Application of the Ordinary Kriging method for prediction of the positive spread of Covid-19 in West Java

B N Ruchjana1*, A N Falah1 and A S Abdullah2

1Department of Mathematics, Universitas Padjadjaran, Jl. Raya Bandung Sumedang km 21 Jatinangor, Sumedang 45363, Indonesia
2Department of Computer Science, Universitas Padjadjaran, Jl.Raya Bandung Sumedang km 21 Jatinangor, Sumedang 45363, Indonesia

*budi.nurani@unpad.ac.id

Abstract. The Ordinary Kriging method is a method used to predict an observation at an unobserved location based on observed points that are spatially related. Corona Virus Disease 19 (Covid-19) is a contagious disease and viral pathogenic infection caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2). This virus spreads almost all over the world including Indonesia and it gives an influence between location. In this study, a case study was carried out with predictions on the positive distribution data of Covid-19 at 27 districts/cities in West Java. The process of calculating predictions of the positive distribution Covid-19 is started by the determination of the experimental semivariogram model. The Covid-19 data have an Exponential Semivariogram model which is used as an input of the Ordinary Kriging. Furthermore, using the R program for the Ordinary Kriging, we can predict the observation of positive Covid-19 at unobserved locations in West Java.

1. Introduction
Geostatistics is a combination of mining, geology, mathematics, and statistics which was originally developed in the mineral industry to estimate mineral reserves on earth [1]. One of the methods that can be used for prediction in Geostatistics is the Kriging method. The Kriging method is a method used to predict a value in an unobserved location based on spatially related points. The Kriging method consists of several types, namely the Ordinary Point Kriging method and the Ordinary Block Kriging method which assumes an unknown average, the Simple Kriging method which assumes a known and constant average, and the Universal Kriging method which assumes an average is known and not constant [2]. The application of the Ordinary Point Kriging method is used in predicting pollutants in the Meuse river floodplain [3], then the contours of the prediction results are developed with projections to the google map [4], the prediction of these pollutants is developed using the Universal Kriging method [5]. The increase in the spread of Covid-19 that has occurred throughout the country, in this study discussed using the Ordinary Kriging method in predicting the positive spread of Covid-19 in West Java. One of the tools that can be used in the prediction calculation process is R which is an open source software, besides that R provides a special package in the calculation of the Kriging method, namely the gstat package [6].

2. Method
2.1. Experimental Semivariogram
Semivariogram is a function describing the degree of spatial correlation of a spatial random variable Z(x). For example in gold mining, semivariogram give description for two samples taken from the
mining area will fluctuate in gold percentage depending on the distance between those samples. Here is the equation of Experimental Semivariogram [7]:

\[ \hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2. \] (1)

where:
- \( \hat{\gamma}(h) \): experimental semivariogram value with distance \( h \)
- \( Z(x_i) \): observation value in location \( x_i \)
- \( Z(x_i + h) \): observation value in location \( x_i + h \)
- \( N(h) \): the number of point pairs within \( h \)
- \( h \): distance between 2 data.

All possible pairs of \( h \) distances are calculated using the following equation:

\[ |h| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \]

2.2. Theoretical Model

The most commonly used theoretical model is the Spherical model, Gaussian model and Exponential model [7]:

Figure 1. Exponential, Gaussian, and spherical models

1. Exponential Model

The exponential model function increases exponentially as the distance between points increases up to range \( a \) (see Figure 1). This exponential model is continuous but not differentiable at the origin. The range parameter \( a \) determines how quickly the covariance decreases with distance.

\[ \gamma(h) = c \left[ 1 - \exp \left( -\frac{h}{a} \right) \right]. \] (2)

2. Gaussian Model

The Gaussian model increases to a constant asymptotically. The range parameter \( a \) indicates the rate at which the function increases with distance (see Figure 1). The Gaussian model is very smooth because it can be infinitely differentiable. This property is unusual in practice and thus the Gaussian model is infrequent used in real applications.

\[ \gamma(h) = c \left[ 1 - \exp \left( -\frac{h^2}{a^2} \right) \right]. \] (3)
3. Spherical Model
This model is commonly used in practical applications. The function increases almost constantly with distance from the origin, and then stopping at a sill distance (see Figure 1). This behavior often fits practically well the observed data. The spherical model represents a continuous phenomenon, though there are many variations in slope.

\[ \gamma(h) = \begin{cases} 
1.5 \left( \frac{h}{a} \right) - 0.5 \left( \frac{h}{a} \right)^3, & h < a \\
0, & h \geq a
\end{cases} \]  

(4)

2.3. Ordinary Kriging Method
An estimator of Ordinary Kriging method has a Best Linear Unbiased Estimator (BLUE) statistical properties, as follows [8, 9]:

1. Linear
Ordinary Kriging estimator obtained from \( n \) observations of data used to form a linear model, namely:

\[ \hat{Z}(x) = \sum_{i=1}^{n} \lambda_i Z(x_i). \]  

(5)

2. Unbiased
The Ordinary Kriging Estimator is unbiased if it meets the following equation:

\[ E[Z(x) - \hat{Z}(x)] = E\left[ \sum_{i=1}^{n} \lambda_i Z(x_i) - Z(x) \right] = 0 \]

because the mean is assumed to be unknown so \( E[Z(x_i) - Z(x)] = 0 \), the estimator can not be fulfilled by the Ordinary Kriging method.

3. Best
The best point here is that the Ordinary Kriging estimator has a minimum variance estimator, as in equation (5), the variance estimator of the Ordinary Kriging is as follows:

\[ \sigma^2_{ok} = \text{Var}[Z(x) - \hat{Z}(x)] \]

\[ \sigma^2_{ok} = \text{Var}[Z(x)] + \text{Var}[Z(x)] - 2 \text{Cov}[Z(x), Z(x)]. \]  

(6)

To obtain the minimum value of the error variance using the Lagrange Multiplier method with the Lagrange Multiplier parameter \( \mu \), the Lagrange Multiplier equation is stated as follows:

\[ F(\lambda, \mu) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \text{Cov}[Z(x_i), Z(x_j)] + \sigma^2 - 2 \sum_{i=1}^{n} \lambda_i \text{Cov}[Z(x_i), Z(x)] + 2 \mu \left[ \sum_{i=1}^{n} \lambda_i - 1 \right]. \]  

(7)

by deriving the equation for the four variables, so that the kriging weight matrix can be formed as follows:
where:
- \( C_{nn} \): The covariance variance matrix between the observed variable at location \( n \) and the observed variable at location \( n' \).
- \( C_{n0} \): The covariance variance vector between the observed variable at location \( n \) with the variable to be predicted.
- \( \mu \): Lagrange Multiplier Parameters.

The Ordinary Kriging estimator variance equation is obtained as follows:

\[
\sigma_e^2 = \sigma^2 - \sum_{i=1}^{n} \lambda_i \text{Cov}[Z(x_i), Z(x)] - \mu. \tag{9}
\]

The minimum estimator variance is commonly called the Ordinary Kriging estimator variance; thus the best characteristic is the Ordinary Kriging method.

3. Result and Discussion

3.1. Research Data

Based on data obtained from the West Java Province Covid-19 Information and Coordination Center (PIKOBAR), the data used in this study is the data on the positive distribution of Covid-19 in 27 districts/cities in West Java on March 30, 2020. There are 17 districts/cities positively infected with Covid-19 and 10 districts/cities that have not been / are not positively infected with Covid-19, so that 10 districts/cities are used as unsampled locations which are predicted using the Ordinary Kriging method. The data on the positive distribution of Covid-19 can be seen in Table 1.

| No. | District/Cities         | \( x \) (meter) | \( y \) (meter) | Positive Covid-19 |
|-----|------------------------|-----------------|----------------|------------------|
| 1   | Bandung District       | 783878.9        | 9212196        | 5                |
| 2   | Bandung Barat District | 766336.8        | 9239952        | 3                |
| 3   | Bekasi District        | 737892.5        | 9310886        | 18               |
| 4   | Bogor District         | 707864.6        | 9272289        | 7                |
| 5   | Ciamis District        | 211762.7        | 9201105        | 0                |
| 6   | Cianjur District       | 735407.6        | 9246736        | 0                |
| 7   | Cirebon District       | 224759.7        | 9250971        | 2                |
| 8   | Garut District         | 805859.2        | 9182652        | 0                |
| 9   | Indramayu District     | 185429.8        | 9286580        | 0                |
| 10  | Karawang District      | 766998.7        | 9304493        | 6                |
| 11  | Kuningan District      | 223314.2        | 9228029        | 2                |
| 12  | Majalengka District    | 193270.9        | 9241029        | 1                |
| 13  | Pangandaran District   | 223954.2        | 9157489        | 0                |
Experimental Semivariogram Calculation

The semivariogram value or often called is calculated based on all possible distance pairs where the distance function used is the Euclidean distance, a function of the distance $h$ which states the difference between the main variable and the difference in the additional variable $h$, by using equation (1) the semivariogram value and the number of distance pairs. The research data that was sampled consisted of 17 districts/cities, making the manual calculation process for the semivariogram value of the positive distribution of Covid-19 difficult. Program assistance is needed to simplify the process of calculating the semivariogram value, in the R Studio program to calculate the semivariogram value, you can use the variogram function so that the calculation results of the semivariogram value are obtained as in Table 2.

Table 2. Experimental semivariogram calculation for positive Covid-19 distribution

| No | The number of data pairs that are equally distance | Distance   | Experimental Semivariogram |
|----|-----------------------------------------------|------------|-----------------------------|
| 1  | 2                                             | 9404.704  | 50.5                        |
| 2  | 1                                             | 14418.25  | 8                           |
| 3  | 7                                             | 22594.07  | 79.42857                    |
|    |                                               | ...       | ...                         |
| 23 | 2                                             | 126908.8  | 153                         |
| 24 | 1                                             | 134131.9  | 112.5                       |
| 25 | 1                                             | 140253.2  | 392                         |

Table 2 shows that there are 25 criteria for the number of data pairs that are the same distance from 17 districts/cities as observed locations. Furthermore, the plot in Figure 2 is obtained from the semivariogram value to the distance ($h$) where all the information is synthesized in one point per distance class, which then the plot is used to fit the best theoretical model.
3.3. Theoretical Model Fitting

The semivariogram fitting process using a theoretical model approach can be carried out in two stages [10]:

1. Perform the calculation process of the theoretical model on the three models used, then make a plot and it can be seen in plain view based on the plot of the fitting results.
2. Determined based on the minimum Sum Squares Error (SSE) of the three models.

After the process of calculating the theoretical model, fitting between the semivariogram and the theoretical model is carried out using the fit function. Variograms are then plotted and can be seen in Figure 3.
Figure 3 shows in plain view that of the three theoretical model approaches that are following the experimental semivariogram plot is the Exponential model, for more accurate results the SSE is calculated from the three models and can be seen in Table 3. It obtains that the Exponential model has a minimum SSE of 6.023146e-05, so the Exponential model is used as input in the calculation process of the Ordinary Kriging method.

### Table 3. SSE theoretical model

| Model       | SSE         |
|-------------|-------------|
| Exponential | 6.023146e-05|
| Gaussian    | 6.343123e-05|
| Spherical   | 6.216507e-05|

3.4. **Prediction of ordinary kriging for prediction of positive Covid-19 distribution**

After obtaining the Exponential model from the fitting results, the model is used as input in prediction calculations using the Ordinary Kriging method. The purpose of the Ordinary Kriging method is to predict the unobserved locations based on the sample locations. The calculation results can be seen in Table 4.

### Table 4. Prediction of positive Covid-19 distribution at unobserved locations

| No. | District/Cities         | x (meter) | y (meter) | Prediction for Positive Covid-19 |
|-----|-------------------------|-----------|-----------|---------------------------------|
| 1   | Ciamis District         | 211762.7  | 9201105   | 6.389463                        |
| 2   | Cianjur District        | 735407.6  | 9246736   | 6.296900                        |
| 3   | Garut District          | 805859.2  | 9182652   | 7.620536                        |
| 4   | Indramayu District      | 185429.8  | 9286580   | 7.554535                        |
| 5   | Pangandaran District    | 223954.2  | 9157489   | 7.543700                        |
| 6   | Subang District         | 799538.8  | 9274599   | 7.158450                        |
| 7   | Tasikmalaya District    | 184520.8  | 9165875   | 6.884230                        |
| 8   | Banjar City             | 227324.3  | 9184594   | 7.273130                        |
| 9   | Cirebon City            | 230267.7  | 9255425   | 4.645638                        |
| 10  | Sukabumi City           | 712957.2  | 9233271   | 4.901836                        |

4. **Conclusion**

The process of calculating the prediction of the positive distribution of Covid-19 using the functions in the R Program for the Ordinary Kriging method produces the best theoretical model, the Exponential model as input in prediction calculations in unsampled locations. Based on the results of prediction calculations in 10 districts/cities, the average positive spread of Covid-19 is 6 people.

**Acknowledgments**

The authors thank to Rector Universitas Padjadjaran who provided financial assistance for the dissemination of research through Library and Online Data Research Grant 2020 with contract number: 1735/UN6.3.1/LT/2020. Also, thanks to the reviewers for the valuable review of this paper.

**References**

[1] Cressie N A 2015 *Statistics for spatial data Revised Edition* (New York: John Wiley & Sons, Inc.)

[2] Ruchjana B N, Abdullah A S, and Falah A N 2017 *Application of R-based GStat for geostatistics with a spatial data mining approach* (Jatinangor: ALQAPRINT)

[3] Falah A N, Abdullah A S, Parmikanti K, and Ruchjana B N 2017 Prediction of cadmium pollutant with ordinary point kriging method using GStat-R *AIP Conference Proceedings: The 2nd International Conference on Applied Statistics (ICAS II 2016)* 1827 020019
[4] Gunawan A A, Falah A N, Faruk A, Lutero D S, Ruchjana B N, and Abdullah A S 2016 Spatial data mining for predicting of unobserved zinc pollutant using ordinary point kriging *International Workshop on Big Data and Information Security* 83-88 (New Jersey: IEEE)

[5] Falah A N, Subartini B, and Ruchjana B N 2017 Application of universal kriging for prediction pollutant using GStat R *Journal of Physics: Conference Series* 893 012022

[6] Bivand R S, Pebesma E, and Rubio V G 2013 *Applied spatial data analysis with R Second Edition* (New York: Springer)

[7] Youkuo C, Yongguo Y, and Wangwen W 2015 Coal seam thickness prediction based on least squares support vector machines and kriging method *EJGE* 20 167-176

[8] Olea R A 1999 *Geostatistics for engineers and earth scientist* (New York: Kluwer Academic Publishers)

[9] Margaret A 1998 *Basic Linear Geostatistics* (Berlin: Springer-Verlag)

[10] Ruchjana B N 2002 A Generalized Space-Time Autoregressive Order One Model and Its Application to Oil Production Data *Unpublished Ph.D. Dissertation* (Bandung: ITB)