JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 10, No. 1, January 2026, pp. 1385-1399

Geostatistical Co-Kriging Approach for Estimating Total Coliform Bacteria in the Rivers of DKI Jakarta

Salwa Salsabila[1*], Dwi Agustin Nuriani Sirodj[1]
[1]Statistics Program Study, Universitas Islam Bandung, Indonesia
salwasalsabila764@gmail.com

Article History:
Received : 28-08-2025
Revised : 12-11-2025
Accepted : 14-11-2025
Online : 01-01-2026

Keywords:
Co-Kriging;
Spatial Interpolation;
Total Coliform Bacteria.

ABSTRACT

Spatial statistics and geostatistics are essential for analyzing spatially distributed data, particularly in environmental studies where data gaps are prevalent. However, limited studies have applied multivariate geostatistical approaches, particularly Co-Kriging (CK), to assess microbial contamination in tropical urban river systems, where pollution patterns are highly variable and data gaps are frequent. This study employs CK, a multivariate geostatistical interpolation technique, to estimate Total Coliform Bacteria concentrations in the rivers of DKI Jakarta, Indonesia. Total Coliform Bacteria served as the primary variable, with Biochemical Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) incorporated as secondary variables. A total of 120 sampling points were analyzed, with data collected by Dinas Lingkungan Hidup DKI Jakarta during the second monitoring period in June 2022. Semivariogram modelling identified the Gaussian model as the best fit, yielding the lowest root mean square error (RMSE) of 11.468, which performed better than both the Spherical and Exponential models. Model performance was further evaluated through Leave-One-Out Cross-Validation (LOOCV), in which one data point was systematically removed and re-estimated in multiple iterations to calculate the residuals and assess model accuracy. The CK analysis was performed using RStudio software. CK predictions closely matched observed concentrations, demonstrating strong model performance. At unsampled locations, the estimated mean Total Coliform Bacteria concentration was 7.711×10^6 MPN/100 ml with a standard deviation of 4.406×10^6 MPN/100 ml. The high variance indicates substantial spatial heterogeneity, likely driven by data outliers, weak spatial autocorrelation in COD, and low correlations between Total Coliform–COD and BOD–COD pairs. These findings highlight the potential of geostatistical CK to provide reliable spatial predictions of microbial contamination in urban river systems, thereby supporting evidence-based water quality monitoring and management in densely populated regions. The insights generated in this study can help environmental authorities in DKI Jakarta optimize monitoring strategies, prioritize pollution hotspot interventions, and strengthen urban river health management to protect public health and guide sustainable urban water governance.

Crossref
https://doi.org/10.31764/jtam.v10i1.34391

—————————— ◆ ——————————

A. INTRODUCTION

Spatial statistics is a branch of statistics that analyzes data with geographic locations, helping to identify spatial patterns and relationships. Geostatistics, a subfield of spatial statistics, is particularly useful for estimating values at locations where data are missing. One of the main challenges in spatial analysis is incomplete or sparse data, which can reduce estimation accuracy. To address this, spatial interpolation methods are used, which predict values at unsampled locations based on nearby sampled data (Conolly, 2020). This technique encompasses non-geostatistical methods, such as Inverse Distance Weighted (IDW), and geostatistical methods, such as Kriging. Compared to IDW method, Kriging provides higher accuracy due to its interpolation weight accounting not only for the distance between observation and estimation point but also for the spatial correlation among these points. The studies conducted by Singh & Verma (2019), Usowicz et al. (2021), Dewana et al. (2022) indicate that Kriging provides more accurate estimations, yielding lower Root Mean Square Error (RMSE) values compared to IDW.

Among interpolation methods, Kriging is widely used because it accounts for both the distance between points and their spatial correlation. Ordinary Kriging (OK) estimates a single variable assuming a constant mean across the study area. However, OK is limited when multiple correlated variables are available. Co-Kriging (CK) improves estimation accuracy by incorporating one or more secondary variables that are correlated with the target variable, allowing better predictions in areas with sparse data (Belkhiri et al., 2020; Bogunovic et al., 2021; Rostami et al., 2020).

In DKI Jakarta, water quality monitoring data show that approximately 70% of rivers are heavily polluted, with an average Total Coliform concentrations of 2.46×10^7 MPN/100 mL according to Dinas Lingkungan Hidup DKI Jakarta, indicating serious public health risks. While previous studies in Jakarta and other urban river systems have primarily focused on descriptive water quality monitoring, they have not applied Co-Kriging to estimate microbial contamination, compared different semivariogram models, or incorporated correlated water quality variables such as BOD and COD to enhance spatial prediction accuracy (Djembarmanah & Salsabila, 2024; Pratama et al., 2020). Therefore, the novelty of this study lies in applying Co-Kriging to model Total Coliform Bacteria in a tropical urban river system while integrating BOD and COD as secondary variables due to their influence on microbial growth and their practicality for routine monitoring.

Total Coliform Bacteria is chosen as the primary indicator of water quality because it reflects microbial contamination and potential health risks from fecal pollution(Rusdiyanto et al., 2021). The presence of these bacteria indicates the possible presence of pathogenic microorganisms that can cause illnesses such as gastrointestinal infections (diarrhea, nausea, vomiting), urinary tract infections, and skin and eye infections. These risks are particularly higher for vulnerable groups, including children, the elderly, and individuals with weakened immune systems.

For the purpose of this study, water quality monitoring data were collected from 120 sampled points spread across the main rivers of DKI Jakarta. Nevertheless, several locations remain unsampled in the monitoring process, necessitating detailed spatial estimation. Therefore, this study aims to identify the best semivariogram model for estimating the spatial distribution of Total Coliform Bacteria and to estimate the Total Coliform Bacteria concentrations at unsampled river locations in DKI Jakarta in 2022 using the Co-Kriging method with BOD and COD as secondary variables, in order to produce more accurate and detailed spatial estimates to support optimal river water quality management.

B. METHODS

This study employs secondary data on river water quality in DKI Jakarta Province for the year 2022, obtained from Dinas Lingkungan Hidup DKI Jakarta through the Satu Data Jakarta portal. The dataset includes 120 sampled points and 12 unsampled points, with variables of Total Coliform Bacteria (MPN/100 ml), Biochemical Oxygen Demand (BOD) (mg/L), and Chemical Oxygen Demand (COD) (mg/L). The data were collected by Dinas Lingkungan Hidup DKI Jakarta during the second monitoring period in June 2022. The spatial extent of the study covers the entire Jakarta region, excluding the Thousand Islands (*Kepulauan Seribu*). The research stages are illustrated in the flowchart shown in Figure 1.

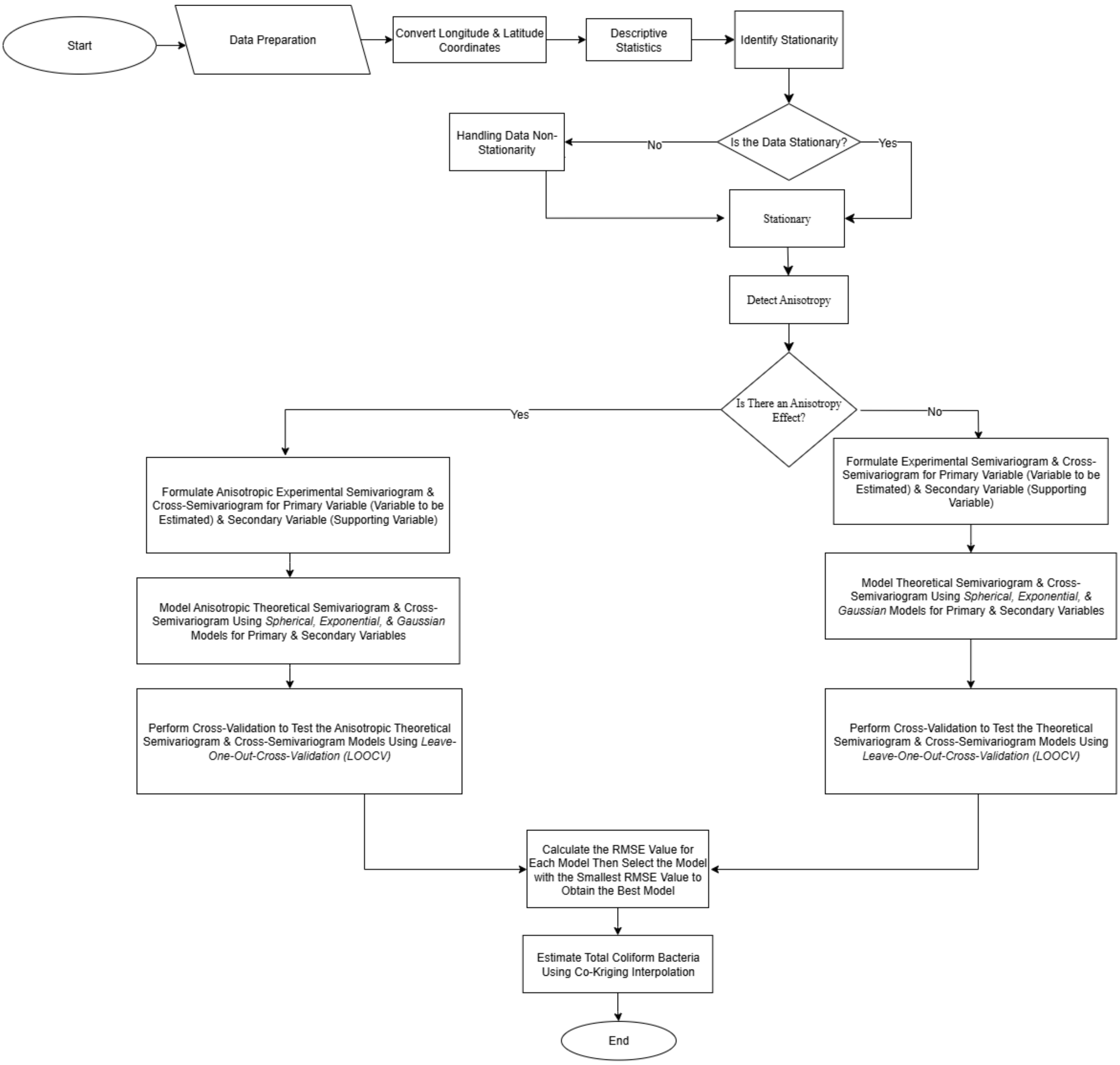

Figure 1. Research Flow Chart

1. Stationarity

Spatial dataset is considered stationary when it exhibits a random pattern without any discernible trend (Sikder & Züfle, 2020). The examination of stationary can be performed by plotting the data values against their spatial coordinates.

2. Isotropic and Anisotropic

A semivariogram that models the distance function while considering direction is called an anisotropic semivariogram, whereas a semivariogram that models distance without considering direction is called an isotropic semivariogram (Abbasnejadfard et al., 2021). If anisotropy is present in the data, distances between points are calculated by taking the main spatial direction into account. Conversely, if the data are isotropic, distances are calculated simply based on the coordinates of each point.

3. Experimental Semivariogram

The experimental semivariogram is calculated from the measurement data and plotted as a function of separation distance (Schiappapietra et al., 2022; Usowicz et al., 2021). It quantifies how the similarity between data points changes with distance and serves as the basis for fitting a theoretical semivariogram model for subsequent spatial interpolation.

4. Theoretical Semivariogram

There are several types of theoretical semivariogram models, including the spherical, exponential, and Gaussian models (Ding et al., 2018). These models are used to fit the experimental semivariogram and describe how spatial correlation changes with distance. Key parameters include the nugget effect (C_0), representing variability at very short distances; the partial sill (C), reflecting the increase in variance due to spatial correlation; and the range (A), which is the distance at which spatial autocorrelation becomes negligible.

5. Cross-Semivariogram

The cross-semivariogram is used to quantify the spatial relationship between two variables (Addis et al., 2023; Payares-garcia et al., 2023). It measures how the values of one variable co-vary with another variable over space, helping to improve the accuracy of multivariate spatial interpolation like Co-Kriging.

6. Cross-validation

Cross-validation was conducted to evaluate the performance of the candidate models(Lei, 2020). Specifically, Leave-One-Out Cross-Validation (LOOCV) was applied, where one data point is sequentially removed from the dataset and the model is recalibrated to predict that point (Liu et al., 2025). The removed data point is then estimated using the theoretical semivariogram model in the Co-Kriging method. This process is repeated for all data points, producing residuals derived from the differences between the estimated and observed values. The accuracy of each model was assessed using the Root Mean Square Error (RMSE), with lower RMSE values indicating better predictive performance (Pham et al., 2019).

7. Co-Kriging (CK)

Co-Kriging is a variant of Kriging that utilizes more than one correlated variable to enhance prediction accuracy (Xiao et al., 2018). The linear relationship for the estimated value at an unsampled points $\widehat{\boldsymbol{Z}}(s_0)$ in CK can be expressed in Equation (1) as follows (Dowd & Pardo-igúzquiza, 2024).

$$\left(\widehat{\boldsymbol{Z}}(s_0)\right) = \sum_{i=1}^{n}\sum_{j=1}^{k} \lambda_{ji} Z_j(s_i) \quad (1)$$

where $Z_j(s_i)$ is the variable used at the i-th location, while λ_{ji} represents the weight for the variable used at the i-th location. The weights can be calculated using Equation (2) as follows.

$$\boldsymbol{\lambda}_{+}(s_0) = \boldsymbol{C}_{+}^{-1}\boldsymbol{c}_{+}(s_0) \quad (2)$$

where $\boldsymbol{C}_{+}^{-1}$ is the inverse of the combined semivariance matrix and $\boldsymbol{c}_{+}(s_0)$ represents the semivariance vector for all variables at all locations with respect to s_0 (the unsampled location). Then, Equation (3) presents the minimum variance estimator in CK.

$$\boldsymbol{\sigma}_k^2(s_0) = \gamma_{11}(s_0, s_0) - \sum_{i=1}^{n}\sum_{j=1}^{k} \lambda_{ji}\, \gamma_{1j'}(s_0, s_i) + m_1 \quad (3)$$

C. RESULT AND DISCUSSION

1. Descriptive Statistics

This study utilized river water quality data from the Special Capital Region (DKI) of Jakarta in 2022. Figure 2 presents the distribution of the variables used in this study, namely Total Coliform Bacteria (Z_1), Biochemical Oxygen Demand (BOD, Z_2), and Chemical Oxygen Demand (COD, Z_3).

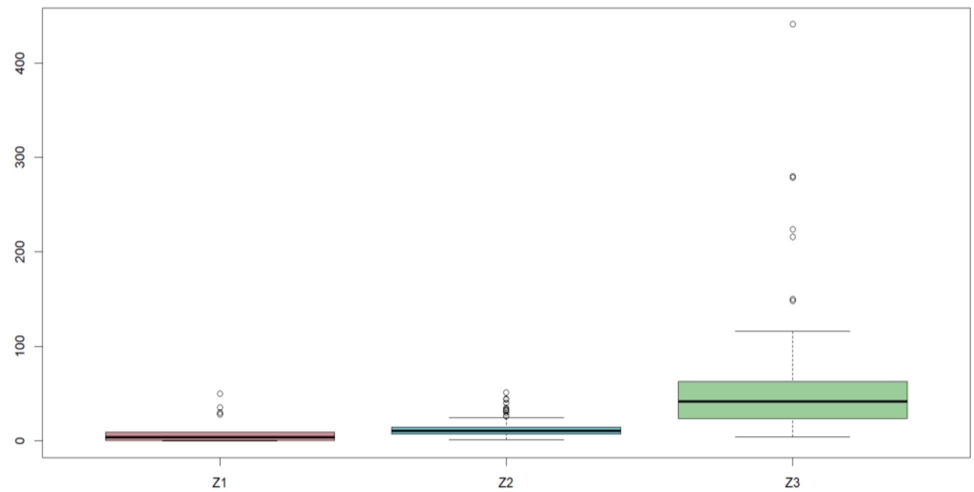

Figure 2. Distribution of Total Coliform Bacteria ($\boldsymbol{Z_1}$), Biochemical Oxygen Demand ($\boldsymbol{Z_2}$), and Chemical Oxygen Demand ($\boldsymbol{Z_3}$)

Figure 2 shows that each variable contains several extreme values (outliers), with Z_1 having 23 outliers, Z_2 having 13 outliers, and Z_3 having 8 outliers. These extreme values may result from multiple interacting factors, including: (1) localized point-source pollution, where specific river reaches receive concentrated domestic or industrial effluents; (2) spatial heterogeneity in river geomorphology and hydrodynamics, such as low-velocity zones, high sediment deposition, or stagnant sections that facilitate bacterial accumulation; (3) variation in anthropogenic pressures, particularly in densely populated and industrial corridors that exhibit pronounced temporal and spatial fluctuations in water quality; and (4) natural environmental drivers, including intense rainfall events, tidal influence, and surface runoff, which can abruptly elevate microbial and organic contamination levels(Bojarczuk et al., 2018).

These outliers are not removed, as Co-Kriging utilizes the full variability of the data to produce more accurate spatial estimates, including in areas with high pollution levels. Removing or treating these outliers would result in the loss of information and alter the characteristics of the data. The descriptive statistics of Total Coliform Bacteria (Z_1), Biochemical Oxygen Demand (Z_2), and Chemical Oxygen Demand (Z_3) from 120 sampling points are presented in Table 1 below.

Table 1. Descriptive Statistics of the Variables

Variable	Mean	Variance	Standard Deviation	Minimum	Maximum
Z1*	10.560	264.385	16.260	0.005	50.000
Z2	12.834	93.005	9.644	1.330	51.130
Z3	57.510	3656.041	60.465	4.30	441.000

*expressed in millions

Table 1 shows that the average value of Z_1 is 10.56 × 10^6 MPN/100 ml, indicating a high level of microbiological pollution since many rivers have exceeded the water quality standard set by Dinas Lingkungan Hidup DKI Jakarta (DLH) of 5,000 MPN/100 ml. The maximum value of 50 × 10^6 MPN/100 ml was found in the Sunter, Buaran, Patukangan, Cideng, and Kalibaru Timur rivers, all of which are located in densely populated downstream areas, while the minimum value of 0.005 × 10^6 MPN/100 ml was observed in the Tarum Barat and Angke rivers. The high variance (264.385 × 10^6 MPN/100 ml) and standard deviation (16.260 × 10^6 MPN/100 ml) indicate large variations across sampling points, likely influenced by differences in population density, river location (upstream or downstream), sanitation infrastructure, and domestic waste management, for Z_2, the mean is 12.834 mg/L, far above the DLH standard of 3 mg/L, with the highest value (51.130 mg/L) in the Mampang River, which is surrounded by dense settlements, traditional markets, and food stalls, while the lowest value (1.330 mg/L) occurs in the Tarum Barat River, which is not directly used as a domestic wastewater channel. The high variance (93.005) and standard deviation (9.644) suggest that organic pollution levels vary considerably among rivers. Meanwhile, Z_3 shows an average of 57.51 mg/L, also exceeding the DLH standard of 25 mg/L, with a maximum of 441.000 mg/L found in the Buaran River again located in a densely populated downstream area and the lowest value in the Tarum Barat River. The large variance (3,656.041) and standard deviation (60.465) further confirm the substantial variability of chemical pollution levels across sampling locations.

2. Stationarity Assumption

The Co-Kriging interpolation method assumes that the dataset used must be stationary, meaning that its statistical properties (such as mean and variance) do not vary systematically across space. To verify this assumption, spatial distribution plots were generated for Total Coliform Bacteria, Biochemical Oxygen Demand (BOD), and Chemical Oxygen Demand (COD), as shown Figure 3.

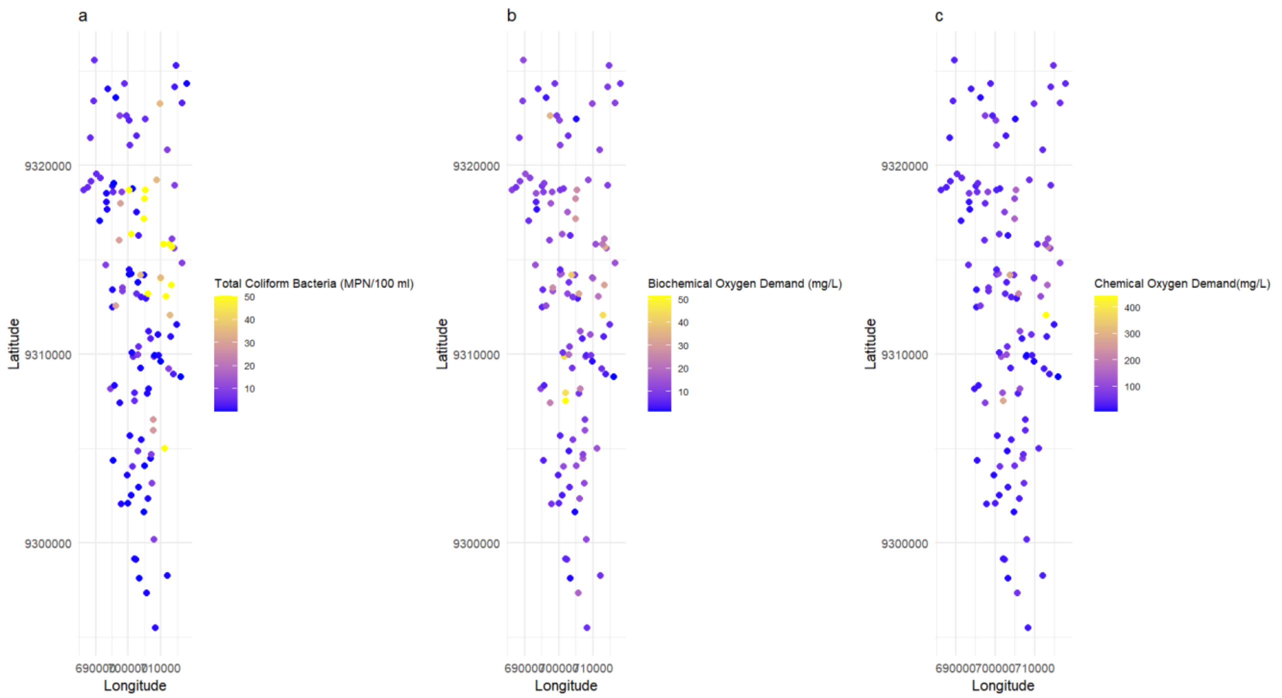

Figure 3. Plots of (a) Total Coliform Bacteria, (b) BOD, and (c) COD

As illustrated in Figure 3, the spatial distribution of colours for the three parameters appears randomly scattered, with no visible systematic gradients or directional trends. This randomness indicates that the variables exhibit relatively constant statistical behaviour across the study area. Therefore, the data for Total Coliform Bacteria, BOD, and COD can be considered stationary, satisfying the fundamental assumption required for the Co-Kriging interpolation process.

In this stage of the research, data exploration and validation of statistical assumptions were carried out to ensure the reliability of the spatial modelling process. By confirming that the datasets meet the stationarity requirement, the analysis can proceed confidently to the variogram modelling and Co-Kriging interpolation stages. The results obtained in this section form an essential foundation for accurate and unbiased spatial prediction of river water quality parameters across the study area.

3. Examination of Anisotropic Semivariogram

The anisotropic semivariogram analysis is conducted to assess directional dependence in the spatial variability of the examined variables. Anisotropy can be evaluated through the variogram map (Vmap) as well as by comparing the sill, range, and nugget values of each variable across different directions. The results are summarized in Table 2 and illustrated in Figure 4.

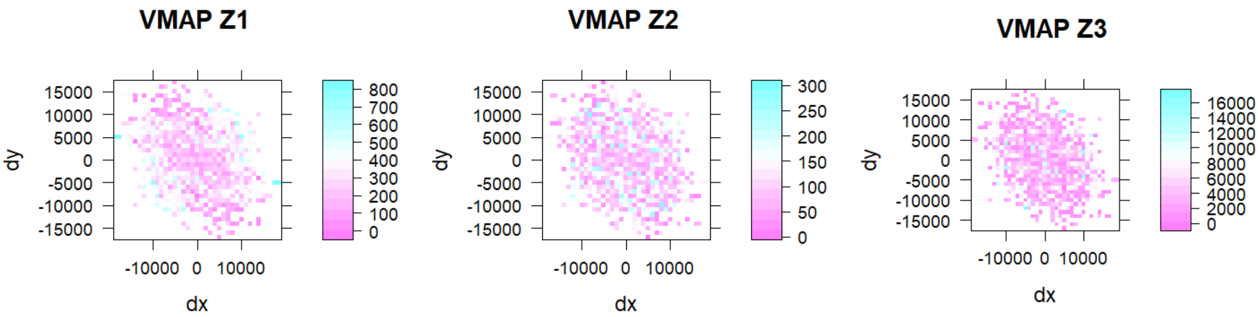

Figure 4. Vmap Z_1, Z_2 dan Z_3

Figure 4 shows that no dominant direction is evident in Z_1, Z_2, or Z_3. The relatively uniform spatial continuity suggests that the data exhibit isotropy rather than anisotropy.

Table 2. Directional Semivariogram

	Z_1				Z_2				Z_3			
						Spherical						
	0°	45°	90°	135°	0°	45°	90°	135°	0°	45°	90°	135°
Range	1169	1169	1169	1169	4752	4752	4752	4752	1499	1499	1499	1499
Sill	303	303	303	303	107	107	107	107	3992	3992	3992	3992
Nugget	155	155	155	155	39	39	39	39	1051	1051	1051	1051
						Exponential						
	0°	45°	90°	135°	0°	45°	90°	135°	0°	45°	90°	135°
Range	11359	11359	11359	11359	1941	1941	1941	1941	315	315	315	315
Sill	397	397	397	397	108	108	108	108	4157	4157	4157	4157
Nugget	242	242	242	242	46	46	46	46	0	0	0	0
						Gaussian						
	0°	45°	90°	135°	0°	45°	90°	135°	0°	45°	90°	135°
Range	5523	5523	5523	5523	1626	1626	1626	1626	531	531	531	531
Sill	308	308	308	308	107	107	107	107	4755	4755	4755	4755
Nugget	141	141	141	141	36	36	36	36	2317	2317	2317	2317

Table 2 shows that the variables Z_1, Z_2, and Z_3 have identical sill, range, and nugget values across all directions. This indicates that the data are isotropic. Therefore, Co-Kriging interpolation can be performed by modeling only the distance function without considering directional effects.

4. Experimental Semivariogram and Cross-Semivariogram

In the Co-Kriging interpolation method, the initial step prior to performing the interpolation is the construction of experimental semivariograms and cross-semivariograms for all variables. Figure 5 presents the experimental semivariogram and cross-semivariogram plots.

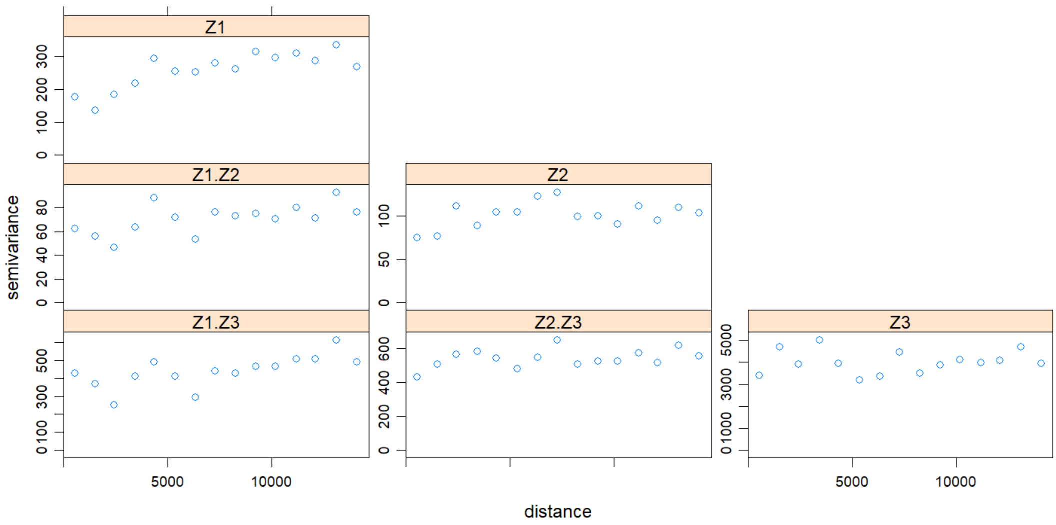

Figure 5. Experimental Semivariogram and Cross-Semivariogram Plots

Figure 5 presents the semivariogram and cross-variogram for three spatial variables (Z_1, Z_2, and Z_3), which are utilized in geostatistical analysis, particularly in the Co-Kriging method. The X-axis represents the distance between points (lag distance), while the Y-axis indicates the semivariance values that describe the degree of data variation across locations. After obtaining the experimental semivariogram and cross-semivariogram, the next step is to fit them with the theoretical semivariogram and cross-semivariogram models.

5. Theoretical Semivariogram and Cross-Semivariogram

The experimental semivariograms and cross-semivariograms were fitted to theoretical models to identify the best representation of spatial dependence. In this study, the spherical, exponential, and gaussian models were evaluated as candidate theoretical semivariogram models.

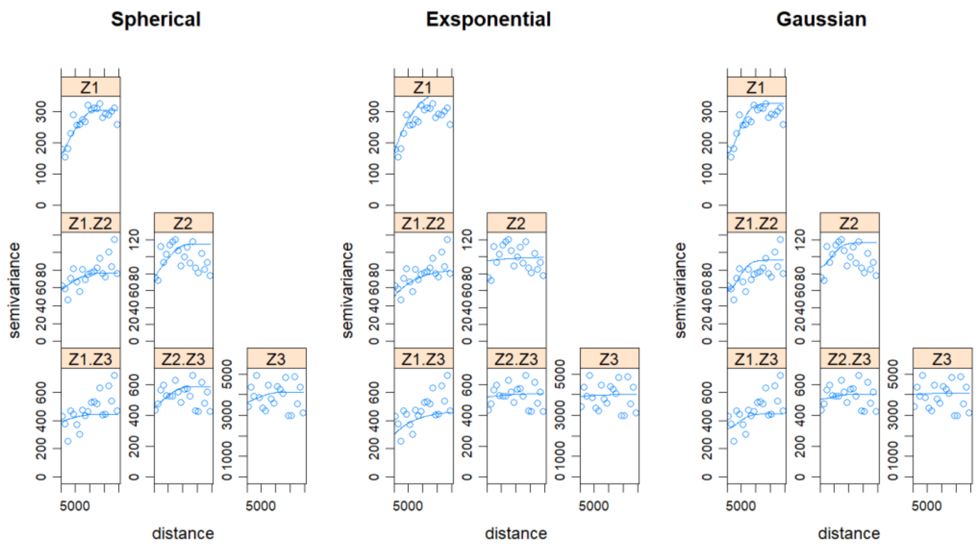

Figure 6. Spherical, Exponential, and Gaussian Semivariogram and Cross-Semivariogram Models

Based on the theoretical semivariogram and cross-semivariogram plots in Figure 6, the parameter estimates of the nugget effect (C_0), sill (C_0 + C), and range (A) were derived. These estimates are summarized in Table 3.

Table 3. Theoretical Semivariogram and Cross-Semivariogram Models

Model	Variable	*Nugget effect (C0)*	*Sill (C0 + C)*	*Range (A)*	Nugget To Sill Ratio *(C0/(C0 + C)%*
Spherical	Z_1	158.238	303.813	11893.150	52%
	Z_2	74.995	114.855	11893.150	65%
	Z_3	3607.696	4121.720	11893.150	87%
	$Z_1 * Z_2$	57.895	76.371	11893.150	76%
	$Z_1 * Z_3$	393.003	444.245	11893.150	88%
	$Z_2 * Z_3$	445.225	585.440	11893.150	76%
Exponential	Z_1	165.808	400.623	7993.272	41%
	Z_2	95.056	99.680	7993.272	95%
	Z_3	3900.475	4029.298	7993.272	97%
	$Z_1 * Z_2$	50.060	82.577	7993.272	61%
	$Z_1 * Z_3$	300.467	473.933	7993.272	63%
	$Z_2 * Z_3$	518.199	542.274	7993.272	96%
Gaussian	Z_1	165.542	325.053	5457.888	51%
	Z_2	87.993	116.691	5457.888	75%
	Z_3	4007.563	4076.177	5457.888	98%
	$Z_1 * Z_2$	56.498	91.335	5457.888	62%
	$Z_1 * Z_3$	343.847	446.306	5457.888	77%
	$Z_2 * Z_3$	505.236	535.299	5457.888	94%

Spatial autocorrelation based on the Nugget To Sill Ratio can be classified into three categories: strong (≤ 25%), moderate (25%–75%), and weak (> 75%)(Sharma & Sood, 2022). Table 3 indicates that under the spherical model, the primary variable Z_1 and the secondary variable Z_2 exhibit moderate spatial autocorrelation, whereas Z_3 shows weak spatial autocorrelation. The semivariogram plot of the Z_1 and Z_2 pair demonstrates an increasing and stabilizing pattern, suggesting strong spatial correlation between these two variables. In contrast, the Z_1 and Z_3 pair displays a relatively flat pattern, indicating weak or even absent spatial correlation. Meanwhile, the Z_2 and Z_3 pair shows an increasing and stabilizing trend, which implies strong spatial correlation.

The exponential model, the results reveal that Z_1 demonstrates moderate spatial autocorrelation, while both Z_2 and Z_3 exhibit weak autocorrelation. Nevertheless, the semivariogram of the Z_1 and Z_2 pair again shows a rising and stabilizing pattern, suggesting strong spatial correlation. The Z_1 and Z_3 pair also indicates the presence of spatial correlation, although not as strong as that of the Z_1 and Z_2 pair. In contrast, the Z_2 and Z_3 pair presents a relatively flat pattern, reflecting weak or negligible correlation.

Under the Gaussian model, both Z_1 and Z_2 display moderate spatial autocorrelation, whereas Z_3 continues to show weak autocorrelation. Similar to the spherical and exponential models, the Z_1 and Z_2 pair maintains an increasing and stabilizing trend, indicating strong spatial correlation. The Z_1 and Z_3 pair also suggests the existence of spatial correlation, albeit weaker than that observed between Z_1 and Z_2. Conversely, the Z_2 and Z_3 pair reveals a flat pattern, indicating weak or absent spatial correlation.

Subsequently, after fitting the experimental semivariogram and cross-semivariogram with their theoretical models, the next step is to select the most suitable model among the three theoretical semivariogram models (spherical, exponential, and Gaussian) for estimating the Total Coliform Bacteria using Co-Kriging (CK) interpolation. The selection of the best-fitting model is carried out through cross-validation using the Leave-One-Out Cross-Validation (LOOCV) approach.

6. Cross-Validation

The selection of the best-fitting theoretical semivariogram and cross-semivariogram model was carried out through cross-validation using the Leave-One-Out Cross-Validation (LOOCV) method. The evaluation criterion was based on the Root Mean Square Error (RMSE) for each model, with the model yielding the lowest RMSE considered the most optimal, as shown in Table 4.

Table 4. Root Mean Square Error (RMSE) Values for Each Model

Model	RMSE
Spherical	11.543
Exsponential	11.549
Gaussian	11.468

Based on Table 4, the Gaussian model produced the lowest RMSE value of 11.468 compared to the Spherical and Exponential models. Therefore, it can be concluded that the Gaussian model is the most suitable theoretical semivariogram for estimating Total Coliform Bacteria in DKI Jakarta in 2022 using Co-Kriging interpolation.

7. Unsampled Point Interpolation Using Co-Kriging

After identifying the best-fitting theoretical semivariogram model, the next step is to estimate Total Coliform Bacteria concentrations at unsampled locations in the rivers of DKI Jakarta for 2022 using the Co-Kriging method with the Gaussian theoretical semivariogram model. Figure 7 below shows the spatial distribution map of the estimated values for both unsampled and sampled locations in the rivers of DKI Jakarta in 2022. Sampled locations are represented by circles, while unsampled locations are indicated by triangles.

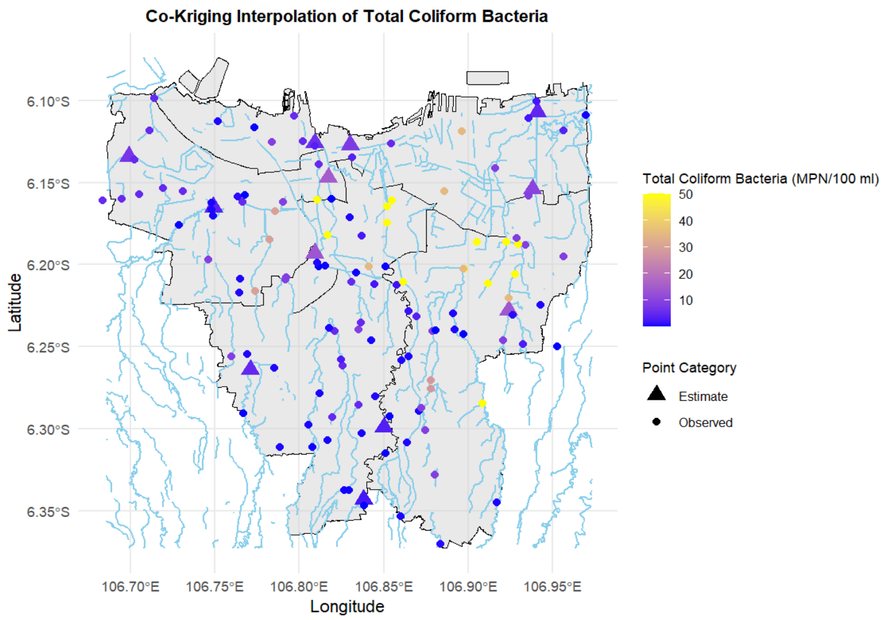

Figure 7. Co-Kriging Interpolation of Total Coliform Bacteria

Table 5 below presents the descriptive statistics of the estimated Total Coliform Bacteria at unsampled locations in DKI Jakarta in 2022.

Table 5. Descriptive Statistics of Estimated Total Coliform Bacteria at Unsampled Locations

Mean	Variance	Std. Deviation	Min	Max
7.711*	19.409*	4.406*	2.382*	14.965*

* expressed in millions

Based on Table 5, the mean estimated value of Total Coliform Bacteria at unsampled locations is 7.711×10^6 MPN/100ml. This is lower than the mean value observed at sampled points, which is 10.560×10^6 MPN/100ml. The variance and standard deviation of the estimates at unsampled locations are also smaller than those at sampled points, being 19.409×10^6 MPN/100ml and 4.406×10^6 MPN/100ml, respectively. Furthermore, the range between the maximum and minimum estimated values is narrower compared to the sampled points, spanning from 2.382×10^6 106 MPN/100ml to 14.965×10^6 MPN/100ml. The lowest estimated Total Coliform Bacteria concentration was observed in the upper reach of the Ciliwung River at geographic coordinates (106.838; −6.343), which corresponds to areas with relatively higher elevation, better flow velocity, and lower population density. These conditions typically promote better dilution and self-purification processes, resulting in reduced bacterial concentrations. In contrast, the highest concentration was found in another section of the Ciliwung River at coordinates (106.817; −6.147), which lies within the downstream area of Jakarta. This segment is characterized by dense residential settlements, industrial activities, and frequent domestic wastewater discharge directly into the river. The limited water circulation and accumulation of organic materials in this area further enhance microbial proliferation.

These spatial differences highlight the significant influence of both anthropogenic and natural factors on bacterial distribution within the Ciliwung River. Upstream regions tend to exhibit better water quality due to reduced human pressure and faster flow, whereas downstream sections suffer from pollutant accumulation, leading to high Total Coliform levels and increased health and environmental risks. This observation is consistent with the spatial distribution map of the estimates presented in Figure 7, where the estimated values closely match those of neighbouring locations. The variance of the estimates is relatively high, indicating that the Total Coliform Bacteria counts vary considerably across different river points. In this study, the high variance may be attributed to the presence of outliers in the dataset and the inclusion of a secondary variable, Chemical Oxygen Demand (COD), which exhibits weak spatial autocorrelation. Additionally, the correlation between Total Coliform Bacteria and COD is also weak.

These spatial prediction results underscore the severe microbial pollution affecting river systems in DKI Jakarta, particularly the Ciliwung River, which flows through highly populated areas, informal settlements, and neighborhoods with limited domestic wastewater treatment infrastructure. The pronounced Total Coliform hotspots observed in downstream sections of the Ciliwung reflect cumulative pollutant loads originating upstream and the reduced natural purification capacity as the river enters denser urban cores. This condition aligns with findings by Pamurda et al. (2023), who reported that population density and poor sanitation infrastructure significantly influence water quality deterioration along the Ciliwung River, especially in Jakarta's urbanized regions. Similarly, Priyono et al. (2021) observed that the water quality in the Bogor segment of the Ciliwung River declines due to increased domestic waste input and runoff from surrounding residential areas, further supporting the observed spatial patterns of Total Coliform distribution.

The smoother Co-Kriging estimates at unsampled locations indicate stable spatial trends shaped by hydrological connectivity and the continuous downstream transport of fecal contaminants. Environmentally, these findings highlight the vulnerability of the Ciliwung River as a critical urban water corridor and underscore significant public health risks for communities residing along its banks or engaging in informal water contact activities. Accordingly, this study provides scientific support for prioritizing sanitation improvements, strengthening wastewater management infrastructure, and implementing targeted monitoring interventions along key reaches of the Ciliwung to mitigate microbial exposure and promote sustainable urban river management in Jakarta.

D. CONCLUSION AND SUGGESTIONS

This study demonstrates the capability of Co-Kriging to predict the spatial distribution of Total Coliform Bacteria in the rivers of DKI Jakarta by incorporating BOD and COD as secondary variables. The approach effectively identified microbial contamination hotspots, particularly in densely populated downstream areas, highlighting the influence of anthropogenic activities on urban river water quality.

Despite its promising performance, the study acknowledges limitations related to extreme outliers, weak correlations with some auxiliary variables, and the use of a single-year dataset, which may constrain temporal interpretation. Future research should explore robust or hybrid Co-Kriging approaches, integrate additional water quality indicators with stronger spatial relationships, and apply multi-year datasets to capture temporal dynamics.

The novelty of this work lies in applying a multivariate Co-Kriging framework with BOD and COD as auxiliary predictors and systematically evaluating semivariogram models to improve predictive reliability. This study advances geostatistical applications for microbial water quality assessment in developing megacities and provides practical insights for supporting targeted monitoring and data-driven river management in urban tropical environments.

REFERENCES

Abbasnejadfard, M., Bastami, M., & Fallah, A. (2021). Investigating the spatial correlations in univariate random fields of peak ground velocity and peak ground displacement considering anisotropy. *Geoenvironmental Disasters*, *8*(1), 24. https://doi.org/10.1186/s40677-021-00196-w

Addis, H. K., Ayalew, B., Gebretsadik, M., Abera, A., Getu, L. A., & Addis, A. K. (2023). Cross-Correlation of Soil Moisture and Stone Content and Their Spatial Pattern Across the Different Slope Aspects and Soil Depth. *Turkish Journal of Agriculture - Food Science and Technology*, *11*(4), 625–633. https://doi.org/10.24925/turjaf.v11i4.625-633.5279

Belkhiri, L., Tiri, A., & Mouni, L. (2020). Spatial distribution of the groundwater quality using kriging and Co-kriging interpolations. *Groundwater for Sustainable Development*, *11*(August), 100473. https://doi.org/10.1016/j.gsd.2020.100473

Bogunovic, I., Filipovic, L., Filipovic, V., & Pereira, P. (2021). Spatial mapping of soil chemical properties using multivariate geostatistics. A study from cropland in eastern croatia. *Journal of Central European Agriculture*, *22*(1), 201–210. https://doi.org/10.5513/JCEA01/22.1.3011

Bojarczuk, A., Jelonkiewicz, Ł., & Lenart-Boroń, A. (2018). The effect of anthropogenic and natural factors on the prevalence of physicochemical parameters of water and bacterial water quality indicators along the river Białka, southern Poland. *Environmental Science and Pollution Research*, *25*(10), 10102-10114. https://link.springer.com/article/10.1007/s11356-018-1212-2

Conolly, J. (2020). Spatial Interpolation. *Archaeological Spatial Analysis*, 118–134. https://doi.org/10.4324/9781351243858-7

Dewana, B. R., Prasetyo, S. Y. J., & Hartomo, K. D. (2022). Comparison of IDW and Kriging Interpolation Methods Using Geoelectric Data to Determine the Depth of the Aquifer in Semarang, Indonesia. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, *8*(2), 215. https://doi.org/10.26555/jiteki.v8i2.23260

Ding, Q., Wang, Y., & Zhuang, D. (2018). Comparison of the common spatial interpolation methods used to analyze potentially toxic elements surrounding mining regions. *Journal of Environmental Management*, *212*, 23–31. https://doi.org/10.1016/j.jenvman.2018.01.074

Djembarmanah, R. S., & Salsabila, G. Analysis of the Water Quality of the River in West Java as the Raw Water for Drinking Water. *Jurnal Presipitasi: Media Komunikasi dan Pengembangan Teknik Lingkungan*, *21*(3), 802-811. https://doi.org/10.14710/presipitasi.v21i3.802-811

Dowd, P. A., & Pardo-igúzquiza, E. (2024). The Many Forms of Co-kriging : A Diversity of Multivariate Spatial Estimators. *Mathematical Geosciences*, *56*(2), 387–413. https://doi.org/10.1007/s11004-023-10104-7

Lei, J. (2020). Cross-Validation With Confidence. *Journal of the American Statistical Association*, *115*(532), 1978–1997. https://doi.org/10.1080/01621459.2019.1672556

Liu, Z., Van Niekerk, J., & Rue, H. (2025). Leave-group-out cross-validation for latent gaussian models. *Sort*, *49*(1), 121–146. https://doi.org/10.57645/20.8080.02.25

Pamurda, A., Mahendra, D., Pratama, M. A., Moersidik, S. S., Rahmawati, S., & Iresha, F. M. (2023). *S*patial dynamics of microplastic pollution in water and sediments of the Ciliwung river along with conditions of water Quality field parameters and population density. *Journal of Ecological Engineering, 24*(8). 296–309. http://dx.doi.org/10.12911/22998993/166311

Payares-garcia, D., Osei, F., Mateu, J., & Stein, A. (2023). A Poisson cokriging method for bivariate count data. *Spatial Statistics*, *57*, 100769. https://doi.org/10.1016/j.spasta.2023.100769

Pham, T. G., Kappas, M., Huynh, C. Van, Hoang, L., & Nguyen, K. (2019). Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of central Vietnam. *ISPRS International Journal of Geo-Information*, *8*(3), 147.

https://doi.org/10.3390/ijgi8030147

Pratama, M. A., Immanuel, Y. D., & Marthanty, D. R. (2020). A multivariate and spatiotemporal analysis of water quality in Code River, Indonesia. *The Scientific World Journal*, *2020*(1), 8897029.https://doi.org/10.1155/2020/8897029

Priyono, A., Rushayati, S. B., & Sujati, A. B. (2021). Water Quality Characteristic of Ciliwung River at Bogor Botanical Garden Segmen, Bogor. *Media Konservasi*, *22*(2), 111-117. https://doi.org/10.11648/j.ijees.20210605.12

Rostami, A. A., Karimi, V., Khatibi, R., & Pradhan, B. (2020). An investigation into seasonal variations of groundwater nitrate by spatial modelling strategies at two levels by kriging and co-kriging models. *Journal of Environmental Management*, *270*(June). https://doi.org/10.1016/j.jenvman.2020.110843

Rusdiyanto, E., Sitorus, S. R., Noorachmat, B. P., & Sobandi, R. (2021). Assessment of the actual status of the Cikapundung river Waters in the Densely-Inhabited Slum area, Bandung City. *Journal of Ecological Engineering*, *22*(11), 198-208. https://doi.org/10.12911/22998993/142916

Schiappapietra, E., Stripajová, S., Pažák, P., Douglas, J., & Trendafiloski, G. (2022). Exploring the impact of spatial correlations of earthquake ground motions in the catastrophe modelling process: a case study for Italy. *Bulletin of Earthquake Engineering*, *20*(11), 5747–5773. https://doi.org/10.1007/s10518-022-01413-z

Sharma, R., & Sood, K. (2022). Characterization of Spatial Variability of Soil Parameters in Apple Orchards of Characterization of Spatial Variability of Soil Parameters in Apple Orchards of Himalayan Region Using Geostatistical Analysis. *Communications in Soil Science and Plant Analysis*, *51*(8), 1065–1077. https://doi.org/10.1080/00103624.2020.1744637

Sikder, A., & Züfle, A. (2020). Augmenting geostatistics with matrix factorization: a case study for house price estimation. *ISPRS International Journal of Geo-Information*, *9*(5), 288. https://doi.org/10.3390/ijgi9050288

Singh, P., & Verma, P. (2019). A comparative study of spatial interpolation technique (IDW and Kriging) for determining groundwater quality. *GIS and Geostatistical Techniques for Groundwater Science*, 43–56. https://doi.org/10.1016/B978-0-12-815413-7.00005-5

Usowicz, B., Lipiec, J., Łukowski, M., & Słomiński, J. (2021). Improvement of spatial interpolation of precipitation distribution using cokriging incorporating rain-gauge and satellite (SMOS) soil moisture data. *Remote Sensing*, *13*(5). https://doi.org/10.3390/rs13051039

Xiao, M., Zhang, G., Breitkopf, P., Villon, P., & Zhang, W. (2018). Extended Co-Kriging interpolation method based on multi-fidelity data. *Applied Mathematics and Computation*, *323*, 120–131. https://doi.org/10.1016/j.amc.2017.10.055