BAREKENG: Journal of Mathematics and Its Applications
June 2024 Volume 18 Issue 2 Page 1283–1296
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss2pp1283-1296

SPATIAL INSIGHTS INTO EARTHQUAKE STRENGTH: A SULAWESI CASE STUDY USING ORDINARY AND ROBUST KRIGING METHODS

Nanda Lailatul Humairah[1], Achmad Fauzan[2*]

[1,2]*Statistics Department, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia*
Jl. Kaliurang KM 14.5, Sleman-Yogyakarta, 55584, Indonesia

*Corresponding author's e-mail: * **achmadfauzan@uii.ac.id***

Article History:

Received: 14th January 2024
Revised: 25th February 2024
Accepted: 3rd May 2024
Published: 1st June 2024

Keywords:

Earthquake;
Interpolation;
Ordinary Kriging;
Robust Kriging.

ABSTRACT

The data from the Meteorology, Climatology and Geophysics Agency (BMKG) in the last 22 years shows that there have been 230 destructive earthquakes in Indonesia with the highest incidence in 2021. One of the islands frequently hit by earthquakes is Sulawesi Island. According to the 2020 Disaster Risk Index Book (IRBI), 63 of the 81 regencies/cities on Sulawesi Island have a high category earthquake risk index. Based on this, information is needed as a first step in disaster mitigation so that the government can take preventive and anticipatory actions to reduce risks associated with earthquakes and ensure the safety of people on the island of Sulawesi, one of which is obtained through spatial interpolation. In this study, the Kriging methods of interpolation, Ordinary Kriging (OK), and Robust Kriging (RK) were used. From the analysis with OK and RK, the best theoretical semivariogram model is the Exponential model with nugget, sill, and range values of respectively 0.40, 0.70, and 6.50 for OK and 0.35, 0.90, and 9.50 for RK. Both methods produced the results that most areas of Sulawesi Island have the potential for shallow earthquakes with a magnitude of around 3.2 to 4.0 on the Richter scale. The potential for earthquakes with high strength is more common around the seas to the east and north of Central Sulawesi Province. The highest estimation results are at the coordinates of 120,029° East Longitude and 1.159° North Latitude, namely in the sea north of South Dampal. According to the results of K-Fold Cross Validation and Leave One Out Cross Validation, the more accurate method for estimating earthquake strength on Sulawesi Island is the RK method because the RMSE and MAPE values in the RK method are smaller than the OK method.

How to cite this article:

N. L. Humairah and A. Fauzan., "SPATIAL INSIGHTS INTO EARTHQUAKE STRENGTH: A SULAWESI CASE STUDY USING ORDINARY AND ROBUST KRIGING METHODS," *BAREKENG: J. Math. & App.*, vol. 18, iss. 2, pp. 1283-1296, June, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

Indonesia is a country located at the boundary of four tectonic plates: the Eurasian Plate, the Australian Plate, the Indian Plate, and the Pacific Plate. This country is also situated within the Ring of Fire region, resulting in high seismic activity. According to the World Risk Report of year 2021, the disaster risk index in Indonesia increased by 0.28 when compared to the index from the previous year, placing Indonesia at 38th out of 181 countries with the highest risk against natural disasters in the world [1]. Throughout 2021, the National Agency for Disaster Countermeasure (BNPB) recorded 3,092 incidences of natural disasters with a total of 665 deaths, and 119 of the cases were earthquakes [2]. According to the Ministry of Energy and Mineral Resources (ESDM) RI (2022), earthquakes are natural disasters that hit Indonesia often. Within the past 22 years (2000-2021), a total of 230 destructive earthquakes were recorded, with the highest number of earthquakes recorded in 2021, at a total of 26 based on data published by the Meteorology, Climatology and Geophysics Agency (BMKG).

Banda Aceh City’s Regional Agency for Disaster Countermeasure (2018) stated that earthquakes causing extensive damage usually have a shallow hypocenter just 60 km or less below the Earth’s surface. According to the book of Indonesia’s Disaster Risk Index (IRBI) for 2020, 63 out of 81 or 77.78% of regions/cities on the Island of Sulawesi have a high-risk index (BNPD, 2021), as the island is located at the boundary of three large tectonic plates; the Indo-Australian Plate, the Pacific Plate, the Eurasian Plate, and a smaller tectonic plate, the Philippine Plate [3]. For the reasons mentioned above, information is needed as the first step in disaster mitigation, which will allow the government to take preventive and anticipative measures to lower the risk posed by earthquakes and guarantee the survival of the community.

One approach to obtaining this information is through spatial interpolation. Spatial interpolation is used to predict the potential magnitude of earthquakes at certain points based on the hypocenters of past earthquakes. Spatial interpolation is a method that utilizes the magnitudes at sampled points to predict values at locations outside the sample [4]. The types of spatial interpolation are Zero-order interpolation, Thiessen polygons, Voronoi polygons, Dirichlet cells, Inverse distance, Kriging, and Spline [5]. The Kriging method was harnessed in this study as it is one of the methods of spatial interpolation. Predicted values from Kriging interpolation come close to the values of the interpolated sample, even when the sample is enlarged to a size that approaches infinity [6]. This estimation method also considers factors that affect the estimation accuracy, which are the number of samples, sample positions, distance between samples and the estimated point, spatial continuity, and the variables involved.

Kriging is one of the interpolation methods that exploits spatial values at the sampled locations to predict values at different locations outside the sample and/or locations that have not been included in the sample [7]. Kriging will produce predicted values which are less pressive when the stationarity assumption is violated and outliers are present. The most popular method of Kriging interpolation is Ordinary Kriging (OK), which does not accommodate outliers. Then, the OK method was further developed into Robust Kriging (RK), which transforms the variogram weight to form robustness against outliers. Some previous studies related to Robust kriging in handling outliers include Zhou [8] improvement in spatial interpolation with a case of seabed terrain, Qu [9] improves the distribution of points and sources of heavy metals in soil using robust geostatistics and a strong spatial receptor model with categorical soil type data, and Syukur [10] conducted research on robust kriging on rainfall data in the South Sulawesi region to estimate the presence of outliers caused by missing data. Based on the aforementioned circumstances, the two spatial interpolation methods are compared to estimate the magnitude of earthquakes in Sulawesi Island based on the hypocenters of preceding earthquakes within the year 2021. Then, the two methods are compared at differing percentages of outliers.

2. RESEARCH METHODS

2.1 Data and Location

The population used in this study is earthquake magnitudes on the island of Sulawesi. The research sample is the magnitudes of shallow earthquakes in Sulawesi Island that happened within the period of January 1st, 2021, and December 31st, 2021, with as many as 1880 points, and the magnitude varies in the domain of 1.0 to 6.3 on the Richter scale. The data was retrieved from BMKG's official website with the following URL http://repogempa.bmkg.go.id/. Figure 1 is an illustration of the research area collected from the official website of the United States Geological Survey (USGS) with the following URL http://earthquake.usgs.gov/.

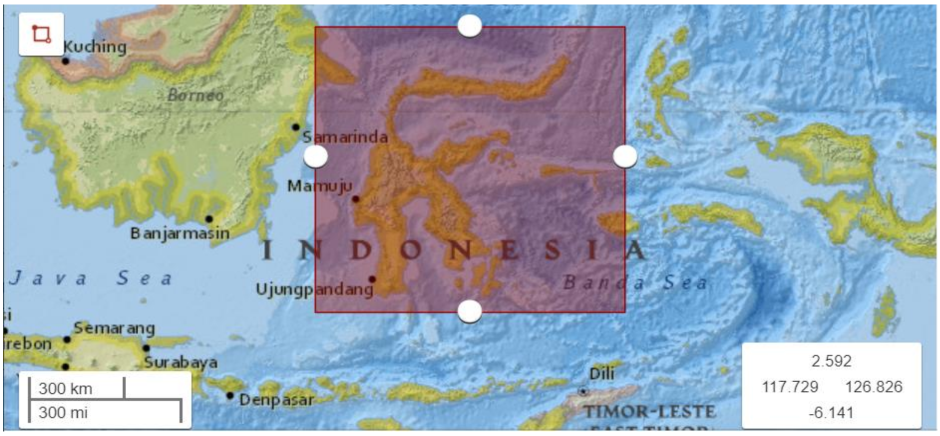

Figure 1. **Map of the Region Studied in This Research**

2.2 Experimental Semivariogram

Experimental semivariogram is a semivariogram obtained from measurement data or samples [11]. The experimental semivariogram can indicate the extent to which a point is not related to other points [12]. The first step in calculating the experimental semivariogram is to calculate the Euclidean distance to get the distance between the two locations, then proceed with calculating the squared difference between each paired location value. The results of the semivariant points formed are quite numerous, resulting in a congested plot, and not all of them can be interpreted. Then a binning process is carried out to reduce the number of points in the experimental semivariogram. The binning process is the process of grouping location pairs based on their distance from each other [13]. After the binning process is carried out, the experimental semivariogram for OK can be calculated using Equation (1) [14].

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

where $\gamma(h)$: semivariogram value and $N(h)$: the number of pairs of sample points that have a distance h, $Z(x_i)$: observed value at location x_i, $z(x_i + h)$: observed value at location $x_i + h$, and h: location distance between samples or lags.

2.3 Theoretical Semivariogram

The curve formed from the experimental semivariogram must be equipped with a mathematical function or model so that it can be used to estimate the semivariance at any distance [15]. The smooth curve fitted to the experimental semivariogram is called the theoretical semivariogram [16]. There are three parameters of the semivariogram: sill, nugget, and range. Sill is the value of the semivariogram when the distance is constant, and its value can be equal to the variance (convergent). Nuggets represent variations at very small distances (lag), including errors in measurement. Range is the distance when the semivariogram achieves sill [17]. After obtaining the values of the three parameters on the semivariogram, the theoretical semivariogram values are calculated. The value obtained from the theoretical semivariogram will be compared with the experimental semivariogram or commonly called structural analysis. Next, is to choose which model has the smallest error value, which will later be used to estimate spatial data. In this study, three theoretical semivariogram models will be used which are most often applied as a comparison for experimental semivariograms, which are the Spherical, Exponential and Gaussian models [18]. The formula used to calculate the theoretical semivariogram of the spherical model as in Equation (2) [19].

$$\gamma(h) = \begin{Bmatrix} C_0 + C\left[\left(\frac{3h}{2a}\right) - \left(\frac{h}{2a}\right)^3\right] & \text{for } 0 < h \leq a \\ C_0 + C & \text{for } h > a \end{Bmatrix}, \tag{2}$$

where h: location distance between samples, C_0: nugget effect, C: partial sill, and a: range. The formula used to calculate the theoretical semivariogram of the exponential model is as **Equation (3) [20]**.

$$\gamma(h) = C_0 + C\left[1 - \exp\left(-\frac{h}{a}\right)\right]. \tag{3}$$

The formula used to calculate theoretical semivariogram of the Gaussian model is **Equation (4) [21]**.

$$\gamma(h) = C_0 + C\left[1 - \exp\frac{-h^2}{a^2}\right] \tag{4}$$

2.4 Ordinary Kriging

The Ordinary Kriging (OK) method is one of the commonly used Kriging methods when the average is unknown and constant. The OK method produces an estimator that is the Best Linear Unbiased Estimator (BLUE). This means that the estimator has the smallest variance compared to other estimators. The data used in the OK method is spatial data with an unknown population average and is assumed to be stationary **[12]**. The statistical model for OK is defined as **Equation (5) [22]**.

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

where $\hat{z}(x)$: predicted value at x, w_i : weight that sets the distance between points, i: 1,2, …., n the number of samples used for estimation, and $z(x_i)$: the i-th actual value of the x variable.

2.5. Robust Kriging

The common OK method does not accommodate the presence of outliers in the data, so the OK method is developed into Robust Kriging (RK), which transforms the weight of the variogram so that it becomes a variogram that is robust to outliers **[7]**. In contrast to the classical semivariogram calculation, to accommodate the existence of outliers in the spatial data, a robust semivariogram is used, as defined in **Equation (6) [23]**.

$$\gamma(h) = \frac{\left\{\frac{1}{N(h)}\sum_{i=1}^{N(h)}|z(x_i) - z(x_i + h)|^{\frac{1}{2}}\right\}^4}{2\left(0{,}457 + \frac{0{,}494}{N(h)} + \frac{0{,}045}{N(h)^2}\right)}, \tag{6}$$

where $\left(0{,}457 + \frac{0{,}494}{N(h)} + \frac{0{,}045}{N(h)^2}\right)$ was bias correction factor $N(h)$ the number of point pairs that have a distance h. From the results of the experimental semivariogram and theoretical semivariogram analysis, structural analysis was then carried out, and the theoretical semivariogram model with the smallest error value was selected to determine the Kriging weights to be used in RK estimation. Estimates on RK are presented in **Equation (7) [21]**.

$$\hat{z}(x) = \sum_{i=1}^{n} w_i z(x_i)\omega(z(x_i)), \tag{7}$$

where $\omega(z(x_i))$ was semivariogram weight transformation to reduce extreme values.

2.6 Cross Validation

Cross validation is one of the most widely used data resampling methods to estimate error values or errors from prediction results and prevent overfitting **[24]**. The commonly used cross validation method is K-fold cross validation. K-fold is a popular cross validation method by folding k data and repeating (iterating) the experiment k times **[25]**. This method can be used if a limited amount of data is available **[26]**. The recommended number of folds for selecting the best model is 10 because the 10-fold cross validation method tends to provide a less biased estimate of accuracy compared to other cross validations [27]. The procedure of 10-fold cross-validation may be described as follows. (1) Randomly rearrange all of the samples. (2) Divide the samples into 10 sub-folds. (3) Within the divided 10 sub-folds: a. Select one-fold to be used as a holdout or test set, b. Use the remaining 9 (from $10 - 1$) folds as the training set, c. Preserve the assessment score and reject the model. (4) Continue the iteration until each individual fold has been used as a testing set. Calculate the mean score of the recorded scores [28]. An illustration of the K-fold cross validation algorithm is presented in Figure 2.

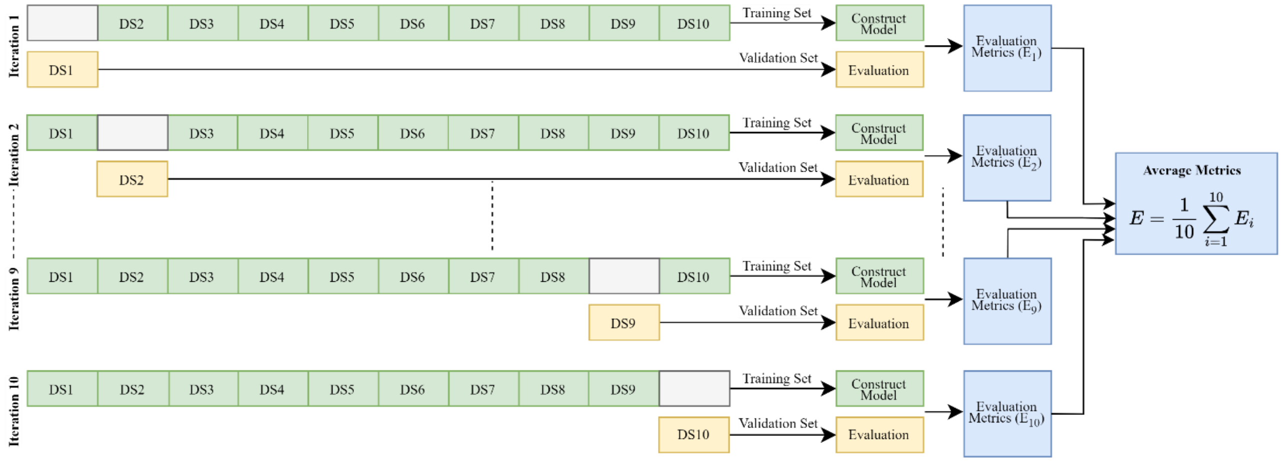

Figure 2. Illustration of the K-Fold Cross-Validation Algorithm [28]

DS is a dataset and average metrics can be the average of Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), or the average value of other metrics. Another method besides K-fold cross validation that is often used is the Leave One Out Cross-Validation (LOOCV) method. LOOCV is a type of cross validation approach where each observation is considered as a testing dataset and other observations $(n - 1)$ are considered as a training dataset or the number of folds is equal to the number of observations $(K = n)$. In LOOCV, when the model formation is completed with the training dataset, predictions are calculated using the testing data. This is repeated n times for each observation from the testing dataset. This method helps reduce bias and randomization. This method aims to reduce error rates and prevent overfitting [29]. While the process is generally user-friendly and does not require any specific configuration, there are instances where it is not recommended, such as when dealing with a somewhat big dataset or a computationally intensive model [30]. Figure 3 is a schematic illustration of the LOOCV technique.

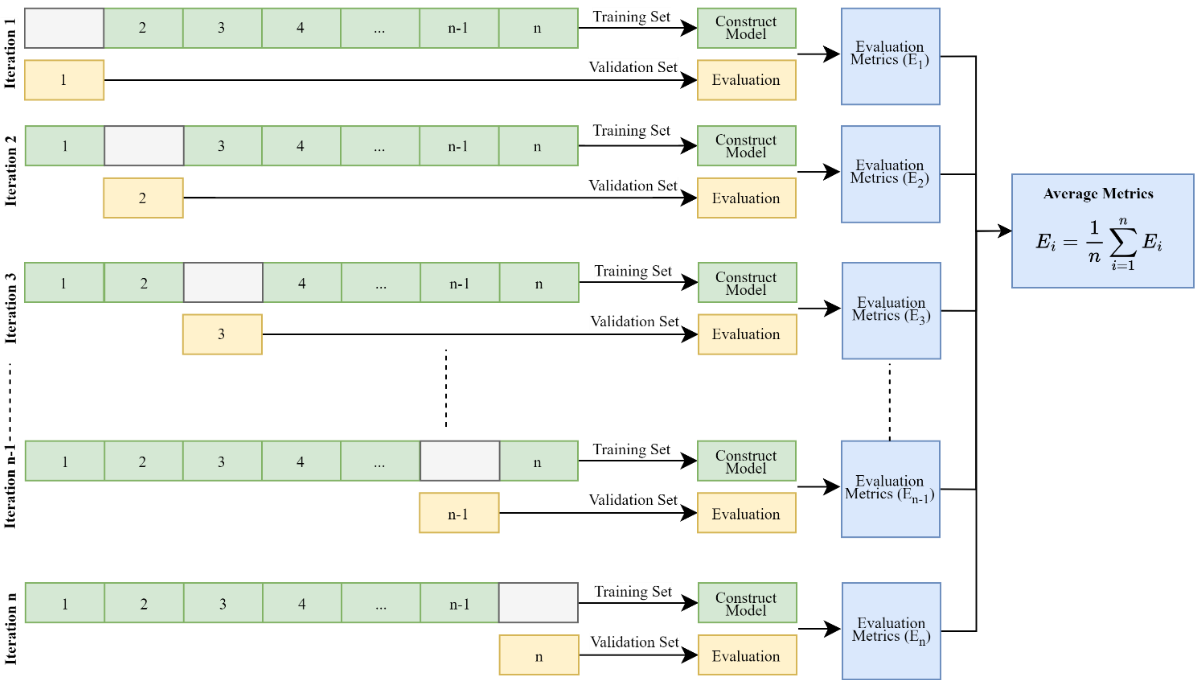

Figure 3. Illustration of the Leave-One-Out Cross-Validation (LOOCV) Technique [31]

2.7 Prediction Accuracy

Evaluation of prediction accuracy serves to measure how well the prediction method performs in predicting data. A measure of the accuracy of prediction results is a measure used to compare prediction models and monitor predictive activity to ensure that the predicted activity operates properly and produces accurate values [32]. The following are several types of measurements of the accuracy of prediction methods [33]: (1) Root Mean Squared Error (RMSE), (2) Mean Absolute Error (MAE), and (3) Mean Absolute Percentage Error (MAPE). The equations of RMSE, MAE, and MAPE are presented in Equation (8).

$$\begin{gathered} RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} \\ MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|^2 \\ MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y_i}}{y_i}\right| \times 100\% \end{gathered} \quad (8)$$

where y_i: actual data, $\hat{y_i}$ = predicted data, n: number of data.

2.8 Research Flow

The stages of data organization and analysis in this study is described by the flowchart on Figure 4.

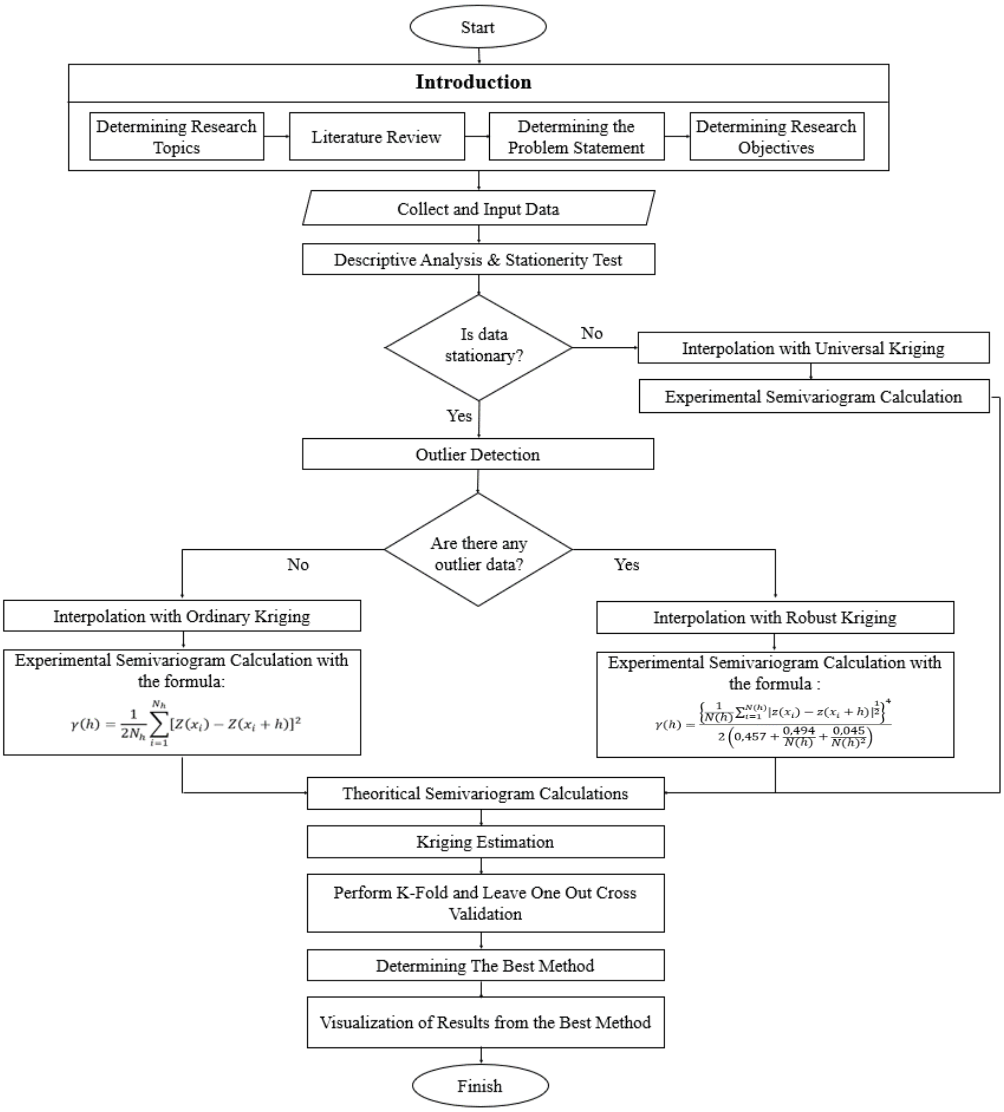

Figure 4. **Research Flow Chart**

According to **Figure 4**, the first step entails a literature review involving the exploration of topics relevant to the current reality, coupled with an association with various scientific articles and news sources. Subsequently, data collection and input for the analytical process follow, specifically focusing on the seismic strength data of shallow earthquakes in Sulawesi Island, acquired from the official BMKG website. Post-data acquisition, a descriptive analysis is conducted to elucidate the general depiction or synopsis of the utilized dataset.

Subsequently, the dataset is partitioned into two segments, namely training data and testing data. Following this, assessments for stationarity and outlier detection are conducted to ensure compliance with the assumptions inherent in the kriging method. The third step involves the spatial interpolation process utilizing the Ordinary Kriging (OK) method. The phases in OK encompass: (1) computation of experimental semivariogram values employing the OK method and **Equation (1)**, (2) determination of theoretical semivariogram values utilizing the nugget, sill, and range parameters in the OK method and **Equation (2)** - **Equation (4)**. Subsequent to this, structural analysis is undertaken, entailing a comparison between the values of the experimental semivariogram and the theoretical semivariogram model, based on the smallest RMSE, MAE, and MAPE values, as determined by **Equation (8)**. The concluding phase in the OK process involves estimation via the OK method utilizing the most suitable semivariogram model.

In spatial interpolation using the Robust Kriging (RK) method, the process begins with the calculation of experimental semivariogram values using **Equation (6)**. Subsequently, the next step involves computing the values of the theoretical semivariogram using the nugget, sill, and range parameters in the RK method, utilizing **Equations (2)** to **Equation (4)**. The subsequent steps mirror those of the OK method, involving structural analysis and the determination of the best-fitted theoretical semivariogram model by selecting the smallest values of RMSE, MAE, and MAPE in the RK method. Finally, estimation is carried out using the RK method with the best-fitted theoretical semivariogram model. As an innovative approach, cross validation is implemented in two (2) ways: (1) K-fold cross validation and LOOCV. The results of the interpolation are then presented in the form of simple maps using the best-fitted method with the assistance of the Quantum Geographic Information System (QGIS) software.

3. RESULTS AND DISCUSSION

3.1 Descriptive Analysis

The following is a histogram of earthquake magnitudes in Sulawesi Island in 2021. Notice from **Figure 5** that most earthquakes occurring in Sulawesi Island and its surroundings are measured between magnitudes 3.0 and 4.0 on the Richter scale. Based on the data, the average magnitude of earthquakes that occurred in Sulawesi Island in 2021 was 3.203, with an average depth of 15.150 km. There are assumptions that must be met when using the Kriging method; according to Krause & Krivoruchko **[4]**, the main statistical assumption in the Kriging method is the stationarity test because if the data is stationary, then the average and semivariogram of the data are the same in all locations within the data coverage (spatial homogeneity). Stationarity testing can be done by plotting, as shown in **Figure 5**.

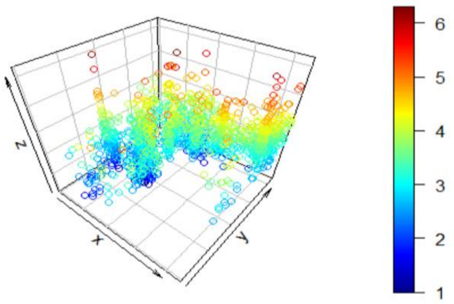

Figure 5. Data Stationarity Plot of Earthquake Magnitudes

In **Figure 5**, the x axis represents longitude, the y axis represents latitude, and z is the magnitude of the earthquake at that location. The 3-dimensional plot shows that earthquake magnitudes are randomly distributed, which can be recognized from the absence of the discrimination of colors by any particular region. So, based on this it can be concluded that visually there is no upward or downward trend in earthquake magnitudes on the island of Sulawesi, in other words, the data is already stationary. Outlier detection is done using a boxplot. Outliers within the 2021 earthquake magnitude data from Sulawesi Island are displayed as a boxplot on **Figure 6**.

Boxplot of Earthquake Strength Data on the Island of Sulawesi in 2021

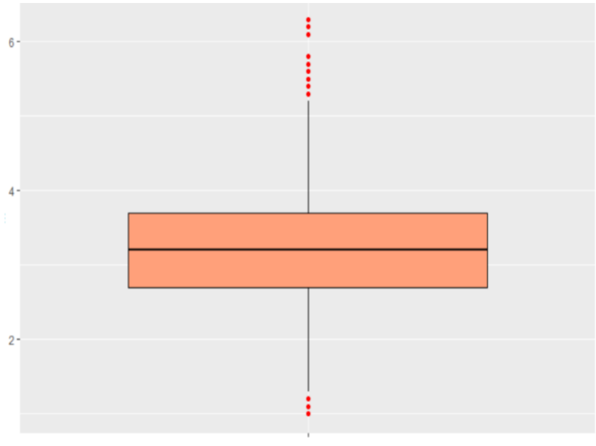

Figure 6. Boxplot of Earthquake Magnitudes in Sulawesi Island

Figure 6 reveals that the earthquake magnitude dataset on Sulawesi Island contains outliers. There are 12 outlier data that are clearly visible in the boxplot, so the RK method can be used to interpolate earthquake magnitude data in Sulawesi Island and can be compared with the OK method. Nevertheless, in this study, in addition to the RK method, results of the analysis using OK will also be presented to determine the efficiency of the RK method compared to OK.

3.2 Estimation of Earthquake Magnitudes Using Ordinary Kriging

3.2.1. Experimental and Theoretical Semivariogram Analysis

After checking the assumptions on the Kriging method, the next step is to determine the experimental semivariogram for earthquake strength data on Sulawesi Island. After obtaining the experimental semivariogram value, the next step is to determine the theoretical semivariogram model. Some theoretical semivariogram models that are commonly used are Spherical, Exponential, and Gaussian models. However, before constructing the theoretical semivariogram model, the parameter values of the semivariogram, sill, nugget and range, are first determined. Then, from several sets of parameters, one set of values is selected for each theoretical semivariogram model based on the smallest MAPE value.

Table 1. Iteration of Theoretical Semivariogram Parameter Values

Iteration	Parameter			MAPE		
	Sill	Nugget	Range	Spherical	Exponential	Gaussian
1	0.5	0.4	3	18.89270916	20.74935674	20.40008205
2	0.5	0.35	3	19.06149509	23.22167004	23.0037326
3	0.6	0.4	5	11.02272699	14.20291025	14.25510615
4	0.6	0.35	5	11.04285636	17.47242333	18.2610484
5	0.7	0.4	6.5	11.16093009	9.068551441	10.7326935
6	0.7	0.35	6.5	10.37745693	12.64529599	15.64316019

According to **Table 1,** it can be inferred that the smallest MAPE in the Spherical model is 10,338 on the 6th iteration of the model, the smallest MAPE in the Exponential model is 9,069 on the 5th iteration and the smallest MAPE in the Gaussian model is 10.733 on the 5th iteration.

3.2.2. Structural Analysis

Examining the best theoretical semivariogram model can be done using structural analysis. Structural analysis is the process of matching the experimental semivariogram with the theoretical semivariogram model. Visually, structural analysis can be carried out using plots followed by the calculation of the error value in each theoretical semivariogram model. A semivariogram modeling plot can be drawn in **Figure 7**.

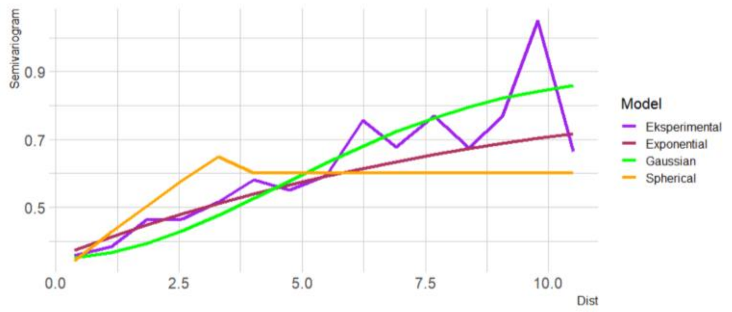

Figure 7. Semivariogram Modelling of Ordinary Kriging

Figure 7 is a plot visualizing the theoretical semivariogram model as compared to the experimental semivariogram. The purple line represents the experimental semivariograms, while the red, green, and orange lines represent the theoretical semivariograms: the Exponential, Gaussian, and Spherical models. The error values in the theoretical semivariogram model are presented in **Table 2** below.

Table 2. Theoretical Semivariogram Error Values

Model	RMSE	MAE	MAPE
Spherical	0.0875	0.0662	10.3775
Exponential	0.0928	0.0639	9.0686
Gaussian	0.0913	0.0704	10.7327

The smallest MAE and MAPE values are found in the Exponential model. Therefore, the theoretical semivariogram model on OK used in estimating earthquake strength data on Sulawesi Island is the Exponential model.

3.2.3. Ordinary Kriging Estimation Results

Estimating the magnitude of earthquakes in the island of Sulawesi using the OK method uses the Exponential model as a suitable theoretical semivariogram. Then the semivariogram parameter values were obtained. The nugget with a value 0.40, the sill with a value of 0.70 and the range with a value of 6.5. Interpolation by the OK method to estimate the magnitude of earthquakes throughout Sulawesi Island is carried out with a grid spacing of 0.1 degrees, which means that for every 0.1 degree distance, an estimate of the strength of an earthquake is carried out based on 1,880 points of earthquakes that have occurred throughout 2021 as a sample. Using the R software, 8,008 location points were generated, which were estimated and visualized in **Figure 8**.

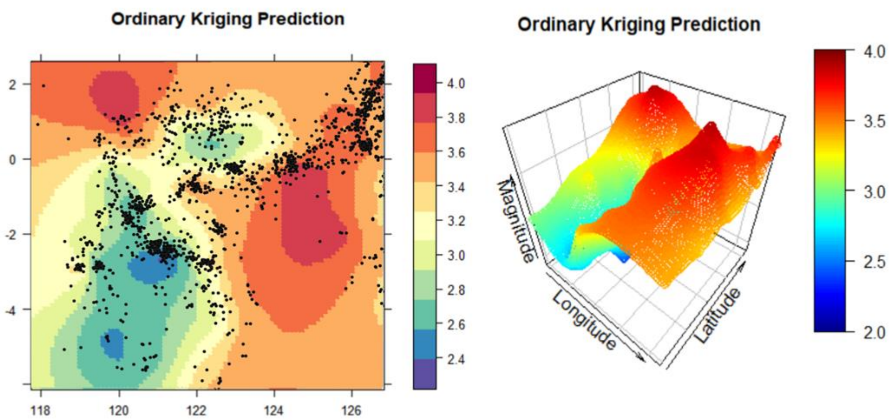

Figure 8. 3-Dimensional Plot of the Results of Estimating Earthquake Magnitudes

Figure 8 is the result of estimating the magnitude of earthquakes on the Island of Sulawesi using the OK method. The black dots that are scattered randomly indicate the location of earthquakes throughout 2021, while the purple to red gradations show the results of earthquake magnitude estimation using OK. The redder the visualization results, the higher the potential for earthquake strength. The visualization results with OK show that most of Sulawesi Island is dominated by light green to reddish-orange gradations, which means that the estimated potential for the strength of most earthquakes on Sulawesi Island is around 3.2 to 4.0 magnitude.

In addition to generating an estimated value of the earthquake magnitudes at unsampled location points, the OK method also generates the variance value of the estimation results. According to Safira [20] , the variance of the estimation results can show the error value in each estimate made. Visualization of the variance of the estimation results are presented in Figure 9 below.

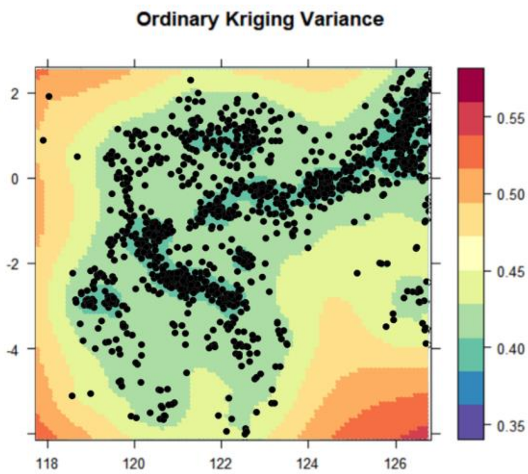

Figure 9. Variance of Estimation Results Using Ordinary Kriging

Figure 9 shows that the variance of the estimation results ranges from 0.4 to 0.56, where the redder the visualization, the greater the error value of the estimation results. So, based on this visualization, we can interpret that the error value from the estimation results with the OK method is dominated by green to reddish orange with an error value of around 0.4 to 0.55.

3.3 Estimation of Earthquake Magnitudes Using Robust Kriging

3.3.1. Experimental and Theoretical Semivariogram Analysis

After carrying out spatial interpolation with the Ordinary Kriging (OK) method, the next step is to interpolate with the Robust Kriging (RK) method. Just like the OK method, the RK method also begins by determining an experimental semivariogram for earthquake magnitude data on the island of Sulawesi. After obtaining experimental semivariogram values with the RK method, the next step is to determine a theoretical semivariogram model similar to that of the OK method. Based on the RK experimental semivariogram, several possible sill, nugget, and range values are obtained and are shown in Table 3.

Table 1. Iteration of Theoretical Semivariogram Parameter Values

Iteration	Parameter			MAPE		
	Sill	Nugget	Range	Spherical	Exponential	Gaussian
1	0.6	0.3	4	15.12571	16.55707736	16.98223
2	0.6	0.35	4	20.88621	14.63202946	13.34472
3	0.8	0.3	6.5	21.97652	8.65071317	15.19393
4	0.8	0.35	6.5	35.53143	10.38991993	9.944033
5	0.9	0.3	9.5	24.13174	11.06940867	22.33296
6	0.9	0.35	9.5	26.32237	8.086130168	16.09552

Based on Table 3, it can be seen that the smallest MAPE in the Spherical model is 15.126 in the 1st iteration model, the smallest MAPE in the Exponential model is 8.086 in the 6th iteration model, and the smallest MAPE in the Gaussian model is 9.944 in the 4th iteration.

3.3.2. Structural Analysis

As with the OK method, assessing for the best theoretical semivariogram model in the RK method also uses structural analysis, both visually, with plots, and by calculating the error values for each theoretical semivariogram model.

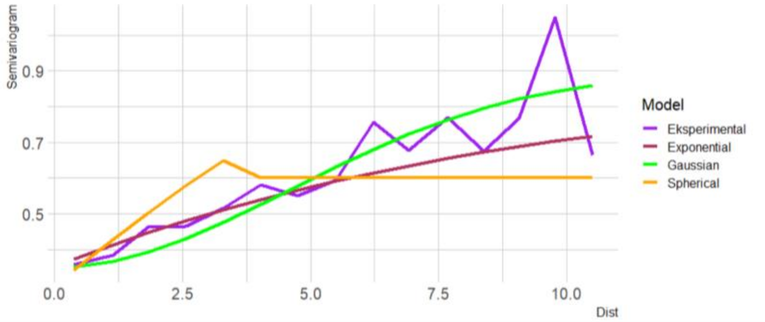

Figure 10. Modelling the Robust Kriging Semivariogram

Figure 10 depicts a plot illustrating the comparison between the theoretical semivariogram model and the experimental semivariogram using the RK approach. The experimental semivariograms are depicted by the purple lines, while the theoretical semivariograms, namely the Exponential, Gaussian, and Spherical models, are represented by the red, green, and orange lines, respectively. The Exponential model, represented by the red line, is the most accurate match for the experimental semivariogram. The RK approach employs RMSE, MAE, and MAPE values to measure the error value for each theoretical semivariogram. The theoretical semivariogram model has three error values, which may be found in **Table 4.**

Table 2. Error Values from Theoretical Semivariogram

Model	RMSE	MAE	MAPE
Spherical	0.1498	0.1052	20.36854
Exponential	0.1059	0.0614	14.92471
Gaussian	0.0896	0.0663	19.43228

Based on the results of computing the error values in **Table 4**, we can perceive that the smallest MAE and MAPE values are in the Exponential model. Then the theoretical semivariogram model in the RK method used in estimating earthquake strength data on Sulawesi Island is the Exponential model.

3.3.3. Results of Estimation Using the Robust Kriging Method

Estimating the magnitude of earthquakes on the island of Sulawesi using the RK method harnesses the Exponential model as a suitable theoretical semivariogram. Then, the semivariogram parameter values were obtained: the nugget with a value of 0.35, the sill with a value of 0.9, and the range with a value of 9.5. Interpolation with the RK method to estimate the strength of earthquakes throughout Sulawesi Island is carried out with the same grid spacing as the OK method, which is 0.1 degrees. This means that for every 0.1 degree distance, an estimate of the strength of an earthquake is carried out based on 1,880 points of earthquake occurrence as a sample. Using the R software, 8,008 estimated location points were produced, and the visualization results are shown in **Figure 11**.

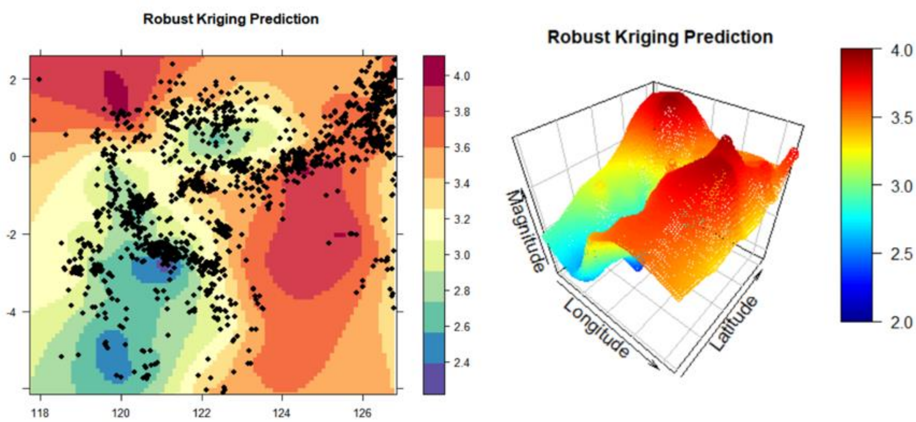

Figure 11. 3-Dimensional Plot of Estimated Earthquake Magnitudes

Figure 11 is the result of estimating the magnitude of earthquakes on Sulawesi Island using the RK method. Scattered black dots indicate the location of earthquake occurrences in 2021, while purple to red gradations indicate the results of earthquake strength estimation using the RK method. The redder the visualization results, the higher the earthquake strength is, and the closer it is to purple, the smaller it is. Based on the visualization with RK, it is shown that most of Sulawesi Island is dominated by yellow to red color gradations, which means the estimation of the potential strength of earthquakes that often occur on Sulawesi Island based on the RK method has a magnitude of around 3.2 to more than 4.0 Magnitude. Similar to OK, the RK method also produces variance or error values from the estimation results. Visualization of the variance of the estimation results is presented in **Figure 12** below.

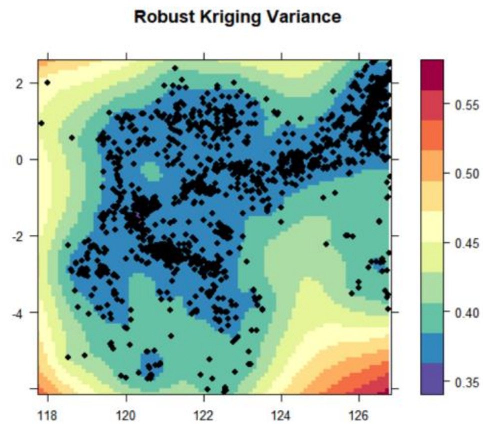

Figure 12. Variance of the Estimation Results Using Robust Kriging

Figure 12 illustrates that the variance of the estimation results in the RK method ranges from 0.35 to 0.55, where the redder the visualization, the greater the error value of the estimation results. So, based on this visualization, it can be seen that the error value from the estimation results with the RK method is dominated by blue to green colors with an error value of around 0.35 to 0.45. So, when compared with the error value from the estimation results with the OK method, the RK method produces a smaller error.

3.4 The Best Method for Estimating Earthquake Magnitudes

Selection of the best method for estimating earthquake strength data on Sulawesi Island, it is done by assessing the RMSE and MAPE values in the 10-fold cross validation and LOOCV. **Table 5** shows the result of the RMSE and MAPE values for the OK and RK methods, respectively.

Table 1. Error Values of Ordinary and Robust Cleaning

Methods	10-Fold CV		LOOCV	
	RMSE	MAPE	RMSE	MAPE
Ordinary Kriging	0.622	16.188	0.619	16.126
Robust Kriging	0.618	16.061	0.617	16.041

According to the RMSE and MAPE values in **Table 5** with the 10-fold Cross Validation and LOOCV methods, the error values for the two methods are not much different. This could happen because the data used in this study contained only a few outliers, just 12 out of 1880 data or around 0.6%. Although the prediction error values are not much different, it can be seen that the RMSE and MAPE with the two Cross Validation methods are smaller in the RK method, so it is reasonable to conclude that the more accurate spatial interpolation method for estimating earthquake magnitudes in Sulawesi Island is the RK method. Based on the results of the best spatial interpolation for estimating the magnitude of earthquakes on the island of Sulawesi that have been obtained by the RK method, a visualization is drawn in the form of a simple map to encourage understanding and is shown in **Figure 11**. As described in Section 3.1, this study not only employs the RK method but also presents an analysis using the OK method to assess its effectiveness. Despite the presence of a minimal percentage of outliers (0.6%) in the dataset, the RK method is found to be superior to the OK method, albeit the difference is not statistically significant. Future research could involve simulations to determine the minimal threshold of outliers' quantity and quality necessary to produce a statistically significant difference in metric values.

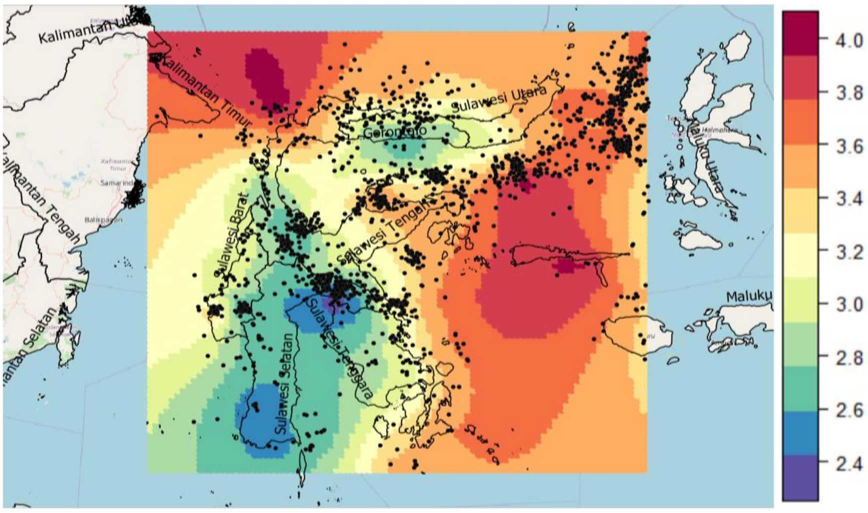

Figure 13. **Visualization of the Results of Interpolation of Earthquake Strength in Sulawesi Using Robust Kriging**

Figure 13 is a visualization of the interpolation results using the RK method. The scattered black dots indicate the location of earthquake occurrences in 2021, while purple to red gradations indicate the results of earthquake magnitudes estimation using the RK method. Based on the visualization in **Figure 13**, it is revealed that earthquakes with large magnitudes are more likely to occur around the sea to the east and north of Central Sulawesi Province, which are shown in gradations of red to dark red. The largest estimation result is at coordinates 120.029° East, 1.159° South Latitude, which is located in the sea north of South Dampal, Tolitoli Regency, Central Sulawesi, while the smallest estimate is at coordinates 121.129° East Longitude, -2.741° South Latitude, specifically at Malili, East Luwu Regency, South Sulawesi, shown in blue to purple.

4. CONCLUSIONS

The most frequent magnitude of earthquakes that occurred on Sulawesi Island and its surroundings in 2021 was between magnitudes 3.0 and 4.0 with an average of 3,203 magnitude. Then the largest earthquake strength was 6.3 Magnitude which occurred in Tojo Una-Una, Central Sulawesi on July 26 2021. Then the results of interpolation using Ordinary Kriging (OK) and Robust Kriging (RK) obtained the best theoretical semivariogram model to estimate earthquake strength, between shows the Exponential model. The estimation results with the two methods are also in magnitudes, namely that most of Sulawesi Island has the potential to experience earthquakes with a magnitude of around 3.2 to 4.0 magnitude. The potential for earthquakes with large magnitudes is more likely to occur around the sea to the east and the sea to the north of Central Sulawesi Province. The largest estimation result is at coordinates 120.029° East, 1.159° South Latitude, at the sea north of Dampal Selatan, Tolitoli Regency, Central Sulawesi while the smallest estimate is at coordinates 121.129° East Longitude, -2.741° South Latitude, namely in Malili, East Luwu Regency, South Sulawesi. The RK method is a more accurate method for estimating the strength of earthquakes on Sulawesi Island because the results of calculations using 10-fold cross validation and LOOCV on the RK method error values, namely RMSE and MAPE, which are smaller than the RMSE and MAPE values from the OK method.

REFERENCES

[1] M. Aleksandrova *et al.*, *The World Risk Index 2021*. Bündnis Entwicklung Hilft, 2021.

[2] BNPB, “Catatan Refleksi Akhir Tahun Penanggulangan Bencana 2021,” https://bnpb.go.id/berita/catatan-refleksi-akhir-tahun-penanggulangan-bencana-2021.

[3] A. Sompotan, *Struktur Geologi Sulawesi*. Bandung: Institut Teknologi Bandung, 2012.

[4] E. Krause and K. Krivoruchko, “Concepts and applications of Kriging,” 2022.

[5] I. Al Aswant, “Analisis perbandingan metode interpolasi untuk pemetaan ph air pada sumur bor di kabupaten Aceh besar berbasis SIG,” Final Task, Universitas Syiah Kuala, Aceh, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:127975385

[6] Suprajitno, *Pengantar Geostatistik*. Jakarta: Universitas Indonesia, 2005.

[7] Darmanto and Soepraptini, *Robust Kriging Untuk Interpolasi Spasial Pada Data Spasial Berpencilan (Outlier)*. Malang: Universitas Brawijaya, 2009.

[8] P. Zhou, J. Chen, and S. Wang, "A Dual Robust Strategy for Removing Outliers in Multi-Beam Sounding to Improve Seabed Terrain Quality Estimation," *Sensors*, vol. 24, no. 5, p. 1476, Feb. 2024, doi: 10.3390/s24051476.

[9] M. Qu, J. Chen, B. Huang, and Y. Zhao, "Enhancing apportionment of the point and diffuse sources of soil heavy metals using robust geostatistics and robust spatial receptor model with categorical soil-type data," *Environmental Pollution*, vol. 265, p. 114964, Oct. 2020, doi: 10.1016/j.envpol.2020.114964.

[10] M. I. Syukur, "Penerapan metode robust kriging pada data curah hujan wilayah sulawesi selatan untuk mengestimasi adanya outlier yang disebabkan oleh data hilang," Final Task, Universitas Hasanuddin, Makassar, 2023.

[11] A. D. R. Bahtiyar, A. Hoyyi, and H. Yasin, "Ordinary Kriging dalam estimasi curah hujan di kota Semarang," *Jurnal Gaussian*, vol. 3, no. 2, pp. 151–159, 2014, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[12] A. N. Alfiana, "Metode Ordinary Kriging pada Geostatistika," Universitas Negeri Yogyakarta, Yogyakarta, 2010.

[13] ESRI, "ArcGIS 9.2 Desktop Help: Binning the Empirical Semivariogram," http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Binning_the_empirical_semivariogram.

[14] N. A. C. Cressie, *Statistics for Spatial Data*, Revised edition. 1990.

[15] L. Wang, "Spatial Interpolation," http://www.fresnostate.edu/csm/ees/documents/facstaff/wang/gis200/lecture-notes/gis/chap15.pdf.

[16] H. Mohebzadeh, "Comparison of methods for fitting the theoretical variogram to the experimental variogram for estimation of depth to groundwater and its temporal and spatial variations," *American-Eurasian Journal of Agricultural & Environmental Sciences (AEJAES)*, vol. 18, no. 2, pp. 64–76, 2018, doi: 10.5829/idosi.aejaes.2018.64.76.

[17] I. Y. H. Fikliani, "Estimasi luasan serangan penyakit bulai pada tanaman jagung di kabupaten jombang dengan metode robust kriging," Final Task, Institut Teknologi Sepuluh Nopember, Surabaya, 2016.

[18] M. Mälicke, S. K. Hassler, M. Weiler, T. Blume, and E. Zehe, "Exploring hydrological similarity during soil moisture recession periods using time dependent variograms," *Hydrology and Earth System Sciences Discussions*, vol. 2018, pp. 1–25, 2018, doi: 10.5194/hess-2018-396.

[19] G. Babish, *Geostatistics Without Tears: A Practical Guide to Geostatistics, Variograms and Kriging*. 2000.

[20] M. C. Safira, A. Fauzan, and M. A. S. Adhiwibawa, "Interpolasi polutan Nitrogen Dioksida (NO2) di kota Yogyakarta dengan pendekatan Ordinary Kriging dan Inverse Distance Weighted," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 14, no. 2, pp. 55–66, 2022.

[21] A. Rahmasari and Noeryanti, "Prediksi data spasial yang tidak tersampel dan mengandung pencilan menggunakan metode Robust Kriging (studi kasus: kualitas udara NO2 pemukiman di kota Yogyakarta)," *Jurnal Statistika Industri dan Komputasi*, vol. 06, no. 02, pp. 132–140, 2021.

[22] G. Hatfield, *Spatial statistics. In Practical Mathematics for Precision Farming*. USA: Dakota State University, 2018.

[23] A. S. Wahyudi, Sugito, and D. Ispriyanti, "Metode robust kriging untuk mengestimasi data spasial berpencilan (studi kasus: pencemaran udara gas NO2 di kota Semarang)," *Jurnal Gaussian*, vol. 5, no. 3, pp. 321–330, 2016, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[24] D. Berrar, "Cross-Validation," in *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 2019, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.

[25] F. Ratnawati, "Implementasi algoritma Naive Bayes terhadap analisis sentimen opini film pada Twitter," *INOVTEK Polbeng - Seri Informatika*, vol. 3, no. 1, p. 50, Jun. 2018, doi: 10.35314/isi.v3i1.335.

[26] S. Hulu, "Analisis kinerja metode cross validation dan K-Nearest Neighbor dalam klasifikasi data," Master Thesis, Universitas Sumatera Utara, Sumatera Utara, 2020.

[27] A. Wibowo, "10 Fold-Cross Validation," https://mti.binus.ac.id/: https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/.

[28] P. Sokkhey and T. Okazaki, "Hybrid Machine Learning Algorithms for Predicting Academic Performance," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 1, 2020, doi: 10.14569/IJACSA.2020.0110104.

[29] Geeksforgeeks, "LOOCV (Leave One Out Cross-Validation) in R Programming," https://www.geeksforgeeks.org/loocvleave-one-out-cross-validation-in-r-programming/#:~:text=LOOCV(Leave%20One%20Out%20Cross%2DValidation)%20is%20a%20type,considered%20as%20the%20training%20set.

[30] Z. Shao and M. J. Er, "Efficient Leave-One-Out Cross-Validation-based regularized extreme learning machine," *Neurocomputing*, vol. 194, pp. 260–270, Jun. 2016, doi: 10.1016/j.neucom.2016.02.058.

[31] G.-W. Cha *et al.*, "Development of a prediction model for demolition waste generation using a random forest algorithm based on small datasets," *Int J Environ Res Public Health*, vol. 17, no. 19, p. 6997, Sep. 2020, doi: 10.3390/ijerph17196997.

[32] L. Sunarmintyastuti, S. Alfarisi, and F. S. Hasanusi, "Peramalan penentuan jumlah permintaan konsumen berbasis teknologi informasi terhadap produk bordir pada kota Tasikmalaya," *Jurnal Penelitian Pendidikan*, vol. 16, no. 3, pp. 288–296, Jan. 2017, doi: 10.17509/jpp.v16i3.4824.

[33] E. Woschnagg and J. Cipan, *Evaluating Forecast Accuracy*. University of Vienna, 2004.