

## **Identification of New Positive Tuberculosis Cases in Aceh Besar District Using Ordinary Criging**

Zurnila Marli Kesuma<sup>1\*</sup>, Latifah Rahayu<sup>2</sup>, Aja Fatimah Zohra<sup>3</sup>,

<sup>1,2,3</sup> *Statistics Department, Syiah Kuala University, Banda Aceh, 23111, Indonesia,*

<sup>\*)</sup>Corresponding author: kesumaku@yahoo.com

### **Abstract**

Tuberculosis is a contagious infectious disease caused by *Mycobacterium tuberculosis*. The risk level of transmission of pulmonary tuberculosis with positive smear of acid-resistant bacteria (BTA). BTA is greater high risk than pulmonary tuberculosis with negative smear. This study aims to predict areas of the number of new BTA positive tuberculosis cases in Aceh Besar district spatially using the kriging method. The data used are secondary data coordinate the point of the number of new cases of positive smear tuberculosis in each puskesmas in the province of Aceh in 2015-2017. The experimental semivariogram calculation is the first step in estimating using ordinary kriging, which is a reference for getting the parameters to be used in theoretical semivariogram calculations. In this case, the semivariogram suitable for new cases of smear positive tuberculosis is the one of the Spherical model. The sill value used is 126.530, the nugget=62, while the range used is based on the distance of each class = 0.158. The results of the analysis showed that the regions with the lowest positive BTA Tuberculosis prediction cases were around Pulo Aceh Sub-District, Montasik Sub-District, and Indrapuri Sub-District, with less than 15 cases. Whereas the region with the most positive predictions of new cases of BTA Tuberculosis was around Baitussalam Sub-District and Jantho City, with more than 30 cases.

**Keywords:** criging method, semivariogram, Tuberculosis

## **Introduction**

According to WHO (2015), India, Indonesia and China are the countries with the highest prevalence of Tuberculosis cases in the world, each (23%, 10% and 10%) of the total events in the world. In Indonesia, tuberculosis is the most deadly disease among other infectious diseases and is the third leading cause of death after heart disease and acute respiratory disease. The number of Tuberculosis cases in Indonesia in 2015 was 330,910 cases, 4,023 of which were new cases of positive smear in Aceh Province (Aceh Health Office, 2018)

The number of new cases of positive smear tuberculosis in each district / city health center in Aceh Province can be analyzed spatially. Spatial data is data presented in the geographical position of an object, related to location, shape and relationship in the space of the earth. One method in spatial analysis is kriging. Kriging is a division of the continuous interpolation method, where the kriging method is developed to solve problems related to geostatistics. In particular, kriging is a spatial statistical method developed with the aim of making it easier in the field of estimation, which is used to handle regionalized variables (Gelfand, A. E 2010).

The kriging method is also called an estimation method that is used to estimate the value at a point. In other words, this method is used to estimate the magnitude of the characteristic value ( $Z$ ) at the non-sampled point based on information from the sampled points around it. According to Gelfand, A. E 2010, if the spatial data with an average ( $\mu$ ) is known and is of constant value, the Simple Kriging method is used. If the spatial data with an average ( $\mu$ ) is unknown, does not contain outliers and does not contain trends, then the Ordinary Kriging method is used. However, if the spatial data contains trends, the Universal Kriging method is used.

Based on the above background, a study were conducted to analyze the data on the number of new cases of positive smear tuberculosis in Aceh Besar District using the Ordinary Kriging method.

## **Materials**

The first step before estimation was conducting a descriptive analysis of data on new cases of positive smear tuberculosis with plots and thematic maps. The following is the thematic map of Aceh Besar District that was used in this study



Figure 1. Thematic Map of Aceh Besar District

Secondly, we check the plot of data for the number of new cases of positive smear tuberculosis and coordinates of points (longitude and latitude) of puskesmas in the District of Aceh Besar that contain outliers or not. If there is, the outliers are discarded or not included during data processing or analysis. The estimation value produced is not correct if the spatial

data that is estimated is spatial outlier. In visualization, the way to find out outliers are using boxplot.

Testing stationary assumptions were done by using stationary data. The resulting plot does not contain a specific trend or by looking at the distribution of varied data or randomness of colors that do not cause certain color gradations. The description of stationary level can be seen in Figure 2.

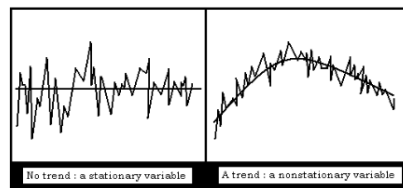


Figure 2. Plot stationer and non-stationer (source: Endra Angen Laksana, 2010)

The next step is Calculate experimental semivariogram. The parameters that will be used for theoretical semivariogram calculations can be calculated after the experimental semivariogram value. This value can be calculated using equation (1).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(s_i) - Z(s_i + h)]^2 \quad (1)$$

With:

$N(h)$  = the number of pairs of data points having h distance

$Z(s_i)$  = the value of observations on location  $s_i$

$Z(s_i + h)$  = value of observations on location  $(s_i + h)$

The parameters that will be used for theoretical semivariogram calculations can be calculated after the experimental semivariogram value is obtained. Some parameters used to find values in a theoretical semivariogram are *Nugget effect* ( $C_0$ ) is the effect is the approach of the semivariogram value at a distance of about zero. Nugget reflects small scale data that are not correlated. *Sill* is a semivariogram value when there is no significant increase (when semivariogram tends to reach a stable value). *Sill* values generally approach the range of data and do not change for unlimited h. *Range* (a) is the maximum (h) distance where there is still a correlation between data.

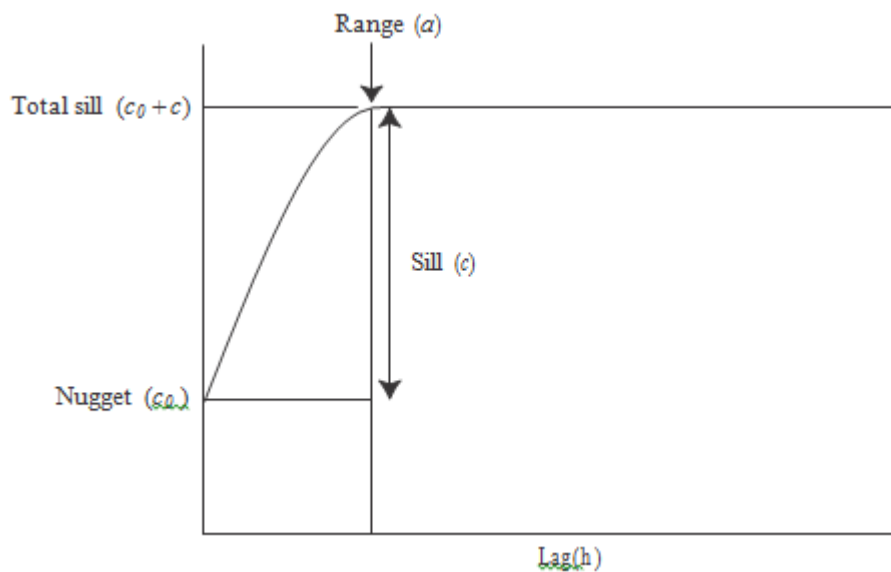


Figure 3. Bounded semivariogram (Source: Christopher D. Lloyd, 2011)

## Methods

Data were obtained from Health Office of Aceh Besar district and the Ministry of Health of the Republic of Indonesia in the form of point coordinates and data on the number of new cases of positive smear tuberculosis in each puskesmas in 2015-2017. The coordinates of the puskesmas point used in this study are longitude 950 BT and latitude 50 LU.  $Z$  is the number of new cases of positive smear tuberculosis,  $X$  is a decimal fraction longitude and  $Y$  is decimal fraction latitude.

A descriptive analysis of data on new cases of positive smear tuberculosis with summary data and thematic maps were conducted. We checked the spatial outlier and test *stationarity assumption*.

The parameters that will be used for theoretical semivariogram calculations can be calculated after the experimental semivariogram value. Structural analysis, were found by comparing experimental semivariogram values with the three theoretical semivariogram models. The final step is calculating the estimated number of new cases of positive smear tuberculosis and calculate the error variation using Ordinary Kriging estimation.

## Results and Discussion

Table 1. Tuberculosis Description in Aceh Besar District

|  | Sex       |           | Total Cases |
|--|-----------|-----------|-------------|
|  | Male      | Female    |             |
| Number of new cases of TBC BTA positif (%) | 532 (63%) | 309 (37%) | 841 (100%)  |
| Number of case TBC (%)                     | 612 (64%) | 349 (36%) | 961 (100%)  |

From Table 1, we found that males are more prone to smear positive TB than females. This is due to the habit of men who often go out of the house, so they are likely to be exposed to the air and inhaled droplets containing *M. tuberculosis*.

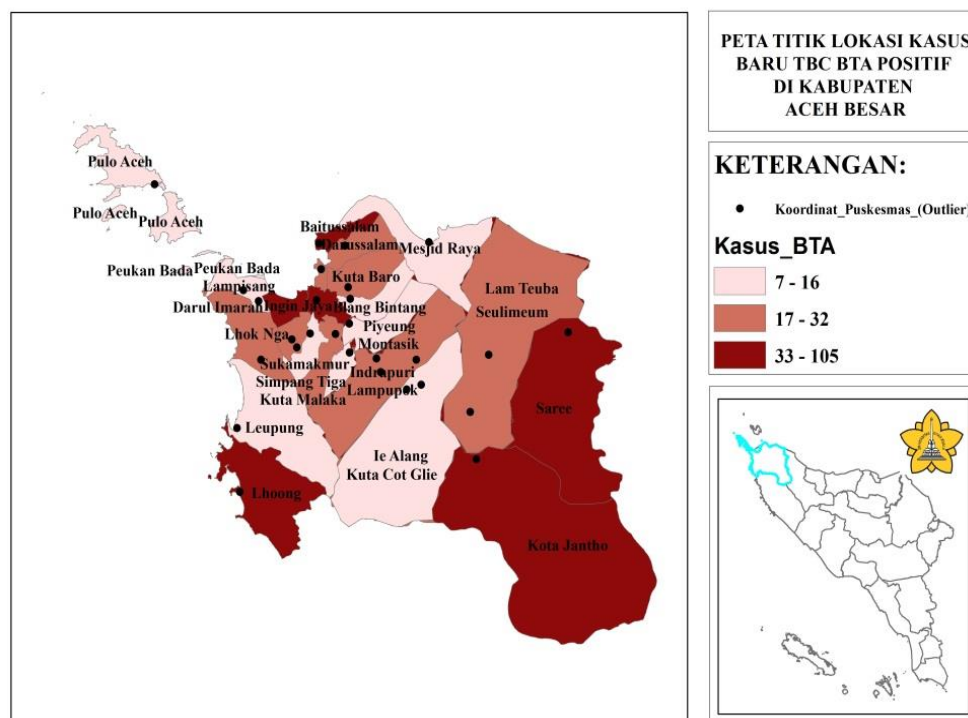


Figure 5. Distribution of the number of new cases of positive smear tuberculosis

Figure 5 Shows the spread of new cases of positive smear tuberculosis in the District of Aceh Besar based on color gradations. If the color on the map gets closer to red, the number of

health centers with the most number of new cases of BTA positive smear tuberculosis is increasing. They are Kota Jantho, Sare, Lhoong, Ingin Jaya, Darul Imarah, and Baitussalam. The pink color shows the health centers that have the fewest positive cases of BTA TB, between the Pulo Aceh, Blang Bintang, Ie Alang, Kuta Malacca, Lampupok, Leupung, Mesjid Raya, Peukan Bada, Piyeung and Simpang Tiga. The number of cases is classified using quantile.

Stationarity test were be done by looking at the randomness of the color in the plot or the plot does not form a particular color gradation. The following figure shows a plot of data on the number of new cases of smear positive tuberculosis.

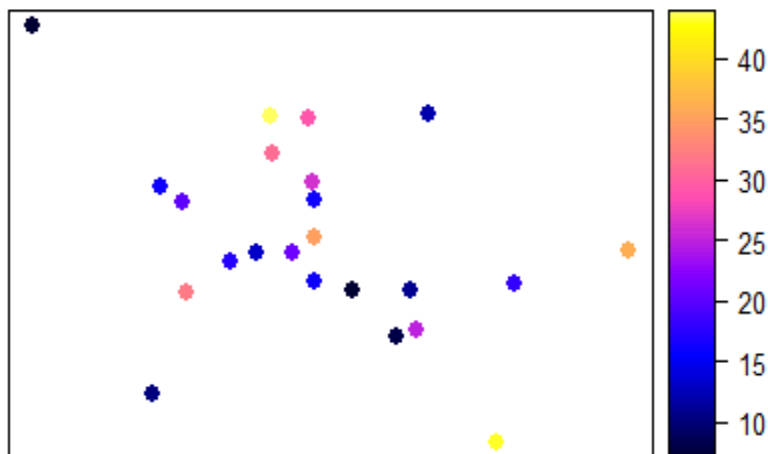


Figure 6. Plot data on the number of new cases of smear positive tuberculosis

Figure 6 shows the colors contained in the plot do not form a particular color gradation, and it is concluded that the assumption of stationarity is fulfilled. The data can be estimated using ordinary kriging.

### **Experimental Semivariogram**

Experimental semivariogram values obtained from actual data are used as a reference to get the parameters that will be used in theoretical semivariogram calculations.

Table 2. Experimental semivariograms of new cases of smear positive tuberculosis

| Class | Distance (h) | Number of data pairs (N(h)) | Semivariogram (γ(h)) |
|-------|--------------|-----------------------------|----------------------|
| 1     | 0,023        | 5                           | 61,700               |
| 2     | 0,034        | 5                           | 81,500               |
| 3     | 0,049        | 14                          | 63,250               |
| 4     | 0,073        | 17                          | 84,706               |
| 5     | 0,105        | 34                          | 97,044               |
| 6     | 0,141        | 44                          | 133,830              |
| 7     | 0,176        | 28                          | 107,518              |
| 8     | 0,218        | 33                          | 160,303              |

In terms of distance, semivariograms form pairs of data that have similar characteristics and relatively close distances between one location and another. From Table 2, it showed that most data pairs are at a distance of 0.141, (44 data pairs) with a semivariogram value is 133.830.

### *Theoretical Semivariograms*

The parameters used to determine the theoretical semivariogram model are nuggets, sills and ranges. The nugget value used in this study was 62. The sill value was obtained from the data variant value of the number of new cases of positive smear tuberculosis which was 126.530 and the range value was obtained from the middle value of "distance" which semivariogram value was close to the sill value, = 0.158.

From the results of structural analysis, it was obtained that a suitable semivariogram for a new case of positive smear tuberculosis was a Spherical model semivariogram. The theoretical Spherical semivariogram model and its images obtained based on the parameters of the model,cal model as follows:

$$\gamma(h) = \begin{cases} 126,530 \left[ \left( \frac{3h}{2(0,158)} \right) - \left( \frac{h}{2(0,158)} \right)^3 \right] & \text{untuk } 0 \leq h \leq a \\ 126,530 & \text{untuk } h \geq a \end{cases}$$



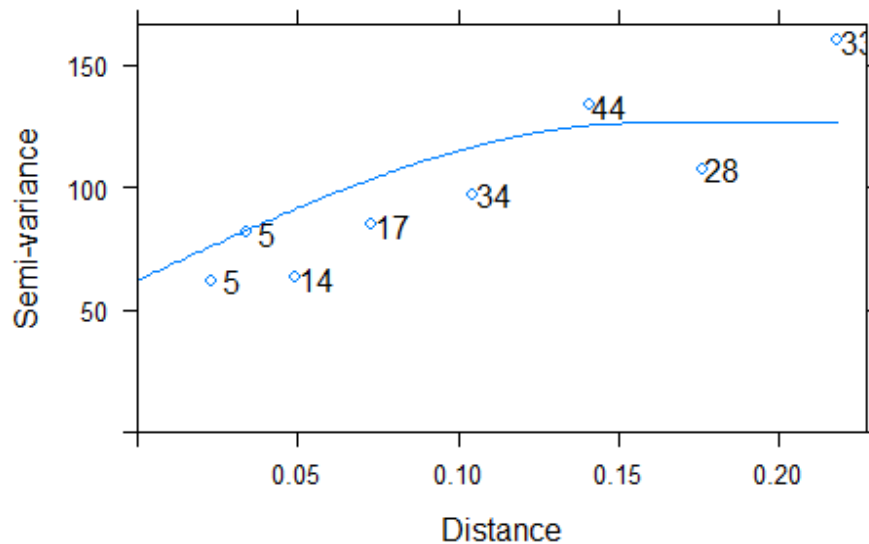


Figure 7. Spherical model in new cases of positive smear tuberculosis

Figure 7. shows the Spherical model in a new case of positive smear tuberculosis being stable after reaching an average range of 0.158 with a sill value of 126.530 and a nugget of 62. Data points on the plot show semivariograms based on experimental data, and lines showing semivariograms based on theoretical (Spherical model). The more stable semivariogram value will produce a small correlation or it can be said there is no correlation.

### ***Estimation of New Positive Tuberculosis Case***

A suitable theoretical semivariogram is used to estimate new cases of positive smear tuberculosis. Based on 23 data used, there were 5000 locations estimated results. Predictions of new cases of positive smear tuberculosis can be seen in the following figure:

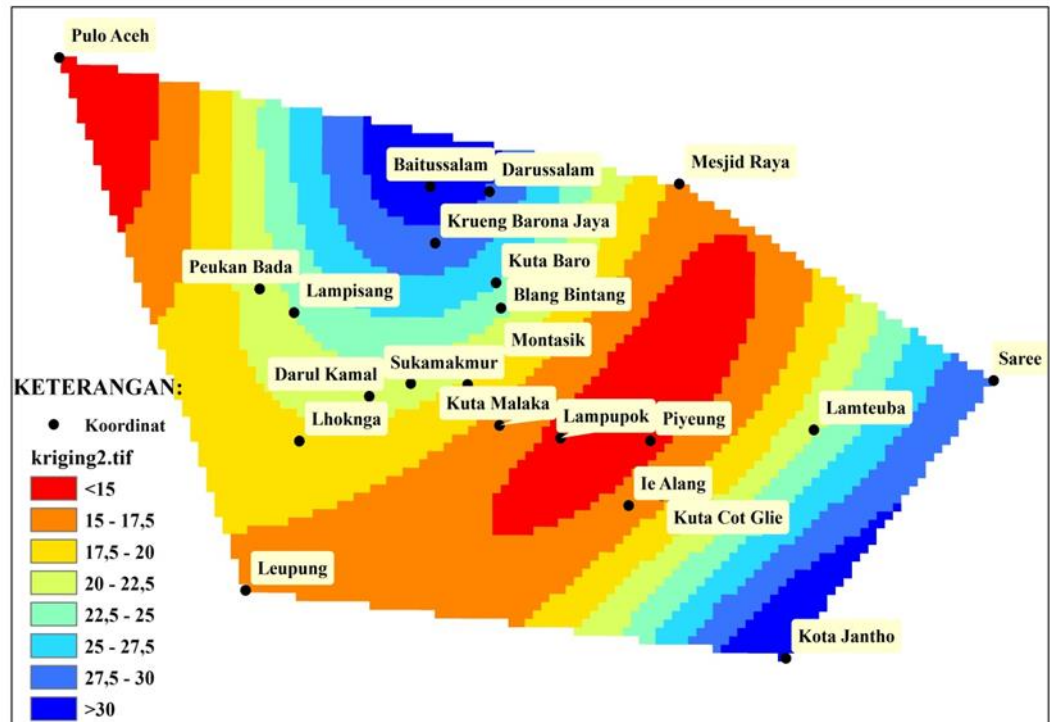


Figure 8. Plot prediction of a new case of positive smear tuberculosis

Figure 8 shows the orange-colored region is the prediction area with the lowest positive tuberculosis cases (<15 cases), the area is around Pulo Aceh Sub-District, Montasik Sub-District, and Indrapuri Sub-District. Whereas the dark blue color is the region with the highest prediction of new positive tuberculosis cases (> 30 cases) the area is located around Baitussalam Sub-District and Jantho City. Based on the gradation of color, the area approaching orange is an area where the number of new cases of smear positive TB is getting lower. Whereas the area that is approaching dark blue is an area with a higher number of new cases of smear positive TB.

Table 8 Predicted areas with the highest cases in Aceh Besar District

| No | Cases Interval | Sub-District  |
|----|----------------|---|
| 1  | 20 – 22,5      | Darul Kamal, Simpang Tiga, Peukan Bada dan Seulimum |
| 2  | 22,5 – 25      | Blang Bintang                                       |
| 3  | 25 – 27,5      | Kuta Baro   |
| 4  | 27,5 – 30      | Krueng Barona Jaya, Darussalam, Lembah Seulawah     |
| 5  | >30            | Baitussalam dan Kota Jantho                         |

## **Conclusion**

The areas with the lowest prediction of new positive Tuberculosis cases were around Pulo Aceh Sub-District, Montasik Sub-District, and Indrapuri Sub-District. Whereas the region with the highest prediction of new positive Tuberculosis cases is located around Baitussalam Sub-district and Jantho City.

## **Acknowledgment**

We would like to thanks to Hibah Lab Funding for supporting this research and publication.

## **References**

- Aceh Health Office. (2018). Aceh Health Profile 2017. Banda Aceh: Aceh Health Office
- Gelfand, A. E., Diggle, P. J., Fuentes, M., & Guttorp, P. (2010). *Handbook of Spatial Statistics*. New York: Taylor and Francis Group.
- Laksana, E. A. (2010). Geostatistical Data Analysis with Universal Kriging [Thesis]. Yogyakarta: Yogyakarta State University.
- LLoyd, D. C. (2011). *Local Models for Spatial Analysis* (Second Edition). New York: Taylor and Francis Group.