



Comparison of Ordinary Kriging and Cokriging for Spatial Estimation Based on Simulated Data

Siti Mutiah^{1*}, Muhammad Nur Aidi², Asep Saefuddin³, and Fitrah Ernawati⁴

^{1,2,3}*Department of Statistics and Data Science, School of Data Science, Mathematics, and Informatics, IPB University, Indonesia*

⁴*National Research and Innovation Agency (BRIN), Indonesia*

*Corresponding author's email: smutiah842@gmail.com

Abstract

This study compares the performance of Ordinary Kriging (OK) and Cokriging (CK) methods in spatial estimation based on simulated data. Twelve scenarios are arranged based on a combination of sample sizes (50, 250, 500) and correlation levels between variables ($\rho = 0.1, 0.6, 0.9$), with each scenario repeated 30 times. Spatial data are randomly generated within the geographical boundaries of Indonesia, and variables are modeled using spherical variograms with nugget ($C_0 = 0$), sill ($C_1 = 10$), and range ($a = 10$). Model evaluation is performed using Leave-One-Out Cross Validation (LOOCV) with RMSE and R^2 as evaluation metrics. The results show that Cokriging consistently produces more accurate estimates than Ordinary Kriging in all scenarios. In the best configuration (CK, $n = 500$), RMSE = 1.04 and $R^2 = 0.945$, whereas the best performance of OK only achieved RMSE = 1.06 and $R^2 = 0.873$. All levels of correlation in Cokriging yielded good results, especially with sufficient data. Therefore, Cokriging is recommended as a superior spatial interpolation method in multivariate spatial contexts, particularly when relevant secondary information is available.

Keywords: Cokriging, estimation, geostatistical interpolation, ordinary kriging, spatial simulation.

Copyright © 2025 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)

1 Introduction

Spatial data is data with geographic location attributes, allowing the identification of patterns, trends, and relationships between phenomena based on the place or location of observation, where values collected from different locations are related to each other, and the relationship tends to weaken as the distance between locations increases [1, 2]. Spatial estimation is one of the main topics in geostatistics that is highly relevant in various fields, such as natural resource mapping, environmental risk analysis, and regional management. In this context, the main challenge is making predictions at unobserved locations using information from limited observation points while maintaining high accuracy. It is necessary to perform interpolation to overcome the lack of data in unsampled areas [3]. Various interpolation approaches have been developed, but the Kriging method is consistently the primary choice because of its statistically optimal nature. It produces unbiased predictions, and the sum of its weights produces more accurate estimates than other interpolation methods [4].

In the spatial statistics framework, the estimation process considers the observed values and the spatial dependencies represented through the covariogram function. This distinguishes geostatistical methods such as Kriging from conventional interpolation because of the probabilistic basis in producing estimates. Therefore, accurate spatial structure modeling is a crucial component in ensuring the

reliability of spatial predictions. Kriging is a geostatistical technique to predict and interpolate data at unsampled locations [5]. The Kriging interpolation method is generally used to predict unsampled areas by optimizing weights and having minor mean square errors and residuals; this model can be used to identify the distribution of the high-risk regions [6, 7, 8, 9].

Two widely used methods in the Kriging framework are Ordinary Kriging (OK) and Cokriging. Ordinary Kriging estimates a variable value at a certain point by observing similar data in other areas [10]. Cokriging is a kriging method used to calculate data at a location, using the primary variable and one or more other variables that have a spatial correlation with the variable [4]. Cokriging can improve prediction accuracy at unmeasured locations and minimize estimation errors [11]. Therefore, a systematic study is needed to evaluate both methods' performance in a controlled environment.

Most previous studies have focused on applying these methods in empirical cases. Although useful, these approaches often have limitations in terms of validation since the actual values at unobserved locations are unknown. To overcome this, simulation-based approaches offer distinct advantages. Models and simulations are often used to help make more effective and efficient decisions [12]. With fully controlled data, it is possible to objectively assess the accuracy of the estimates and observe how spatial characteristics and relationships between variables affect the performance of each method.

This study uses simulated data to compare the performance of Ordinary Kriging and Cokriging in the context of spatial estimation. The primary focus lies on the quantitative evaluation of each method's prediction accuracy in various correlation configurations and the number of observation points. The results are expected to provide methodological contributions in selecting spatial estimation techniques appropriate to the data conditions faced.

In addition to methodological contributions, this study has practical relevance for fields that rely on spatial data with limited secondary information availability. The choice between OK and Cokriging is often based on intuition or data availability without adequate quantitative evaluation. Therefore, the results of this study are expected to provide evidence-based guidance in determining the most efficient and appropriate spatial estimation method.

2 Methods

2.1 Research Design

This research is designed as a simulation-based experimental study to evaluate and compare the performance of two popular geostatistical methods, Ordinary Kriging and Cokriging. This design allows control over spatial conditions, such as the number of observation points and the level of correlation between variables, resulting in an objective and measurable performance assessment of data with known ground truth.

Simulation data also plays a vital role in this study. Simulation data is generated based on a specific algorithm in accordance with the mathematical model being analyzed [13]. Simulation, modeling, and evaluation were conducted using the R software running on the Google Colab platform to streamline computing time and modeling processes. The packages used include `gstat`, `sp`, and `geoR`, while visualization was performed using `ggplot2`. A total of 12 simulation scenarios were designed from a combination of methods (OK and Cokriging), data size (50, 250, and 500 points), and correlation between variables (0.1, 0.6, 0.9 for Cokriging). Each scenario was repeated 30 times. Table 1 shows the details of the simulation design.

Table 1: Simulation Scenario

Method	Data Size	Correlation
Ordinary Kriging	50	-
	250	-
	500	-
Cokriging	50	0.1
	50	0.6
	50	0.9
	250	0.1
	250	0.6
	250	0.9
	500	0.1
	500	0.6
	500	0.9

2.2 Ordinary Kriging and Cokriging Methods

This study uses the spatial interpolation approach of kriging. This geostatistical technique can estimate values at a location not sampled based on information from surrounding points. Kriging utilizes the spatial relationship between observation points represented through a variogram model. This study applies two types of kriging: ordinary kriging and cokriging.

Ordinary kriging is used to estimate the value of a variable based on information from the same variable in surrounding locations, assuming that the average spatial process is unknown but constant throughout the research area. The mathematical model of ordinary kriging is formulated as follows [1]:

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (1)$$

with $\sum \lambda_i = 1$, where $Z(x_0)$ is the estimated value at location x_0 , $Z(x_i)$ is the value at the i -th observation point, and λ_i is the weight determined based on the variogram model.

In addition to ordinary kriging, this study applies the cokriging method as a spatial estimation technique that uses information from the main variables and one or more secondary variables with spatial correlation. Estimation is carried out with a combination of the two types of variables formulated by [14] as follows:

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z_x(x_i) + \sum_{j=1}^n \mu_j Z_y(y_j) \quad (2)$$

Where $Z_x(x_i)$ is the estimate of the primary variable, $Z_y(y_j)$ is the estimated value of the secondary variable, λ_i and μ_j are each estimated weights obtained based on the spatial structure between points. Secondary variables are selected based on theoretical considerations and empirical spatial correlations, which can be analyzed through scatterplots and cross-variograms [15].

2.3 Simulation Data Processing Procedure

The following are the steps for analyzing simulation data using the ordinary kriging and cokriging methods.

1. Determine the coordinates of observation points (x_i, y_i) , with latitude and longitude ranges adjusted to the geographical area of Indonesia.
2. Calculate the distance between points using the Euclidean Distance formula:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3)$$

3. Determine the center point, and choose one center point (x_0, y_0) as the reference point in the kriging process.

4. Generate data using the spherical variogram. According to [14], the parameters used are: nugget $C_0 = 0$, sill $C_1 = 10$, and range $a = 10$. The spherical variogram model is:

$$\gamma(h) = \begin{cases} 10 \left(1.5 \frac{h}{10} - 0.5 \left(\frac{h}{10} \right)^3 \right), & \text{if } h < 10 \\ 10, & \text{if } h \geq 10 \end{cases}$$

5. Determine the estimated value using the kriging method and the variogram to calculate the estimated value $Z(x_0)$ at the location x_0 using both the ordinary kriging and cokriging methods.

6. Cross Validation

Conduct cross-validation and test the theoretical semivariogram model using Leave-One-Out Cross Validation (LOOCV):

- Suppose $Z(x_i)$ is the N th observation point data.
- Temporarily delete $Z(x_i)$, the N th observation data from the observation data set.
- Perform testing using the Kriging method for each spherical variogram model on the remaining $N - 1$ observation data.
- Compare the estimated value of $\hat{Z}(x_i)$ with $Z(x_i)$. Calculate the estimation error:

$$e^* = \hat{Z}(x_i) - Z(x_i)$$

e^* is called the estimation error.

- Repeat step (1) for each $i = 1, 2, \dots, n$.

7. Compute model performance:

- Root Mean Squared Error (RMSE) is usually used to measure the accuracy of the estimation. A considerable RMSE value indicates the estimation was inaccurate. The RMSE formula is as follows [16].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (z(x_i) - \hat{z}(x_i))^2}$$

- Coefficient of determination R^2 , the R^2 value functions to show how well the estimation results explain the data variation, where a value approaching 1 indicates the model has strong spatial prediction capabilities.:

$$R^2 = 1 - \frac{\sum_{i=1}^N (z(x_i) - \hat{z}(x_i))^2}{\sum_{i=1}^N (z(x_i) - \bar{z})^2}$$

8. Select the model with the lowest RMSE and the highest R^2 as the estimator that performs best.

3 Results and Discussion

Data analysis begins with generating spatial simulation data based on 12 scenarios, a combination of the number of observation points (50, 250, and 500) and the level of correlation between variables ($\rho = 0.1, 0.6$, and 0.9). The longitude and latitude coordinates of the observation points are generated randomly within the geographical boundaries of Indonesia. The value of the primary variable (Z_1) is generated through a Gaussian process with a spherical variogram model, while the secondary variable (Z_2) is generated by considering the linear correlation to Z_1 . Estimation is carried out using the Ordinary Kriging and Cokriging methods, and validation is conducted using the Leave-One-Out Cross-Validation (LOOCV) approach. The performance of each method is evaluated by two primary metrics: Root Mean Squared Error (RMSE) and coefficient of determination (R^2).

3.1 Analysis of the Relationship Between Primary and Secondary Variables

The relationship between the primary variable (Z_1) and the secondary variable (Z_2) is an essential aspect in assessing the potential contribution of the secondary variable in the spatial model. Figure 1 shows that the magnitude of the correlation value ρ greatly influences the relationship pattern between these two variables. In the scenario with $\rho = 0.1$, the distribution of scatterplot points is randomly distributed and does not show a particular pattern, indicating that Z_2 does not carry significant spatial information about Z_1 . On the other hand, in the scenarios with $\rho = 0.6$ and $\rho = 0.9$, the points appear to be increasingly dense following the identity line, indicating a strong and positive linear relationship between the two variables.

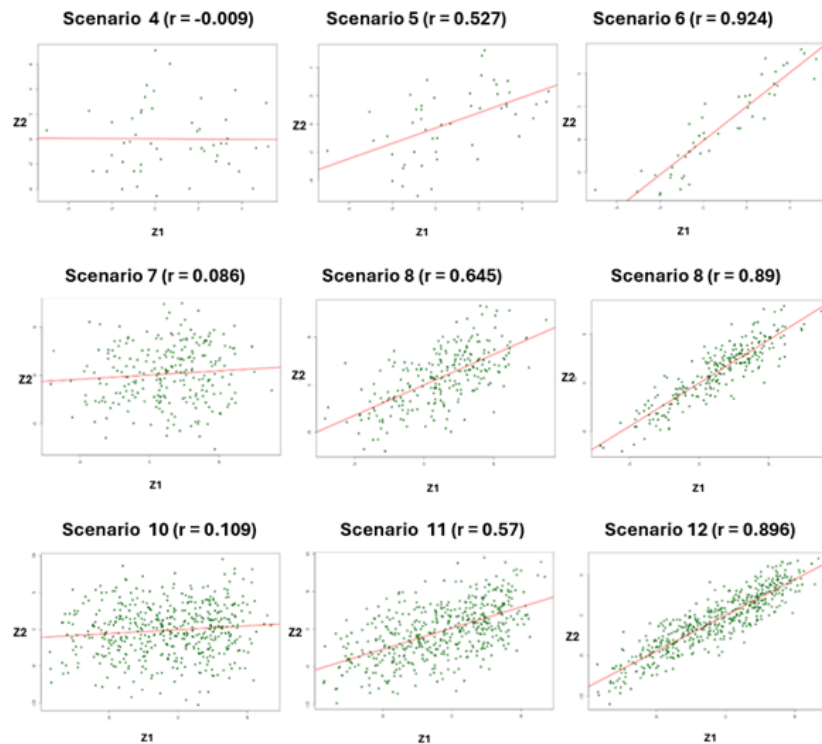


Figure 1: Correlation of primary and secondary variables

Figure 1 visually shows the changes in the relationship pattern between Z_1 and Z_2 among various correlation values. The increase in compactness of the data distribution at higher ρ values indicates the great potential of Z_2 to be utilized in spatial estimation using the Cokriging approach. This is in line with the basic principle of Cokriging, which assumes the existence of additional relevant information from secondary variables. This finding is also supported by previous literature. The effectiveness of Cokriging is highly dependent on the strength of the spatial correlation between primary and secondary variables [17].

3.2 Variogram Structure and Cross-Variogram

A variogram is a key tool in geostatistics to characterize the spatial structure of a variable. Figure 2 shows the empirical variogram and the spherical model fitting results for the primary and secondary variables in several scenarios. The variogram model can be fitted well to the data, as seen from the model line (fit) that follows the semivariance pattern at various lag distances. The success of this fitting indicates that the spatial structure of the simulated data is accurately captured by the spherical model, which is the most common variogram model used in kriging.

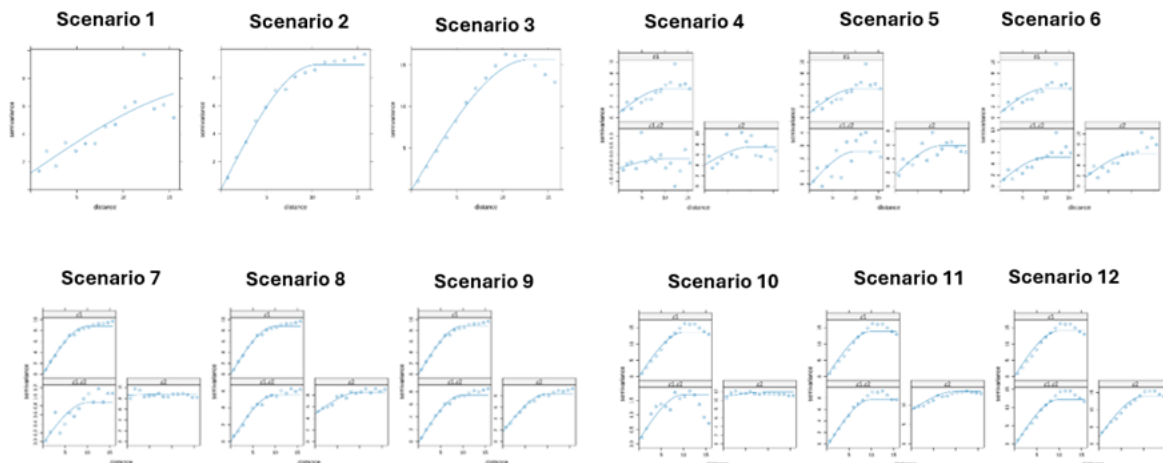


Figure 2: Variogram of each scenario

In the Cokriging scenario, the cross-variogram structure between variables Z_1 and Z_2 also shows a stable pattern, especially in medium to high correlation scenarios. This can be seen from the smooth and symmetrical shape of the cross-variogram, reflecting a strong spatial relationship between the two variables. A good cross-variogram indicates that Cokriging can effectively utilize information from secondary variables.

Figure 2 shows examples of variogram and cross-variogram fitting for several scenarios. The shape of the variogram follows theoretical expectations: semivariance increases with distance until it reaches a threshold (sill), then stabilizes. In the secondary variables, the variogram fitting is slightly more varied but remains within reasonable limits. The cross-variogram between Z_1 and Z_2 shows spatial structure coherence, which is stronger as ρ increases.

Theoretically, good variogram fitting is essential in building a stable and accurate kriging model [1]. The inconsistency of the variogram structure is often the leading cause of spatial prediction failure. Therefore, the successful fitting across these scenarios indicates that the simulation has successfully mimicked realistic spatial conditions and is reliable for method evaluation.

3.3 Visualization of Spatial Prediction Results

Visualization of spatial prediction results provides a qualitative picture of how the model maps values at unobserved locations. Figure 3 shows the prediction result maps for several scenarios using Ordinary Kriging and Cokriging. It appears that Cokriging produces a smoother and more unified spatial pattern, especially in scenarios with high correlations between variables. The resulting spatial contours are more defined, with soft color gradations and natural transitions between regions.

In contrast, the prediction results from Ordinary Kriging in scenarios with limited data appear to have a rougher surface, with spatial variations that tend to be sharp and inconsistent. This indicates the model's limitations in filling spatial information gaps without the support of additional variables.

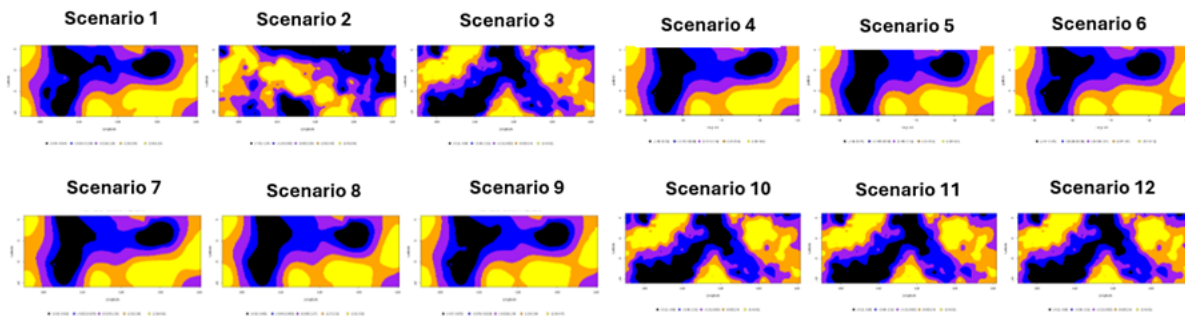


Figure 3: Visualization of each scenario's predictions

Figure 3 confirms that Cokriging is better able to produce realistic and stable prediction surfaces visually. Research by [18] supports this finding, stating that integrating secondary variables in spatial predictions allows the model to absorb additional information that can smooth the prediction contours.

3.4 Prediction Performance: Prediction vs Actual Visualization

This visualization not only adds an interpretative aspect but also supports the claim that Cokriging is visually superior when the correlation requirement between variables is met. To evaluate the model's estimation accuracy visually, Figure 4 presents a scatterplot between the actual and predicted values for each scenario. This plot allows direct observation of how close the predicted results are to the true values. Ideally, the points will form a straight line parallel to the identity line $y = x$, indicating perfect prediction.

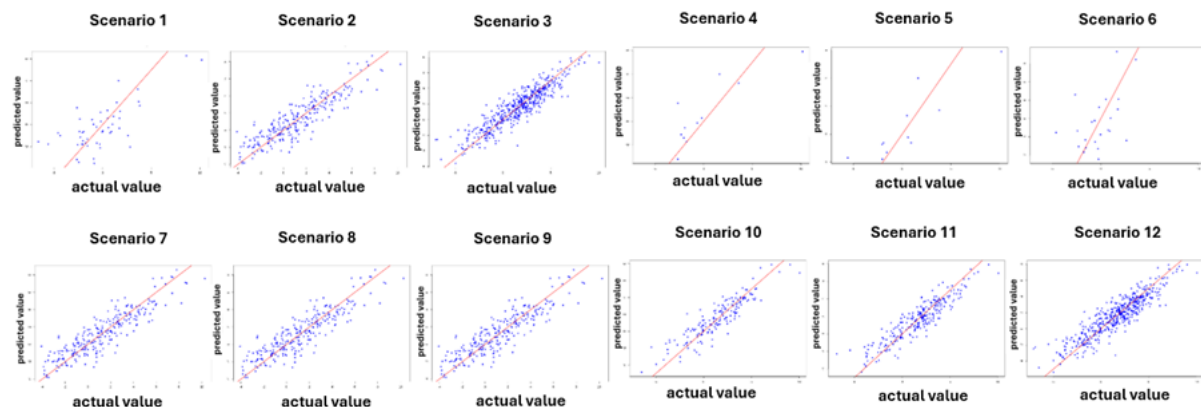


Figure 4: Visualization of predictions for each scenario

Figure 4 shows that in scenarios with high data volume and strong correlation, such as Scenario 12 ($n = 500$, $\rho = 0.9$), the prediction points tend to cluster near the identity line, indicating high accuracy. In contrast, in early scenarios such as Scenario 1 ($n = 50$, OK), the distribution of points appears much more spread out, and many predictions deviate significantly from their actual values.

Figure 4 illustrates the striking differences in performance between scenarios. Cokriging produces more consistent and precise predictions, especially in highly correlated scenarios. In contrast, Ordinary Kriging shows greater prediction fluctuations when information is limited, particularly without the support of secondary variables.

3.5 Numerical Evaluation: RMSE and R^2

Based on the evaluation results of 12 scenarios, each tested 30 times, the Ordinary Kriging (OK) method showed an increase in performance as the amount of data increased. In the OK scenario with

$n = 50$, the RMSE value was recorded at 1.99, with the coefficient of determination R^2 only 0.492. However, when the number of points increased to $n = 500$, the RMSE dropped significantly to 1.06, and R^2 increased to 0.873. This indicates that the addition of observation data has a positive effect on the stability and accuracy of the model. However, the results are still below the performance of the Cokriging method.

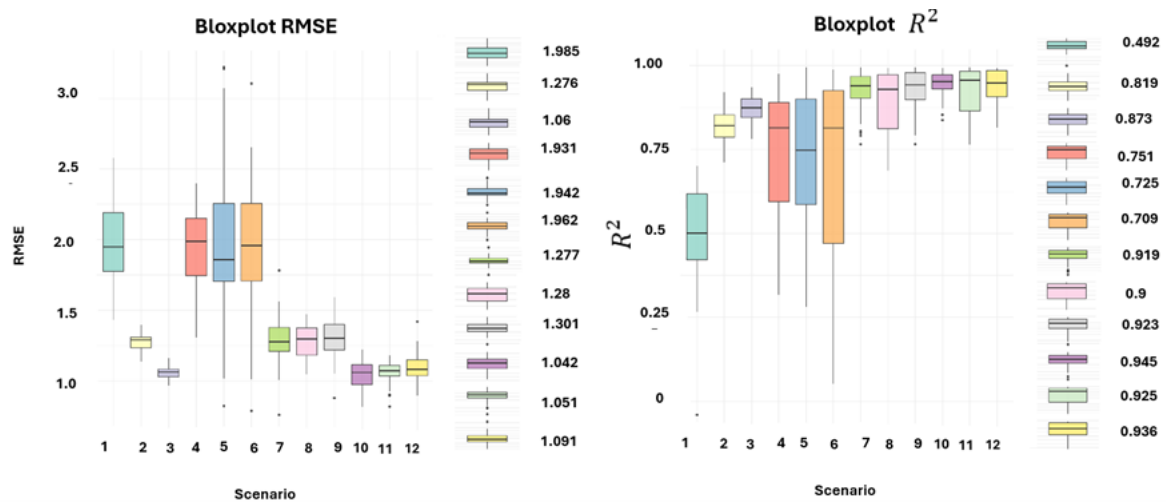


Figure 5: Visualization of each scenario's predictions

Figure 5 illustrates the distribution of RMSE and R^2 . The Cokriging (CK) method shows consistent superiority over Ordinary Kriging (OK) in all scenario combinations. In the scenario with $n = 500$, Cokriging produces very low and stable RMSE values: 1.04 for low correlation, 1.05 for medium correlation, and 1.09 for high correlation. The R^2 values in the three configurations are also high and relatively even: 0.945, 0.925, and 0.936, respectively. Although the RMSE increases slightly at high correlation, all Cokriging configurations at $n = 500$ still show more accurate and predictive performance than Ordinary Kriging. This is in line with previous studies showing that the Cokriging method provides more accurate estimates compared to the Ordinary Kriging method or other interpolation methods [19, 20, 21].

These findings provide a strong quantitative basis for the conclusion that Cokriging provides better prediction performance than Ordinary Kriging. Therefore, selecting the best method should consider both the number of observation points and the availability of informative secondary variables. Based on all combinations of results, the best scenario in this study is Cokriging with $n = 500$ at all correlation levels, having the lowest RMSE (1.04) and the highest R^2 (0.945). This scenario provides the most optimal balance between prediction accuracy and spatial stability.

4 Conclusion

The results of this study indicate that the Cokriging method has a consistent statistical advantage over Ordinary Kriging in terms of spatial prediction accuracy. The evaluation was carried out on 12 scenarios that varied based on the number of observation points and the level of correlation between variables, each tested through 30 repetitions. The method was evaluated using two primary metrics: Root Mean Squared Error (RMSE) as an indicator of estimation error and the coefficient of determination (R^2) as an indicator of predictive strength. In all configurations with $n = 500$, Cokriging recorded excellent predictive performance, with RMSE reaching 1.04 and R^2 reaching 0.945. In contrast, Ordinary Kriging, although showing improved performance as the number of points increased, was only able to achieve a minimum RMSE of 1.06 and a maximum R^2 of 0.873. These results strengthen the understanding that using additional information from spatially correlated secondary variables can significantly improve the accuracy of the interpolation model. Cokriging proved to be more efficient in minimizing errors and explaining the structure of data variability, especially in large data configurations commonly found in

spatial mapping and environmental geostatistical studies. Therefore, Cokriging is recommended as a superior spatial estimation method for multivariate data with a strong spatial structure.

CRediT Authorship Contribution Statement

Siti Mutiah contributed to methodology, software development, formal analysis, data curation, original draft writing, and visualization. Muhammad Nur Aidi, as the main supervisor, contributed to conceptualization, supervision, validation, and review & editing of the manuscript. Asep Saefuddin contributed to conceptualization, supervision, and manuscript review. Fitrah Ernawati contributed to conceptualization, supervision, validation, and manuscript review.

Declaration of Generative AI and AI-assisted technologies

The author used the generative AI tool ChatGPT (version 4, OpenAI) to support idea development and exploration in code writing and LaTeX formatting. In addition, Grammarly was used to assist with grammar checking and refinement of the English text. All outputs generated by these tools were manually reviewed and adjusted to ensure accuracy, originality, and academic integrity.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research was supported by the Department of Statistics and Data Science, School of Data Science, Mathematics, and Informatics (SSMI), IPB University. The author also acknowledges financial support from the Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of the Republic of Indonesia. The author thanks Muhammad Nur Aidi, Asep Saefuddin, and Fitrah Ernawati for their valuable guidance and supervision.

Data Availability

Data Availability

The data and R scripts used in this study were generated and executed in Google Colab. The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

- [1] N. Cressie, *Statistics for Spatial Data*. New York: John Wiley & Sons, INC, revised ed., 1993.
- [2] N. Nur Rohma, "Pendugaan metode ordinary kriging," *Jurnal Penelitian Ilmu Sosial dan Eksakta*, vol. 2, no. 1, pp. 21–29, 2022.
- [3] O. Arnilia, M. N. Aidi, Indahwati, and F. Ernawati, "Prediction of pneumonia patient proportions using kriging and inverse distance weighted methods," *International Journal of Scientific Research in Science, Engineering and Technology*, vol. 4099, pp. 391–399, 2023.

- [4] S. M. Meliyana R. and A. S. Ahmar, "Analysis of the distribution of diarrhea patients in bogor regency month using the ordinary kriging method," *ARRUS Journal of Social Sciences and Humanities*, vol. 3, no. 2, pp. 156–162, 2023.
- [5] D. Ardianti, H. Pramodyo, and N. Nurjannah, "Distance and areas weighting of gwr kriging for stunting cases in east java," *CAUCHY Jurnal Matematika Murni dan Aplikasi*, vol. 6, no. 4, pp. 212–217, 2021.
- [6] T. M. Tebeje, B. L. Seifu, K. U. Mare, and Y. S. Asgedom, "Geospatial determinants and spatio-temporal variation of early initiation of breastfeeding and exclusive breastfeeding in ethiopia," *BMC Public Health*, vol. 24, p. 2011, 2024.
- [7] Y. Flores, A. Chaves, and G. Suza, "Prediction of edapho-climatic parameters in the incidence of campylobacter spp. in northwestern mexico," *PLOS Global Public Health*, vol. 2016, pp. 1–22, 2024.
- [8] S. Y. Kassie, H. S. Ngusie, A. W. Demsash, and T. D. Alene, "Spatial distribution of short birth interval and associated factors among reproductive age women in ethiopia," *BMC Pregnancy and Childbirth*, vol. 23, no. 1, pp. 1–14, 2023.
- [9] C. Yuan and X. Wang, "Source apportionment and health risk assessment of heavy metals in soils of old industrial areas – a case study," *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, p. 2395, 2023.
- [10] N. N. Rohma, "Perbandingan pendugaan metode ordinary kriging dan metode ordinary kriging dengan teknik jackknife," *Mathematics Applied Journal (MAJ)*, vol. 4, no. 2, pp. 101–111, 2023.
- [11] M. R. N. N. D. Hendra Perdana, "Metode cokriging untuk mengestimasi kandungan air di kota pontianak," *Bimaster Buletin Ilmiah Matematika Statistika dan Terapannya*, vol. 8, no. 4, pp. 933–942, 2019.
- [12] M. A. Lasaiba, "Pengolahan data spasial dalam perencanaan penggunaan lahan yang berkelanjutan," *Geoforum*, vol. 2, no. 1, pp. 1–12, 2023.
- [13] H. N. Fajri, R. Johar, and M. Ikhsan, "Peningkatan kemampuan spasial dan self-efficacy siswa melalui model discovery learning berbasis multimedia," *Jurnal Tadris Matematika*, vol. 9, no. 2, pp. 180–196, 2016.
- [14] E. H. Isaaks and R. M. Srivastava, *Applied Geostatistics*. Oxford University Press, 1990.
- [15] C. D. Rosita, "Kemampuan penalaran dan komunikasi matematis: Apa, mengapa, dan bagaimana ditingkatkan pada mahasiswa," *Jurnal Euclid*, vol. 1, no. 1, pp. 33–46, 2014.
- [16] T. P. Robinson and G. Metternicht, "Testing the performance of spatial interpolation techniques for mapping soil properties," *Computers and Electronics in Agriculture*, vol. 50, no. 2, pp. 97–108, 2006.
- [17] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*. Oxford University Press, 1997.
- [18] H. Wackernagel, *Multivariate Geostatistics: An Introduction with Applications*. New York: Springer, 1995.
- [19] 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 dan Teknologi Lingkungan*, vol. 8, no. 1, pp. 73–86, 2022.
- [20] C. Maharza and N. F. Rahmah, "Pengaruh number of neighbors terhadap perhitungan ordinary kriging dan cokriging," *ReTH Journal*, pp. 170–176, 2020.

- 312 [21] A. P. Nugraha, A. G. Irwan, and L. Nugraha, “Perbandingan geostatistik metode kriging dan co-
313 kriging menggunakan estimasi point kriging,” *ReTH Journal*, pp. 177–181, 2020.