D. Kusnandar, and N. Miftahul Huda, "Analisis Autokorelasi Spasial Kasus Positif Covid-19 Menggunakan Indeks Moran dan LISA," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, vol. 11, no. 3, pp. 483–492, 2022.

[15] A. A. Rodrigues, T. M. Siqueira, T. L. C. Beskow, and L. C. Timm, "Ordinary Cokriging applied to generate intensity-duration-frequency equations for Rio Grande do Sul State, Brazil," *Theor Appl Climatol*, vol. 155, no. 3, pp. 2365–2378, Mar. 2024, doi: 10.1007/s00704-024-04829-6

[16] D. N. Gujarati and D. C. Porter, *Basic Econometrics*, Fifth Edition. McGraw-Hill Education, 2008.

[17] E. A. Varouchakis, "Gaussian Transformation Methods for Spatial Data," *Geosciences (Basel)*, vol. 11, no. 5, May 2021, doi: 10.3390/geosciences

[18] A. Tsanawafa, D. A. Kusuma, and B. N. Ruchjana, "Penerapan Model Spatial Autoregressive Exogenous pada Data Penetapan Warisan Budaya Takbenda di Pulau Jawa," *Jurnal Matematika Integratif*, vol. 19, no. 2, p. 137, Dec. 2023, doi: 10.24198/jmi.v19.n2.46526.137-147

[19] J. LeSage and R. K. Pace, *Introduction to Spatial Econometrics*, 1st ed. New York: Chapman and Hall/CRC, 2009, doi: 10.1201/9781420064254

[20] R. Yendra and R. R. Risman, "Penerapan Metode Ordinary Kriging pada Pendugaan Kriminalitas di Kota Pekanbaru Riau," *Jurnal Sains Matematika dan Statistika*, vol. 5, no. 1, 2019.

[21] A. Muhtar, D. Ningrum, and R. M. A. Hutagaol, "Penerapan Time Series Forecasting untuk Memprediksi Pertumbuhan Ekonomi Indonesia 2024," *Data Sciences Indonesia (DSI)*, vol. 3, no. 2, pp. 79–89, May 2024, doi: 10.47709/dsi.v3i2.3263

[22] M. Beenstock and D. Felsenstein, *The Econometric Analysis of Non-Stationary Spatial Panel Data*. Springer International Publishing, 2019, doi: 10.1007/978-3-030-03614-0

[23] M. C. Safira, A. Fauzan, M. Alfafisurya, and S. Adhiwibawa, "Interpolasi Polutan Nitrogen Dioksida (NO_2) di Kota Yogyakarta dengan Pendekatan Ordinary Kriging dan Inverse Distance Weighted," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 14, no. 2, pp. 55–66, Dec. 2022, doi: https://doi.org/10.34123/jurnalasks.v14i2.359

[24] N. N. Rohma, "Pendugaan Metode Ordinary Kriging (Studi Kasus Data Curah Hujan di Malang Raya)," *Jurnal Penelitian Ilmu Sosial dan Eksakta*, vol. 2, no. 1, pp. 21–29, Sep. 2022, doi: 10.47134/trilogi.v2i1.33

[25] C. O. Wilke, *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*, 1st ed., vol. 1. Taiwan: O'Reilly Media, Inc., 2019.

[26] J. Cao, C. Li, Q. Wu, and J. Qiao, "Improved Mapping of Soil Heavy Metals Using a Vis-NIR Spectroscopy Index in an Agricultural Area of Eastern China," *IEEE Access*, vol. 8, pp. 42584–42594, 2020, doi: 10.1109/ACCESS.2020.2976902

[27] X. Zhang, L. Lian, and F. Zhu, "Parameter fitting of variogram based on hybrid algorithm of particle swarm and artificial fish swarm," *Future Generation Computer Systems*, vol. 116, pp. 265–274, Mar. 2021, doi: 10.1016/j.future.2020.09.026

[28] A. Seddiki and S. Dehimi, "Effect of choosing a variogram model to predict salinity and its impact on the environment and geotechnical structures," *Technium Social Sciences Journal*, vol. 39, pp. 860–872, Jan. 2023. [Online]. Available: www.techniumscience.com

[29] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press, 1997.

[30] B. Usowicz and J. Lipiec, "Spatial Variability of Saturated Hydraulic Conductivity and its Links with other Soil Properties at the Commune Scale," 2021, doi: 10.21203/rs.3.rs-138269/v1

[31] K. Srinivasan et al., "An efficient implementation of artificial neural networks with K-fold cross validation for process optimization," *Journal of Internet Technology*, vol. 20, no. 4, pp. 1213–1225, 2019, doi: 10.3966/160792642019072004020

[32] X. Zhang and C.-A. Liu, "Model Averaging Prediction by K-Fold cross validation," 2022, doi: https://doi.org/10.1016/j.jeconom.2022.04.007Getrightsandcontent

[33] Y. Widyaningsih, G. P. Arum, and K. Prawira, "Aplikasi K-Fold Cross Validation dalam Penentuan Model Regresi Binomial Negatif Terbaik," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 15, no. 2, pp. 315–322, Jun. 2021, doi: 10.30598/barekengvol15iss2pp315-322

[34] A. Antal, P. M. P. Guerreiro, and S. Cheval, "Comparison of spatial interpolation methods for estimating the precipitation distribution in Portugal," *Theor Appl Climatol*, vol. 145, no. 3–4, pp. 1193–1206, Aug. 2021, doi: 10.1007/s00704-021-03675-0

[35] P. A. Dowd and E. Pardo-Igúzquiza, "The Many Forms of Co-kriging: A Diversity of Multivariate Spatial Estimators," *Math Geosci*, vol. 56, no. 2, pp. 387–413, Feb. 2024, doi: 10.1007/s11004-023-10104-7

[36] K. D. Lawrence, R. K. Klimberg, and S. M. Lawrence, *Fundamentals of Forecasting using Excel*. Industrial Press, 2009.

[37] Badan Pusat Statistik, "Sistem Informasi Geospasial," Accessed: Mar. 10, 2025. [Online]. Available: https://sig.bps.go.id/