MIMS Majalah Ilmiah Matematika dan Statistika 25 (2), 2025, 47 – 63
eISSN 2722-9866 pISSN 1411-6669
https://mims.jurnal.unej.ac.id/index.php/MIMS/index

Spatial modeling of hotel prices in the Yogyakarta city area using ordinary kriging and cokriging approaches

Hannura Adriana, Achmad Fauzan*

Department of Statistics, Faculty of Mathematics and Natural Sciences,
Universitas Islam Indonesia, Sleman, 55584, Indonesia.
*Corresponding author: achmadfauzan@uii.ac.id

Received: 05-04-2025, accepted: 22-09-2025

Abstract

Yogyakarta, renowned as one of Indonesia's most prominent tourist destinations, owes its appeal to its natural beauty, well-preserved environment, and rich cultural and historical heritage. Its reputation as a safe and comfortable destination has led to a consistent annual increase in tourist arrivals. Consequently, there is a growing demand for hotel accommodations that offer competitive pricing to avoid financial losses while meeting tourist expectations. Tourists often rely on ratings and reviews of hotel services and facilities, making these factors significant determinants of pricing strategies. This study aims to provide spatially informed pricing recommendations for potential hotel developments in Yogyakarta using kriging spatial interpolation methods. Two kriging approaches were employed: Ordinary Kriging (OK) and Cokriging (CK), incorporating hotel price and rating data as primary variables. The analysis identified the Exponential semivariogram model as optimal for OK and the Spherical semivariogram model as optimal for CK. Both methods predicted hotel prices around Yogyakarta City to range from IDR 300,000 to IDR 400,000 for locations farther from the city center. Among the two methods, CK demonstrated superior predictive performance, yielding lower Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) values compared to OK. These findings highlight the potential of CK for providing accurate and actionable insights into hotel pricing strategies, offering valuable guidance for stakeholders considering investments in Yogyakarta's thriving hospitality industry.

Keywords: Cokriging, hotel price, interpolation, ordinary kriging

MSC2020: 62H11

1. Introduction

Yogyakarta is one of the leading tourist destinations in Indonesia and the international arena. This reputation is supported by its commitment to preserving nature and the environment, as well as maintaining cultural heritage sites and historical landmarks. Additionally, the people of Yogyakarta uphold traditional Javanese values, which are reflected in customs, language, social interactions, and arts. The region's relatively high level of safety and comfort makes Yogyakarta a favorite destination for tourists, leading to an annual increase in visits from both domestic and international travelers [1]. The high number of tourists visiting Yogyakarta underscores the importance of accommodation needs, particularly hotels. Setting the selling prices offered by hotels must be carried out with precision and careful consideration.

In addition to hotel prices, hotel ratings also play a significant role in strategies to attract tourists. Hotel ratings are customer evaluations typically represented in the form of stars, reflecting their opinions on the services or other aspects they experienced during their stay. These ratings also serve as a form of word-of-mouth communication to share a customer's experience [2], [3]. Customer satisfaction contributes to the assessment of the quality of services provided. Customer reviews can also provide valuable feedback for hotels to improve their services and facilities, thereby attracting more customers in the future [4]. Based on this, determining competitive hotel pricing is a crucial aspect for management in securing customers.

The spatial interpolation approach serves as an alternative for predicting hotel prices at specific locations based on existing hotel price data. Spatial interpolation is a method used to estimate values at locations without data by utilizing information from surrounding locations with known values [5], [6]. Essentially, spatial interpolation assumes that attribute data are continuous across space [7], [8] and is widely used to generate continuous data when observations are collected at discrete locations within a Geographic Information System (GIS). One of the most popular and reliable methods in spatial interpolation is kriging [9], [10].

Kriging is a geostatistical interpolation method used to estimate values at locations without data by utilizing the spatial relationships or correlations between nearby locations [11], [12]. Ordinary Kriging (OK) is a specific Kriging method that estimates values at unsampled locations under the assumption that the mean value in the area is unknown but constant [13], [14]. Several studies have explored the application of OK, including Yanto's research, which identified landslide-prone areas using hard soil depth in a small mountainous region of western Central Java, an area frequently affected by landslides [15]. Similarly, Pirestani (2024) utilized OK to predict the distribution of soil properties, providing insights for sustainable agricultural planning and soil resource management [16]. Chalaba [17] applied OK to analyze the spatial distribution of Soil Organic Carbon (SOC), producing maps that serve as guides for soil management and optimizing soil sampling strategies.

In addition to Ordinary Kriging (OK), Cokriging (CK) extends the approach by incorporating secondary variables to improve prediction accuracy [18], [19]. The CK approach is based on the assumption that the primary variable is correlated with the secondary variable, enabling more precise interpolation estimates [20]. Several previous studies have applied this method, including Eko (2021), who used CK to analyze chloride content and pH, finding that the estimated chloride content was below the maximum safe limit for consumption [21]. Anik (2019) applied CK to investigate monthly rainfall variables in West Java, comparing isotropic and anisotropic variograms. The isotropic variogram yielded the smallest prediction errors, with Root Mean Square Error (RMSE) values ranging from 0.54 to 1.46 [22]. Similarly, Eka [23] applied CK using oil production as the primary variable and natural gas production as the secondary variable across nine well locations. Meanwhile, Naomi (2019) utilized CK to study total dissolved solids and color content in water. Among the three cross-semivariogram models analyzed, the spherical model provided the best results [24].

Despite its significant potential, the application of spatial interpolation methods for hotel price estimation remains underexplored, particularly when involving multiple variables rather than a single variable. This issue becomes even more pertinent in the Yogyakarta region, which is notably one of the top tourist destinations in Indonesia [25]. Therefore, examining spatial interpolation techniques, such as Ordinary Kriging (OK) for single-variable analysis and Co-Kriging (CK) for multi-variable analysis, constitutes a compelling and essential area of research to support the development of the tourism sector in this region. This study focuses on comparing two geostatistical methods, namely OK and CK to predict hotel prices in the Yogyakarta region. The primary objective of this analysis is to determine which method provides the best predictive results, considering spatial aspects and the correlation of additional variables that influence prices. The findings of this study are expected to benefit various stakeholders, such as hotel managers and tourism businesses in developing competitive pricing strategies, local governments and urban planners in supporting tourism development and regional planning, as well as investors and property developers in identifying strategic locations for investment. Furthermore, this study also contributes to the academic and research community by enriching the literature on the application of geostatistical methods, while indirectly helping tourists understand hotel price variations in the region.

2. Methods

In this study, the data used are secondary data. Data collection was conducted on March 24, 2023, covering the area around Yogyakarta City, located at coordinates 110.344°–110.43° E and 7.826°–7.745° S, as visualized in Figure 1. The variables analyzed include the primary variable, which is hotel price, the secondary variable, which is hotel rating, and distance variables represented by the longitude (east) and latitude (south) coordinates of each hotel location.

The steps involved in the estimation process using the OK and CK methods are generally presented in the flowchart in Figure 2. In this study, QGIS was applied for data preprocessing and basic visualization [26], whereas the subsequent analytical procedures were carried out using R [27]. The integration of these two software tools ensured both efficient spatial data handling and rigorous statistical analysis.

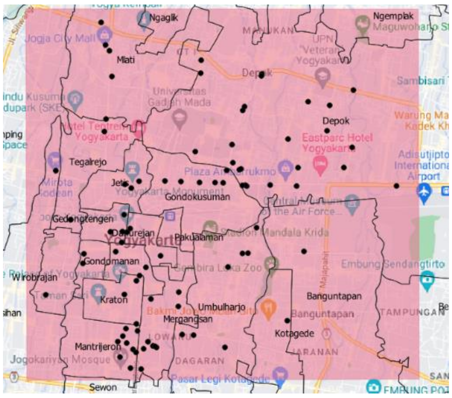

Figure 1. Map of observation point distribution around Yogyakarta city

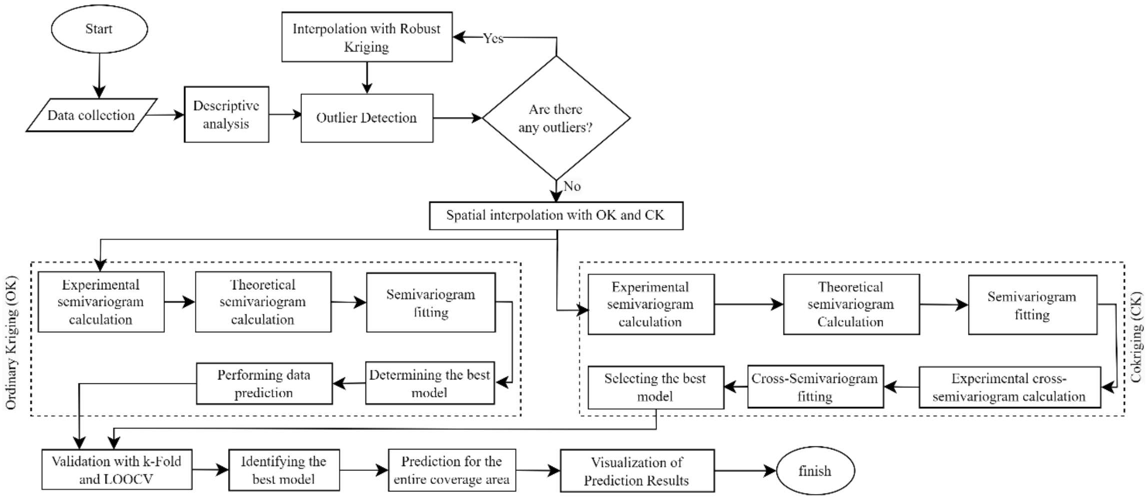

Figure 2. Research flowchart

2.1 Ordinary Kriging

Based on Figure 2, the stages of Ordinary Kriging (OK) are as follows.

1. The initial step involved conducting descriptive statistical analysis to present an overview of the data distribution and characteristics.
2. The second stage involves determining the experimental semivariogram values and selecting the theoretical semivariogram model. The experimental semivariogram is derived from observational data, which is then visualized as a plot function of distance. The experimental semivariogram is used to measure the extent to which values at one point are no longer correlated with those at another point [28]. The semivariogram calculation is conducted using Equation (1) [29].

$$\gamma(h) = \frac{1}{2N_h}\sum_{i=1}^{N(h)}[z(x_i) - z(x_i + h)]^2 \tag{1}$$

$\gamma(h)$: semivariogram value, h: distance between sample location, $N(h)$: number of sample point pairs with distance h, $z(x_i)$: observed value at location x_i, $z(x_i + h)$: observed value at location $x_i + h$. The experimental semivariogram derived from data often exhibits irregular patterns, making it challenging to interpret and unsuitable for direct analysis. Therefore, a structural analysis is necessary to fit the experimental semivariogram to a theoretical semivariogram model, ensuring a smooth and continuous covariance pattern [30]. In developing theoretical semivariograms, three commonly used models are: (1) spherical, (2) exponential, and (3) gaussian. These models include three key parameters: (1) sill, representing the maximum variance, (2) range, indicating the distance where spatial correlation becomes negligible, and (3) nugget, accounting for measurement errors or microscale variations [31]. The equations for these theoretical semivariogram models are summarized in Table 1 [32].

Table 1. Theoretical semivariograms

Number	Theoretical Semivariograms	Semivariogram Calculation
1	Spherical	$\gamma(h) = \begin{cases} C_0 + C\left[\left(\frac{3h}{2a}\right) - \left(\frac{h}{2a}\right)^3\right], & 0 < h \le a \\ C_0 + C, & h > a \end{cases}$
2	Exponential	$\gamma(h) = C_0 + C\left[1 - \exp\left(-\frac{h}{a}\right)\right]$
3	Gaussian	$\gamma(h) = C_0 + C\left[1 - \exp\frac{-h^2}{a^2}\right]$

C_0: nugget effect, C: partial sill, dan a: range.

3. The third stage involves fitting the experimental semivariogram with the theoretical semivariogram. This process is followed by selecting the best theoretical semivariogram model based on the smallest Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), as defined in Equation (3) and Equation (4) [33]. This stage aims to identify the most accurate model for representing the spatial pattern.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{4}$$

y_i: actual data, $\hat{y}_i$: prediction data, and n: data size.

4. The final step is to perform predictions using the best-fitting theoretical semivariogram model. This prediction is carried out by applying Equation (5), resulting in the most accurate estimations based on the previously identified spatial relationships [34].

$$\hat{z}(x) = \sum_{i=1}^{n} w_i(z(x_i)) \tag{5}$$

$\hat{z}(x)$: predicted value for variable x, w_i: weighting to determine the distance between points, $i = 1,2,\dots,n$: the number of sample points used in the interpolation for predicting the value at location x, and $z(x_i)$: actual value of variable x in the i^{th} data point. The weights used in the OK method are determined using Equation (6).

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \varphi \end{pmatrix} = \begin{pmatrix} \gamma(s_1, s_1) & \cdots & \gamma(s_1, s_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(s_n, s_1) & \cdots & \gamma(s_n, s_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(s_0, s_n) \\ \vdots \\ \gamma(s_0, s_n) \\ 1 \end{pmatrix} \tag{6}$$

s shows the location of row-*n* and column-*n*. or can be expressed as Equation (7).

$$\mathbf{w} = \mathbf{A}^{-1}\mathbf{b} \tag{7}$$

w: weight that determine the distance between points, **A**: covariance matrix between actual observations, and **b**: covariance matrix between actual observations and predictions.

2.2 Cokriging

The next step involves spatial interpolation analysis using Cokriging (CK). The interpolation process with CK consists of several stages, which are described in detail in the subsequent section.

1. First, calculate the experimental semivariograms for the hotel price and rating variables to understand the spatial relationships of each variable.
2. Subsequently, theoretical semivariograms are computed using the same three models as in OK.
3. Fitting the experimental semivariograms to the theoretical models for both variables, hotel price and rating, to determine the most suitable semivariogram model for each variable.
4. Calculating the cross-semivariogram for both variables using Equation (8).

$$\hat{\gamma}_{12}(h) = \frac{1}{2N(h)} \sum_{N(h)} [(Z_1(s) - Z_1(s+h))(Z_2(s) - Z_2(s+h)] \tag{8}$$

$Z_1(s), Z_1(s+h)$: measured value of the primary variable at the sample location *s* and *s*+*h*, $Z_2(s), Z_2(s+h)$: measured value of the secondary variable at the sample location *s* and *s*+*h*, $N(h)$: number of data pairs at *s* and *s*+*h* with the same distance difference.

5. The subsequent steps in data analysis using the CK method involve several critical stages. First, the theoretical cross-semivariogram for the two variables, hotel price and rating, is calculated using the same models as those applied in the theoretical semivariogram calculation for each variable. Second, the experimental and theoretical cross-semivariograms for the two variables are fitted, followed by selecting the best model based on the smallest RMSE and MAPE values. Finally, hotel price prediction is performed using the best-fitting theoretical cross-semivariogram model by applying Equation (9) [35].

$$\left(\hat{Z}(s_0)\right) = \sum\nolimits_{i=1}^{n} \alpha_i Z_1(s_i) + \sum\nolimits_{j=1}^{m} \beta_j Z_2(s_j) \tag{9}$$

$\hat{Z}(s_0)$: predicted value, $Z_1(s_i)$: primary variable data, $Z_2(s_i)$: secondary variable data, α_i: weight for the primary variable, and β_i: weight for the secondary variable. The weights used in the CK method are determined using Equation (10).

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \\ \beta_1 \\ \vdots \\ \beta_m \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \gamma_{11}(s_1,s_1) & \cdots & \gamma_{11}(s_n,s_1) & \gamma_{21}(s_1,s_1) & \cdots & \gamma_{21}(s_m,s_1) & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{11}(s_1,s_n) & \cdots & \gamma_{11}(s_n,s_n) & \gamma_{21}(s_1,s_n) & \cdots & \gamma_{21}(s_m,s_n) & 1 & 0 \\ \gamma_{12}(s_1,s_1) & \cdots & \gamma_{12}(s_n,s_1) & \gamma_{22}(s_1,s_1) & \cdots & \gamma_{22}(s_m,s_1) & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{12}(s_1,s_m) & \cdots & \gamma_{12}(s_n,s_m) & \gamma_{22}(s_1,s_m) & \cdots & \gamma_{22}(s_m,s_m) & 0 & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_{11}(s,s_1) \\ \vdots \\ \gamma_{11}(s,s_n) \\ \gamma_{12}(s,s_1) \\ \vdots \\ \gamma_{12}(s,s_m) \\ 1 \\ 0 \end{bmatrix} \quad (10)$$

After the OK and CK models are developed, the two models are compared by calculating RMSE and MAPE using k-Fold Cross Validation and Leave-One-Out Cross Validation (LOOCV) techniques. The k-Fold Cross Validation works by dividing the data into k folds and conducting the analysis iteratively for k iterations [36]. LOOCV, on the other hand, tests each data point individually, with the remaining data used as training data [37]. A smaller MAPE value indicates better model performance, as it reflects a lower average percentage error in predictions compared to actual values [38], while a smaller RMSE value signifies greater model accuracy [39]. The final step is to visualize the predicted hotel prices based on the best-performing method.

3. Results and Discussion

3.1 Descriptive Statistics

Based on the collected data, hotel prices in Yogyakarta City vary, with the lowest price recorded at Rp178,189 and the highest at Rp785,949. The average hotel price in this area is Rp383,358. For the hotel rating variable, the lowest rating is 3.7, while the highest rating is 4.9. This wide price range indicates that Yogyakarta accommodates a diverse market segment, from budget travelers to those seeking more comfortable accommodations. The average hotel price serves as a market benchmark, appealing to general travelers, while hotels priced below the average cater to budget-conscious visitors.

The high range of ratings (3.7–4.9) reflects the overall good quality of service provided by hotels in Yogyakarta. A minimum rating of 3.7 indicates satisfactory service, while a maximum of 4.9 suggests premium service standards offered by some hotels. This data also opens the opportunity to analyze the correlation between price and rating, providing insights into whether higher prices correspond to better service quality. Hotels with below-average prices but high ratings could leverage this combination as a competitive advantage in marketing, while hotels with high prices and lower ratings may need to improve their services to meet customer expectations. Overall, this analysis highlights the segmentation of the tourism market in Yogyakarta and the potential for strategic improvements in the hospitality industry.

3.2 Correlation Between Research Variables

The primary variable used is hotel price, while hotel rating serves as the secondary variable. In CK, a correlation test between the two variables is necessary to ensure an adequate relationship. The results of the correlation test, using the Pearson method to measure the relationship between hotel price and rating, are presented in Table 1. This test aims to determine the spatial relationship level between the two variables used in the analysis. Based on the Pearson correlation test, the calculated t-value ($t_{calculate}$ = 8.336) is greater than the critical t-value (t_{table} = 1.663), indicating that the variables have a statistically significant relationship, meeting the requirements of the CK method.

3.3 Hotel Price Prediction Using Ordinary Kriging

Following the flowchart illustrated in Figure 2, the initial step in kriging interpolation involves calculating the experimental semivariogram for the hotel price variable. This calculation encompasses the number of data pairs within each class, the distances between data pairs, and the corresponding semivariogram values. The process begins with determining the distances between observation points, forming a distance matrix. These distances are subsequently categorized into discrete classes, each containing a specific number of data pairs. The experimental semivariogram is then computed for each class based on the observed values at the respective points and the number of data pairs. The results of these calculations are detailed in Table 2.

Table 2. Results of the experimental semivariogram for the hotel price variable

Class	Data pairs	Distance (degrees)	Experimental semivariogram
1	145	0.0033	0.0003
2	240	0.0080	0.0004
3	282	0.0134	0.0004
4	325	0.0187	0.0005
5	414	0.0241	0.0005
6	399	0.0292	0.0005
7	370	0.0348	0.0006
8	374	0.0399	0.0006
9	309	0.0452	0.0005
10	231	0.0506	0.0005
11	182	0.0559	0.0006
12	159	0.0614	0.0006
13	143	0.0661	0.0006
14	64	0.0715	0.0006
15	16	0.0765	0.0005

After deriving the experimental semivariogram, the next stage entails fitting the theoretical semivariogram using three principal models: spherical, exponential, and Gaussian. These models are characterized by three fundamental parameters: sill, range, and nugget. The parameter values are optimized through an iterative process to achieve the lowest MAPE. The results of the MAPE are summarized in Table 3.

Table 3. Theoretical semivariogram models for the hotel price variable

Model	**Parameter**			**MAPE**
	Sill	**Range**	**Nugget**	
Spherical	0.00057	0.05	0.0003	7.2095
Exponential	0.00059	0.02	0.0003	6.0096
Gaussian	0.00056	0.019	0.0003	6.2557

Based on Table 3, the Exponential model demonstrates the lowest MAPE value compared to the Spherical and Gaussian models. Therefore, the Exponential model is selected as the best-fitting model for spatial interpolation using the OK method. The fitting plot between the experimental semivariogram and the theoretical semivariogram based on the Exponential model is shown in Figure 3. This plot illustrates the alignment between the actual data and the generated model, supporting the accuracy of predictions derived from the model.

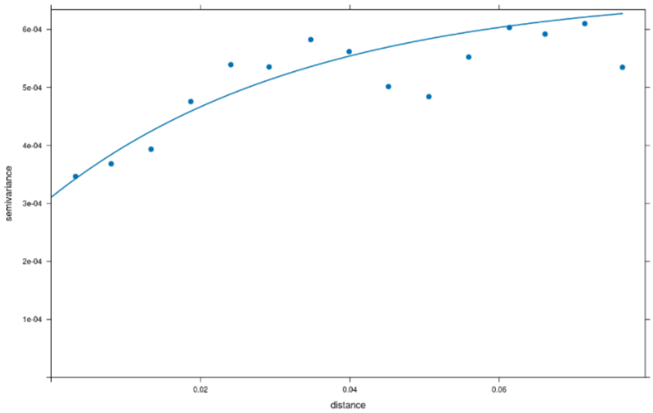

Figure 3. Fitting of the hotel price semivariogram

After obtaining the experimental and theoretical semivariograms using the Exponential model, the next step is to predict hotel prices using the OK methods. This prediction is performed by calculating the total product of hotel price data and the associated weights. The predicted hotel prices are visualized in the form of a map, as shown in Figure 4.

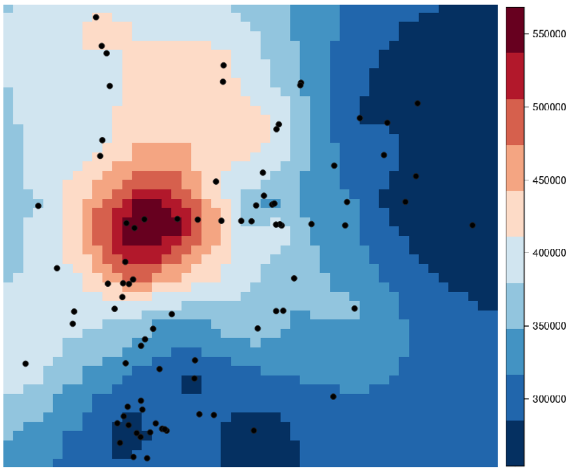

Figure 4. Visualization of predicted hotel prices using ordinary kriging

Figure 4 provides a spatial representation of the distribution of hotel prices within the study area. The black dots represent the locations of the hotels, while the color gradient from blue to red illustrates the predicted hotel prices using the OK method. Colors closer to red indicate higher hotel prices, whereas colors leaning towards blue represent lower prices. Based on this visualization, the majority of the map area is dominated by shades of dark blue to light blue, indicating that the predicted hotel prices in Yogyakarta and its surroundings are predominantly in the range of Rp 300,000 to Rp 400,000. This distribution suggests a significant clustering of moderately priced hotels, aligning with Yogyakarta's reputation as an affordable travel destination catering to a wide range of domestic and international tourists. The lack of extensive red zones may indicate a limited number of premium-priced hotels, highlighting opportunities for investment in upscale accommodations. Conversely, the prevalence of blue zones reflects a strong supply of budget-friendly options, catering to cost-sensitive travelers. This insight is valuable for hotel managers and investors in identifying market segments, as well as for local government and urban planners in balancing the development of tourism infrastructure with affordability and accessibility.

3.4 Spatial Interpolation Using Cokriging

In spatial interpolation using the CK method, two types of variables are required: primary and secondary variables. In this study, the primary variable is hotel price, while the secondary variable is hotel rating. Since the CK method involves two variables, it is necessary to calculate the experimental semivariogram for the hotel rating variable in the same manner as for the hotel price variable. The results of the experimental semivariogram calculation for the hotel rating variable are presented in Table 4. This semivariogram serves as the basis for understanding the spatial relationship of the secondary variable used in the analysis. In addition to calculating the experimental semivariogram for each variable, the CK method also requires the computation of the cross-semivariogram values between the hotel price and hotel rating variables.

Table 4. Results of the experimental semivariogram for the hotel rating variable

Class	Data pairs	Distance (degrees)	Experimental Semivariogram
1	145	0.0033	0.0357
2	240	0.0080	0.0323
3	282	0.0134	0.0301
4	325	0.0187	0.0343
5	414	0.0241	0.0368
6	399	0.0292	0.0393
7	370	0.0348	0.0418
8	374	0.0399	0.0467
9	309	0.0452	0.0350
10	231	0.0506	0.0352
11	182	0.0559	0.0377
12	159	0.0614	0.0387
13	143	0.0661	0.0445
14	64	0.0715	0.0395
15	16	0.0765	0.0406

Generally, the process for calculating the cross-semivariogram is similar to the steps used for calculating the experimental semivariogram. The difference lies in the data pairs, where the data used consist of combinations of hotel prices and hotel ratings. The results of the experimental cross-semivariogram calculation are presented in Table 5, which will be used to understand the spatial correlation between the two variables in the CK analysis.

Table 5. Results of the experimental cross-semivariogram

Class	Data pairs	Distance (degrees)	Experimental Semivariogram
1	290	0.0033	0.0023
2	480	0.0080	0.0021
3	564	0.0134	0.0023
4	650	0.0187	0.0025
5	828	0.0241	0.0032
6	798	0.0292	0.0028
7	740	0.0348	0.0035
8	748	0.0399	0.0038
9	618	0.0452	0.0027
10	462	0.0506	0.0025
11	364	0.0559	0.0027
12	318	0.0614	0.0023
13	286	0.0661	0.0026
14	128	0.0715	0.0033
15	32	0.0765	0.0043

Similar to spatial interpolation using the OK method, the CK method also requires selecting the best-fitting theoretical cross-semivariogram model to predict hotel prices. This process involves evaluating several models, such as spherical, exponential, and Gaussian, based on their resulting parameters. Table 6 presents the optimal parameter values for the theoretical cross-semivariogram of each model, including sill, range, and nugget. The model with the best performance will be utilized to predict hotel prices using the CK method.

Table 6. Results of the theoretical cross-semivariogram

Model	Parameters	Hotel Price	Hotel Rating	Cross Semivariogram	MAPE
Spherical	Sill	0.0006	0.0370	0.0030	11.623
	Range	0.0500	0.0700	0.0427	
	Nugget	0.0003	0.0300	0.0022	
Exponential	Sill	0.0006	0.0450	0.0033	13.9249
	Range	0.0210	0.0600	0.0347	
	Nugget	0.0003	0.0300	0.0022	
Gaussian	Sill	0.0006	0.0390	0.0031	13.7175
	Range	0.0190	0.0210	0.0219	
	Nugget	0.0030	0.0300	0.0022	

Based on Table 6, the spherical model exhibits the lowest MAPE value. Therefore, the spherical model is selected as the best-fitting model for performing spatial interpolation using the CK method. The fitting plot between the experimental and theoretical cross-semivariograms is presented in Figure 5, illustrating the alignment between the actual data and the theoretical model used for hotel price prediction.

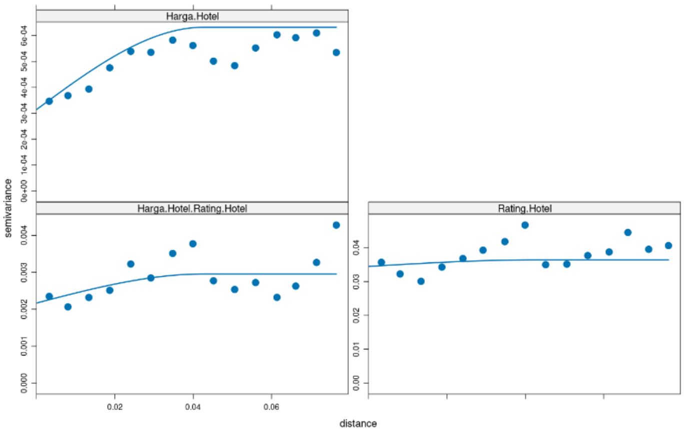

Figure 5. Cross-semivariogram fitting plot

After obtaining the experimental and theoretical cross-semivariograms using the Spherical model, the next step is to predict hotel prices using the CK method. The prediction process in the CK method is similar to that in the OK method, but there is a difference in the calculation of weights. In the CK method, the weight matrix includes the theoretical semivariogram values for the hotel price variable, the hotel rating variable, and the theoretical cross-semivariogram values that link the two variables. The visualization of the predicted hotel prices generated using the CK method is presented in Figure 6, providing a spatial representation of the predicted hotel price distribution. In Figure 6, the black dots represent hotel locations, while the color gradient from blue to red illustrates the predicted hotel prices using the CK method. The closer the color is to red, the higher the predicted hotel price, while blue represents lower prices. From the visualization in Figure 6, the majority of the area is dominated by shades of dark blue to light blue. This indicates that the predicted hotel prices in Yogyakarta City and its surroundings are predominantly within the range of Rp300,000 to Rp400,000.

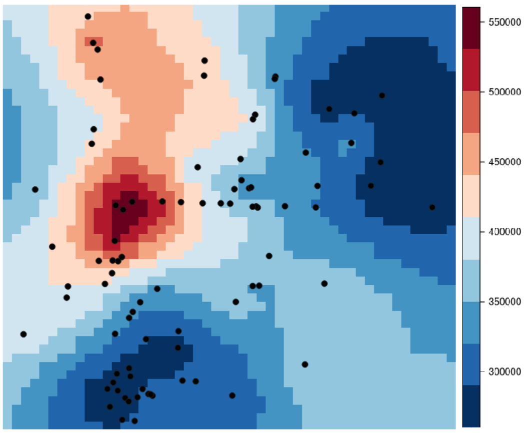

Figure 6. Visualization of predicted hotel prices using cokriging

This trend reflects Yogyakarta's position as a destination with an abundance of budget-friendly accommodations, catering to the majority of domestic and international tourists who are cost-conscious. The limited presence of red areas suggests fewer high-end or premium hotels, which could represent an opportunity for investors seeking to tap into the luxury travel market. Furthermore, the spatial distribution highlights specific areas within the city where higher-priced hotels are concentrated, possibly near major tourist attractions or central business districts. For hotel managers, this insight could guide pricing strategies, identifying areas where competitive pricing could attract a larger share of the market. Urban planners and local authorities could also use this information to support balanced tourism development, ensuring a diverse range of accommodations to meet varying traveler needs. Lastly, this visualization serves as a resource for tourists, helping them better understand the distribution of hotel prices and make informed decisions about where to stay.

3.5 Selection of the Best Model for Hotel Price Estimation

The results of RMSE and MAPE calculations from the 10-Fold Cross-Validation (CV) and Leave-One-Out Cross-Validation (LOOCV) approaches are presented in Table 7. The comparison of these results is used to determine the interpolation method that provides the most accurate hotel price predictions.

Table 7. Error values of the OK and CK methods

Methods	**10-Fold CV**		**LOOCV**	
	RSME	**MAPE**	**RMSE**	**MAPE**
Ordinary Kriging (OK)	132626	27.92694	133273.8	28.12133
Cokriging (CK)	94722.12	19.31397	94812.8	19.20952

Based on the comparison of RMSE and MAPE error values, the CK method demonstrates significantly better performance compared to the OK method. The smaller error values in the CK method indicate that hotel price predictions in Yogyakarta City are more accurate when utilizing CK, which incorporates an additional variable in the form of hotel ratings. The CK method has also proven effective in reducing error values during the spatial interpolation process. This improved performance highlights the value of integrating secondary variables that are closely correlated with the primary variable. By incorporating hotel ratings, the CK method captures additional nuances in the spatial distribution of hotel prices, which the OK method might overlook. This suggests that incorporating relevant supplementary data into spatial analyses can significantly enhance prediction accuracy.

4. Conclusion

Based on the analysis and discussion, the Exponential model was identified as the best-fitting theoretical semivariogram for the Ordinary Kriging (OK) method, whereas the Spherical model proved to be the most suitable cross-semivariogram for the Cokriging (CK) method. Among the two interpolation approaches, CK consistently demonstrated higher predictive accuracy in estimating hotel prices across Yogyakarta City and its surroundings. This superior performance, validated through both 10-Fold Cross-Validation and Leave-One-Out Cross-Validation, underscores the importance of incorporating secondary variables such as hotel ratings into spatial interpolation models. By leveraging this correlation, CK effectively captures additional spatial variability, resulting in more robust and precise predictions compared to single-variable approaches. Beyond its methodological contribution, this study enriches the growing body of literature on geostatistical methods applied in the tourism and hospitality context, an area that remains relatively underexplored. The findings also carry important practical implications: hotel managers can use CK-based predictions to set more competitive and market-aligned pricing strategies; investors and property developers can identify promising areas for new hotel development, especially in zones with limited premium accommodations; and policymakers or urban planners can incorporate these insights to design balanced and sustainable tourism infrastructure that accommodates diverse traveler needs. Ultimately, this study demonstrates how multivariate geostatistical modeling can bridge methodological rigor with practical decision-making, offering both academic value and actionable guidance for stakeholders in the tourism sector

Acknowledgement

We would like to express our gratitude to the Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, for providing research facilities and support throughout this study.

References

[1] D. P. D. I. Yogyakarta, *Statistik Kepariwisataan 2020*, 2021. [Online]. [CrossRef]

[2] Q. Ye, R. Law, B. Gu, and W. Chen, "The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings," *Comput Human Behav*, vol. 27, no. 2, pp. 634–639, Mar. 2011. [CrossRef]

[3] Z. Liu and S. Park, "What makes a useful online review? Implication for travel product websites," *Tour Manag*, vol. 47, pp. 140–151, Apr. 2015. [CrossRef]

[4] K. A. Fachrudin, D. L. Tarigan, and M. F. Iman, "Analisis rating dan harga kamar hotel bintang lima di Indonesia," *Jurnal Akuntansi, Keuangan, dan Manajemen (Jakman)*, vol. 3, no. 3, pp. 207–223, 2022. [CrossRef]

[5] A. Sekulić, M. Kilibarda, G. B. M. Heuvelink, M. Nikolić, and B. Bajat, "Random Forest Spatial Interpolation," *Remote Sens (Basel)*, vol. 12, no. 10, p. 1687, May 2020. [CrossRef]

[6] C. Munyati and N. I. Sinthumule, "Comparative suitability of ordinary kriging and Inverse Distance Weighted interpolation for indicating intactness gradients on threatened savannah woodland and forest stands," *Environmental and Sustainability Indicators*, vol. 12, p. 100151, Dec. 2021. [CrossRef]

[7] A. Akkala, V. Devabhaktuni, and A. Kumar, “Interpolation techniques and associated software for environmental data,” *Environ Prog Sustain Energy*, vol. 29, no. 2, pp. 134–141, Jul. 2010. [CrossRef]

[8] R. Flowerdew and M. Green, “Developments in areal interpolation methods and GIS,” in *Geographic Information Systems, Spatial Modelling and Policy Evaluation*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, pp. 73–84. [CrossRef]

[9] G. S. Bhunia, P. K. Shit, and R. Maiti, “Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC),” *Journal of the Saudi Society of Agricultural Sciences*, vol. 17, no. 2, pp. 114–126, Apr. 2018. [CrossRef]

[10] E. A. Varouchakis and D. T. Hristopulos, “Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins,” *Environ Monit Assess*, vol. 185, no. 1, pp. 1–19, Jan. 2013. [CrossRef]

[11] Z. Feng, X. Zong, T. Xie, and X. Zhang, “Kriging Model Averaging Based on Leave-One-Out Cross-Validation Method,” *J Syst Sci Complex*, vol. 37, no. 5, pp. 2132–2156, Oct. 2024. [CrossRef]

[12] M. Zaresefat, R. Derakhshani, and J. Griffioen, “Empirical Bayesian Kriging, a Robust Method for Spatial Data Interpolation of a Large Groundwater Quality Dataset from the Western Netherlands,” *Water (Basel)*, vol. 16, no. 18, p. 2581, Sep. 2024. [CrossRef]

[13] B. N. Ruchjana, A. N. Falah, and A. S. Abdullah, “Application of the Ordinary Kriging method for prediction of the positive spread of Covid-19 in West Java,” *J Phys Conf Ser*, vol. 1722, no. 1, p. 012026, Jan. 2021. [CrossRef]

[14] G. Heuvelink, E. Pebesma, G. B. M. Heuvelink, and E. J. Pebesma, “Is the ordinary kriging variance a proper measure of interpolation error?,” in *Proceedings of the fifth international symposium on spatial accuracy assessment in natural resources and environmental sciences*, Melbourne: RMIT University, 2002. [Online]. [CrossRef]

[15] Yanto, A. Apriyono, P. B. Santoso, and Sumiyanto, “Landslide susceptible areas identification using IDW and Ordinary Kriging interpolation techniques from hard soil depth at middle western Central Java, Indonesia,” *Natural Hazards*, vol. 110, no. 2, pp. 1405–1416, Jan. 2022. [CrossRef]

[16] N. pirestani, M. Ahmadi Nadoushan, M. H. Abolhasani, and R. Zamani Ahmadmahmoudi, “Mapping soil characteristics: spatio-temporal comparison of land use regression and ordinary kriging in an arid environment,” *Journal of the Indian Society of Remote Sensing*, vol. 52, no. 1, pp. 79–93, Jan. 2024. [CrossRef]

[17] L. M. Chalaba, A. Mulolwa, and O. Lungu, “Application of ordinary kriging in mapping soil organic carbon in Zambia,” *Pedosphere*, vol. 27, no. 2, pp. 338–343,

Apr. 2017. [CrossRef]

[18] 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. [CrossRef]

[19] R. Giraldo, V. Leiva, and C. Castro, "An overview of kriging and cokriging predictors for functional random fields," *Mathematics*, vol. 11, no. 15, p. 3425, Aug. 2023. [CrossRef]

[20] J. Cao, C. Li, Q. Wu, and J. Qiau, "Improved mapping of soil heavy metals using a vis-nir spectroscopy index in an agricultural area of eastern China," *IEEE Access*, vol. 8, pp. 42584–42595, 2020. [CrossRef]

[21] E. P. Putra, Goenjantoro, and Suyitno, "Penaksiran kandungan klorida di sungai Mahakam wilayah Samarinda tahun 2017 dengan metode cokriging," *EKSPONENSIAL*, vol. 11, no. 2, pp. 175–180, 2020. [CrossRef]

[22] A. Djuraidah, S. Rahardiantoro, and A. Desiwari, "Penerapan metode cokriging dengan variogram isotropi dan anisotropi dalam memprediksi curah hujan bulanan Jawa Barat," *Jurnal Meteorologi dan Geofisika*, vol. 20, no. 1, 2019. [CrossRef]

[23] E. O. Romaji, I. N. Latra, and Sutikno, "Estimasi produksi minyak dan gas bumi di Kalimantan Utara menggunakan metode cokriging," *Jurnal Sains dan Seni ITS*, vol. 5, no. 2, pp. 426–431, 2016. [CrossRef]

[24] N. N. Debataraja, D. Kusnandar, N. Imro'ah, and M. Rachmadiar, "Penerapan metode cokriging untuk mengestimasi jumlah zat padat terlarut pada air di permukiman kota Pontianak," *Jurnal Matematika Sains dan Teknologi*, vol. 20, no. 2, 2019. [CrossRef]

[25] Badan Pusat Statistik, "Perkembangan pariwisata desember 2023," Feb. 2024. [CrossRef]

[26] F. Ramdani, *Exploring the Earth with QGIS*. Cham: Springer Nature Switzerland. 2023. [CrossRef]

[27] K. Okoye and S. Hosseini, *R Programming*. Singapore: Springer Nature Singapore, 2024. [CrossRef] doi: 10.1007/978-981-97-3385-9.

[28] N. A. Alfiana, *Metode ordinary kriging pada geostatistika*. 2010. [CrossRef]

[29] A. A. Awali, H. Yasin, and R. Rahmawati, "Estimasi kandungan hasil tambang menggunakan ordinary indicator kriging," *Jurnal Gaussian*, vol. 2, no. 1, 2013. [CrossRef]

[30] T. C. Bailey and A. C. Gatrell, "Interactive spatial data analysis," *Interactive spatial data analysis*, 1995. [CrossRef]

[31] S. J. Pririzki, N. Sari, B. Agustin, and Y. R. E. Simbolon, "Penerapan model

semivariogram eksperimental pada curah hujan bulanan di Indonesia," *Fraction: Jurnal Teori dan Terapan Matematika*, vol. 2, no. 1, 2022. [CrossRef]

[32] G. Rozalia, H. Yasin, and D. Ispriyanti, "Penerapan metode ordinary kriging pada pendugaan kadar No2 di udara (studi kasus: pencemaran udara di kota Semarang)," *JURNAL GAUSSIAN*, vol. 5, no. 1, 2016. [CrossRef]

[33] L. S. Ihzaniah, A. Setiawan, and R. W. N. Wijaya, "Perbandingan kinerja metode regresi k-nearest neighbor dan metode regresi linear berganda pada data Boston housing," *Jambura Journal of Probability and Statistics*, vol. 4, no. 1, 2023. [CrossRef]

[34] G. Hatfield, "Spatial statistics," in *Practical Mathematics for Precision Farming*, 2018. [CrossRef]

[35] F. Ratnawati, "Implementasi algoritma naive bayes terhadap analisis sentimen opini film pada twitter," *INOVTEK Polbeng - Seri Informatika*, vol. 3, no. 1, 2018. [CrossRef]

[36] P. P. Sengupta *et al.*, "Proposed requirements for cardiovascular imaging-related machine learning evaluation (prime): a checklist: reviewed by the american college of cardiology healthcare innovation council," 2020. [CrossRef]

[37] S. Bates, T. Hastie, and R. Tibshirani, "Cross-Validation: What Does It Estimate and How Well Does It Do It?," *J Am Stat Assoc*, vol. 119, no. 546, pp. 1434–1445, Apr. 2024. [CrossRef]

[38] C. Yao *et al.*, "Two-stage detection algorithm for plum leaf disease and severity assessment based on deep learning," *Agronomy*, vol. 14, no. 7, p. 1589, Jul. 2024. [CrossRef]

[39] Y. Zhang, S. Li, Y. Wang, Y. Yan, J. Zhao, and Z. Gao, "Self-adaptive enhanced learning differential evolution with surprisingly efficient decomposition approach for parameter identification of photovoltaic models," *Energy Convers Manag*, vol. 308, p. 118387, May 2024. [CrossRef]