International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET2310257

Prediction of Pneumonia Patient Proportions Using Kriging and Inverse Distance Weighted Methods

Okma Arnilia[1], Muhammad Nur Aidi[2], Indahwati[3], Fitrah Ernawati[4]
[1,2,3]Department of Statistics, IPB University, Bogor, Indonesia
[4]National Research and Innovation Agency, Bogor, Indonesia

ARTICLEINFO

Article History:

Accepted: 01 April 2023
Published: 10 April 2023

Publication Issue
Volume 10, Issue 2
March-April-2023

Page Number
391-399

ABSTRACT

Pneumonia is a respiratory tract disease caused by bacteria, viruses, or fungi. In Indonesia, pneumonia is one of the diseases with the second highest number of cases after malaria from 2007 to 2015. Mapping of regions with pneumonia prevalence needs to be done so that the government can pay more attention to areas with high pneumonia rates, and the public can be more aware of areas that are prone to pneumonia. Spatial methods that can be used to predict the distribution of pneumonia patients are ordinary kriging and inverse distance weighted. The research results showed that the best method chosen for predicting the proportion of pneumonia patients in Sumatra, Java, Kalimantan, and Sulawesi islands is the ordinary kriging method with its theoretical semivariogram model being exponential, and the obtained RMSE value is 0.0178. The average proportion of pneumonia patients in Sumatra Island is 0.01655, with the highest proportion value of 0.02504 in Simalungun District, North Sumatra. Then, the average proportion of pneumonia patients in Java Island is 0.01672, with the highest proportion value of 0.02443 in Garut District, West Java. The proportion of pneumonia patients in Kalimantan Island is 0.01657, with the highest proportion value of 0.01835 in Balikpapan District, East Kalimantan. The average proportion of pneumonia patients in Sulawesi Island is 0.0166, with the highest proportion value of 0.02397 in South Minahasa District, North Sulawesi.
Keywords: Inverse Distance Weighted, Ordinary Kriging, Pneumonia, Semivariogram

I. INTRODUCTION

The rapid development of information technology is currently widely utilized in almost all government and private institutions, including the Ministry of Health. One of the information technologies that is widely used is the Geographic Information System (GIS) as clear spatial information with precise location measurement. The utilization of GIS in the health sector is driven by the orientation of efficiency and effectiveness in both the process and the desired results. GIS is a tool that can show the health condition of the community, especially health problems based on regions or areas.

Currently, mapping public health issues, which were previously done manually, is being abandoned since the development of digital mapping technology for health has been widely used in health institutions. In Indonesia, the development of health GIS began in the 1990s, known as PPM GIS (Geographic Information System for Infectious Disease Control). In the 2000s, health GIS or ArcView was developed for public health (Malaria, TB, Immunization), as most sources of a disease occurrence are related to environmental factors. One disease that is related to environmental aspects is pneumonia. Pneumonia is a respiratory tract infection caused by bacteria, viruses, or fungi. In Indonesia, pneumonia is one of the diseases with the highest number of cases and is one of the countries with a high burden of pneumonia disease.

Based on the data from the Ministry of Health of Indonesia in 2020, during a period of six years from 2009, the coverage of pneumonia detection in toddlers ranged from 20% to 30%. Then, from 2016-2019, the coverage of pneumonia detection increased to 50%-65%. The highest coverage of pneumonia detection in 2019 was in West Papua Province and DKI Jakarta. In 2019, the mortality rate due to pneumonia in toddlers was 0.12%.

Mapping the areas that have a distribution of pneumonia disease needs to be done so that the government can pay more attention to areas with a high incidence of pneumonia cases, and the community can be more aware of areas that are prone to pneumonia disease. Mapping areas requires a large amount of data, where spatial data is one of the data that is difficult to obtain, thus requiring a significant amount of cost to obtain it. Therefore, to obtain data in unsampled areas, interpolation can be performed.

Several spatial methods that can be used for interpolation include Trend, Spline, Inverse Distance Weighted (IDW), and Kriging. Each method will produce different interpolations. The Inverse Distance Weighted method is an interpolation method that has a simple formulation, which is to determine the value of an unknown point using a linear combination of weights from a set of sampled points. Additionally, the IDW method provides reasonably accurate results, making its usage widespread in various fields of study, including GIS. Unlike IDW, the Kriging method provides measures of error and confidence. This method uses a semivariogram that represents the spatial differences and values between all pairs of sample data. The semivariogram also shows weights in interpolation.

Pramono (2008) conducted a study on the comparison of interpolation results using IDW and kriging methods and found that the IDW method was more accurate in interpolating the distribution of suspended sediment. Prasetiyowati et al. (2018) predicted the distribution pattern of DHF disease using IDW and kriging methods and found no significant difference between the IDW and kriging methods. Solihah et al. (2021) compared several spatial interpolation methods and found that the spline with tension method was the best 2D spatial interpolation method, although the most significant method was ordinary kriging. Munyati et al. (2021) also conducted spatial interpolation of the savanna forest gradient and dense forest gradient using ordinary kriging and IDW methods and found that ordinary kriging was more accurate in interpolating the gradient of dense forest, while IDW was more accurate in interpolating the gradient of savanna forest.

Based on previous studies, the distribution of the disease was only focused on a certain area, resulting in several regions not being identified. Therefore, an analysis focusing on the distribution of the disease in several major islands in Indonesia is needed. This study used the kriging and IDW methods for interpolating the proportion of pneumonia patients in unsampled districts/cities in Sumatra, Java, Kalimantan, and Sulawesi Islands.

II. METHODS AND MATERIAL

This study uses secondary data, demographic data, and disease data obtained from the Basic Health Research (Riskesdas) in 2013 conducted by the National Institute of Health Research and Development (NIHRD). The data consists of 15,045 individual samples taken from 124 districts in 24 provinces spread across Sumatra, Java, Kalimantan, and Sulawesi islands. The data used is in the form of the proportion of pneumonia cases in all age groups and the coordinate location data (x and y) per district in the four islands. Prediction will be carried out on each island using the Kriging and IDW interpolation methods.

Prediction of Pneumonia Prevalence

The modeling of the proportion of pneumonia patients was conducted with the aim of obtaining the distribution of pneumonia in the study area. The spatial models used in this study were Interpolation Kriging and IDW. Based on the characteristic pattern of pneumonia patient proportions, the appropriate Kriging interpolation method used in this study was Ordinary Kriging. The IDW method is a simple deterministic method that considers nearby points (NCGIA 1997). The interpolation value at an unsampled point will be similar to the sample data located around that point, and the closer the distance between the sample point and the point to be estimated, the greater the weight, by using the IDW method (Watson et al 1985).

$$\hat{\boldsymbol{Z}}(\boldsymbol{x_0}) = \frac{\sum_{i=1}^{n} Z(x_i)\frac{1}{d_i^k}}{\sum_{i=1}^{n}\frac{1}{d_i^k}} \qquad 1$$

With $\hat{\boldsymbol{Z}}(\boldsymbol{x_0})$ as the estimated value at point $\boldsymbol{x_0}$, $\boldsymbol{Z}(\boldsymbol{x_i})$ the value of z at point i, with i = 1, 2, 3, ..., n, $\boldsymbol{d_i}$ the distance between point $\boldsymbol{x_i}$ and point $\boldsymbol{x_0}$, and k is the exponent that determines the weight value for each prediction.

The prediction of the proportion of pneumonia cases at point $\boldsymbol{x_0}$ for Ordinary Kriging is obtained from (Sawyer 2003):

$$\hat{\boldsymbol{Z}}(\boldsymbol{x_0}) = \sum_{i=1}^{n} \lambda_i \boldsymbol{Z}(\boldsymbol{x_i}) \qquad 2$$

Where the constraint $\sum \lambda_i = 1$, $\boldsymbol{Z}(\boldsymbol{x_i})$ is the sum of pneumonia proportion in the sampled districts/cities and n is the number of samples.

Determination of the Best Model

The accuracy of the pneumonia proportion prediction model using Kriging and IDW methods is measured using the RMSE (Root Mean Square Error) statistical method. The selection of the optimal IDW weight and the best variogram for Kriging is done by analyzing the residuals between the observed data and the estimated data using the RMSE criterion. The RMSE value can be calculated using the formula (M Suprajitno 2005).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(\hat{Z}(x_0) - Z(x_i)\right)^2}{n}} \qquad 3$$

With $\hat{\boldsymbol{Z}}(\boldsymbol{x_0})$ as the estimated predicted value, $\boldsymbol{Z}(\boldsymbol{x_i})$ as the observed value at point i, and n as the number of samples used.

Analysis Design for Kriging and IDW

The process of interpolating to predict the pattern of pneumonia prevalence distribution using Ordinary Kriging and IDW methods is explained in Figure 1. The stages of analysis conducted in this study are as follows:

1. Input sample data

The input data consist of location coordinates (longitude and latitude) and the regional variable of pneumonia prevalence proportion.

2. Interpolation Process

A. Ordinary Kriging method

1. Calculation of experimental semivariogram

Prediction of values in unsampled districts is calculated based on nearby sampled districts at a certain distance, to calculate weights that represent spatial differences in each data pair, it is necessary to calculate the semivariogram.

2. Fitting theoretical semivariogram model

The fitting is done by comparing the results of the experimental semivariogram plot and the theoretical sesmivariogram, then selecting the theoretical semivariogram model and inputting the range, sill, and nugget obtained from the experimental semivariogram.

3. Validation of the theoretical semivariogram model

The validation of the selected semivariogram model is carried out by considering the smallest RMSE value.

4. Estimation of pneumonia proportion

After obtaining the theoretical semivariogram model, the estimation of pneumonia proportion in unsampled areas is carried out.

5. Interpolation Kriging

After the estimation process, Kriging interpolation is carried out with mapping on contour maps.

B. IDW method

The deterministic interpolation technique can be divided into two groups, namely global interpolation and local interpolation. Prediction with global technique is carried out using all data sets, while prediction with local technique is carried out using points in the surrounding area.

Calculate the predicted value using the Inverse Distance Weighted method in equation 2.

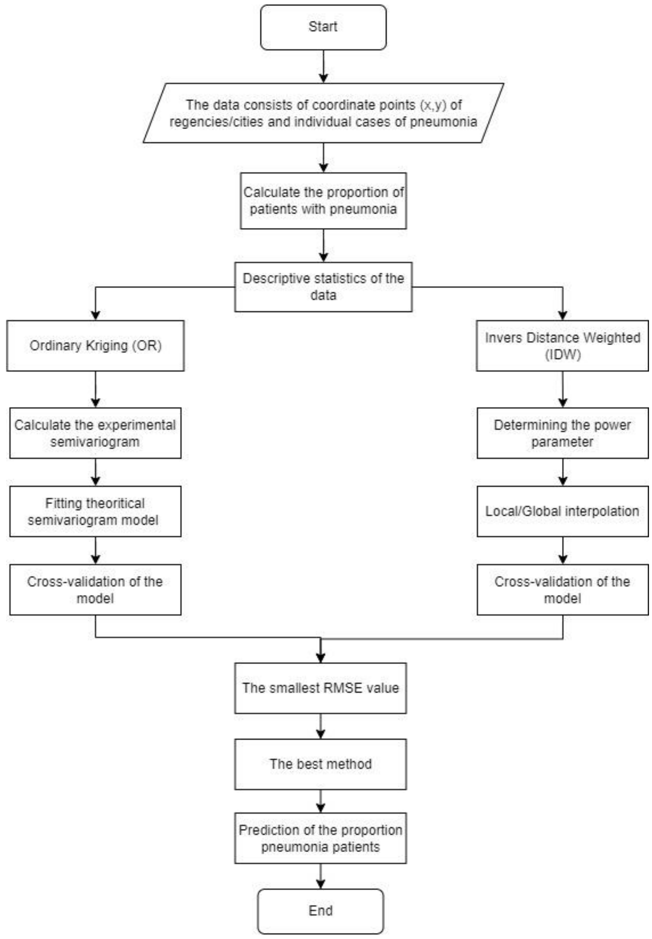

Figure 1. Diagram of the distribution pattern of Pneumonia Incidence

III. RESULTS AND DISCUSSION

Exploratory data analysis

Identification of pneumonia cases based on the results of interviews, with the operational diagnosis of pneumonia by health workers and/or with pneumonia symptoms in the last 12 months, out of a total of 2,697 individual samples, there were only 32 people who suffered from pneumonia on Sumatra Island, out of a total of 10,440 individual samples, there were only 202 people who suffered from pneumonia on Java Island, out of a total of 737 individual samples, there were only 10 people who suffered from pneumonia on Kalimantan Island, and out of a total of 1,171 individual samples, there were only 24 people who suffered from pneumonia on Sulawesi Island.

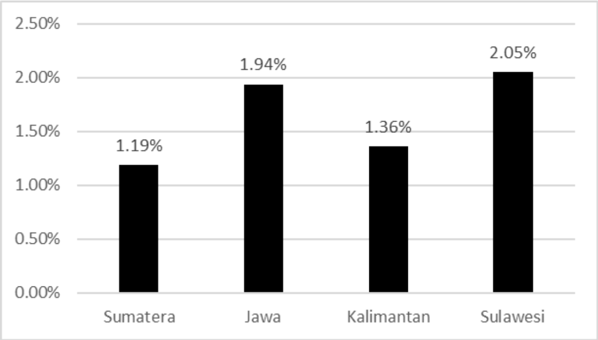

Figure 2. Percentage of Pneumonia cases in Sumatra, Java, Kalimantan, and Sulawesi

The percentage of the proportion of pneumonia patients in Sumatra, Java, Kalimantan, and Sulawesi is relatively low (Figure 2), and the small number of pneumonia patients is due to the lack of identification of individuals who suffer from pneumonia, because out of 24 provinces, only two provinces have health centers in all of their districts/cities that conduct standard Pneumonia examination and management, that is DKI Jakarta and Banten (Ministry of Health of the Republic of Indonesia, 2020).

The prevalence of pneumonia is calculated using the formula, prevalence = ∑ pneumonia cases (diagnosed and/or symptomatic) / ∑ lifetime ART (Riskesdas 2013). The period prevalence and prevalence in 2013 were 1.8% and 4.5%, respectively. Three provinces in Sumatra, Java, Kalimantan, and Sulawesi with the highest period prevalence and prevalence of pneumonia for all ages are Central Sulawesi (2.3% and 5.7%), West Sulawesi (3.1% and 6.1%), and South Sulawesi (2.4% and 4.8%).

The Basic Health Research (Riskesdas) reported that the incidence of pneumonia in the last month (period prevalence) increased from 2.1 ‰ in 2007 to 2.7 ‰ in 2013.

Prediction of the proportion of pneumonia patients using kriging

In the analysis using the kriging method, it is necessary to test the assumption of stationarity, and this can be done in two ways. The first is by looking at the coordinate point plot of each location against the proportion of pneumonia patients, which will result in a 3-dimensional plot. The second method is to perform a formal test, namely the run test. Based on the run test, the p-value of the proportion of pneumonia patients is 0.523. Therefore, it can be concluded that the proportion of pneumonia patients is random. Based on the three-dimensional plot in Figure 3, it can be seen that the proportion of pneumonia patients does not form a specific pattern, indicating that the regional pneumonia variable is stationary.

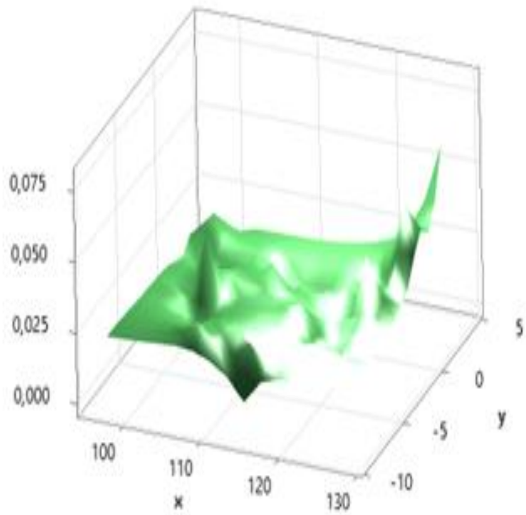

Figure 3. 3D Plot of Proportion of Pneumonia Patients

Before the kriging interpolation method is applied, the calculation of the semivariogram is first performed to determine the most suitable theoretical semivariogram model. In this study, four options of theoretical semivariogram models are used, namely Linear, Spherical, Exponential, and Gaussian. Table 1 presents the experimental semivariogram values of the variable.

Table 1. Experimental semivariogram values of regional proportion of pneumonia variable.

Variable	Number of Pairs	Distance	Semivariogram
Proportion of Pneumonia	2	2.527	2.339950e-04
	4	13.671	8.087567e-05
	22	18.132	2.609783e-04
	28	24.845	3.811346e-04
	32	31.812	2.422011e-04
	42	38.827	2.309855e-04
	36	45.736	2.082957e-04
	4	49.530	1.672064e-04

There are 8 distance classes obtained, with the smallest distance of 2.527 and the largest distance of 49.530. For the regional proportion of pneumonia variable, for data with the smallest distance of 2.527, there are 2 pairs of data with the value of experimental semivariogram of 2.339950e-04, and for data with the largest distance of 49.530, there are 4 pairs of data with the value of experimental semivariogram of 1.672064e-04. Furthermore, there are 8 pairs of data with a distance of 24.845 having the highest experimental semivariogram value of 3.811346e-04, and there are 4 pairs of data with a distance of 13.671 having the smallest semivariogram value of 8.087567e-05. Figure 3 shows the plot graph of the experimental semivariogram results for the proportion of pneumonia cases.

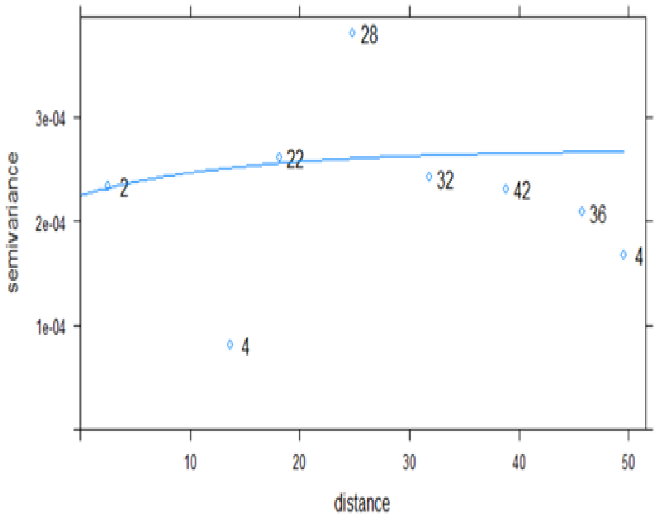

Figure 4. Experimental semivariogram graph of exponential model

The blue line on the experimental semivariogram graph in Figure 4 represents the line plot of the fitted theoretical semivariogram model of the exponential type. From the experimental semivariogram, the parameter values of nugget, sill, and range will be obtained.

Table 2. The values of nugget, sill, and range for the theoretical semivariogram.

Variable		Model			
		Spherical	Exponential	Linear	Gaussian
Proportion of Pneumonia	Nugget	0.00001	0.001	0.0001	0.0001
	Sill	0.00025	0.00025	0.000135	0.00014
	Range	2	2	2	2

Based on the results of fitting the theoretical semivariogram in the table above for the proportion of pneumonia variable, the linear model has the smallest sill value of 0.000135 with nugget and range values of 0.0001 and 2, respectively. Meanwhile, the exponential model has a sill value of 0.00025 with nugget and range values of 0.001 and 2, respectively. The selection of the best theoretical semivariogram model is based on the smallest Root Mean Square Error (RMSE) value. The RMSE values for each theoretical semivariogram model for ordinary kriging analysis are presented in table 3.

Table 3. The RMSE values for each theoretical semivariogram model

Model	RMSE
Linear	0.0179883
Spherical	0.01806036
Exponential	0.01788122
Gaussian	0.01799073

Based on the table above, it can be concluded that the best theoretical semivariogram model for the proportion of pneumonia regional variable is the exponential model. Therefore, the prediction of the proportion of pneumonia patients is done using ordinary kriging method with exponential theoretical variogram model.

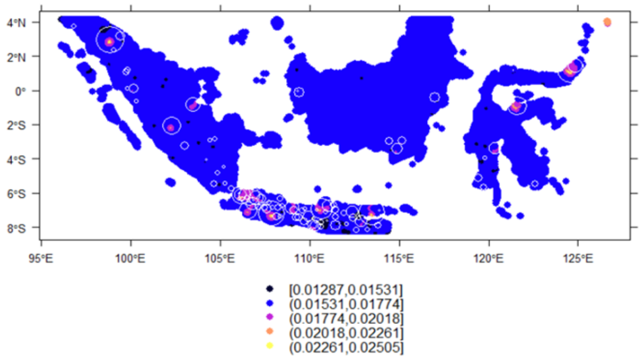

Figure 5. Prediction of the proportion of pneumonia patients in Sumatra, Java, Kalimantan, and Sulawesi islands using Ordinary Kriging

Figure 5 shows the proportion of pneumonia patients in Sumatra, Java, Kalimantan, and Sulawesi islands. The interpolation results show that most of the areas are dominated by blue color, indicating that the proportion of pneumonia patients in the four islands is relatively low.

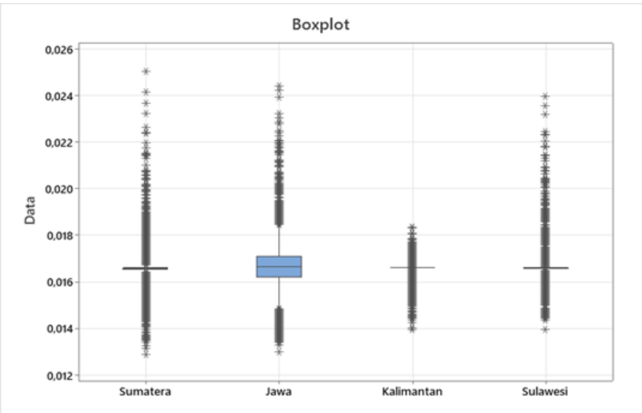

Figure 6. Boxplot of predicted proportion of pneumonia patients in Sumatra, Java, Borneo, and Sulawesi

The boxplot of the prediction results of the proportion of pneumonia patients in Sumatra, Java, Kalimantan, and Sulawesi shows that the average proportion of pneumonia patients in Sumatra is 0.01655 with the highest proportion value of 0.02504 located in Simalungun District, North Sumatra. Meanwhile, the average proportion of pneumonia patients in Java is 0.01672 with the highest proportion value of 0.02443 located in Garut District, West Java. The average proportion of pneumonia patients in Kalimantan is 0.01657 with the highest proportion value of 0.01835 located in Balikpapan District, East Kalimantan. Lastly, the average proportion of pneumonia patients in Sulawesi is 0.0166 with the highest proportion value of 0.02397 located in South Minahasa District, North Sulawesi.

Prediction of pneumonia prevalence using IDW

The result of pneumonia proportion interpolation on unsampled districts/cities in Sumatra, Java, Kalimantan, and Sulawesi using inverse distance weighted method is shown in Figure 7.

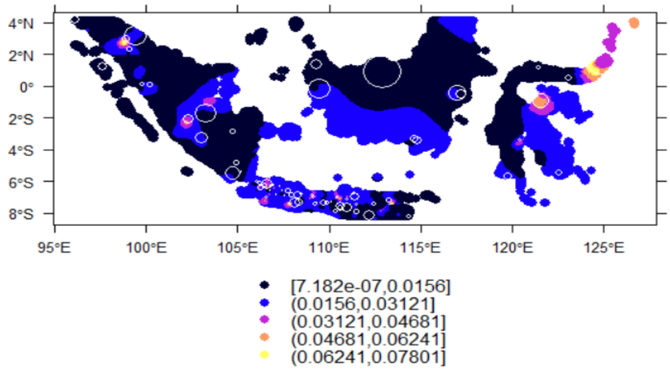

figure 7. Predictions of pneumonia proportion in Sumatra, Java, Kalimantan, and Sulawesi islands using Inverse Distance Weighted (IDW) method

Based on the results of the prediction of the proportion of pneumonia cases in unsampled districts/cities in Sumatra, Java, Kalimantan, and Sulawesi, it shows a diverse color pattern. For Sumatra and Kalimantan, they are dominated by black color, which represents the smallest proportion of pneumonia cases. On the other hand, Java and Sulawesi are dominated by blue color, which represents the second smallest proportion of pneumonia cases.

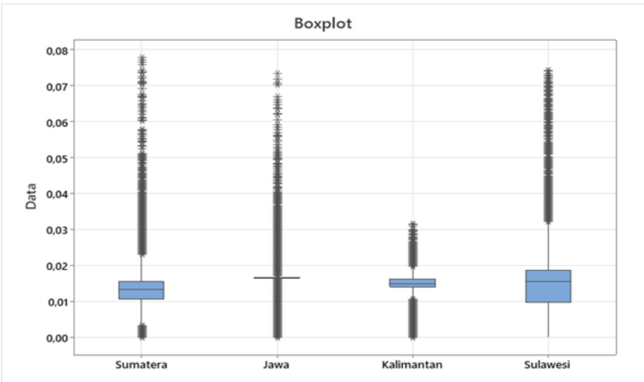

Figure 8. Boxplot of the predicted proportion of pneumonia patients in Sumatera, Java, Kalimantan, and Sulawesi

The boxplot of the predicted proportion of pneumonia patients in Sumatra, Java, Kalimantan, and Sulawesi shows that the average proportion of pneumonia patients in Sumatra is 0.0141, with the highest proportion of 0.078 found in Simalungun Regency, North Sumatra. Meanwhile, the average proportion of pneumonia patients in Java is 0.017, with the highest proportion of 0.0735 found in Pamekasan Regency, East Java. For Kalimantan, the average proportion of pneumonia patients is 0.0147, with the highest proportion of 0.0317 found in Banjar Regency, South Kalimantan. Lastly, the average proportion of pneumonia patients in Sulawesi is 0.0172, with the highest proportion of 0.0746 found in South Minahasa Regency, North Sulawesi.

Comparison of RMSE between Ordinary Kriging and IDW

To determine which method provides the best prediction results between the two methods, a comparison of RMSE values is used. The RMSE value for ordinary kriging is 0.0179 and the RMSE value for inverse distance weighted is 0.0192. The RMSE value for ordinary kriging is smaller than the RMSE value for inverse distance weighted, indicating that the best method for predicting pneumonia patient proportions is ordinary kriging with its best theoretical semivariogram model being the exponential model.

IV. CONCLUSION

The RMSE value for ordinary kriging method is 0.0179 and for inverse distance weighted method is 0.0192. Since the RMSE value of ordinary kriging is smaller than the RMSE value of inverse distance weighted, it can be concluded that the best method for predicting the proportion of pneumonia cases is ordinary kriging with the best theoretical semivariogram model, which is the exponential model.

The prediction results of ordinary kriging in the four major islands of Indonesia, Sumatra, Java, Kalimantan, and Sulawesi, show that the average proportion of pneumonia cases in Sumatra is 0.01655 with the highest proportion value of 0.02504 in Simalungun District, North Sumatra. The average proportion of pneumonia cases in Java is 0.01672 with the highest proportion value of 0.02443 in Garut District, West Java. The average proportion of pneumonia cases in Kalimantan is 0.01657 with the highest proportion value of 0.01835 in Balikpapan District, East Kalimantan. The average proportion of pneumonia cases in Sulawesi is 0.0166 with the highest proportion value of 0.02397 in South Minahasa District, North Sulawesi.

V. REFERENCES

[1] D. Watson and P G M 1985 A Refirement of Inverse Distance Weighted Interpolation Geo-Processing 2 pp 315 – 327

[2] Kementerian Kesehatan Republik Indonesia. 2013. Riset kesehatan dasar 2013. Jakarta: Badan Penelitian dan Pengembangan Kesehatan, Kementerian Kesehatan Republik Indonesia

[3] Kementerian Kesehatan RI. 2020. Profil Kesehatan Indonesia Tahun 2019. Jakarta: Kementerian Kesehatan RI 2020

[4] M Suprajitno. 2005. Pengantar Geostatistik. Jakarta: Universitas Indonesia

[5] Munyati C, Sinthumule NI. 2021. 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 12 100151

[6] NCGIA. 2007. Interpolation: Inverse Distance Weighted. http://www.ncgia.ucsb.edu/pubs/spherekit/inverse

[7] Pramono GH. 2008. Akurasi Metode Idw Dan Kriging Untuk Interpolasi Sebaran Sedimen Tersuspensi. Forum Geografi. Vol. 22, No. 1: 97 – 110

[8] Prasetyowati SS, Sibaroni Y. 2018. Prediction of DHF Disease Spreading Patterns Using Inverse Distance Weighted, Ordinary and Universal Kriging. Journal of Physics: Conf. Series 971 012010. doi :10.1088/1742-6596/971/1/012010
[9] Sawyer CN. 2003. Chemistry For Environmental Engineering and Science, 5th ed.Mc- Graw Hill
[10] Solihah KI, Martono DN, Haryanto B. 2021. Identifying the Best Spatial Interpolation Method for Estimating Spatial Distribution of PM2.5 in Jakarta. Series: Earth and Environmental Science. doi:10.1088/1755-1315/893/1/012043

Cite this article as :

Okma Arnilia, Muhammad Nur Aidi, Indahwati, Fitrah Ernawati, "Prediction of Pneumonia Patient Proportions Using Kriging and Inverse Distance Weighted Methods", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 2, pp. 391-399, March-April 2023. Available at doi : https://doi.org/10.32628/IJSRSET2310257
Journal URL : https://ijsrset.com/IJSRSET2310257